Research Paper
Computational Methods for Differential Equations
http://cmde.tabrizu.ac.ir
Vol. 14, No. 1, 2026, pp. 421-469
DOI:10.22034/cmde.2025.63161.2814

The Kolmogorov forward equations, information theory, and mathematical modeling of the Ising model

Mehdi Shams* **and Mohammad Ali Mirzaie**
Department of Mathematics, University of Kashan, Kashan, Iran.

Abstract

In this paper, one of the applications of the Kolmogorov forward equations in solving the equations in the Ising model will be analyzed. Then, the limiting distribution and the stationary distribution of the corresponding birth and death process will be calculated. The Ising model is one of the famous physical models that is used in other sciences. In this paper, the three special cases of the Ising model, i.e., the square $n \times n$ grid, horizontal, and circular systems, will be analyzed and examined. We will show that, in general, Ising models based on the Boltzmann stationary distribution, the maximum likelihood estimator, and a method of moments strategy to estimate the reverse temperature characteristic of the heat baths are the same. The equation related to finding the maximum likelihood estimator and a method of moments strategy cannot be solved analytically, so these estimates will be calculated using the technique of fitting a linear regression model. By the way, it will be observed that the entropy values for the $n \times n$ grid system are smaller than the corresponding values in the horizontal and circular systems. Then, a general case for the convexity of the entropy of the Boltzmann probability function will be introduced. Also, considering an Ising model on two and three points, where the points take values independently and follow a Markov chain, the stationary distribution as well as the Boltzmann probability function, its induced probability function, and maximum likelihood estimator for the parameter will be calculated. Finally, we will review the Metropolis-Hastings algorithm and Gibbs sampling to simulate the one-dimensional and two-dimensional Ising model and also the Potts model in the **R** software. Finally, we will compare the $n \times n$ grid, horizontal, and circular systems. The induced probability vectors by system energies will be found for the three systems. The entropy of these three induced probability vectors and the entropy of Boltzmann probability and their two-by-two Kullback-Leibler divergences will be plotted and compared as a function of the parameter.

Keywords. The Kolmogorov forward equation, Stationary distribution, Limiting distribution, Entropy, Kullback-Leibler divergence.
1991 Mathematics Subject Classification. 62B10; 60J28; 60J75; 60G10.

1. Introduction

Stochastic processes have applications in many physical/natural phenomena. Robert Brown, the Scottish botanist (1773,1858) had seen moments of suspended particles in water under a microscope in 1828 [35]. He thought moments of particles were random and permanent. After that Einstein and Wiener continued research on this matter, and after that, this process was named honor of Brown and Wiener, well known as 'Brownian motion' or 'Wiener process'. According to the applications of stochastic process in physics, readers can refer to references [10, 20, 41] to get familiar with some literature of statistical physics and statistical thermodynamics. In some papers, the investigation of physical models using differential equations and numerical methods has been studied, including [1, 2, 23, 29] and in others, stochastic differential equations [38] and Brownian motion [4] are accelerated by physics. In both cases, numerical analysis can help solve ordinary or stochastic differential equations.

The Kolmogorov forward equations are a type of partial differential equation (PDE) used in Markov processes. The master equation is a result of the Chapman-Kolmogorov and Kolmogorov forward equations, which is analyzed in [3]. The master equation has been proposed by [16, 39] for spin systems [5] and applied to phase separation [11], polymers

Received: 27 August 2024; Accepted: 19 April 2025.
* Corresponding author. Email: mehdishams@kashanu.ac.ir.

[32], and crystal growth [15]. In [8, 40, 42], various topics about the Ising model have been discussed. In [24], the Ising model with quantum feedback was investigated. The Ising model has been used for solving phase separation and wetting/dewetting, lattice-based liquid-gas model and spin glasses, etc. An elementary Ising spin models are used in industry to represent cascading failures in networks [21]. The Ising models can also be used in neural networks [13]. The Ising model is also used in economic networks and exogenous shocks [22].

One of the properties of a Markov stochastic process is its stationarity, meaning that the distribution of all states of the stochastic process is the same as the initial distribution of the process. This property is also used in applied branches of stochastic processes, such as queuing theory and time series. Now the question arises as to what the relationship is between the limiting distribution of a stochastic process and its stationary distribution, which is examined in this paper for the Ising model.

Two important concepts that are studied in the branch of information theory are the concepts of entropy and the Kullback-Leibler divergences, the former is used to indicate uncertainty and the latter is used to calculate the degree of closeness of two distributions. The concept of entropy was introduced by [37] and the concept of Kullback-Leibler divergence by [25]. In studies such as [31, 36], the relationship between these two concepts in information theory and the concepts of statistical physics can be explored. In this paper, these concepts will be used in the Ising model.

In this paper, at first, we will review the time-continuous jump process and birth and death process in section 2. Also, we will bring some applications in practical problems like Poisson process, gambler's ruin, queue model, and Markov Chain Mont Carlo (MCMC) simulation.

After that, in section 3, we will show that the master equations are solved according to the Chapman-Kolmogorov equation and modeling the Ising model with the Kolmogorov forward and backward equations. Then we calculate the entropy of the distribution of a stochastic process and its Kullback-Leibler divergence distance with the stationary distribution. One of the interesting examples is the Ising model on the square $n \times n$ grid, horizontal line, and circle. We prove that, in general Ising models, the maximum likelihood estimator (MLE) and the estimator using the moment method (MME) for the β (parameter in the Boltzmann probability function) coincide. The equation related to finding MLE and MME cannot be solved analytically, so these estimates are calculated using linear regression models. Then the entropy is calculated in three grids, horizontal, and circular systems. We observe that, the entropy values for the horizontal and circular systems are smaller than corresponding values in the $n \times n$ grid. In the following, we show that the entropy of the Boltzmann probability function is a convex function w.r.t. the parameter in some region, and the conditions for its establishment are presented in a general Ising model. At the end of the section, considering an Ising model on two points (three points on a horizontal line or a triangle), where the points take values independently and follow a Markov chain, the stationary distribution as well as the Boltzmann probability function, its induced probability function and maximum likelihood estimator for parameter are calculated.

Then, in section 4, we will simulate the Ising model in **R** software. For this purpose, at first, we discuss Metropolis-Hastings algorithm and generate a random sample from the normal distribution. After that, we will observe that the Gibbs sampler for simulating a two-dimensional Ising model. Also, we will use Metropolis-Hastings for simulating the two-dimensional Ising model. At the end of this section, we simulate the Potts model in lattice with 3 and 4 neighbors and calculate the empirical probability of spin from sample.

In section 5, the $n \times n$ grid, horizontal, and circular Ising models are compared. The induced probability vectors by system energies and their entropies are found for the $n \times n$ grid, horizontal and circular models. The entropy of Boltzmann probability in these three models and their two-by-two Kullback-Leibler divergences are plotted and compared as a function of the parameter. In this section, we will show that, in for the $n \times n$ grid and horizontal models, the induced entropy of Boltzmann probability and entropy of Boltzmann probability are even functions w.r.t. β. In section 6, the results, suggestions, and challenges ahead for researchers are discussed.

2. Time-continuous jump process and birth and death process

Let time-continuous stochastic process $\{X_t : t \geq 0\}$ with state space E. Suppose $X_0 = x_0$ and the process in random time such as T_0 stops in the state x_0 and after that, jumps to state x_1 and this process are continued. Let $T_{-1} = 0$,

so we can write this process as

$$X_t = x_i, \quad i = 0, 1, \ldots, \quad \sum_{j=-1}^{i-1} T_j \leq t < \sum_{j=0}^{i} T_j.$$

The process must jump to another state at the time $\sum_{j=0}^{i} T_j$, $j = 0, 1$. For a sequence of the continuous and nonnegative random variables T_x, $x = 0, 1, \ldots$ with (probability) density function f_x, Q_{xy} is defined by the probability of a jump to y from x at the end of the stopping time. We can easily see that $Q_{xx} = 0$ and $\sum_{y \neq x} Q_{xy} = 1$. In this state, if the process jumps to x at the time t_0, we have $X_t = x$, $t_0 \leq t < t_0 + T_x$ and $X_{t_0+T_x} \neq x$, $t_0 + T_x < \infty$. Hence, the stochastic process with jump probability Q_{xy} goes to $y \in E$ and we call $\{X_t : t \geq 0\}$ as 'time-continuous jump chain' [7].

We say that a stochastic process $\{N_t : t \geq 0\}$ is counting if N_t is the number of events that have occurred in $(0, t]$. We call a counting stochastic process $\{N_t : t \geq 0\}$ the Poisson process with parameter (rates) λ, if all the following conditions are true [34]:

(i) $N_0 = 0$, (i.e., the number of events at zero instant is equal to zero);
(ii) N_t has independent increments (i.e., the numbers of arrivals to the system in disjoint intervals are independent);
(iii) $P\left(N_{t+h} - N_t = k\right) = \frac{e^{-\lambda h}(\lambda h)^k}{k!}$.

Equivalently, the counting stochastic process $\{N_t : t \geq 0\}$ is called a Poisson process with rates (parameter) λ, if [34]:

(i) $N_0 = 0$;
(ii) N_t has independent and stationary increments (N_t has stationary increments, if for all $t_2 > t_1 \geq 0$, $N_{t_2} - N_{t_1}$ has the same distribution as $N_{t_2-t_1}$, in other words $N_{t_2} - N_{t_1} \overset{d}{=} N_{t_2-t_1}$);
(iii) $P\left(N_h = 0\right) = 1 - \lambda h + o(h)$; $P\left(N_h = 1\right) = \lambda h + o(h)$; $P\left(N_h \geq 2\right) = o(h)$.

One of the important properties that exist in some stochastic processes is the Markov property. The stochastic process has Markov property, if for any $0 \leq t_1 < \ldots < t_n < t_{n+1}$ [34]:

$$P(X_{t_{n+1}} = j \,|\, X_{t_1} = i_1, \ldots, X_{t_{n-1}} = i_{n-1}, X_{t_n} = i) = P(X_{t_{n+1}} = j \,|\, X_{t_n} = i).$$

This property shows that the future and the past of the process are independent of each other under the condition of having the present. We can prove that a time-continuous process has Markov property if and only if for all $x \in E$, T_x has exponential distribution with parameter q_x and probability density function $f_{T_x}(t) = q_x e^{-q_x t}$, $t > 0$ [7]. For example, Poisson process $\{N_t : t \geq 0\}$ has Markov property, i.e., for any $0 = t_0 < t_1 < \ldots < t_n$ and nondecreasing natural numbers $k_i \geq 0$, $i = 1, \ldots, n$, $n \geq 2$, $P(N_{t_n} = k_n \,|\, N_{t_1} = k_1, \ldots, N_{t_{n-1}} = k_{n-1}) = P(N_{t_n} = k_n \,|\, N_{t_{n-1}} = k_{n-1})$, because for all $x \in E$, the stopping time T_x has exponential distribution. To prove this fact, if the Poisson process has entered the state x at the moment t_0 and T_x is the duration of the process stopping in the state x, we have

$$P(T_x > t) = P(N_{t_0+t} - N_{t_0} = 0) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t}.$$

Therefore, the distribution function of T_x obtained as following $F_{T_x}(t) = P(T_x \leq t) = 1 - e^{-\lambda t}$, i.e. T_x has exponential distribution with parameter λ.

For time-continuous stochastic process $\{X_t : t \geq 0\}$, $\pi_x(t) = P(X_t = x)$ is the probability of the system being in state x at time t and $P_{xy}(t) = P(X_{t+s} = y \,|\, X_s = x) = P(X_t = y \,|\, X_0 = x)$ is a probability of jump to y from x at time t [34]. The recent equality is due to the homogeneity of the process. It is obvious that $\sum_{x \in E} \pi_x(t) = 1$, $\sum_{y \in E} P_{xy}(t) = 1$. Also, for $t = 0$ we have

$$P_{xy}(0) = \begin{cases} 1, & x = y, \\ 0, & x \neq y, \end{cases} = \delta_{xy},$$

where δ_{xy} is a Kronecker delta [34]. A probability distribution of all possible states $E = \{x_1, x_2, \ldots\}$ forms a vector is given by $\boldsymbol{\pi} = (\pi_{x_1}, \pi_{x_2} \ldots)$. A distribution $\boldsymbol{\pi}$ is called stationary distribution, if $\boldsymbol{\pi} = \boldsymbol{\pi} P_t$, $t \geq 0$ where $P_t = (P_{xy}(t))$ is a transition probability matrix [7]. The infinitesimal parameters of stochastic process$\{X_t : t \geq 0\}$ written as $q_{xy} = P'_{xy}(t)|_{t=0}$, $x, y \in E$ i.e., q_{xy} is the rate that a transition from x to y [34].

One can easily see that, all homogeneous transition matrices $P_t = (P_{xy}(t))$ have the exponential form e^{tQ}. In the limit of continuous time, the transition matrix $Q = (q_{xy})$ (also known as a transition-rate matrix or an intensity matrix or an infinitesimal generator matrix) of an infinitesimal time difference is expanded as follows:

$$P_t = 1 + tQ + o(t) = \lim_{\varepsilon \to 0} (1 + \varepsilon Q)^{\frac{t}{\varepsilon}} = e^{tQ}.$$

If we let $\boldsymbol{\pi}(t) = (\pi_{x_1}(t), \pi_{x_2}(t)....)$, then $\boldsymbol{\pi}(t+\varepsilon) = \boldsymbol{\pi}(t)(I+\varepsilon Q)+o(\varepsilon)$. This implies that $\boldsymbol{\pi}(t+\varepsilon)-\boldsymbol{\pi}(t) = \varepsilon\boldsymbol{\pi}(t)Q+o(\varepsilon)$ and then

$$\boldsymbol{\pi}'(t) = \frac{d}{dt}\boldsymbol{\pi}(t) = \lim_{\varepsilon \to 0} \frac{\boldsymbol{\pi}(t+\varepsilon)-\boldsymbol{\pi}(t)}{\varepsilon} = \lim_{\varepsilon \to 0} \boldsymbol{\pi}(t)Q + \lim_{\varepsilon \to 0} \frac{o(\varepsilon)}{\varepsilon} = \boldsymbol{\pi}(t)Q,$$

which has the solution $\boldsymbol{\pi}(t+s) = \boldsymbol{\pi}(t)e^{Qs}$ where $s > 0$. Also e^{Qs} defined by its Taylor expansion, $e^{Qs} = \sum_{k=0}^{\infty} \frac{s^k}{k!}Q^k = \sum_i e^{\eta_i s}\mathbf{v}_i\mathbf{w}_i^t$ where η_i are the eigenvalues of Q and $\mathbf{v}_i$ ($\mathbf{w}_i$) are the right (left) eigenvectors of Q. If Q is a irreducible process (a Markov chain is said to be irreducible if all states communicate with each other, and the similarly, the infinitesimal generator matrix Q is irreducible, if there is a positive-rate path q_{xy} from any state to any other state), then the stochastic process is ergodic and there is one eigenvalue equal to 0.

The Chapman-Kolmogorov equation [34] is for a Markov chain with countable E, i.e., for all $s, t > 0$ and $x, y \in E$, $P_{xy}(t+s) = \sum_{z \in E} P_{xz}(t)P_{zy}(s)$ or in matrix form, $P_{t+s} = P_tP_s$, $t, s \geq 0$. By taking the derivative w.r.t. s from both sides of the Chapman-Kolmogorov equation and interchanging the derivative and the sum (due to the finiteness of E), we have $\frac{\partial}{\partial s}P_{xy}(t+s) = \sum_{z \in E} P_{xz}(t)\frac{\partial}{\partial s}P_{zy}(s)$. Now, take $s = 0$ and we get $\frac{\partial}{\partial s}P_{xy}(t) = \sum_{z \in E} P_{xz}(t)P'_{zy}(0)$. By substituting $q_{zy} = P'_{zy}(0)$ in the above equation, the Kolmogorov forward equations [34] are obtained as follows $P'_{xy}(t) = \sum_{z \in E} P_{xz}(t)q_{zy}$. In natural sciences, the Kolmogorov forward equations is also known as the master equation which is mentioned in section 3.

Proving the Kolmogorov backward equation is more difficult, and to obtain it, the following equation [19] must be derived w.r.t. t: $P_{xy}(t) = \delta_{xy}e^{-q_x t} + q_x e^{-q_x t}\int_0^t e^{q_x s}\left(\sum_{z \in E} Q_{xz}P_{zy}(s)\right)ds$. Hence for all $x, y \in E$ s.t. $x \neq y$ we have $\delta_{xy} = 0$ and thus

$$\begin{aligned}
P'_{xy}(t) &= -q_x e^{-q_x t}\left[\int_0^t e^{q_x s}\left(\sum\nolimits_{z \in E} Q_{xz}P_{zy}(s)\right)ds + q_x e^{q_x t}\left(\sum\nolimits_{z \in E} Q_{xz}P_{zy}(t)\right)\right] \\
&= -q_x\left[q_x e^{-q_x t}\int_0^t e^{q_x s}\left(\sum\nolimits_{z \in E} Q_{xz}P_{zy}(s)\right)ds - \sum\nolimits_{z \in E} Q_{xz}P_{zy}(t)\right] \\
&= -q_x P_{xy}(t) + \sum\nolimits_{z \in E} \frac{q_x}{q_{xz}}P_{zy}(t) \\
&= q_x P_{xy}(t) + \sum\nolimits_{z \in E} q_x P_{zy}(t) = \sum\nolimits_{z \in E} q_x P_{zy}(t),
\end{aligned}$$

since $-\frac{q_x}{q_{xx}} = 1$. So, the Kolmogorov backward equation are obtained as follows $P'_{xy}(t) = \sum_{z \in E} q_{xz}P_{zy}(t)$.

By the Chapman-Kolmogorov equation and Taylor expand, we have $P_{xy}(t+s) = \sum_{z \in E} P_{zy}(s)\left(\delta_{xz} + tq_{xz}\right)$. After taking the derivative of t, the above equation is transformed into:

$$\frac{\partial}{\partial t}P_{xy}(t)\,|_{t=s} = \sum\nolimits_{z \in E-\{x\}} P_{zy}(s)q_{xz} + P_{xy}(s)(q_{xx}) = \sum\nolimits_{z \in E-\{x\}} P_{zy}(s)q_{xz} - P_{xy}(s)\sum\nolimits_{z \in E-\{x\}} q_{xz}.$$

Hence, we have a partial differential equation of the distribution, based on the transition rate

$$\frac{\partial}{\partial t}\pi_x(t) = \sum\nolimits_{y \neq x}\left\{\pi_y(t)\,q_{yx} - \pi_x(t)\,q_{xy}\right\}. \tag{2.1}$$

The Equation (2.1) is called the master equation. In some texts, $\sum_{y \neq x} \pi_y(t)\,q_{yx}$ is called gain term and $\sum_{y \neq x} \pi_x(t)\,q_{xy}$ is called loss term. One example from biology is given below:

$$\pi'_n(t) = (n-1)\,\beta\pi_{n-1}(t) - n\beta\pi_n(t), \quad n = 1, 2, ...,$$

where $\pi_a(0) = P(N_0 = a) = 1$. This equation is applied to model population growth with birth, where n is the population index, with reference the initial population, β is the birth rate, and finally $\pi_n(t) = P(N_t = n)$, i.e. the

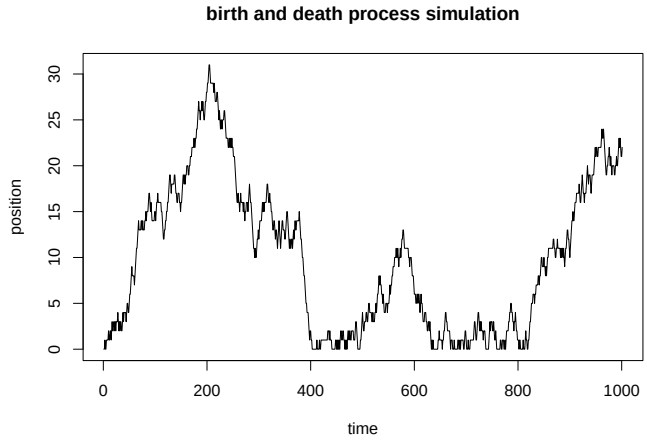


FIGURE 1. Simulate the birth and death process from 0 to 1000 step.

probability of achieving a certain population size. The analytical solution is $\pi_n(t) = (n-1)\beta e^{-n\beta t}\int_0^t \pi_{n-1}(s)\, e^{n\beta s} ds$ or the solution is the negative binomial distribution

$$\pi_n(t) = \binom{n-1}{a-1}\left(e^{-\beta t}\right)^a\left(1-e^{-n\beta t}\right)^{n-a}, \quad n = a, a+1, \ldots.$$

The birth and death process are use in modeling many problems. The pure jump Markov stochastic process $\{X_t : t \geq 0\}$ is called the birth and death process, if the infinitesimal parameters are equal to

$$q_{xy} = \begin{cases} \lambda_x, & y = x+1, \\ \mu_x, & y = x-1, \\ 0, & \text{O.w.} \end{cases}$$

It means that, when the process is in x, we have a birth (death) with rate λ (μ) [34]. With statistical software, like **R** we can simulate the birth and death process. Assume the probability of birth (death) is 0.3 (0.3) and the start time of the process is 0. In Figure 1, we will simulate the process from 0 to 1000 step with **R** software as following:

```
> birth = 0.3
> death = 0.3
> no_event = 1-birth-death
> steps = 1000
> positions = 0
> for(i in 1:steps){
+     if(tail(positions, 1) == 0){
+     move = sample (c(0,1),size = 1, prob =
+     c(no_event/(birth+no_event), birth/(birth+no_event)))
+          positions = c(positions, tail(positions, 1) + move)  next }
+     if(tail(positions, 1) > 0){
+     move = sample(c(-1,0,1), size = 1, prob = c(death, no_event, birth))
+         positions = c(positions, tail(positions, 1) + move) } }
> plot(positions, type = "l", main = "birth and death process simulation",
+         xlab = "time", ylab = "position")
```

As can be seen from Figure 1, when the line goes up, we have the birth and when it comes down, we have the death.

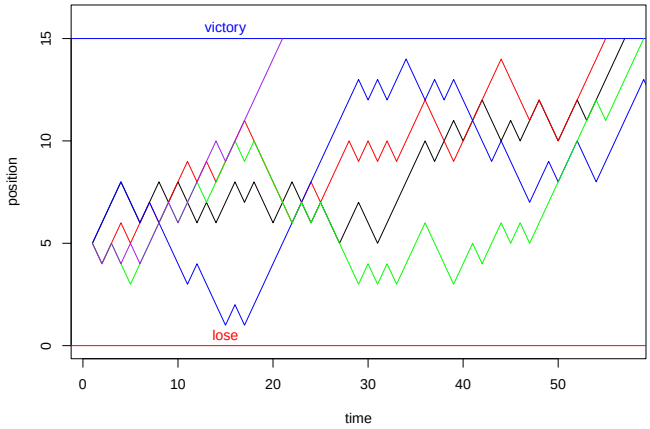


FIGURE 2. Simulate the Gambler's ruin.

One of the famous problems in probability is Gambler's Ruin. Two players, one and two, are playing a game with the probability of winning $0 < p < 1$ for player one and the probability of winning $0 < q = 1 - p < 1$ for player two. Each player has an initial wealth. Both players will play until one of them has lost all their initial wealth and thus cannot play anymore. In every step of the process, we have the birth or death process. Assume player one (two) has 5 (10) score and the probability of winning is 0.6 (0.4). The process will end when player one reaches 0 or 15. In **R** software, we will simulate this process as following:

```
> ###gambelers ruin
> birth = 0.6
> death = 1 - birth
> positions = 5
> while(T){
+   positions = c(positions, tail(positions, 1) + sample(c(-1, 1),
+                 size = 1, prob = c(death, birth)))
+   if(tail(positions, 1) == 0 | tail(positions, 1) == 15) break}
> plot(positions, type = "l", main = "Gambler's Ruin simula-tion",
+       xlab = "time", ylab = "position", ylim = c(0,16))
> abline(h = 15, col = "blue"); abline(h = 0, col = "red")
> text(15, 15.5, "victory", col = "blue")
> text(15, 0.5, "loss", col = "red")
```

If we repeat this process 5 times, we observe that 4 results of this process reach to win. We can say the probability of winning before 60 trials is %80 (Figure 2). One method to calculate the probability of some experience is to simulate for many times. This result becomes from the Strong Law of Large Numbers (SLLN) [35]. Also, we can simulate this process with the package 'gamblers.ruin.gameplay' in the **R** software. With this package, we can calculate the empirical probability of winning and losing for any process as follows:

```
> library(gamblers.ruin.gameplay)
> t1 = Sys.time()
> n = 10000
> win = 0; loos = 0; r = rep(0, n)
```

```
> for(i in 1:n){
+   a = grp.gameplay(5, 0.6, 15)$data
+   if(tail(a$capital, 1) == 15){
+     win = win + 1
+     r[i] = 1
+   }
+   if(tail(a$capital, 1) == 0)
+     loos = loos + 1
+ }
> t2 = Sys.time()
> difftime(t2, t1)
Time difference of 2.75812 mins
> print(paste('probability of wining is ', win/n))
[1] "probability of wining is  0.8727"
> print(paste('probability of loosing is ', loos/n))
[1] "probability of loosing is  0.1273"
```

In the **R** codes above, first, we repeat this process 10,000 times and observe that the probability of winning is 0.8727 and the probability of losing is 0.1273.

For a fixed Markov chain $\{X_t : t \geq 0\}$ with stationary distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, ...)$ and for all bounded function γ, which $X \sim \boldsymbol{\pi}$ (X has distribution $\boldsymbol{\pi}$) we have almost everywhere, $\lim_{n\to\infty} \frac{\gamma(X_1)+\cdots+\gamma(X_n)}{n} = \boldsymbol{E}(\gamma(X))$ where $\boldsymbol{E}(\gamma(X)) = \sum_j \gamma(j)\pi_j$ is an expected value of the random variable $\gamma(X)$ [17].

One of the applications of the birth and death process is in queuing theory. Assume that customers in a bank comes to the system with the birth rate 0.5 and exit from the system with the death rate 0.5. We can simulate this system:

```
> ### bank queue
> birth = 0.3
> death = 0.3
> no_event = 1-birth-death
> steps = 120
> positions = 5
> for(i in 1:steps){
+    if(tail(positions, 1) == 0){
+      move = sample(c(0,1), size = 1, prob = c(no_event/(birth+no_event)
+                     ,birth/(birth+no_event)))
+      positions = c(positions, tail(positions, 1) + move) next }
+    if(tail(positions, 1) > 0){
+      move = sample(c(-1, 0, 1), 1 , p = c(death, no_event, birth))
+      positions = c(positions, tail(positions, 1) + move) } }
> plot(positions, type = "l", main = "bank queue in one day", xlab = "time",
+          ylab = "number of customers", ylim = c(0, max(positions)))
> abline(h = max(positions), col = "blue"); abline(h = 0, col = "red")
```

In Figure 3, suppose that a customer enters in a bank and sees that there are 5 people in the bank. Figure 3, shows the number of customers in the each step. We observe a queue in the bank for two hours. As we see, there is some step that the bank is empty.

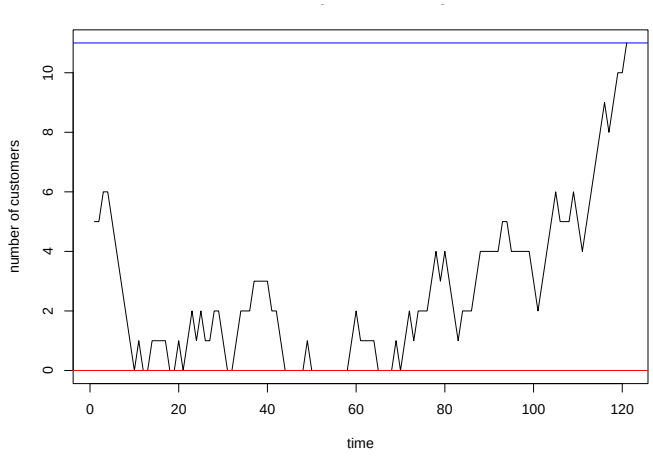


FIGURE 3. Number of customers in bank from 0 to 120.

3. The Kolmogorov forward equations, the master equation and the Ising model

The Kolmogorov forward equation (or the master equation in natural sciences [20]) for a Markov chain is obtained in the Eq. (2.1). In [18], the way to derivate the master equation is specified through the Chapman-Kolmogorov equation. Note that, since $q_{xx} = -\sum_{y\neq x} q_{xy}$, we can now rewrite Eq. (2.1) as following

$$\begin{aligned}\frac{d}{dt}\pi_x(t) &= \sum\nolimits_{y\neq x}\Big(\pi_y(t)q_{yx} - \pi_x(t)q_{xy}\Big) = \Big(\sum\nolimits_{y\neq x}\pi_y(t)q_{yx}\Big) + \pi_x(t)\Big(-\sum\nolimits_{y\neq x} q_{xy}\Big)\\ &= \Big(\sum\nolimits_{y\neq x}\pi_y(t)q_{yx}\Big) + \pi_x(t)q_{xx} = \sum\nolimits_{y\in E}\pi_y(t)q_{yx}.\end{aligned}$$

It is obvious that the above formulation can be rewritten as a matrix equation, $\boldsymbol{\pi}' = \boldsymbol{\pi}Q$. Since π_x is stationary distribution of process, we have global balance equations as following

$$0 = \sum\nolimits_{y\neq x}\pi_y q_{yx} - \pi_x q_{xy}. \tag{3.1}$$

Another proof of Eq. (3.1) using the Kolmogorov forward (backward) equation is as follows. At first, we prove that, $\boldsymbol{\pi} = (\pi_x)$ is a stationary distribution for a continuous Markov chain $\{X_t : t \geq 0\}$ if and only if it satisfies $\boldsymbol{\pi}Q = \mathbf{0}$ where $\mathbf{0}$ is a zero matrix. Now we prove this fact. If $\boldsymbol{\pi}$ is a stationary distribution, then $\boldsymbol{\pi} = \boldsymbol{\pi}P_t$. Differentiating both sides and using matrix form of Kolmogorov backward equation, i.e., ${P'}_t = QP_t$, and then we obtain $\mathbf{0} = \frac{d}{dt}[\boldsymbol{\pi}P_t] = \boldsymbol{\pi}{P'}_t = \boldsymbol{\pi}QP_t$. Now, let $t = 0$ and remember that $P_0 = I$, the identity matrix. We obtain $\mathbf{0} = \boldsymbol{\pi}QP_0 = \boldsymbol{\pi}QI = \boldsymbol{\pi}Q$.

Conversely, let $\boldsymbol{\pi}$ be a probability distribution on state space E that satisfies $\boldsymbol{\pi}Q = \mathbf{0}$. Then, by multiplying both sides of Kolmogorov backward equations, ${P'}_t = QP_t$ by $\boldsymbol{\pi}$, we obtain $(\boldsymbol{\pi}P_t)' = \boldsymbol{\pi}{P'}_t = \boldsymbol{\pi}QP_t = \mathbf{0}\times P = \mathbf{0}$. Thus, we conclude $\boldsymbol{\pi}P_t$ does not depend on t. In particular, for any $t \geq 0$, we have $\boldsymbol{\pi}P_t = \boldsymbol{\pi}P_0 = \boldsymbol{\pi}I = \boldsymbol{\pi}$. Therefore, $\boldsymbol{\pi}$ is a stationary distribution.

From the above fact, i.e., $\boldsymbol{\pi}Q = \boldsymbol{\pi}$, this important result is obtained:

$$\begin{aligned}0 &= \sum\nolimits_{y\in E}\pi_y\, q_{yx} = \Big(\sum\nolimits_{y\neq x}\pi_y\, q_{yx}\Big) + \pi_x\, q_{xx}\\ &= \Big(\sum\nolimits_{y\neq x}\pi_y\, q_{yx}\Big) - \pi_x \sum\nolimits_{y\neq x} q_{xy} = \sum\nolimits_{y\neq x}\Big(\pi_y\, q_{yx} - \pi_x q_{xy}\Big),\end{aligned}$$

and relation Eq. (3.1) is obtained. For a continuous-time Markov chain, if π_x can be found s.t. for every pair of states $x, y \in E$,

$$\pi_y q_{yx} = \pi_x q_{xy}, \tag{3.2}$$

holds, then by summing over y, the global balance equations (i.e. Eq. (3.1)) are satisfied and π_x is the stationary distribution of the process. If Eq. (3.1) can be found, the resulting equations are usually much easier than directly solving Eq. (3.1). Many algorithms, such as the Metropolis-Hastings algorithm and the Glauber dynamics are based on Eq. (3.2) (see section 4). In MCMC algorithm, we move from the old state x to a new one y, then the acceptance probability of this move $P^a(x \to y)$ is calculated, a random number is generated to decide whether this move is accepted. As suggested in Eq. (3.2), in all reversible dynamics: $\pi_y P^a(y \to x) = \pi_x P^a(x \to y)$. Based on this equation, we have the Metropolis dynamics. As $P^a \geqslant 1$, we have $P^a_{Metro}(x \to y) = \min\left\{1, \frac{\pi_y f_{X|Y=y}(x)}{\pi_x f_{Y|X=x}(y)}\right\}$ where $f_{y|x}(y)$ is the probability density function of proposing the new state y starting from the state x, which $\int_{-\infty}^{+\infty} f_{Y|X=x}(y)dy = 1$. By choosing $f_{Y|X=x}(y)$ and sampling the next state y, high rejection rate can be avoided in the simulations. This is known as the Metropolis-Hastings algorithm. In case $f_{X|Y=y}(x) = f_{Y|X=x}(y)$, we recover the normal Metropolis algorithm [26].

Note that $H(\boldsymbol{\pi}) = -\sum_{x\in E} \pi_x \ln \pi_x$ be the entropy of distribution $\boldsymbol{\pi}$ [35]. The following lemma gives the derivative of the entropy of a distribution w.r.t. time.

Lemma 3.1. *For any distribution $\boldsymbol{\pi}$, we have $\frac{\partial}{\partial t} H(\boldsymbol{\pi}(t)) = e_p(t) - h_d(t)$ s.t.*

$$e_p(t) = -\tfrac{1}{2}\sum_{x\in E}\left[\sum_{y\neq x}\{\pi_y(t)\, q_{yx} - \pi_x(t)\, q_{xy}\}\right] \times \left[\ln \frac{\pi_y(t)\, q_{yx}}{\pi_x(t)\, q_{xy}}\right],$$

$$h_d(t) = \tfrac{1}{2}\sum_{x\in E}\left[\sum_{y\neq x}\{\pi_y(t)\, q_{yx} - \pi_x(t)\, q_{xy}\}\right] \times \left[\ln \frac{q_{yx}}{q_{xy}}\right].$$

Proof. By definition of entropy, we have

$$\begin{aligned}\frac{\partial}{\partial t} H(\boldsymbol{\pi}(t)) &= -\frac{\partial}{\partial t}\sum_{x\in E}\pi_x(t)\ln(\pi_x(t)) = -\sum_{x\in E}\frac{\partial \pi_x(t)}{\partial t}\ln(\pi_x(t)) - \sum_{x\in E}\frac{\partial}{\partial t}\pi_x(t)\\ &= -\sum_{x\in E}\frac{\partial \pi_x(t)}{\partial t}\ln(\pi_x(t)) - \frac{\partial}{\partial t}\sum_{x\in E}\pi_x(t) = -\sum_{x\in E}\frac{\partial \pi_x(t)}{\partial t}\ln(\pi_x(t)) - \frac{\partial}{\partial t}1\\ &= -\sum_{x\in E}\frac{\partial \pi_x(t)}{\partial t}\ln(\pi_x(t)).\end{aligned}$$

By substituting Eq. (2.1) in the above expression, we have

$$\frac{\partial}{\partial t} H(\boldsymbol{\pi}(t)) = -\sum_{x\in E}\left[\sum_{y\neq x}\{\pi_y(t)\, q_{yx} - \pi_x(t)\, q_{xy}\}\right]\Big(\ln \pi_x(t)\Big).$$

If we exchange the indexes x and y, we can rewrite $\frac{\partial}{\partial t} H(\boldsymbol{\pi}(t))$ as

$$\begin{aligned}\frac{\partial}{\partial t} H(\boldsymbol{\pi}(t)) &= -\frac{1}{2}\sum_{x\in E}\left[\sum_{y\neq x}\Big(\pi_y(t)q_{yx} - \pi_x(t)q_{xy}\Big)\right] \times \left[\ln \frac{\pi_y(t)q_{yx}}{\pi_x(t)q_{xy}}\right]\\ &\quad + \frac{1}{2}\sum_{x\in E}\left[\sum_{y\neq x}\Big(\pi_y(t)q_{yx} - \pi_x(t)q_{xy}\Big)\right] \times \left[\ln \frac{q_{yx}}{q_{xy}}\right]\\ &= e_p(t) - h_q(t).\end{aligned}$$

where $e_p(t)$ is the instantaneous entropy production rate and $h_d(t)$ is the rate dissipation heat [14]. □

For a birth and death process we got from Eq. (2.1), $\frac{\partial}{\partial t}\pi_n(t) = \lambda_{n-1}\pi_{n-1}(t) + \mu_{n+1}\pi_{n+1}(t) - [\lambda_n + \mu_n]\pi_n(t)$ which $q_{n-1,n} = \lambda_{n-1}$, $q_{n+1,n} = \mu_{n+1}$, $q_{n,n+1} = \lambda_n$ and $q_{n,n-1} = \mu_n$. Let $J_n(t) = \mu_n\pi_n(t) - \lambda_{n-1}\pi_{n-1}(t)$. Hence, we can rewrite Eq. (2.1) as $\frac{\partial}{\partial t}\pi_n(t) = J_{n+1}(t) - J_n(t)$.

In physical research, for equilibrium state that $\boldsymbol{\pi}$ is stationary state, Equation (3.1) holds. One of the important stationary distributions in physics is

$$\xi_x = \frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}, \tag{3.3}$$

where E_x is energy of state x and the normalization constant β is the reverse temperature characteristic of the heat baths. The Boltzmann distribution can be used to describe the states of physical particles such as atoms or molecules, or biological systems. The Boltzmann distribution is based on how the states of the system are filled, depending on their energy. the probability of the system to be in state x with energy E_x is proportional to $e^{-\beta E_x}$, where $\beta = \frac{1}{k_B T}$, T is the temperature which is measured in Kelvin and k_B is the Boltzmann constant [6]. After long time, the system approaches thermal equilibrium and stationary distribution not dependent to time, and equation Eq. (3.1) when transition rate q_{xy} satisfying in condition $q_{xy} = q_{yx} e^{-\beta E_y + \beta E_x}$ holds. Because

$$\begin{aligned}\sum\nolimits_{y\neq x} \left(\xi_y q_{yx} - \xi_x q_{xy}\right) &= \sum\nolimits_{y\neq x} \left(\xi_y q_{yx} - \xi_x q_{yx} e^{-\beta E_x + \beta E_y}\right) \\ &= \sum\nolimits_{y\neq x} \left(\frac{e^{-\beta E_y}}{\sum_{z\in E} e^{-\beta E_z}} q_{yx} - \frac{e^{-\beta E_x}}{\sum_{z\in E} e^{-\beta E_z}} q_{yx} e^{-\beta E_x + \beta E_y} \right) = 0.\end{aligned}$$

The transition rate across (x, z) is given by $q_{xz} = e^{\beta E_x - \beta E_z}$. The stochastic process is in detailed balance (or time reversible) i.e., there exists a distribution $\pi_x = e^{-\beta E_x}$ s.t. $\pi_y q_{yx} = \pi_x q_{xy}$ for any edge z with joining a pair of vertices x and y. Note that, $\pi_x q_{xy}$ is the probability flux across (x, y). Also $\pi_y q_{yx} = \pi_x q_{xy}$ says that the pair of states x and y are in equilibrium along z [6].

In stationary distributions given by Eq. (3.3), we have

$$\frac{\partial}{\partial t} H\left(\boldsymbol{\xi}(t)\right) = -\sum\nolimits_{x\in E} \left[\sum\nolimits_{y\neq x} \left\{\xi_y(t)\, q_{yx} - \xi_x(t)\, q_{xy}\right\}\right] \left(\ln \xi_x(t)\right) = 0,$$

and thus, the entropy of stationary distribution given by Eq. (3.3), i.e., $H\left(\boldsymbol{\xi}(t)\right)$ does not depend on time. The reason for entropy becoming zero is the stationarity of distribution given by Eq. (3.3), which is because the distribution of the stochastic process is the same in all stages, entropy takes zero value.

We can consider the Kullback-Leibler divergence metric (or distance) or relative entropy between two distributions $\boldsymbol{\pi}$ and $\boldsymbol{\xi}$ given by $D\left(\boldsymbol{\pi} \parallel \boldsymbol{\xi}\right) = \sum_{x\in E} \pi_x(t) \ln\left(\frac{\pi_x(t)}{\xi_x(t)}\right)$ [35].

Theorem 3.2. *For fixed arbitrary distribution $\boldsymbol{\pi}$ and stationary distribution $\boldsymbol{\xi}$ given by Eq. (3.3), we have*

$$\begin{aligned}\frac{\partial}{\partial t} D\left(\boldsymbol{\pi} \parallel \boldsymbol{\xi}\right) &= h_d(t) - e_p(t) + \beta \sum\nolimits_{x\in E} E_x \left[\sum\nolimits_{y\neq x} \left\{\pi_y(t)\, q_{yx} - \pi_x(t)\, q_{xy}\right\}\right], \\ \frac{\partial}{\partial t} D\left(\boldsymbol{\xi} \parallel \boldsymbol{\pi}\right) &= -\frac{\sum_{x\in E} e^{-\beta E_x} \sum_{y\neq x} \left\{\frac{\pi_y(t)}{\pi_x(t)}\, q_{yx} - q_{xy}\right\}}{\sum_{y\in E} e^{-\beta E_y}}.\end{aligned}$$

Proof. By definition, we have $D\left(\boldsymbol{\pi} \parallel \boldsymbol{\xi}\right) = \sum_{x\in E} \pi_x(t) \ln\left(\frac{\pi_x(t)}{\xi_x(t)}\right) = -H\left(\boldsymbol{\pi}(t)\right) - \sum_{x\in E} \pi_x(t) \ln\left(\xi_x(t)\right)$. Hence by Lemma 3.1,

$$\begin{aligned}\frac{\partial}{\partial t} D(\boldsymbol{\pi} \| \xi) &= -\frac{\partial}{\partial t} H(\boldsymbol{\pi}(t)) - \frac{\partial}{\partial t} \sum\nolimits_{x\in E} \pi_x(t) \ln\left(\xi_x(t)\right) \\ &= h_d(t) - e_p(t) - \sum\nolimits_{x\in E} \frac{\partial \pi_x(t)}{\partial t} \ln\left(\xi_x(t)\right) - \sum\nolimits_{x\in E} \frac{\frac{\partial}{\partial t}\xi_x(t)}{\xi_x(t)} \pi_x(t) \\ &= h_d(t) - e_p(t) - \sum\nolimits_{x\in E} \frac{\partial \pi_x(t)}{\partial t} \ln\left(\frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}\right) - 0 \\ &= h_d(t) - e_p(t) + \beta \sum\nolimits_{x\in E} E_x \frac{\partial \pi_x(t)}{\partial t} + \sum\nolimits_{x\in E} \frac{\partial \pi_x(t)}{\partial t} \ln\left(\sum\nolimits_{y\in E} e^{-\beta E_y}\right) \\ &= h_d(t) - e_p(t) + \beta \sum\nolimits_{x\in E} E_x \left[\sum\nolimits_{y\neq x} \left\{\pi_y(t)\, q_{yx} - \pi_x(t)\, q_{xy}\right\}\right] \\ &\quad + \ln\left(\sum\nolimits_{y\in E} e^{-\beta E_y}\right) \frac{\partial}{\partial t} \sum\nolimits_{x\in E} \pi_x(t) \\ &= h_d(t) - e_p(t) + \beta \sum\nolimits_{x\in E} E_x \left[\sum\nolimits_{y\neq x} \left\{\pi_y(t)\, q_{yx} - \pi_x(t)\, q_{xy}\right\}\right],\end{aligned}$$

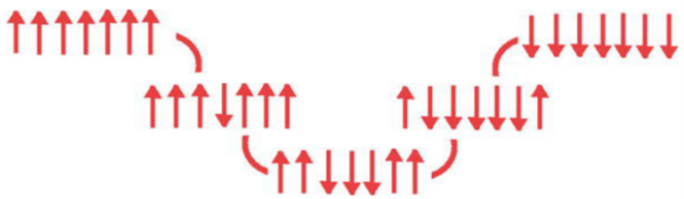


FIGURE 4. The Ising model with random spin in one-dimensional system.

and we conclude similarly that

$$\begin{aligned}\frac{\partial}{\partial t}D\left(\boldsymbol{\xi}\parallel\pi\right)&=\frac{\partial}{\partial t}\left[\sum\nolimits_{x\in E}\xi_x(t)\ln\left(\frac{\xi_x(t)}{\pi_x(t)}\right)\right]=-\frac{\partial}{\partial t}H\left(\boldsymbol{\xi}(t)\right)-\frac{\partial}{\partial t}\sum\nolimits_{x\in E}\xi_x(t)\ln\left(\pi_x(t)\right)\\&=-\sum\nolimits_{x\in E}\frac{\partial}{\partial t}\xi_x(t)\ln\left(\pi_x(t)\right)-\sum\nolimits_{x\in E}\xi_x(t)\frac{\frac{\partial}{\partial t}\pi_x(t)}{\pi_x(t)}\\&=-\sum\nolimits_{x\in E}\left[\frac{e^{-\beta E_x}}{\sum_{y\in E}e^{-\beta E_y}}\right]\times\frac{\sum_{y\neq x}\left\{\pi_y(t)q_{yx}-\pi_x(t)q_{xy}\right\}}{\pi_x(t)}\\&=-\frac{\sum_{x\in E}e^{-\beta E_x}\sum_{y\neq x}\left\{\frac{\pi_y(t)}{\pi_x(t)}q_{yx}-q_{xy}\right\}}{\sum_{y\in E}e^{-\beta E_y}}.\end{aligned}$$

□

The Ising model is a mathematical model of ferromagnetism in statistical mechanics. Beside in physic, it works in many branches like machine learning, image processing, and economics. Ernst (Ernest) Ising (1900-1998) had first shown that no phase transition to a ferromagnetic ordered state occurs in one dimension at any temperature [39]. One good reference that analyzes Ising's life and works is [12]. In [30], the Ising models in ferromagnetic and other physical phenomena are reviewed. In the Ising model system, we have two sides, up (down), in the direction of the field (against the direction of the field) (see Figure 4). Probability of each side showed respectively $\pi_+(t)=\pi_1(t)$ and $\pi_-(t)=\pi_0(t)$, and a probability of jump to + from – (– from +) at time t is $P_{10}(t)\,(P_{01}(t))$, and transition rate from + to – (– to +) showed as $q_{+-}=q_{10}=\mu$ $(q_{-+}=q_{01}=\lambda)$. It's obvious $P_{11}(t)=1-P_{10}(t)$ and $P_{00}(t)=1-P_{01}(t)$. In this memoryless system, the Markov property holds, and in fact, it showed this model is a birth and death process. By using Eq. (2.1), we got:

$$\begin{cases}\frac{d}{dt}\pi_+=-\mu\pi_++\lambda\pi_-,\\ \frac{d}{dt}\pi_-=\mu\pi_+-\lambda\pi_-.\end{cases}$$

Theorem 3.3. *A unique stationary distribution for the Ising model, i.e.,* $\boldsymbol{\pi}=(\pi_0,\pi_1)$ *satisfies* $\pi_0=\frac{\mu}{\lambda+\mu}=1-\pi_1$.

Proof. By using the Kolmogorov forward equation, we got

$$\begin{aligned}P'_{00}(t)&=\sum\nolimits_{z\in E}q_{0z}P_{z0}(t)=q_{00}P_{00}(t)+q_{01}P_{10}(t)=-\lambda P_{00}(t)+\lambda P_{10}(t),\\P'_{10}(t)&=\sum\nolimits_{z\in E}q_{1z}P_{z0}(t)=q_{10}P_{00}(t)+q_{11}P_{10}(t)=\mu P_{00}(t)-\mu P_{10}(t).\end{aligned}$$

Then, we have two equations as following $P'_{00}\left(t\right)=-\lambda P_{00}\left(t\right)+\lambda P_{10}\left(t\right)$, $P'_{10}\left(t\right)=\mu P_{00}\left(t\right)-\mu P_{10}\left(t\right)$. By subtracting the last two equations, we have

$$\begin{aligned}f'(t)&=P'_{00}(t)-P'_{10}(t)=-\lambda P_{00}(t)+\lambda P_{10}(t)-\mu P_{00}(t)+\mu P_{10}(t)\\&=-\lambda\big(P_{00}(t)-P_{10}(t)\big)-\mu\big(P_{00}(t)-P_{10}(t)\big)=-(\lambda+\mu)\big(P_{00}(t)-P_{10}(t)\big)=-(\lambda+\mu)f(t).\end{aligned}$$

where $f(t) = P_{00}(t) - P_{10}(t)$. By solving the homogeneous linear differential equation $f'(t) = -(\lambda+\mu) f(t)$ with initial condition $f(0) = P_{00}(0) - P_{10}(0) = 1 - 0 = 1$, we have $f(t) = e^{-(\lambda+\mu)t}$ or equivalently $P_{00}(t) = P_{10}(t) + e^{-(\lambda+\mu)t}$. Hence, with substitution this expression to $P_{00}(t)$ in equation for $P'_{10}(t)$, we have $P'_{10}(t) = \mu P_{10}(t) + \mu e^{-(\lambda+\mu)t} - \mu P_{10}(t) = \mu e^{-(\lambda+\mu)t}$. So, with integration from equation and according to initial condition $P_{10}(0) = 0$, we got $P_{10}(t) = -\frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} + c$. But $0 = P_{10}(0) = -\frac{\mu}{\lambda+\mu} + c$, so $c = \frac{\mu}{\lambda+\mu}$ and we have $P_{10}(t) = \frac{\mu}{\lambda+\mu}\left(1 - e^{-(\lambda+\mu)t}\right)$. Also $P_{00}(t) = \frac{\mu}{\lambda+\mu}\left(1 - e^{-(\lambda+\mu)t}\right) + e^{-(\lambda+\mu)t} = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}$. Then $P_{01}(t) = 1 - P_{00}(t) = \frac{\lambda}{\lambda+\mu}\left(1 - e^{-(\lambda+\mu)t}\right)$, $P_{11}(t) = 1 - P_{10}(t) = \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t}$.

Limiting distribution calculated as:

$$\lim_{t\to 0} P(t) = \lim_{t\to 0} \begin{bmatrix} P_{00}(t) & P_{01}(t) \\ P_{10}(t) & P_{11}(t) \end{bmatrix} = \lim_{t\to 0} \begin{bmatrix} \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t} & \frac{\lambda}{\lambda+\mu}\left(1 - e^{-(\lambda+\mu)t}\right) \\ \frac{\mu}{\lambda+\mu}\left(1 - e^{-(\lambda+\mu)t}\right) & \frac{\lambda}{\lambda+\mu} + \frac{\mu}{\lambda+\mu} e^{-(\lambda+\mu)t} \end{bmatrix} = \begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{bmatrix},$$

and the stationary distribution for Markov chain satisfies $\pi_0 = \frac{\mu}{\lambda+\mu} = 1 - \pi_1$.

A much simpler way to calculate stationary distribution is to use formula $\boldsymbol{\pi} Q = \mathbf{0}$. We have

$$\mathbf{0} = \boldsymbol{\pi} Q = \begin{bmatrix} \pi_0 & \pi_1 \end{bmatrix} \begin{bmatrix} -\lambda & \lambda \\ \mu & -\mu \end{bmatrix} = \begin{bmatrix} -\lambda\pi_0 + \mu\pi_1, & \lambda\pi_0 - \mu\pi_1 \end{bmatrix}.$$

which results in $\pi_1 = \frac{\lambda}{\mu}\pi_0$. We also need $\pi_0 + \pi_1 = 1$. Solving the above equations, we obtain $\pi_0 = \frac{\mu}{\lambda+\mu} = 1 - \pi_1$.

The above limiting distribution can also be reached with discretization directly. For this purpose, we consider the transition probability matrix as following

$$P = \begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{bmatrix}.$$

Now, we calculate the n^{th} power of the matrix P. By the Hamilton-Cayley theorem [28], we have $P^2 - tr(P).P + \det(P).I = \mathbf{0}$ where $tr(P)$ (trace of a square P) is defined to be the sum of elements on the main diagonal of P and $\det(P)$ is called determinant of P. Since $tr(P) = 1$ and $\det(P) = 0$, Hamilton-Cayley theorem [28] is rewritten as follows $\mathbf{0} = P^2 - P$ or $P^2 = P$, i.e., P is an idempotent matrix, and $P^n = P$. So, $\lim_{n\to\infty} P^n = P$ and after that $\pi_0 = \frac{\mu}{\lambda+\mu} = 1 - \pi_1$. □

Example 3.4. A one-dimensional Ising model is proposed in this example. We consider the Markov Chain $\{X_t : t \geq 0\}$ on state space $E = \{+1, -1\}$ with transition probability $p_{x,y} = \frac{1}{\psi} e^{\rho xy + \theta y + \varphi x}$ where we choose φ and ψ so that $\sum_y p_{x,y} = 1$ and also, θ and ρ are unknown parameter. Hence for $x = 1, -1$ we have following equations:

$$1 = p_{1,1} + p_{1,-1} = \frac{1}{\psi}\left(e^{\rho+\theta+\varphi} + e^{-\rho-\theta+\varphi}\right) = \frac{2e^{\varphi}\cosh(\rho+\theta)}{\psi},$$
$$1 = p_{-1,1} + p_{-1,-1} = \frac{1}{\psi}\left(e^{-\rho+\theta-\varphi} + e^{\rho-\theta-\varphi}\right) = \frac{2e^{-\varphi}\cosh(\rho-\theta)}{\psi},$$

or $\psi = 2e^{\varphi}\cosh(\rho+\theta)$ and $\psi = 2e^{-\varphi}\cosh(\rho-\theta)$. We divide the above two expressions and get the following result $e^{2\varphi} = \frac{\cosh(\rho-\theta)}{\cosh(\rho+\theta)}$. In which case $\varphi = \frac{1}{2}\ln\left(\frac{\cosh(\rho-\theta)}{\cosh(\rho+\theta)}\right)$, $\psi = 2\sqrt{\cosh(\rho+\theta)\cosh(\rho-\theta)}$.

Therefore, the transition probability matrix is rewritten as below

$$P = \begin{pmatrix} \frac{e^{\varphi}e^{\rho+\theta}}{\psi} & \frac{e^{\varphi}e^{-\rho-\theta}}{\psi} \\ \frac{e^{-\varphi}e^{-\rho+\theta}}{\psi} & \frac{e^{-\varphi}e^{\rho-\theta}}{\psi} \end{pmatrix} = \begin{pmatrix} \frac{e^{\rho+\theta}}{2\cosh(\rho+\theta)} & \frac{e^{-\rho-\theta}}{2\cosh(\rho+\theta)} \\ \frac{e^{-\rho+\theta}}{2\cosh(\rho-\theta)} & \frac{e^{\rho-\theta}}{2\cosh(\rho-\theta)} \end{pmatrix}.$$

In addition, when $\rho = 0$, there exist $\lambda, \mu > 0$ s.t. $2\cosh(\theta) = 2\cosh(-\theta) = \sqrt{\frac{\mu}{\lambda}} + \sqrt{\frac{\lambda}{\mu}}$. In this special case, we have

$$P = \begin{pmatrix} \frac{e^{\rho+\theta}}{2\cosh(\rho+\theta)} & \frac{e^{-\rho-\theta}}{2\cosh(\rho+\theta)} \\ \frac{e^{-\rho+\theta}}{2\cosh(\rho-\theta)} & \frac{e^{\rho-\theta}}{2\cosh(\rho-\theta)} \end{pmatrix} = \begin{pmatrix} \frac{e^{\theta}}{\delta} & \frac{e^{-\theta}}{\delta} \\ \frac{e^{\theta}}{\delta} & \frac{e^{-\theta}}{\delta} \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{\mu\lambda^{-1}}}{\delta} & \frac{\sqrt{\lambda\mu^{-1}}}{\delta} \\ \frac{\sqrt{\mu\lambda^{-1}}}{\delta} & \frac{\sqrt{\lambda\mu^{-1}}}{\delta} \end{pmatrix} = \begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{bmatrix},$$

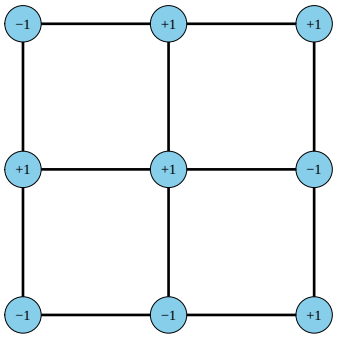


FIGURE 5. Sample configuration x on at the 3×3 grid.

where $\delta = \sqrt{\mu\lambda^{-1}} + \sqrt{\lambda\mu^{-1}}$ and in this case, the model becomes the classic Ising model presented at the Theorem 3.3.

There is another way to find the transition probability matrix. For $\rho = 0$, we have $\varphi = \frac{1}{2}\ln\left(\frac{\cosh(-\theta)}{\cosh(\theta)}\right) = \frac{1}{2}\ln(1) = 0$, $\psi = 2\sqrt{\cosh(\theta)\cosh(-\theta)} = 2\cosh(\theta)$, because $\cosh(\theta)$ is an even and possitive function. Let $2\cosh(\theta) = e^{\theta} + e^{-\theta} = \sqrt{\frac{\mu}{\lambda}} + \sqrt{\frac{\lambda}{\mu}} = \psi$ for some $\lambda, \mu > 0$. Hence $\zeta^2 - \psi\zeta + 1 = 0$ s.t. $\zeta = e^{\theta}$. By solving this equation in terms of ζ, we have

$$\zeta = e^{\theta} = \frac{\psi \pm \sqrt{\frac{\mu}{\lambda} + \frac{\lambda}{\mu} + 2 - 4}}{2} = \frac{\sqrt{\frac{\mu}{\lambda}} + \sqrt{\frac{\lambda}{\mu}} \pm \left(\sqrt{\frac{\mu}{\lambda}} - \sqrt{\frac{\lambda}{\mu}}\right)}{2} = \sqrt{\frac{\mu}{\lambda}}, \sqrt{\frac{\lambda}{\mu}}.$$

One of the answers, namely $\zeta = e^{\theta} = \sqrt{\frac{\mu}{\lambda}}$, leads to transition probability matrix corresponding to the classic Ising model. Because $p_{x,y} = \frac{1}{\psi} e^{\rho xy + \theta y + \varphi x} = \frac{1}{\sqrt{\frac{\mu}{\lambda}} + \sqrt{\frac{\lambda}{\mu}}} e^{\theta y}$ which are the elements of the following matrix

$$P = \begin{pmatrix} p_{1,1} & p_{1,-1} \\ p_{-1,1} & p_{-1,-1} \end{pmatrix} = \begin{pmatrix} \frac{e^{\theta}}{\sqrt{\frac{\mu}{\lambda}} + \sqrt{\frac{\lambda}{\mu}}} & \frac{e^{-\theta}}{\sqrt{\frac{\mu}{\lambda}} + \sqrt{\frac{\lambda}{\mu}}} \\ \frac{e^{\theta}}{\sqrt{\frac{\mu}{\lambda}} + \sqrt{\frac{\lambda}{\mu}}} & \frac{e^{-\theta}}{\sqrt{\frac{\mu}{\lambda}} + \sqrt{\frac{\lambda}{\mu}}} \end{pmatrix} = \begin{bmatrix} \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \\ \frac{\mu}{\lambda+\mu} & \frac{\lambda}{\lambda+\mu} \end{bmatrix}.$$

Example 3.5. We know that the Ising model was used as a magnetism model and image processing in physical research. Consider a graph G consisting of edges, in which each site i is assigned a spin of +1 or –1. Thus $x_i = \pm 1$, for all i. Let a square $n \times n$ grid, where each point is connected to four neighbors. Thus, there are $k = 2^{n^2}$ possible states, so $E = \{E_{x_1}, ..., E_{x_k}\}$. For each state $x \in \{x_1, ..., x_k\}$, the energy is defined as follows $E_x = -\sum_{i\sim j} x_i x_j$, where the sum is over all pairs of sites $i$and j, which are neighbors. For the configuration x in the Figure 5,

$$\begin{aligned} E_x &= -\sum\nolimits_{i\sim j} x_i x_j = -\sum\nolimits_{(m,n)\sim(v,w)} a_{mn} a_{vw} \\ &= -\Big(a_{11}a_{12} + a_{12}a_{13} + a_{11}a_{21} + a_{12}a_{22} + a_{13}a_{23} + a_{21}a_{22} + a_{22}a_{23} + a_{21}a_{31} + a_{22}a_{32} + a_{23}a_{33} + a_{31}a_{32} + a_{32}a_{33}\Big) \\ &= -\Big((-1)(-1) + (-1)(1) + (-1)(1) + (-1)(1) + (1)(-1) + (1)(1) \\ &+ (1)(-1) + (1)(-1) + (1)(1) + (-1)(1) + (-1)(1) + (1)(1)\Big) = 4. \end{aligned}$$

where

$$(x_i) = (-1,-1,1,1,1,-1,-1,1,1) \equiv \Sigma = (a_{ij}) = \begin{pmatrix} -1 & -1 & 1 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{pmatrix}.$$

For the convenience of calculations, we wrote the vector (x_i) as matrix $\Sigma = (a_{ij})$.

We plot this configuration in **R** software in Figure 5. All 512 states can be calculated from one state (for example, the state shown in Figure 5) as follows:

$$(x_1,\dots,x_9) = (-1,-1,1,1,1,-1,-1,1,1)\begin{pmatrix} a_1 \\ \vdots \\ a_9 \end{pmatrix} = a_3 + a_4 + a_5 + a_8 + a_9 - (a_1 + a_2 + a_6 + a_7),$$

where $a_i = \pm 1$, for all i. In the 3×3 grid Ising model, there exists $2^9 = 512$ states and we can calculate the energy of these 512 states as follows:

```
> ### calculating energy in 3*3 network
> sample_space = c(-1,1)
> total_energy = 0
> for(a1 in sample_space){
+ for(a2 in sample_space){
+ for(a3 in sample_space){
+ for(a4 in sample_space){
+ for(a5 in sample_space){
+ for(a6 in sample_space){
+ for(a7 in sample_space){
+ for(a8 in sample_space){
+ for(a9 in sample_space){
+ network = matrix(c(a1, a2, a3, a4, a5, a6, a7, a8, a9), 3, 3, byrow = T)
+ energy_of_network = (-1)*(network[1, 1]*network[1, 2]
+ +network[1, 1] * network[2, 1]
+ +network[1, 2] *  network[1, 3]
+ +network[1, 2] * network[2, 2]
+ +network[1, 3] * network[2, 3]
+ +network[2, 1] * network[2, 2]
+ +network[2, 1] * network[3, 1]
+ +network[2, 2] * network[2, 3]
+ +network[2, 2] * network[3, 2]
+ +network[2, 3] * network[3, 3]
+ +network[3, 1] * network[3, 2]
+ +network[3, 2] * network[3, 3])
+ total_energy = c(total_energy, energy_of_network)}}}}}}}}}
> mat = total_energy[2:513]
> exbar = mean(total_energy); exbar
[1] 0
> plot(total_energy, xlab = 'network', ylab = 'energy')
```

As seen in Figure 6, the average of all energies in the 3×3 grid Ising model is equal to zero, i.e., $\frac{1}{512}\sum_{i=1}^{512} E_{x_i} = 0$. This is the average of all 512 support points, which is always equal to zero in some Ising models such as the 3×3 grid, horizontal and circular models (see Examples 3.10 and 3.11). Now the question is raised, what is the expected

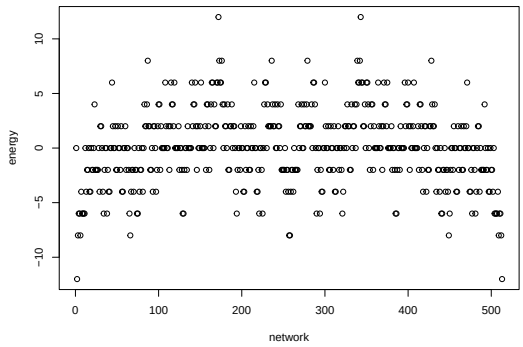


FIGURE 6. All possible energies in the 3×3 grid system containing 9 points.

value for the energy random variable with the corresponding probability function given in Eq. (3.3). This quantity is calculated as follows:

$$\boldsymbol{E}_\beta\left(E_x\right)=\frac{\sum_{x\in E} E_x e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}.$$

Similarly, the variance of energies is calculated as follows:

$$\boldsymbol{Var}_\beta(E_x)=\boldsymbol{E}_\beta(E_x^2)-\Big(\boldsymbol{E}_\beta(E_x)\Big)^2=\sum\nolimits_{x\in E} E_x^2\frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}-\left(\frac{\sum_{x\in E} E_x e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}\right)^2.$$

Note that, the Boltzmann distribution is a mass probability function on E, defined by Eq. (3.3). For $\beta=0$, the distribution given by 3.3 becomes discrete uniform distribution, $DU\{1,...,512\}$ (see Figure 7), i.e., $\xi_x=\frac{1}{512}$, $E_x=E_{x_1},...,E_{x_{512}}$. We calculate this mass probability function in **R** software as follows (The output of **R** codes is shown in Figure 7):

```
> beta=0
> s=exp(-beta*total_energy)/sum(exp(-beta*total_energy))
> plot(s,type='h',col='gray50',xlab='systems',ylab='probabilities')
> abline(h=1/512,col='red')
```

The probability mass function for some $\beta>0$ ($\beta<0$) is plotted in Figure 8 (Figure 9). These figures show how the probability changes for the corresponding energies and for different β values using a bar graph.

Given an arbitrary state x, a site i has uniform distribution. The spin at that site is then updated from the conditional distribution of that site given the other sites of x. Let the sites be $1,\ldots,m$ and x_k be the spin at site k, and $\boldsymbol{x}_{-k}=(x_1,...,x_{k-1},x_{k+1},...,x_m)$ be the other $m-1$ sites of the configuration. For fixed k, write $\boldsymbol{x}^+=(x_1,...,x_{k-1},+1,x_{k+1},...,x_m)$ and $\boldsymbol{x}^-=(x_1,...,x_{k-1},-1,x_{k+1},...,x_m)$. Then,

$$P\left(x_k=+1\,|\,x_{-k}\right)=\frac{\xi_{x^+}}{\xi_{x_{-k}}}=\frac{\xi_{x^+}}{\xi_{x^+}+\xi_{x^-}}=\frac{\frac{e^{-\beta E_{x^+}}}{\sum_y e^{-\beta E_y}}}{\frac{e^{-\beta E_{x^+}}}{\sum_y e^{-\beta E_y}}+\frac{e^{-\beta E_{x^-}}}{\sum_y e^{-\beta E_y}}}=\frac{e^{-\beta E_{x^+}}}{e^{-\beta E_{x^+}}+e^{-\beta E_{x^-}}}=\frac{1}{1+e^{\beta\left(E_{x^+}-E_{x^-}\right)}}.$$

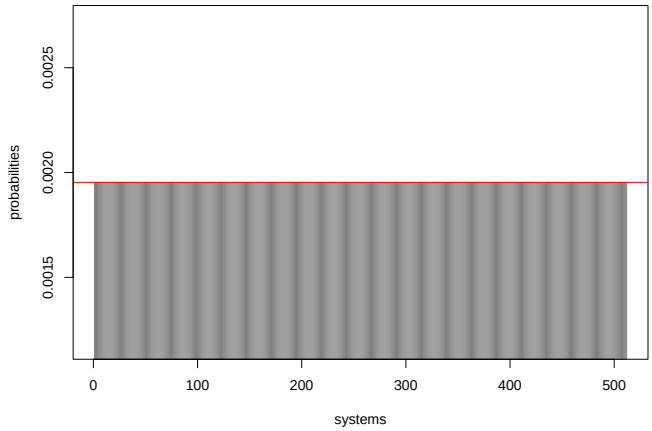


FIGURE 7. The probability mass function ξ_x for $\beta = 0$ in the 3×3 grid system.

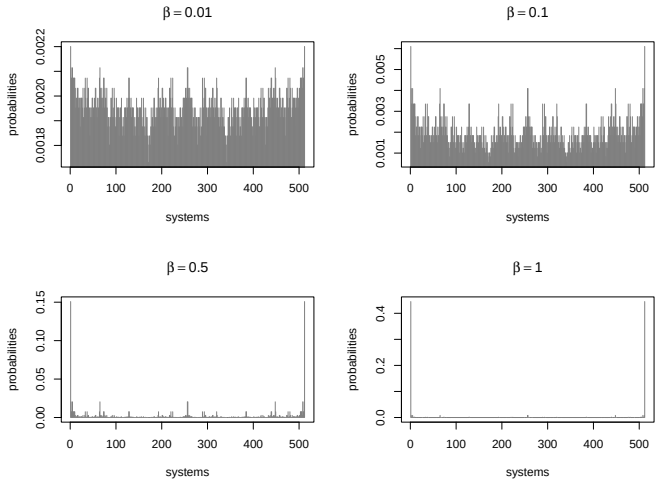


FIGURE 8. The probability mass function ξ_x for some $\beta > 0$ in the 3×3 grid system.

Observe that $E_{\boldsymbol{x}^+} = -\Big(\sum_{\substack{i\sim j\\ i,j\neq k}} x_i x_j + \sum_{i\sim k} x_i\Big)$. Also $E_{\boldsymbol{x}^-} = -\Big(\sum_{\substack{i\sim j\\ i,j\neq k}} x_i x_j - \sum_{i\sim k} x_i\Big)$. This gives $E_{\boldsymbol{x}^+} - E_{\boldsymbol{x}^-} = -2\sum_{i\sim k} x_i$ and after that we have

$$d_k(\beta) = P\left(x_k = +1 \,|\, \boldsymbol{x}_{-k}\right) = \frac{1}{1 + e^{\beta\left(E_{x^+} - E_{x^-}\right)}} = \frac{1}{1 + e^{-2\beta \sum_{i\sim k} x_i}}.$$

Also,

$$\begin{aligned} d'_k(\beta) &= P\left(x_k = -1 \,|\, x_{-k}\right) = 1 - P\left(x_k = +1 \,|\, x_{-k}\right) \\ &= 1 - \frac{1}{1 + e^{-2\beta \sum_{i\sim k} x_i}} = \frac{e^{-2\beta \sum_{i\sim k} x_i}}{1 + e^{-2\beta \sum_{i\sim k} x_i}} = \frac{1}{1 + e^{2\beta \sum_{i\sim k} x_i}}. \end{aligned}$$

For example, in state given by Figure 5 , we have

$$d_5(\beta) = P\left(x_5 = +1 \,|\, \boldsymbol{x}_{-5}\right) = \frac{1}{1 + e^{-2\beta \sum_{i\sim 5} x_i}} = \frac{1}{1 + e^{-2\beta \sum_{(i,j)\sim(2,2)} a_{ij}}}$$

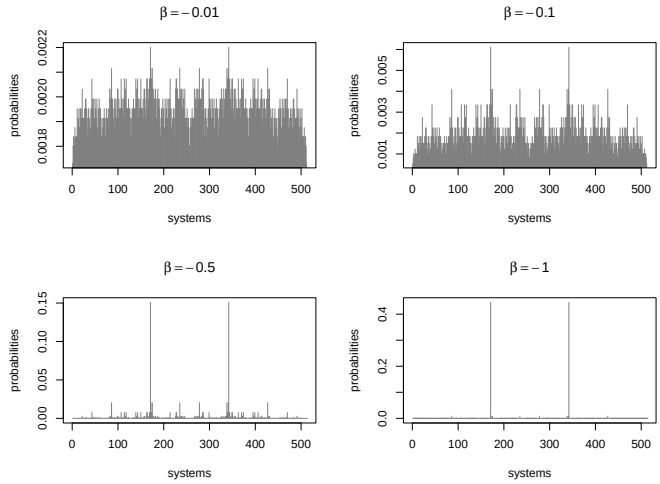


FIGURE 9. The probability mass function ξ_x for some $\beta < 0$ in the 3×3 grid system.

$$= \frac{1}{1+e^{-2\beta(a_{12}+a_{21}+a_{23}+a_{32})}} = \frac{1}{1+e^{-2\beta(-1+1-1+1)}} = \frac{1}{2},$$

$$d_8'(\beta) = P\left(x_8 = -1 \,|\, \boldsymbol{x}_{-8}\right) = \frac{1}{1+e^{2\beta\sum_{i\sim 8} x_i}} = \frac{1}{1+e^{2\beta\sum_{(i,j)\sim(3,2)} a_{ij}}}$$
$$= \frac{1}{1+e^{2\beta(a_{22}+a_{31}+a_{33})}} = \frac{1}{1+e^{2\beta(1-1+1)}} = \frac{1}{1+e^{2\beta}}.$$

Similarly, the conditional probabilities of the rest of the states in Figure 3 will be as follows without calculation:

$$d_k\left(\beta\right) = P\left(x_k = +1 \,|\, x_{-k}\right) = \begin{cases} \frac{1}{2}, & k = 1,5,9, \\ \frac{1}{1+e^{-2\beta}}, & k = 2,8, \\ \frac{1}{1+e^{2\beta}}, & k = 4, \\ \frac{1}{1+e^{-6\beta}}, & k = 6, \\ \frac{1}{1+e^{4\beta}}, & k = 3, \\ \frac{1}{1+e^{-4\beta}}, & k = 7. \end{cases}$$

In Figure 10, we plot $d_k\left(\beta\right)$ for all possible values k from 1 to 9 and some parameter β for -2 to 2. All $d_k\left(\beta\right)$ curves at the $\beta = 0$ have the same value $\frac{1}{2}$, in which case the conditional probabilities at all k points are equal to $\frac{1}{2}$, i.e. $d_k(0) = P(x_k = +1 \mid x_{-k}) = \frac{1}{2}$. The numerical values of the function $d_k(\beta)$ at some points of β are given in Table 1.

Note that, since

$$d_k(-\beta) = \begin{cases} \frac{1}{2}, & k = 1,5,9, \\ \frac{1}{1+e^{2\beta}}, & k = 2,8, \\ \frac{1}{1+e^{-2\beta}}, & k = 4, \\ \frac{1}{1+e^{6\beta}}, & k = 6, \\ \frac{1}{1+e^{-4\beta}}, & k = 3, \\ \frac{1}{1+e^{4\beta}}, & k = 7, \end{cases} \qquad = 1 - d_k(\beta),$$

we conclude that curve $d_k\left(\beta\right)$ has the central symmetry around point $(0, 0.5)$ (also see Figure 10).

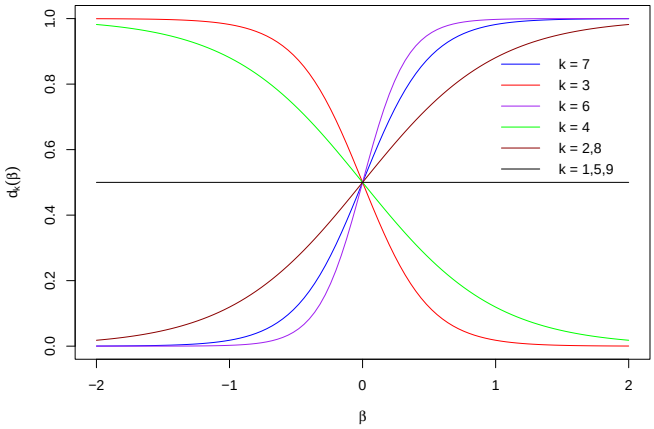


FIGURE 10. Display conditional probability $d_k(\beta)$ based on different values of k.

TABLE 1. Values of $d_k(\beta)$ for $\beta = 0, 0.441, 0.75, -1.5$ and $k = 1, ..., 9$.

k/β	0	0.441	0.75	-1.5
1	0.50	0.50	0.50	0.50
2	0.50	0.71	0.82	0.05
3	0.50	0.15	0.05	≈ 1
4	0.50	0.29	0.18	0.95
5	0.50	0.50	0.50	0.50
6	0.50	0.93	0.99	≈ 0
7	0.50	0.85	0.95	≈ 0
8	0.50	0.71	0.82	0.05
9	0.50	0.50	0.50	0.50

In the calculation of conditional probabilities $d_k(\beta)$, considering that the support of the corresponding random variable has two members $\{-1, +1\}$, the conditional mathematical expectation can be calculated as follows:

$$\begin{aligned} e_k(\beta) &= \boldsymbol{E}_\beta(x_k \,|\, \boldsymbol{x}_{-k}) = \sum\nolimits_{i\in\{-1,1\}} i \times P(x_k = i \,|\, \boldsymbol{x}_{-k}) \\ &= 1 \times d_k(\beta) + (-1) \times d'_k(\beta) = 1 \times d_k(\beta) + (-1) \times (1 - d_k(\beta)) \\ &= 2d_k(\beta) - 1 = \frac{2}{1 + e^{-2\beta \sum_{i\sim k} x_i}} - 1 \\ &= \frac{1 - e^{-2\beta \sum_{i\sim k} x_i}}{1 + e^{-2\beta \sum_{i\sim k} x_i}}. \end{aligned}$$

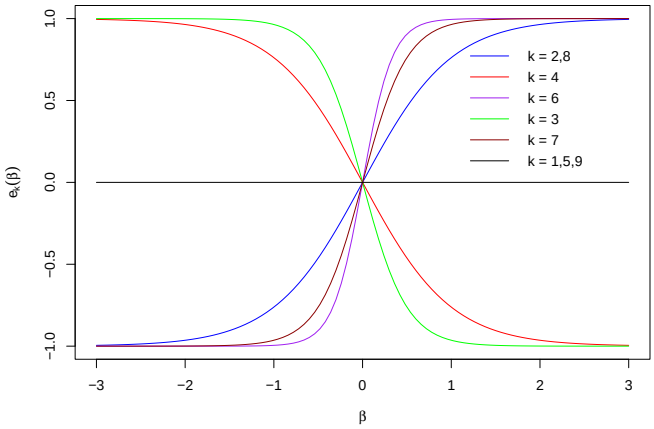


FIGURE 11. Display conditional mathematical expectation $e_k(\beta)$ based on different values of k.

In other words

$$e_k(\beta) = \boldsymbol{E}_\beta(x_k \mid \boldsymbol{x}_{-k}) = \frac{1 - e^{-2\beta\sum_{i\sim k} x_i}}{1 + e^{-2\beta\sum_{i\sim k} x_i}} = \begin{cases} 0 & k = 1, 5, 9 \\ \frac{1-e^{-2\beta}}{1+e^{-2\beta}}, & k = 2, 8, \\ \frac{1-e^{2\beta}}{1+e^{2\beta}}, & k = 4, \\ \frac{1-e^{-6\beta}}{1+e^{-6\beta}}, & k = 6, \\ \frac{1-e^{4\beta}}{1+e^{4\beta}}, & k = 3, \\ \frac{1-e^{-4\beta}}{1+e^{-4\beta}}, & k = 7. \end{cases}$$

In Figure 11, we plot $e_k(\beta) = E_\beta(x_k|\boldsymbol{x}_{-k})$ for all possible values k from 1 to 9 and some parameter β for -2 to 2.

Note that, without using the **R** software, considering that the relation $e_k(\beta) = 2d_k(\beta) - 1$ is established between the two functions, we could draw graph $e_k(\beta)$ with the help of graph $d_k(\beta)$. In this way, the behavior of function $e_k(\beta)$ is the same as the behavior of function $d_k(\beta)$, and the only difference is that function $d_k(\beta)$ is stretched vertically and translated down. It is easily seen

$$e_k(-\beta) = 2d_k(-\beta) - 1 = 2(1 - d_k(\beta)) - 1 = 1 - 2d_k(\beta) = -e_k(\beta),$$

which shows that the function $e_k(\beta)$ is odd, and in other words, it has central symmetry around the origin (also see Figure 11).

Similarly, the conditional variance is also calculated. For this purpose, we have

$$\boldsymbol{E}_\beta\left(x_k^2|x_{-k}\right) = \sum\nolimits_{i\in\{-1,1\}} i^2 \times P(x_k = i|\boldsymbol{x}_{-k}) = d_k(\beta) + d'_k(\beta) = d_k(\beta) + (1 - d_k(\beta)) = 1.$$

The result is

$$v_k(\beta) = \boldsymbol{Var}_\beta(x_k|\boldsymbol{x}_{-k}) = \boldsymbol{E}_\beta\left(x_k^2|\boldsymbol{x}_{-k}\right) - \left(\boldsymbol{E}_\beta\left(x_k|\boldsymbol{x}_{-k}\right)\right)^2$$
$$= 1 - \left(\frac{1-e^{-2\beta\sum_{i\sim k}x_i}}{1+e^{-2\beta\sum_{i\sim k}x_i}}\right)^2 = \frac{4e^{-2\beta\sum_{i\sim k}x_i}}{\left(1+e^{-2\beta\sum_{i\sim k}x_i}\right)^2}$$
$$= \begin{cases} 1 & k\in\{1,5,9\} \\ \frac{4e^{-2\beta}}{(1+e^{-2\beta})^2}, & k\in 2,8, \\ \frac{4e^{2\beta}}{(1+e^{2\beta})^2}, & k=4, \\ \frac{4e^{-6\beta}}{(1+e^{-6\beta})^2}, & k=6, \\ \frac{4e^{4\beta}}{(1+e^{4\beta})^2}, & k=3, \\ \frac{4e^{-4\beta}}{(1+e^{-4\beta})^2}, & k=7. \end{cases}$$

The interesting thing that happens in function $v_k(\beta) = \boldsymbol{Var}_\beta(x_k|\boldsymbol{x}_{-k})$. To prove this fact, we will show that $h(a) = \frac{4e^{a\beta}}{(1+e^{a\beta})^2}$ is an even function, so we proceed as follows:

$$h(a) = \frac{4e^{a\beta}}{(1+e^{a\beta})^2} = \frac{4e^{a\beta}}{1+e^{2a\beta}+2e^{a\beta}} = \frac{4e^{-a\beta}}{e^{-2a\beta}+1+2e^{-a\beta}} = \frac{4e^{-a\beta}}{(1+e^{-a\beta})^2} = h(-a).$$

According to the above identity, we have

$$v_3(\beta) = \frac{4e^{4\beta}}{(1+e^{4\beta})^2} = h(4) = h(-4) = \frac{4e^{-4\beta}}{(1+e^{-4\beta})^2} = v_7(\beta),$$
$$v_4(\beta) = \frac{4e^{2\beta}}{(1+e^{2\beta})^2} = h(2) = h(-2) = \frac{4e^{-2\beta}}{(1+e^{-2\beta})^2} = v_2(\beta) = v_8(\beta).$$

Now we can rewrite function $v_k(\beta) = \boldsymbol{Var}_\beta(x_k|\boldsymbol{x}_{-k})$ as below:

$$v_k(\beta) = \boldsymbol{Var}_\beta(x_k|\boldsymbol{x}_{-k}) = \begin{cases} 1, & k=1,5,9, \\ \frac{4e^{-2\beta}}{(1+e^{-2\beta})^2}, & k=2,4,8, \\ \frac{4e^{4\beta}}{(1+e^{4\beta})^2}, & k=3,7, \\ \frac{4e^{-6\beta}}{(1+e^{-6\beta})^2}, & k=6. \end{cases}$$

Next, in Figure 12, we plot $v_k(\beta) = Var_\beta(x_k|\boldsymbol{x}_{-k})$ for all possible values k from 1 to 9 and some parameter β for -3 to 3. It can be easily observed

$$v_k(-\beta) = 1 - \left(\boldsymbol{E}_{-\beta}\left(x_k|\boldsymbol{x}_{-k}\right)\right)^2 = 1 - \left(e_k(-\beta)\right)^2 = 1 - \left(-e_k(\beta)\right)^2 = v_k(\beta),$$

which shows that the function $e_k(\beta)$ is even, and in other words, it has axial symmetry around the vertical axis (also see Figure 12).

An interesting point that is worth noting is that in the case $\beta = 0$, the three functions $d_k(\beta) = P(x_k = +1|\boldsymbol{x}_{-k})$, $e_k(\beta) = \boldsymbol{E}_\beta(x_k|\boldsymbol{x}_{-k})$ and $v_k(\beta) = \boldsymbol{Var}_\beta(x_k|\boldsymbol{x}_{-k})$ do not depend on the parameter β, and exactly this happens for the cases where k is on the diagonal elements of the 3×3 grid matrix, i.e., points 1, 5, and 9. In other words $d_1(0) = d_5(0) = d_9(0) = \frac{1}{2}$, $e_1(0) = e_5(0) = e_9(0) = 0$ and $v_1(0) = v_5(0) = v_9(0) = 1$.

In Figure 13, we simulate the 60×60 grid Ising model with some parameter β. The Gibbs sampler was run for 100,000 steps. The state space E has 2^{3600} elements. Simulations B and C are for attractive systems, with $\beta > 0$. Note that $\beta = 0.441$ is a critical inverse temperature. In case D, $\beta < 0$ and the system is repelling.

In ***R*** codes below, we observe how the Metropolis-Hastings algorithm works and how it can simulate the Ising model. This structure will also be provided by package in section 4.1:

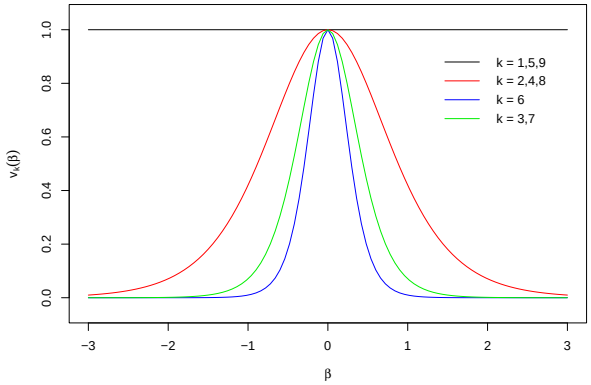


FIGURE 12. Display conditional variance $v_k(\beta)$ based on different values of k.

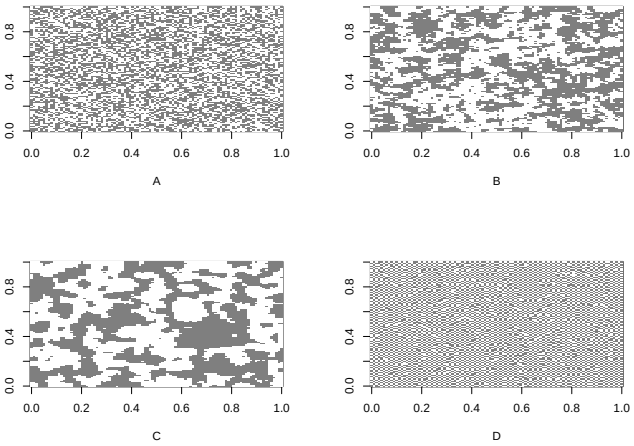


FIGURE 13. Ising simulation on a 60×60 grid. Parameter values are: A: $\beta = 0$, B: $\beta = 0.441$, C: $\beta = 0.75$, D: $\beta = -1.5$.

```
> ### simulation of model ###
> start_time = Sys.time()
> par(mfrow = c(2, 2))
> beta_temp = c(0, 0.441, 0.75, -1.5)
> plot_name = c("A", "B", "C", "D")
> for (L in 1:4) {
+ w = 100
+ beta = beta_temp[L]
+ trial = 100000
+ matr = matrix(sample(c(-1, 1), (w+2)^(2), replace = TRUE), nrow = w + 2)
+ matr[c(1, w + 2), ] = 0
+ matr[ ,c(1, w + 2)] = 0
+ for(K in 1:trial){
```

```
+ Q = sample(2:(w + 1), 1)
+ Y = sample(2:(w+1), 1)
+ degree = matr[Q + 1, Y] + matr[Q - 1, Y] + matr[Q, Y - 1] + matr[Q, Y + 1]
+ probability = (1)/(1 + exp(-beta * 2 * degree))
+ if (runif(1) < probability)
+ {matr[Q, Y] = 1}
+ else
+ {matr[Q, Y] = -1}
+ }
+ at_end = matr[2: (w + 1), 2:(w + 1)]
+ image(at_end, xlab = plot_name[L], col = c(0, 'gray50'))
+ }
> finish_time = Sys.time()
> difftime(finish_time, start_time) Time difference of 6.625356 secs
```

Now let us compute the entropy of distribution given by Eq. (3.3) . By the definition of entropy, we have

$$\begin{aligned} H_\beta(\boldsymbol{\xi}) &= -\sum\nolimits_{x\in E}\xi_x\ln\xi_x = -\sum\nolimits_{x\in E}\frac{e^{-\beta E_{x_i}}}{\sum_{y\in E}e^{-\beta E_y}}\ln\left(\frac{e^{-\beta E_{x_i}}}{\sum_{y\in E}e^{-\beta E_y}}\right)\\ &= \sum\nolimits_{x\in E}\frac{e^{-\beta E_{x_i}}\left(\beta E_{x_i}+\ln\left(\sum_{y\in E}e^{-\beta E_y}\right)\right)}{\sum_{y\in E}e^{-\beta E_y}}\\ &= \beta\sum\nolimits_{x\in E}\frac{E_{x_i}e^{-\beta E_{x_i}}}{\sum_{y\in E}e^{-\beta E_y}}+\ln\left(\sum\nolimits_{y\in E}e^{-\beta E_y}\right)\sum\nolimits_{x\in E}\xi_x\\ &= \beta\boldsymbol{E}_\beta(E_x)+\ln\left(\sum\nolimits_{y\in E}e^{-\beta E_y}\right). \end{aligned}$$

In the conclusion section, $H_\beta(\boldsymbol{\xi})$ will be drawn in the 3×3 grid, horizontal and circular models.

Now the issue arises that the entropy function $H_\beta(\boldsymbol{\xi})$ is convex (or concave) for what values of β. The following proposition answers this question.

Proposition 3.6. *In the Ising model, the entropy of Boltzmann probability given by Eq. (3.3), i.e., $H_\beta(\boldsymbol{\xi})$ is a convex function w.r.t. β, if $\beta\mathbf{Cov}_\beta\left(E_x,E_x^2\right)\geq\mathbf{Var}_\beta\left(E_x\right)\left(2\beta\boldsymbol{E}_\beta\left(E_x\right)+1\right)$.*

Proof. We know $H_\beta(\boldsymbol{\xi})=\beta\boldsymbol{E}_\beta\left(E_x\right)+\ln\left(\sum_{y\in E}e^{-\beta E_y}\right)$. Thus, we have

$$\begin{aligned} H'_\beta(\boldsymbol{\xi}) &= \beta\frac{d}{d\beta}\Big(\boldsymbol{E}_\beta(E_x)\Big)+\boldsymbol{E}_\beta(E_x)-\boldsymbol{E}_\beta(E_x)=\beta\frac{d}{d\beta}\left(\frac{\sum_{x\in E}E_xe^{-\beta E_x}}{\sum_{y\in E}e^{-\beta E_y}}\right)\\ &= -\beta\left(\boldsymbol{E}_\beta\left(E_x^2\right)-\left(\boldsymbol{E}_\beta(E_x)\right)^2\right)=-\beta\mathbf{Var}_\beta\left(E_x\right). \end{aligned}$$

Therefore, the second derivative of $H_\beta(\boldsymbol{\xi})$ w.r.t. β is equal to

$$\begin{aligned} H''_\beta(\boldsymbol{\xi}) &= -\mathbf{Var}_\beta\left(E_x\right)+\beta\left[-\frac{d}{d\beta}\Big(\boldsymbol{E}_\beta\left(E_x^2\right)\Big)+2\boldsymbol{E}_\beta\left(E_x\right)\frac{d}{d\beta}\Big(\boldsymbol{E}_\beta\left(E_x\right)\Big)\right]\\ &= -\mathbf{Var}_\beta\left(E_x\right)-\beta\frac{1}{\left(\sum_{y\in E}e^{-\beta E_y}\right)^2}\times\Big[-\Big(\sum\nolimits_{x\in E}E_x^3e^{-\beta E_x}\Big)\Big(\sum\nolimits_{y\in E}e^{-\beta E_y}\Big)\\ &+\Big(\sum\nolimits_{y\in E}E_ye^{-\beta E_y}\Big)\Big(\sum\nolimits_{x\in E}E_x^2e^{-\beta E_x}\Big)\Big]-2\beta\boldsymbol{E}_\beta\left(E_x\right)\mathbf{Var}_\beta\left(E_x\right)\\ &= -\mathbf{Var}_\beta\left(E_x\right)+\beta\left[\boldsymbol{E}_\beta\left(E_x^3\right)-\boldsymbol{E}_\beta\left(E_x\right)\boldsymbol{E}_\beta\left(E_x^2\right)\right]-2\beta\boldsymbol{E}_\beta\left(E_x\right)\mathbf{Var}_\beta\left(E_x\right) \end{aligned}$$

$$= \beta \mathbf{Cov}_\beta \left(E_x, E_x^2\right) - \mathbf{Var}_\beta \left(E_x\right) \Big(2\beta \boldsymbol{E}_\beta \left(E_x\right) + 1\Big).$$

Because the function $H_\beta(\boldsymbol{\xi})$ is continuous w.r.t. β, its second derivative is non-negative, if

$$\beta \mathbf{Cov}_\beta \left(E_x, E_x^2\right) \geq \mathbf{Var}_\beta \left(E_x\right) \left(2\beta E_\beta \left(E_x\right) + 1\right).$$

Therefore, under this condition, $H_\beta(\boldsymbol{\xi})$ is convex □

The inequality presented in Proposition 3.6 can be written in terms of the third central moment instead of calculating the covariance between energy and its square.

Proposition 3.7. *In the Ising model, the entropy of Boltzmann probability given by Eq. (3.3) , i.e., $H_\beta(\boldsymbol{\xi})$ is a convex function w.r.t. β, if $\beta \boldsymbol{E}_\beta \Big(E_x - \boldsymbol{E}_\beta \left(E_x\right)\Big)^3 \geq \boldsymbol{Var}_\beta \left(E_x\right)$.*

Proof. By Proposition 3.6, we have

$$\begin{aligned}
H''_\beta(\xi) &= \beta \boldsymbol{Cov}_\beta(E_x, E_x^2) - \boldsymbol{Var}_\beta(E_x)\Big(2\beta \boldsymbol{E}_\beta(E_x) + 1\Big) \\
&= \beta \boldsymbol{E}_\beta(E_x^3) - \beta \boldsymbol{E}_\beta(E_x^2)\boldsymbol{E}_\beta(E_x) - 2\beta E_\beta(E_x^2)E_\beta(E_x) + 2\beta\Big(\boldsymbol{E}_\beta(E_x)\Big)^3 - \boldsymbol{Var}_\beta(E_x) \\
&= \beta\left[\boldsymbol{E}_\beta(E_x^3) - 3E_\beta(E_x^2)\boldsymbol{E}_\beta(E_x) + 3E_\beta(E_x)\Big(\boldsymbol{E}_\beta(E_x)\Big)^2 - \Big(\boldsymbol{E}_\beta(E_x)\Big)^3\right] - \boldsymbol{Var}_\beta(E_x) \\
&= \beta \boldsymbol{E}_\beta((E_x - \boldsymbol{E}_\beta(E_x))^3) - \boldsymbol{Var}_\beta(E_x).
\end{aligned}$$

Hence, $H''_\beta(\boldsymbol{\xi}) \geq 0$ if and only if $\beta \boldsymbol{E}_\beta \left((E_x - \boldsymbol{E}_\beta \left(E_x\right))^3\right) \geq \boldsymbol{Var}_\beta \left(E_x\right)$. The continuation of the proof is similar to the proof of Proposition 3.6. □

Suppose that some distribution in an n-membered set has its own probability masses. In this case, there are n possible masses of $\xi_1, \ldots, \xi_n$.

Proposition 3.8. *The only distribution $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n)$ that has the maximum entropy $H(\boldsymbol{\xi}) = -\sum_{i=1}^n \xi_i \ln \xi_i$ under conditions $\sum_{i=1}^n \xi_i = 1$ and $\boldsymbol{E}\left(E_x\right) = \sum_{i=1}^n \xi_i E_i = c$ is the Boltzmann distribution.*

Proof. Consider the Lagrange function as

$$\Psi\left(\xi_1, \ldots, \xi_n, \lambda_1, \lambda_2\right) = -\sum\nolimits_{i=1}^n \xi_i \ln\left(\xi_i\right) + \lambda_1 \left(\sum\nolimits_{i=1}^n \xi_i - 1\right) + \lambda_2 \left(\sum\nolimits_{i=1}^n \xi_i E_i - c\right).$$

By taking the derivative w.r.t $\xi_i, i = 1, \ldots, n$, λ_1 and λ_2 we have

$$\begin{cases}
\dfrac{\partial \Psi\left(\xi_1, \ldots, \xi_n, \lambda_1, \lambda_2\right)}{\partial \xi_i} = -\ln\left(\xi_i\right) - 1 + \lambda_1 + \lambda_2 E_i = 0, \quad i = 1, \ldots, n, \\
\sum\nolimits_{i=1}^n \xi_i = 1, \\
\sum\nolimits_{i=1}^n \xi_i E_i = c.
\end{cases}$$

Thus, by taking $e^{\lambda_1 - 1} = k$ and $\lambda_2 = -\beta$, we have $\xi_i = k e^{-\beta E_i}$. Since $\sum_{i=1}^n \xi_i = 1$, the value of k is calculated as followse $k = \frac{1}{\sum_{i=1}^n e^{-\beta E_j}}$. So $\xi_i = \frac{e^{-\beta E_i}}{\sum_{i=1}^n e^{-\beta E_j}}$. Therefor the Boltzmann distribution is the distribution that maximize the entropy $H(\boldsymbol{\xi}) = -\sum_{i=1}^n \xi_i \ln \xi_i$ w.r.t. two conditions $\sum_{i=1}^n \xi_i = 1$ and $\sum_{i=1}^n \xi_i E_i = c$. □

Now the question arises that in model Eq. (3.3) based on a random sample, MLE and MME for β are the same. Suppose $X_1, ..., X_n \overset{i.i.d.}{\sim} f_\theta(x)$. Note that, MME involves equating sample moments ($\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$) with theoretical moments ($\boldsymbol{E}_\theta\left(X\right) = \sum_x x f_\theta(x)$) [42]. So, we say $\tilde{\theta}$ is MME for θ. In general, if, we consider the likelihood function as follows $\boldsymbol{L}\left(\theta\right) = \prod_{i=1}^n f_\theta(x)$. Thus, $\hat{\theta}$ is a MLE for θ, if $\hat{\theta} = \arg\max_\theta \boldsymbol{L}(\theta)$, i.e., $\boldsymbol{L}(\hat{\theta}) = \max_\theta \boldsymbol{L}(\theta)$ [42].

In the following theorem, we prove that MLE and MLE calculated based on the first moment for β are equal.

Theorem 3.9. *In the Ising model, MLE and MME for β coincide.*

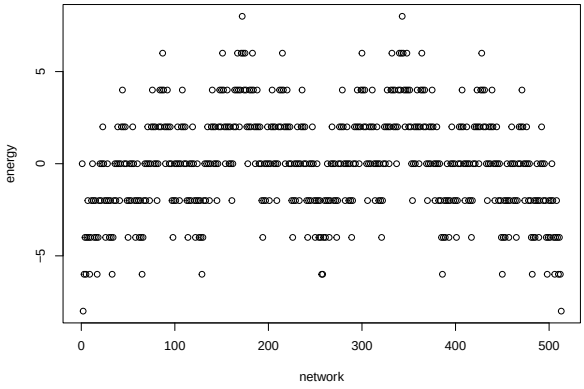


FIGURE 14. All possible energies in the horizontal system containing 9 points.

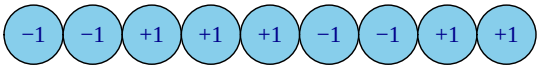


FIGURE 15. Sample configuration on the horizontal system.

Proof. Suppose $E_{x_1},\ldots,E_{x_n} \overset{\text{i.i.d.}}{\sim} \xi_x = \frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}$ denote a random sample of size n from Eq. (3.3). It is worth mentioning, MME is based on matching the first sample moment ($\bar{E}_x = \frac{1}{n}\sum_{i=1}^{n} E_{x_i}$) with the first population moment ($\boldsymbol{E}_\beta(E_x) = \sum_x E_x \xi_x$). In order to get this equality, we must have

$$\boldsymbol{E}_\beta(E_x) = \sum\nolimits_{x\in E} E_x \xi_x = \sum\nolimits_{x\in E} E_x \frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}} = \frac{\sum_{x\in E} E_x e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}} = \bar{E}_x = \frac{1}{n}\sum\nolimits_{i=1}^{n} E_{x_i}, \tag{3.4}$$

which leads to the solution of equation $\bar{E}_x = \boldsymbol{E}_\beta(E_x)$.

On the other hand, to find MLE, we first calculate the likelihood function as follows

$$\boldsymbol{L}(\beta) = \prod_{i=1}^{n} \xi_{x_i} = \prod_{i=1}^{n} \frac{e^{-\beta E_{x_i}}}{\sum_{y\in E} e^{-\beta E_y}} = \frac{e^{-\beta \sum_{i=1}^{n} E_{x_i}}}{\left(\sum_{y\in E} e^{-\beta E_y}\right)^n}.$$

For convenience in calculations, we calculate the natural logarithm of the likelihood function as follows:

$$\boldsymbol{l}(\beta) = \ln(\boldsymbol{L}(\beta)) = -\beta \sum\nolimits_{i=1}^{n} E_{x_i} - n\ln\left(\sum\nolimits_{y\in E} e^{-\beta E_y}\right) = -n\left[\bar{E}_x\beta + \ln\left(\sum\nolimits_{y\in E} e^{-\beta E_y}\right)\right].$$

Thus, taking the derivative w.r.t. β gives

$$\boldsymbol{l}'(\beta) = -n\left[\bar{E}_x - \frac{\sum_{y\in E} E_y e^{-\beta E_y}}{\sum_{y\in E} e^{-\beta E_y}}\right] = -n\left(\bar{E}_x - \boldsymbol{E}_\beta(E_x)\right) = 0.$$

which leads to the solution of equation $\bar{E}_x = \boldsymbol{E}_\beta(E_x)$, which is the same as Equation (3.2). It is noteworthy that, $\boldsymbol{l}''(\beta) = -n\boldsymbol{Var}_\beta(E_x) \le 0$. Hence, $\hat{\beta} = \arg\max_\theta \boldsymbol{L}(\beta)$, that is, the $\hat{\beta}$ obtained from equation $\bar{E}_x = E(E_x)$ is a MLE for β.

This fact shows that MLE and MME coincide, i.e., $\hat{\beta} = \tilde{\beta}$. □

Let $r(\beta) = \boldsymbol{E}_\beta(E_x) - \bar{E}_x = \frac{\sum_{y\in E} E_y e^{-\beta E_y}}{\sum_{y\in E} e^{-\beta E_y}} - \bar{E}_x$. Therefore, the equation $r(\beta) = \boldsymbol{E}_\beta(E_x) - \bar{E}_x = 0$, whose roots were impossible to find analytically, was solved numerically. In the conclusion section, this root will be calculated numerically in the 3×3 grid, horizontal and circular models.

Example 3.10. In the continuation of Example 3.5, consider an Ising model on a horizontal line containing n points where $x_i = \pm 1$, for all i. Similar before, there are 2^n possible states. For $n = 9$, there exist $2^9 = 512$ states and in Figure 14, we have calculated the energy of these 512 states by **R** codes as follows:

```
> sample_space = c(-1, 1)
> total_energy = 0
> for(a1 in sample_space){
+ for(a2 in sample_space){
+ for(a3 in sample_space){
+ for(a4 in sample_space){
+ for(a5 in sample_space){
+ for(a6 in sample_space){
+ for(a7 in sample_space){
+ for(a8 in sample_space){
+ for(a9 in sample_space){
+ network = c(a1, a2, a3, a4, a5, a6, a7, a8, a9)
+ energy_of_network = (-1) * (network[1] * network[2]+
+ network[2] * network[3]+
+ network[3] * network[4]+
+ network[4] * network[5]+
+ network[5] * network[6]+
+ network[6] * network[7]+
+ network[7] * network[8]+
+ network[8] * network[9])
+ total_energy = c(total_energy, energy_of_network)}}}}}}}}}
> vec = total_energy[2:513]
> exbar = mean(total_energy); exbar
[1] 0
> plot(total_energy, xlab = 'network', ylab = 'energy')
```

For example, the configuration x shown in Figure 15 including states $(x_i) = (-1, -1, 1, 1, 1, -1, -1, 1, 1)$ on a horizontal line has following energy:

$$\begin{aligned} E_x &= -\sum\nolimits_{i\sim j} x_i x_j = -\sum\nolimits_{i=1}^{8} x_i x_{i+1} \\ &= -\Big((-1)(-1) + (-1)(1) + (1)(1) + (1)(1) + (1)(-1) + (-1)(-1) + (-1)(1) + (1)(1)\Big) = -2. \end{aligned}$$

Example 3.11. In the continuation of Example 3.5, consider an Ising model on a circle containing n points where $x_i = \pm 1$, for all i. Similar before, there are 2^n possible states. For $n = 9$, there exist $2^9 = 512$ states and in Figure 16, we have calculated the energy of these 512 states.

For example, the configuration shown in Figure 17 including states

$$(x_i) = (-1, -1, 1, 1, 1, -1, -1, 1, 1) \equiv (\sigma_i) = \begin{pmatrix} -1 & -1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 & -1 \end{pmatrix},$$

on a circle has following energy:

$$\begin{aligned} E_x &= -\sum\nolimits_{i\sim j} x_i x_j = -\left(\sum\nolimits_{i=1}^{8} x_i x_{i+1} + x_9 x_1\right) \\ &= -\Big((-1)(-1) + (-1)(1) + (1)(1) + (1)(1) + (1)(-1) + (-1)(-1) + (-1)(1) + (1)(1) + (1)(-1)\Big) = -1. \end{aligned}$$

Note that, (σ_i) is a symbol to show permutation.

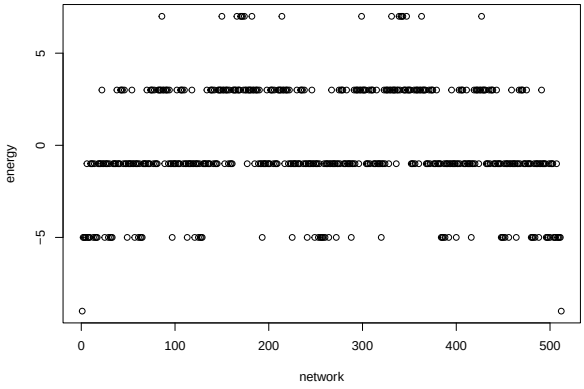


FIGURE 16. All possible energy in the circular system containing 9 points.

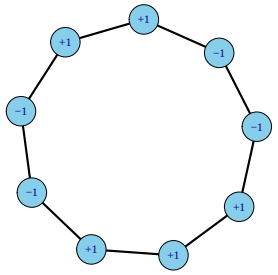


FIGURE 17. Sample configuration x on the circular system.

TABLE 2. The Energy and Boltzmann probability function for four possible configurations in the Ising model with 2 points.

Label	Configuration	Energy	The Boltzmann Probability Function
1	$(x_1) = (-1,-1)$	$E_1 = -1$	$\xi_1 = \frac{e^{\beta}}{4\cosh(\beta)}$
2	$(x_2) = (1,1)$	$E_2 = -1$	$\xi_2 = \frac{e^{\beta}}{4\cosh(\beta)}$
3	$(x_3) = (1,-1)$	$E_3 = 1$	$\xi_3 = \frac{e^{-\beta}}{4\cosh(\beta)}$
4	$(x_4) = (-1,1)$	$E_4 = 1$	$\xi_4 = \frac{e^{-\beta}}{4\cosh(\beta)}$

As you can see in Figure 16, in the circular system, we do not have any state whose energy is equal to zero, But the average of all energies is equal to zero.

In the Conclusion section, broader comparisons about the 3×3 grid, horizontal and circular models are presented.

Example 3.12. Now, Consider an Ising model with 2 points where $x_i = \pm 1$, for all i. Thus, there are 4 possible states and we can calculate the energy of these 4 states as follows (Table 2).

Vector $\boldsymbol{\tau}_1$ can be referred to as induced probability vector by system energies and it is calculated as follows:

$$\boldsymbol{L}(\beta) = \boldsymbol{\tau}_1(i) = P(E_x = i) = \begin{cases} \xi_1 + \xi_2 = \frac{e^{\beta}}{2\cosh(\beta)}, & i = -1, \\ \xi_3 + \xi_4 = \frac{e^{-\beta}}{2\cosh(\beta)}, & i = 1. \end{cases}$$

The induced probability function in the Ising model with 2 points is plotted in terms of different β values in Figure 18.

We observe that, only for $\beta = 0$ we have $\tau_1(-1) = \tau_1(1) = \frac{1}{2}$. If the parameter space is equal to $\Theta = [a, b]$, where $a < b \in \mathbb{R}$, then the maximum likelihood estimator in terms of observed energy will be as follows:

$$\hat{\beta} = \begin{cases} b, & \text{if} \quad E_x = -1, \\ a, & \text{if} \quad E_x = 1, \end{cases}$$

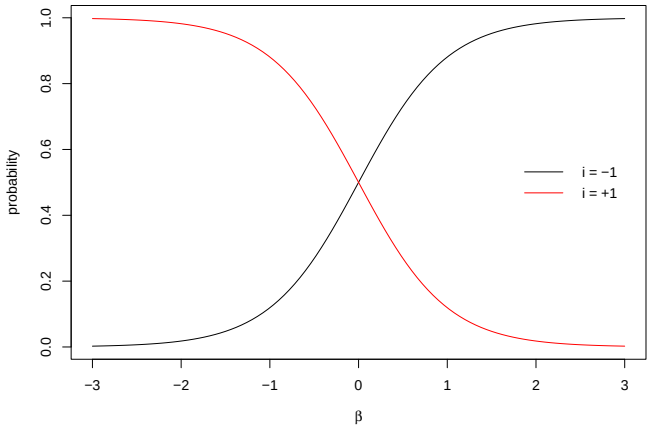


FIGURE 18. The probability function in the Ising model with 2 nudes.

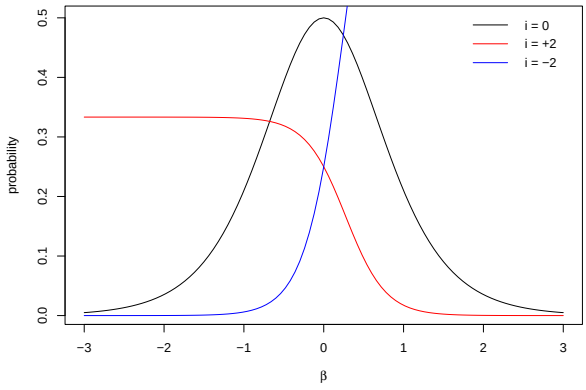


FIGURE 19. The probability function in the Ising model with 3 nudes.

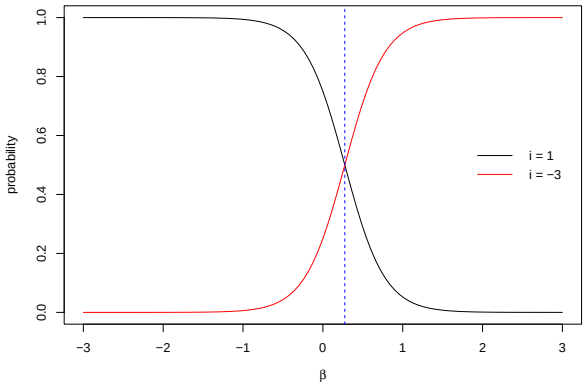


FIGURE 20. The induced probability function in the Ising model on a triangle containing 3 nudes.

because functions $\frac{e^{\beta}}{2\cosh(\beta)}$ and $\frac{e^{-\beta}}{2\cosh(\beta)}$ are increasing and decreasing w.r.t. β, respectively Similarly, Consider an Ising model on a horizontal line (or a triangle) containing 3 points where $x_i = \pm 1$, for all i. Thus, there are 8 possible states and we can calculate the energy of these 8 states in Table 3.

Based on Table 3 , in the horizontal system, induced probability function is calculated as follows:

$$\boldsymbol{L}(\beta) = \boldsymbol{\tau}_2(i) = P(E_x = i) = \begin{cases} \xi_3 + \xi_4 + \xi_7 + \xi_8 = \frac{1}{1+\cosh(2\beta)}, & i = 0, \\ \xi_5 + \xi_6 = \frac{e^{-\beta}}{3e^{-\beta}+e^{3\beta}}, & i = 2, \\ \xi_1 + \xi_2 = \frac{e^{3\beta}}{3e^{-\beta}+e^{3\beta}}, & i = -2. \end{cases}$$

The induced probability function in the Ising model on a horizontal line containing 3 points is plotted in terms of different β values in Figure 19. Note that for $\beta = 0$, we have $\tau_2(0) = 2\tau_2(2) = 2\tau_2(-2) = \frac{1}{2}$. If the parameter space is equal to $\Theta = [a, b]$, where $a < b \in \mathbb{R}$, then MLE in terms of observed energy in the Ising model on a horizontal line

TABLE 3. The Energy and Boltzmann probability function for eight possible configurations in the Ising model on a horizontal line (and a triangle) containing 3 points.

Label	Configuration	Energy	Boltzmann probability function
The horizontal system			
1	$(x_1) = (-1,-1,-1)$	$E_1 = -2$	$\xi_1 = \frac{e^{2\beta}}{4+4\cosh(2\beta)}$
2	$(x_2) = (1,1,1)$	$E_2 = -2$	$\xi_2 = \frac{e^{2\beta}}{4+4\cosh(2\beta)}$
3	$(x_3) = (-1,1,1)$	$E_3 = 0$	$\xi_3 = \frac{1}{4+4\cosh(2\beta)}$
4	$(x_4) = (1,1,-1)$	$E_4 = 0$	$\xi_4 = \frac{1}{4+4\cosh(2\beta)}$
5	$(x_5) = (1,-1,1)$	$E_5 = 2$	$\xi_5 = \frac{e^{-2\beta}}{4+4\cosh(2\beta)}$
6	$(x_6) = (-1,1,-1)$	$E_6 = 2$	$\xi_6 = \frac{e^{-2\beta}}{4+4\cosh(2\beta)}$
7	$(x_7) = (-1,-1,1)$	$E_7 = 0$	$\xi_7 = \frac{1}{4+4\cosh(2\beta)}$
8	$(x_8) = (1,-1,-1)$	$E_8 = 0$	$\xi_8 = \frac{1}{4+4\cosh(2\beta)}$
The triangular system			
1	$\begin{pmatrix}-1 & -1 & -1\\ -1 & -1 & -1\end{pmatrix}$	$E_1 = -3$	$\xi_1 = \frac{e^{3\beta}}{6e^{-\beta}+2e^{3\beta}}$
2	$\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}$	$E_2 = -3$	$\xi_2 = \frac{e^{3\beta}}{6e^{-\beta}+2e^{3\beta}}$
3	$\begin{pmatrix}-1 & 1 & 1\\ 1 & 1 & -1\end{pmatrix}$	$E_3 = 1$	$\xi_3 = \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}$
4	$\begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 1\end{pmatrix}$	$E_4 = 1$	$\xi_4 = \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}$
5	$\begin{pmatrix}1 & -1 & 1\\ -1 & 1 & 1\end{pmatrix}$	$E_5 = 1$	$\xi_5 = \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}$
6	$\begin{pmatrix}-1 & 1 & -1\\ 1 & -1 & -1\end{pmatrix}$	$E_6 = 1$	$\xi_6 = \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}$
7	$\begin{pmatrix}-1 & -1 & 1\\ -1 & 1 & -1\end{pmatrix}$	$E_7 = 1$	$\xi_7 = \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}$
8	$\begin{pmatrix}1 & -1 & -1\\ -1 & -1 & 1\end{pmatrix}$	$E_8 = 1$	$\xi_8 = \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}$

containing 3 points, will be as follows:

$$\hat{\beta} = \begin{cases} b, & \text{if } E_x = -2 \text{ or } (E_x = 0,\ a < b < 0),\\ 0, & \text{if } E_x = 0,\ a < 0 < b,\\ a, & \text{if } E_x = 2 \text{ or } (E_x = 0,\ 0 < a < b), \end{cases}$$

because $\frac{1}{1+\cosh(2\beta)}$ is increasing on $(0,\infty)$ and decreasing on $(-\infty,0)$, $\frac{e^{-\beta}}{3e^{-\beta}+e^{3\beta}}$ is decreasing and $\frac{e^{3\beta}}{3e^{-\beta}+e^{3\beta}}$ is increasing.

Based on Table 3, in triangular system, the induced probability function is calculated as follows:

$$L(\beta) = \tau_3(i) = P(E_x = i) = \begin{cases} \sum_{j=3}^{8} \xi_j = \frac{e^{-\beta}}{e^{-\beta}+\frac{1}{3}e^{3\beta}}, & i = 1,\\ \xi_1 + \xi_2 = \frac{e^{3\beta}}{3e^{-\beta}+e^{3\beta}}, & i = -3. \end{cases}$$

The induced probability function in the Ising model on a triangle containing 3 points is plotted in terms of different β values in Figure 20.

We observe that, only for $\beta = 0.2746402$ we have, $\tau_3(1) = \tau_3(-3) = \frac{1}{2}$. If the parameter space is equal to $\Theta = [a,b]$, where $a < b \in \mathbb{R}$, then MLE in terms of observed energy in the Ising model on a triangle containing 3 points, will be

as follows:

$$\hat{\beta} = \begin{cases} b, & \text{if } E_x = -3, \\ a, & \text{if } E_x = 1. \end{cases}$$

because functions $\frac{e^{3\beta}}{3e^{-\beta}+e^{3\beta}}$ and $\frac{e^{-\beta}}{e^{-\beta}+\frac{1}{3}e^{3\beta}}$ are increasing and decreasing w.r.t. β, respectively.

Define the Markov chain $\{X_n : n \in \mathbb{N}\}$ in such a way that X_n represents the state of the process at the n^{th} stage. In this case, $E_0 = \{-1, 1\}$, and

$$P_{xx} = P(X_1 = x \mid X_0 = x) = p, \quad p \neq 0, 1, \quad x \in \{-1, 1\},$$
$$P_{xy} = P(X_1 = y \mid X_0 = x) = q = 1 - p, \quad q \neq 0, 1, \quad x, y \in \{-1, 1\}, \quad x \neq y.$$

Therefore, in the Ising model containing 2 points where $x_i = \pm 1$, for all i, that the values of these two points are independent of each other, the transition probability matrix will be as follows:

$$P = \begin{pmatrix} p^2 & q^2 & pq & pq \\ q^2 & p^2 & pq & pq \\ pq & pq & p^2 & q^2 \\ pq & pq & q^2 & p^2 \end{pmatrix}.$$

It is worth noting that, the row and column labeling in the above matrix is done as follows (based on Table 3): Label 1 : $(x_1) = (-1, -1)$, Label 2 : $(x_2) = (1, 1)$, Label 3 : $(x_3) = (1, -1)$ and Label 4 : $(x_4) = (-1, 1)$.

A way to calculate a stationary distribution of a Markov chain $\{X_n : n \in \mathbb{N}\}$ is to use formula $\boldsymbol{\pi} = \boldsymbol{\pi} P$ where $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4)$, which leads to the solution of the following equations

$$\begin{cases} \left(p^2 - 1\right)\pi_1 + q^2\pi_2 + pq\left(\pi_3 + \pi_4\right) = q^2\pi_1 + \left(p^2 - 1\right)\pi_2 + pq\left(\pi_3 + \pi_4\right) = 0, \\ pq\left(\pi_1 + \pi_2\right) + \left(p^2 - 1\right)\pi_3 + q^2\pi_4 = pq\left(\pi_1 + \pi_2\right) + q^2\pi_3 + \left(p^2 - 1\right)\pi_4 = 0. \end{cases}$$

The above equation system has solution $\pi_1 = \pi_2$ and $\pi_3 = \pi_4$, when $p, q \neq 0, 1$. We also need $\pi_0 + \pi_1 + \pi_3 + \pi_4 = 1$. Solving the above equations, we obtain $\boldsymbol{\pi} = \left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$ is an unique stationary distribution for a Markov chain $\{X_n : n \in \mathbb{N}\}$, because every irreducible Markov chain $\{X_n : n \in \mathbb{N}\}$ with a finite state space $E_1 = \{(-1, -1), (1, 1), (1, -1), (-1, 1)\}$ is positive recurrent and hence has a stationary distribution. By the way, since $\{X_n : n \in \mathbb{N}\}$ is an aperiodic process, we can calculate limiting distribution as follows

$$\lim_{n\to\infty} P^n = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}.$$

that the columns of the above matrix are the stationary distribution. It should be noted that the vector containing the Boltzmann probability functions, i.e., $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3, \xi_4) = \left(\frac{e^{2\beta}}{4+4\cosh(2\beta)}, \frac{e^{2\beta}}{4+4\cosh(2\beta)}, \frac{e^{-\beta}}{4\cosh(\beta)}, \frac{e^{-\beta}}{4\cosh(\beta)}\right)$ is not stationary.

Similarly, in the Ising model containing 3 points, that the values of these two points are independent of each other, the transition probability matrix will be as follows:

$$P = \begin{pmatrix} p^3 & q^3 & pq^2 & pq^2 & pq^2 & p^2q & p^2q & p^2q \\ q^3 & p^3 & p^2q & p^2q & p^2q & pq^2 & pq^2 & pq^2 \\ pq^2 & p^2q & p^3 & pq^2 & pq^2 & p^2q & p^2q & q^3 \\ pq^2 & p^2q & pq^2 & p^3 & pq^2 & p^2q & q^3 & p^2q \\ pq^2 & p^2q & pq^2 & pq^2 & p^3 & q^3 & p^2q & p^2q \\ p^2q & pq^2 & p^2q & p^2q & q^3 & p^3 & pq^2 & pq^2 \\ p^2q & pq^2 & p^2q & q^3 & p^2q & pq^2 & p^3 & pq^2 \\ p^2q & pq^2 & q^3 & p^2q & p^2q & pq^2 & pq^2 & p^3 \end{pmatrix}.$$

where labeling is according to Table 3. It is easily shown $\boldsymbol{\pi} = \left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right)$ is an unique stationary distribution for a Markov chain $\{X_n : n \in \mathbb{N}\}$, because every irreducible Markov chain $\{X_n : n \in \mathbb{N}\}$ with a finite state space

$$E_2 = \{(-1,-1,-1),\ (1,1,1),\ (-1,1,1),\ (1,1,-1),\ (1,-1,1),\ (-1,1,-1),\ (-1,-1,1),\ (1,-1,-1)\}.$$

$$\left(E_2 = \left\{\begin{pmatrix}-1 & -1 & -1\\ -1 & -1 & -1\end{pmatrix}, \begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}, \begin{pmatrix}-1 & 1 & 1\\ 1 & 1 & -1\end{pmatrix}, \begin{pmatrix}1 & 1 & -1\\ 1 & -1 & 1\end{pmatrix}, \begin{pmatrix}1 & -1 & 1\\ -1 & 1 & 1\end{pmatrix}, \begin{pmatrix}-1 & 1 & -1\\ 1 & -1 & -1\end{pmatrix}, \begin{pmatrix}-1 & -1 & 1\\ -1 & 1 & -1\end{pmatrix}, \begin{pmatrix}1 & -1 & -1\\ -1 & -1 & 1\end{pmatrix}\right\}\right),$$

in the horizontal (triangular) system, is positive recurrent and hence has a stationary distribution. Also, the vector containing the Boltzmann probability functions, in the horizontal (triangular) system, i.e.

$$\xi = \left(\frac{e^{2\beta}}{4+4\cosh(2\beta)}, \frac{e^{2\beta}}{4+4\cosh(2\beta)}, \frac{1}{4+4\cosh(2\beta)}, \frac{1}{4+4\cosh(2\beta)}, \frac{e^{-2\beta}}{4+4\cosh(2\beta)}, \frac{e^{-2\beta}}{4+4\cosh(2\beta)}, \frac{1}{4+4\cosh(2\beta)}, \frac{1}{4+4\cosh(2\beta)}\right),$$

$$\left(\xi = \left(\frac{e^{3\beta}}{6e^{-\beta}+2e^{3\beta}}, \frac{e^{3\beta}}{6e^{-\beta}+2e^{3\beta}}, \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}, \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}, \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}, \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}, \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}, \frac{e^{-\beta}}{6e^{-\beta}+2e^{3\beta}}\right)\right),$$

is not stationary.

4. Simulation of the ISING model in $\boldsymbol{R}$

4.1. **Metropolis-Hastings alghorithm.** The most famous application of the MCMC method appears in the Metropolis-Hastings algorithm. The algorithm is named after Nicholas Metropolis [1915,1999], a physicist, who first proposed this method in 1950, and W.K. Hastings [1930,2016], a Canadian statistician, who extended this algorithm in 1970s.

The algorithm constructs a Markov chain $\{X_t : t \geq 0\}$ with state space E and stationary distribution $\boldsymbol{\pi} = (\pi_1, \pi_2, ...)$. Suppose P is a transition matrix for some irreducible Markov chain with the state space E. Assume that $X_n = i$. The next stage of the process, i.e., X_{n+1} is chosen by the following procedure [9]:

1. Choose a new state from P, i.e., choose j with probability P_{ij};
2. Let $\alpha_{ij} = \frac{\pi_j P_{ji}}{\pi_i P_{ij}}$.

Assume that $X_n = i$ and $U \sim U(0,1)$. Define

$$X_{n+1} = \begin{cases} j, & \text{if } U \leq \alpha_{ij}, \\ i, & \text{if } U > \alpha_{ij}. \end{cases}$$

For example, using random number generator, simulate a standard normal random variable using MCMC with the probability density function $\varphi_t = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}$. For the Metropolis-Hastings algorithm, from state s, the process moves to t, where t has uniform distribution on $(s-1, s+1)$. Hence, $P_{st} = \frac{1}{2}$, $s-1 \leq t \leq s+1$.

The acceptance function is

$$\alpha_{st} = \frac{\pi_t P_{ts}}{\pi_s P_{st}} = \frac{\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\right)\frac{1}{2}}{\left(\frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}}\right)\frac{1}{2}} = e^{-\frac{t^2-s^2}{2}}.$$

Below, the MCMC sequence is repeated one million times. By the way, the overlaid curve is φ_t. With the Kolmogorov-Smirnov goodness of fit test we obtained, the random sample collect from standard normal distribution.

In the Boltzmann probability given by Eq. (3.3), we have

$$\frac{\pi_x}{\pi_y} = \frac{q_{yx}}{q_{xy}} = e^{-\beta(E_x - E_y)} = e^{-\beta \Delta E},$$

where $\Delta E = (E_x - E_y)$. Let

$$q_{xy} = \begin{cases} e^{-\beta \Delta E}, & \text{if } \Delta E > 0, \\ 1, & \text{if } \Delta E < 0. \end{cases}$$

The following algorithm is suggested:

1. Choose some state i;
2. Calculate differential value ΔE;
3. If $\Delta E < 0$, change the spin. If $\Delta E > 0$, then
   a. Generate $0 < r < 1$ as a random number from a uniform distribution;
   b. if $r < e^{-\beta \Delta E}$, change the spin;
4. Go to the next state and go to 2.

The states are generated with a Boltzmann probability given by Eq. (3.3). In **R**, we can generate random sample with package 'bayess', by command 'isinghm'. In Figure 21, we generate random sample for some β via Metropolis-Hastings method, and we observe that, when β become larger, the number of clusters become smaller and cluster get further to each other.

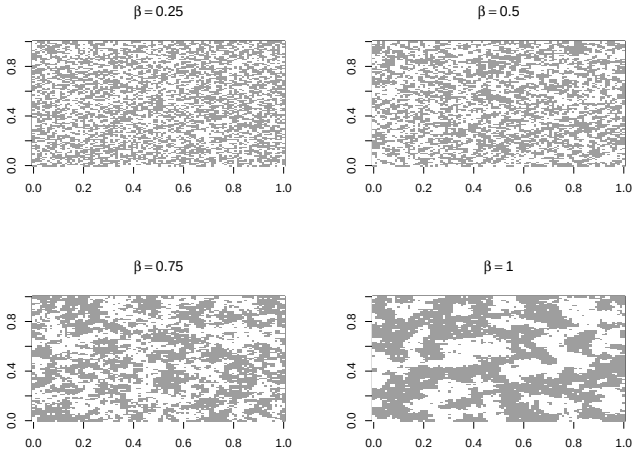


FIGURE 21. Simulating Metropolis-Hastings by $\beta = 0.25, 0.5, 0.75, 1$.

4.2. **GIBBS SAMPLING.** The Gibbs sampling was named by the new researchers in honor of the efforts of the physicist Josiah Willard Gibbs. In this method, the target distribution π is an m-dimensional joint density $\pi(\boldsymbol{x}) = \pi(x_1, ..., x_m)$.

We construct a multivariate Markov chain with limiting distribution π. The generated elements from algorithm are $X^{(0)}, X^{(1)}, ... \left(X_1^{(0)}, ..., X_m^{(0)}\right), \left(X_1^{(1)}, ..., X_m^{(1)}\right), ...$ Clearly, this method is a special case of the Metropolis-Hastings algorithm [27].

In package 'bayess', we can use command 'isingibbs' to generate a random sample.

4.3. **POTTS MODEL.** In the Ising model, every particle can spin in two side. What happened if particles can spin in more than two side?

In this case particles in crystal or topological graph that particles have more than 2 neighbors. The model is named after Renfrey Potts [1925,2005] [33], who described the model near the end of his PHD in 1951. We just simulate this model in **R**.

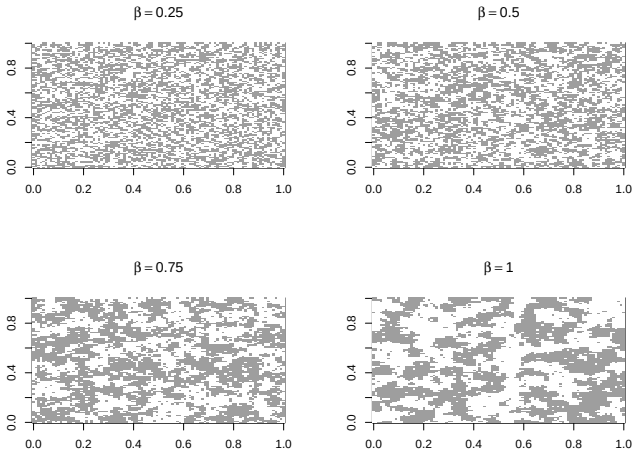


FIGURE 22. Simulating Gibbs by $\beta = 0.25, 0.5, 0.75, 1$.

In package 'bayess' there is two functions 'pottsgibbs' and 'pottshm' that respectively simulate Potts model with Gibbs sampling and Metropolis-Hastings sampling. With Gibbs sampling, we can't choose number of neighbor in lattice and sample space is $E = \{1, 2, 3, 4\}$, but in Metropolis-Hastings sampling, we can select any member of the sample space. Potts model with $E = \{1, 2, 3, 4\}$ shows the fact that we have 4 neighbors. In **R**, for Potts we have to use square matrix (at first Gibbs sampling is done):

```
> library(bayess)
> pots_gibbs = pottsgibbs(niter = 1000, numb = 10, beta = 0)
> table(pots_gibbs); prop.table(table(pots_gibbs))
pots_gibbs
 1  2  3  4
22 31 26 21
pots_gibbs
   1    2    3    4
0.22 0.31 0.26 0.21
> barplot(table(pots_gibbs), col = 'skyblue')
> image(pots_gibbs, col = c('blue', 'red', 'green', 'purple'))
```

In Figure 22, we generate random sample for some β via Gibbs method, and we observe that, when β become larger, the number of clusters become smaller and cluster get further to each other.

In Figure 23 we see Potts model simulation for case. As we observe 22 blue shapes. Also, 1, 2, 3, and 4 represent blue, red, green and purple respectively.

In **R** software, choose the tenth column of the matrix as the first row in the image, the ninth column of the matrix as the second row in the image, and the first column of the matrix as the last row in the image. Now we want to simulate with Metropolis-Hastings sampling. In Metropolis-Hastings method, we can use any matrix (in Gibbs method, we have to use square matrix) and we can use any set of number for sample space:

```
> pots_mh = pottshm(ncol = 7, niter = 10, n = 10, m = 10, beta = 0)
> table(pots_mh); prop.table(table(pots_mh))
pots_mh
```

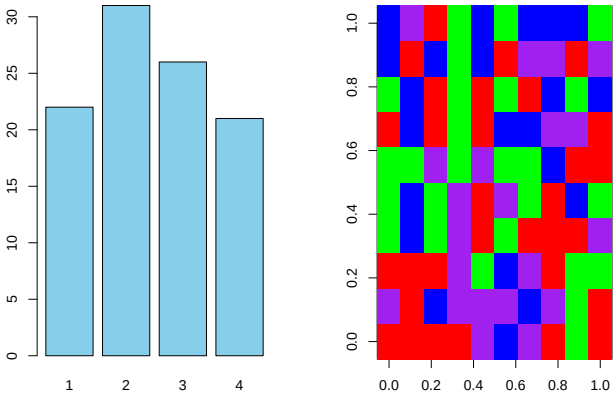


FIGURE 23. Potts model with 4 neighbors ($E = \{1, 2, 3, 4\}$) and bar plot for calculate emprical probability for $\beta = 0$.

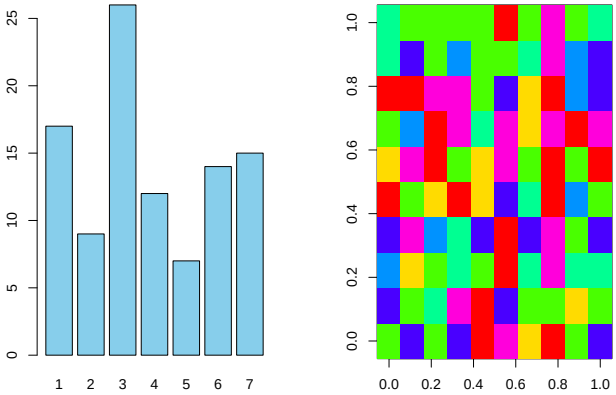


FIGURE 24. Potts model in 10×10 matrix with Metropolis-Hastings and bar plot for calculate emprical probability for $\beta = 0$

```
 1  2  3  4  5  6  7
17  9 26 12  7 14 15
pots_mh
   1    2    3    4    5    6    7
0.17 0.09 0.26 0.12 0.07 0.14 0.15
> image(pots_mh, col = rainbow(7))
> barplot(table(pots_mh), col = 'skyblue')
```

As we see in Figure 24 , the sample space is $E = \{1, 2, 3, 4, 5, 6, 7\}$. Also we can see any parametr of β and any sample space. In this part, the application of Potts models in image processing was presented

TABLE 4. Display frequency table (and probability values) for different energy states (for $\beta = 0$), in the 3×3 grid system.

energy	number	probabilities
-12	2	0.00390625
-8	8	0.01562500
-6	32	0.06250000
-4	46	0.08984375
-2	96	0.18750000
0	144	0.28125000
2	96	0.18750000
4	46	0.08984375
6	32	0.06250000
8	8	0.01562500
12	2	0.00390625

5. STATISTICAL COMPARISON OF MODELS AND DISCUSSION

In this section, we make a statistical comparison between the models presented in three Examples: 3.5, 3.10, and 3.12 for some parameters.

First, we discuss the concentration of the probability masses on different energy states in three models. In Table 4, the frequency of different energy states (and probability values for $\beta = 0$), in the 3×3 grid system is shown. For notation in the following sections, we will denote the vector of probabilities in Table 4 (for case $\beta = 0$) by $\boldsymbol{\tau}_1$. Vector $\boldsymbol{\tau}_1$ can be referred to as the induced probability vector by system energies. For example, the probability of one of the states in Table 4 is calculated as follows:

$$\boldsymbol{\tau}_1(-12) = P(E_x = -12) = \xi_{x_1} + \xi_{x_2} = \frac{1}{512} + \frac{1}{512} = 0.00390625,$$

where

$$x_1 \equiv \Sigma_1 = (a_{ij}) = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix}, \quad x_2 \equiv \Sigma_2 = (a_{ij}) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

Note that in both states, x_1 and x_2 we reach the same energy of -12:

$$\begin{aligned} E_{x_1} = E_{x_2} &= -\sum\nolimits_{i\sim j} x_i x_j = -\sum\nolimits_{(m,n)\sim(v,w)} a_{mn}a_{vw} \\ &= -\big(a_{11}a_{12} + a_{12}a_{13} + a_{11}a_{21} + a_{12}a_{22} + a_{13}a_{23} + a_{21}a_{22} \\ &\quad + a_{22}a_{23} + a_{21}a_{31} + a_{22}a_{32} + a_{23}a_{33} + a_{31}a_{32} + a_{32}a_{33}\big) \\ &= -(12 \times 1) = -12. \end{aligned}$$

As can be seen from Table 4, in 144 states, the energy is zero, and in general, the number of states whose energy is equal to $E_x = 2, 4, 6, 8, 12$ is the same as the number of states whose energy is equal to $E_x = -2, -4, -6, -8, -12$, respectively. In other words, the concentration of the number of energies is symmetrical around zero (for all β), and the probability distribution is also symmetrical around zero (for only $\beta = 0$). Figures 25, 26, and 27 answer this question for $\beta = 0$, $\beta > 0$, and $\beta < 0$, respectively, in the 3×3 grid system. The proof of the fact mentioned in section 3 that in the 3×3 grid system, the average of all energies is zero is written as follows:

$$\begin{aligned} \frac{1}{512}\sum\nolimits_{i=1}^{512} E_{x_i} = \frac{1}{512}\Big[&2(-12) + 8(-8) + 32(-6) + 46(-4) + 96(-2) + 144(0) \\ &+ 96(2) + 46(4) + 32(6) + 8(8) + 2(12)\Big] = 0. \end{aligned}$$

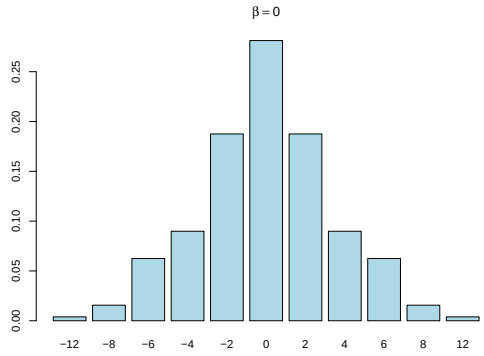


FIGURE 25. The probability of obtaining some energies for $\beta = 0$ in the 3×3 grid system.

In case $\beta = 0$, the entropy of the random variable with support $E = S_{E_x} = \{0, \pm 2, \pm 4, \pm 6, \pm 8, \pm 12\}$ whose the probability mass function $\boldsymbol{\tau}_1$ is given in Table 4 is equal to $H_0(\boldsymbol{\tau}_1) = -\sum_{i \in S_{E_x}} p_i \ln p_i = 1.937360313$.

In Figure 25, the symmetry of probabilities is well observed. In Figure 26, (Figure 27) we observe that, for $\beta > 0$ ($\beta < 0$), the distribution $\boldsymbol{\tau}_1$ corresponding to in the 3×3 grid model puts more mass on low (high) energy states.

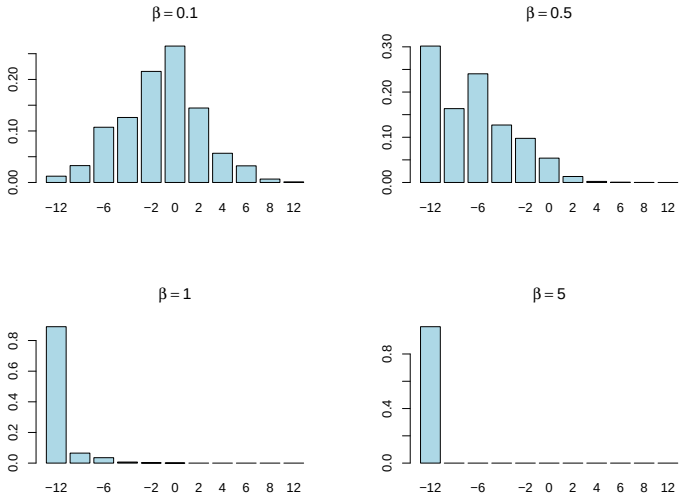


FIGURE 26. The probability of obtaining some energies for $\beta > 0$ in the 3×3 grid system

In Table 5, the frequency of different energy states for $\beta = 0$, in the horizontal model is shown. For notation in the following sections, we will denote the vector of probabilities in Table 5 (for $\beta = 0$) by $\boldsymbol{\tau}_2$. For example, the probability of one of the states in Table 5 is calculated as follows:

$$\boldsymbol{\tau}_2(-8) = P(E_x = -8) = \xi_{x_1} + \xi_{x_2} = \tfrac{1}{512} + \tfrac{1}{512} = 0.00390625,$$

where $x_1 = (-1,-1,-1,-1,-1,-1,-1,-1,-1)$ and $x_2 = (1,1,1,1,1,1,1,1,1)$. Note that in both states x_1 and x_2 we reach the same energy of -8: $E_{x_1} = E_{x_2} = -\sum_{i \sim j} x_i x_j = -\sum_{i=1}^{8} x_i x_{i+1} = -(8 \times 1) = -8$. In the horizontal system, the average of all energies is zero is written as follows:

$$\tfrac{1}{512} \sum_{i=1}^{512} E_{x_i} = \tfrac{1}{512} \left[2(-8) + 16\,(-6) + 56(-4) + 112(-2) + 140(0) + 112(2) + 56(4) + 16(6) + \ 8(8)\right] = 0.$$

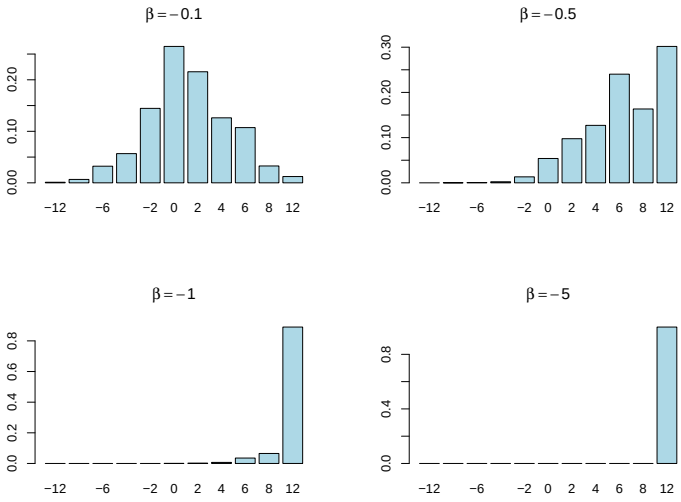


FIGURE 27. The probability of obtaining some energies for $\beta < 0$ in the 3×3 grid system.

In case $\beta = 0$, the entropy of the random variable with support $S_{E_x} = \{0, \pm 2, \pm 4, \pm 6, \pm 8\}$ whose the probability mass function is given in Table 5 is equal to $H_0(\boldsymbol{\tau}_2) = -\sum_{i \in S_{E_x}} p_i \ln p_i = 1.763993465$.

TABLE 5. Display frequency table (and probability values) for different energy states (for $\beta = 0$), in the horizontal system.

energy	number	probabilities
-8	2	0.00390625
-6	16	0.03125000
-4	56	0.10937500
-2	112	0.21875000
0	140	0.27343750
2	112	0.21875000
4	56	0.10937500
6	16	0.03125000
8	2	0.00390625

Similar to the 3×3 grid system, in the horizontal system, the symmetry of concentration around zero (for only $\beta = 0$) can be seen in Table 5. Figures 28, 29, and 30, show the probability distribution for $\beta = 0$, $\beta > 0$, and $\beta < 0$, respectively in the horizontal system.

Again, the symmetry of probabilities for $\beta = 0$ is well observed (see Figure 28). In Figure 29, (Figure 30) we observe that, for $\beta > 0$ ($\beta < 0$), the distribution $\boldsymbol{\tau}_2$ corresponding to in the horizontal model puts more mass on low (high) energy states.

In Table 6, the frequency of different energy states for $\beta = 0$, in the circular system is shown. For notation in the following sections, we will denote the vector of probabilities in Table 6 (for $\beta = 0$) by $\boldsymbol{\tau}_3$. For example, the probability of one of the states in Table 6 is calculated as follows:

$$\boldsymbol{\tau}_3(-9) = P(E_x = -9) = \xi_{x_1} + \xi_{x_2} = \tfrac{1}{512} + \tfrac{1}{512} = 0.00390625,$$

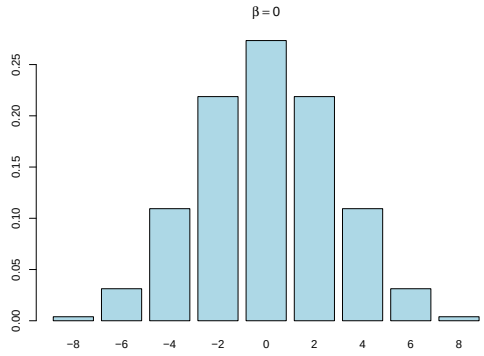


FIGURE 28. The probability of obtaining some energies for $\beta = 0$ in the horizontal system.

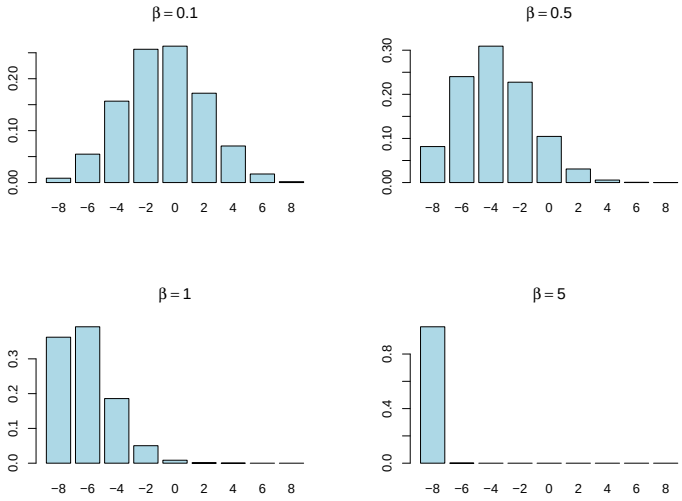


FIGURE 29. The probability of obtaining some energies for $\beta > 0$ in the horizontal system.

where

$$x_1 \equiv (\sigma_i) = \begin{pmatrix} -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & -1 & -1 \end{pmatrix},$$

$$x_2 \equiv (\sigma_i) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}.$$

Note that in both states x_1 and x_2 we reach the same energy of -9:

$$E_{x_1} = E_{x_2} = -\sum\nolimits_{i\sim j} x_i x_j = -\left(\sum\nolimits_{i=1}^{8} x_i x_{i+1} + x_9 x_1\right) = -(9 \times 1) = -9.$$

In the circular system, unlike the 3×3 grid and horizontal systems, symmetry of energy values is not observed (see the Figure 31), but the average of all energies is zero:

$$\tfrac{1}{512} \sum\nolimits_{i=1}^{512} E_{x_i} = \tfrac{1}{512} \left[2(-9) + 72(-5) + 252(-1) + 168(3) + 18(7)\right] = 0.$$

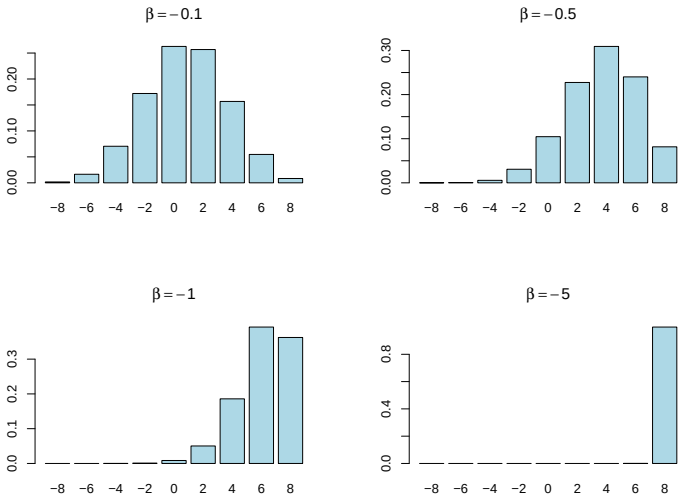


FIGURE 30. The probability of obtaining some energies for $\beta < 0$ in the horizontal system.

TABLE 6. Display frequency table (and probability values) for different energy states (for $\beta = 0$), in the circular system

energy	number	probabilities
-9	2	0.00390625
-5	72	0.14062500
-1	252	0.49218750
3	168	0.32812500
7	18	0.03515625

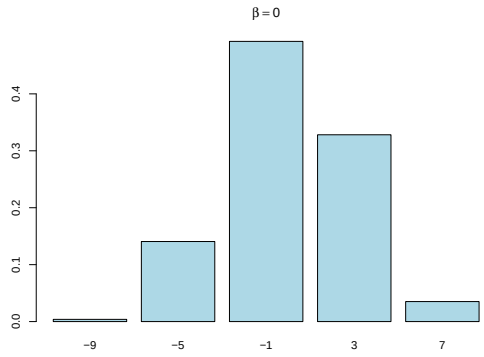


FIGURE 31. The probability of obtaining some energies for $\beta = 0$ in the circular system.

In case $\beta = 0$, the entropy of the random variable with support $S_{E_x} = \{-9, -5, -1, 3, 7\}$ whose the probability mass function is given in Table 6 is equal to $H_0(\boldsymbol{\tau}_3) = -\sum_{i \in S_{E_x}} p_i \ln p_i = 1.129779654$.

Figures 31, 32, and 33 show the probability distribution for $\beta = 0$, $\beta > 0$, and $\beta < 0$, respectively in the circular system. The output of the **R** codes below is shown in Figure 31:

The entropies of the 3×3 grid, horizontal and circular models for $\beta = 0$ are summarized in Table 7.

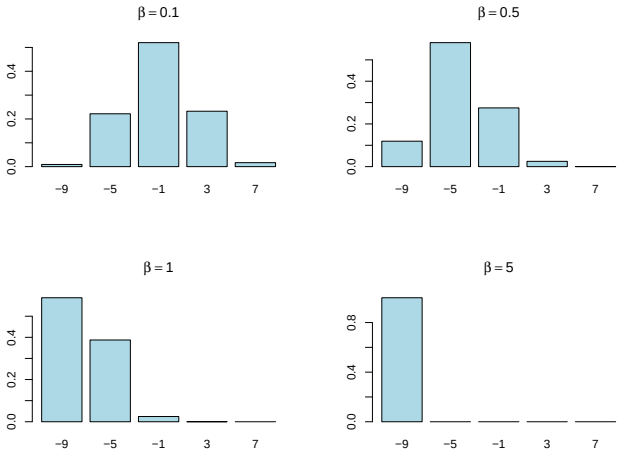


FIGURE 32. The probability of obtaining some energies for $\beta > 0$ in the circular system.

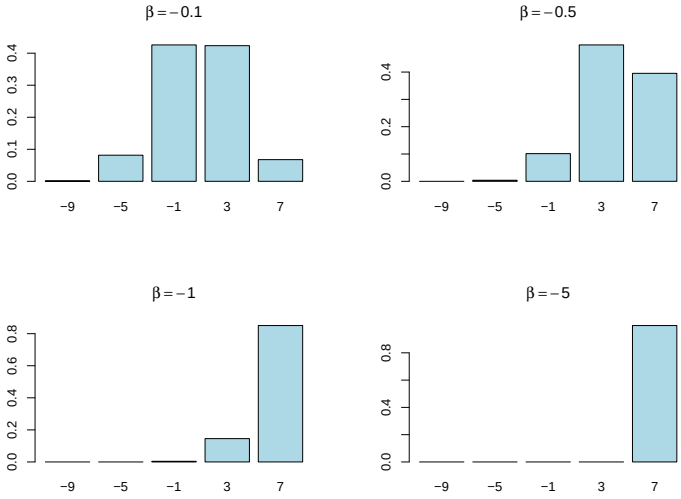


FIGURE 33. The probability of obtaining some energies for $\beta < 0$ in the circular system.

TABLE 7. The entropy for the probability mass functions given in Tabels 4, 5, and 6 for $\beta = 0$.

System Type	Value
grid system	1.93736
horizontal system	1.76399
circular system	1.12978

Note that, by the information given in Table 7, $H_0(\boldsymbol{\tau}_1) > H_0(\boldsymbol{\tau}_2) > H_0(\boldsymbol{\tau}_3)$, i.e., the entropy of induced probability vector of the 3×3 grid, horizontal, and circular models for $\beta = 0$ decreases respectively. Figure 35, compares the entropies of induced probabilities for the general β.

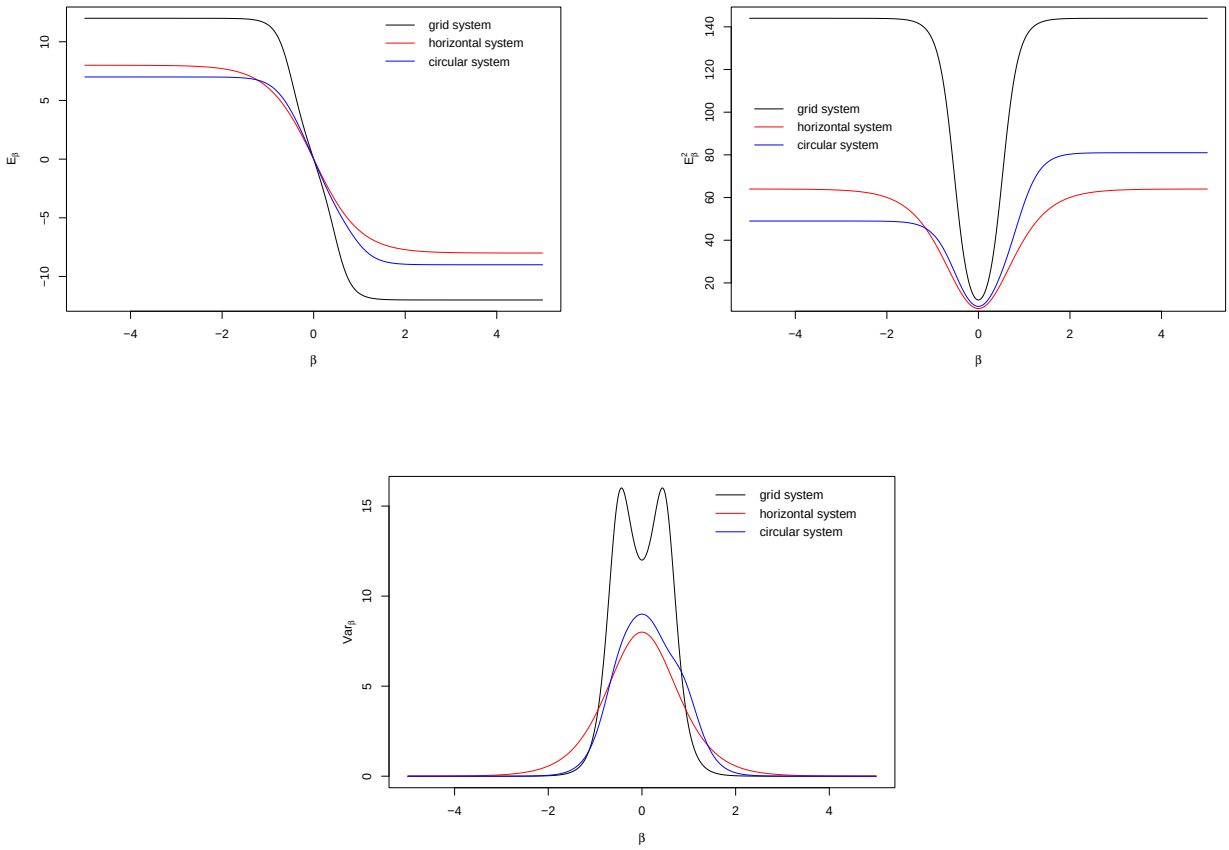


FIGURE 34. The expected value, second moment and variance in corresponding system w.r.t β.

In the following, the mathematical expectation $e(\beta) = \boldsymbol{E}_\beta(E_x) = \frac{\sum_{x\in E} E_x e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}$, the second moment $\boldsymbol{E}_\beta\left(E_x^2\right) = \frac{\sum_{x\in E} E_x^2 e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}$ and the variance $\boldsymbol{Var}_\beta(E_x) = \sum_{x\in E} \frac{E_x^2 e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}} - \left(\frac{\sum_{x\in E} E_x e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}\right)^2$ of the random variable corresponding to Eq. (3.3) in the 3×3 grid, horizontal, and circular models are given along with Figure 34.

From the Figure 34, it can be seen that $\boldsymbol{E}_\beta(E_x)$ is a decreasing function w.r.t β in the 3×3 grid, horizontal and circular models. By the way, $\boldsymbol{E}_\beta\left(E_x^2\right)$ is a decreasing (increasing) function w.r.t β for $\beta<0$ ($\beta>0$) in all mentioned models. Also, according to diagram $\boldsymbol{Var}_\beta(E_x)$, the dispersion is higher for β's that are close to zero (almost for $\beta\in[-2,2]$), and the dispersion is lower for β's that are far from zero on both sides (almost for $|\beta|>2$). By the way, $e(\beta)=\boldsymbol{E}_\beta(E_x)$ is an odd function in the 3×3 grid and horizontal models (but not in the circular model). Since, the concentration of the number of energies in the 3×3 grid and horizontal models is symmetrical around zero (for all β), we have

$$e(-\beta)=\boldsymbol{E}_{-\beta}(E_x)=\frac{\sum_{x\in E} E_x e^{\beta E_x}}{\sum_{y\in E} e^{\beta E_y}}=-\frac{\sum_{x\in E} E_x e^{-\beta(-E_x)}}{\sum_{y\in E} e^{-\beta E_y}}=-\frac{\sum_{x\in E} E_x e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}=-e(\beta). \tag{5.1}$$

Also $e_2(\beta)=\boldsymbol{E}_\beta\left(E_x^2\right)$ and $v(\beta)=\boldsymbol{Var}_\beta(E_x)$ are even functions in the 3×3 grid and horizontal models (but not in the circular model). Since, the concentration of the number of energies in the 3×3 grid and horizontal models is

symmetrical around zero (for all β), we have

$$e_2(-\beta)=\boldsymbol{E}_{-\beta}\left(E_x^2\right)=\frac{\sum_{x\in E}E_x^2e^{\beta E_x}}{\sum_{y\in E}e^{\beta E_y}}=\frac{\sum_{x\in E}E_x^2e^{-\beta(-E_x)}}{\sum_{y\in E}e^{-\beta E_y}}=\frac{\sum_{x\in E}E_x^2e^{-\beta E_x}}{\sum_{y\in E}e^{-\beta E_y}}=e_2(\beta),$$
$$v(-\beta)=\boldsymbol{Var}_{-\beta}\left(E_x\right)=e_2(-\beta)-\left(e(-\beta)\right)^2=e_2(\beta)-\left(e(\beta)\right)^2=v(\beta).$$

Table 7 calculates the entropy corresponding to $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ and $\boldsymbol{\tau}_3$ i.e., the probability distributions given in Tabels 4, 5 and 6 only for case $\beta=0$. Now the question arises, what is the entropy corresponding to Figures 29 to 34 for other β values. In Figure 35, the entropy diagram is drawn in terms of different β.

As we see in Figure 35, the entropy of the 3×3 grid system is bimodal, i.e., $\arg\max_\beta H_\beta(\boldsymbol{\tau}_1)=-0.2252252,\ 0.2252252$, $H_{-0.2252252}(\boldsymbol{\tau}_1)=H_{0.2252252}(\boldsymbol{\tau}_1)=1.952689$, the entropies of the 3×3 grid and horizontal systems are symmetric and the entropy of the circular system is not symmetric. Note that $H_\beta(\boldsymbol{\tau}_1)\geq H_\beta(\boldsymbol{\tau}_2)>H_\beta(\boldsymbol{\tau}_3)$, for $\beta\in[-0.4427,0.4427]$, i.e., the entropy of induced probability vector in the 3×3 grid, horizontal and circular models for $|\beta|\leq0.4427$ decreases respectively. Also, the entropy of induced probability vector in the 3×3 grid and horizontal models is symmetric about $\beta=0$, because by Table 4 (similarly Table 5), we have

$$\begin{aligned}
h_{\tau_1}(\beta)&=H_\beta(\boldsymbol{\tau}_1)=-\sum\nolimits_{i\in S_{E_x}}\boldsymbol{\tau}_1(i)\ln\left(\boldsymbol{\tau}_1(i)\right)\\
&=-\boldsymbol{\tau}_1(0)\ln\left(\boldsymbol{\tau}_1(0)\right)-\sum\nolimits_{i\in\{2,4,6,8,12\}}\boldsymbol{\tau}_1(i)\ln\left(\boldsymbol{\tau}_1(i)\right)-\sum\nolimits_{i\in\{2,4,6,8,12\}}\boldsymbol{\tau}_1(-i)\ln\left(\boldsymbol{\tau}_1(-i)\right)\\
&=-P(E_x=0)\ln\left(P(E_x=0)\right)-\sum\nolimits_{i\in\{2,4,6,8,12\}}\frac{e^{-\beta i}}{\sum_{y\in E}e^{-\beta E_y}}\ln\left(\frac{e^{-\beta i}}{\sum_{y\in E}e^{-\beta E_y}}\right)\\
&-\sum\nolimits_{i\in\{2,4,6,8,12\}}\frac{e^{\beta i}}{\sum_{y\in E}e^{-\beta E_y}}\ln\left(\frac{e^{\beta i}}{\sum_{y\in E}e^{-\beta E_y}}\right)\\
&=\frac{1}{\sum_{y\in E}e^{-\beta E_y}}\ln\left(\sum\nolimits_{y\in E}e^{-\beta E_y}\right)-\frac{1}{\sum_{y\in E}e^{-\beta E_y}}\sum\nolimits_{i\in\{2,4,6,8,12\}}e^{-\beta i}\ln\left(e^{-\beta i}\right)\\
&-\frac{1}{\sum_{y\in E}e^{-\beta E_y}}\sum\nolimits_{i\in\{2,4,6,8,12\}}e^{\beta i}\ln\left(e^{\beta i}\right)\\
&=\frac{1}{\sum_{y\in E}e^{-\beta E_y}}\ln\left(\sum\nolimits_{y\in E}e^{-\beta E_y}\right)+\frac{\beta}{\sum_{y\in E}e^{-\beta E_y}}\sum\nolimits_{i\in\{2,4,6,8,12\}}ie^{-\beta i}-\frac{\beta}{\sum_{y\in E}e^{-\beta E_y}}\sum\nolimits_{i\in\{2,4,6,8,12\}}ie^{\beta i}\\
&=\frac{1}{\sum_{y\in E}e^{-\beta E_y}}\ln\left(\sum\nolimits_{y\in E}e^{-\beta E_y}\right)+\frac{\beta}{\sum_{y\in E}e^{-\beta E_y}}\sum\nolimits_{i\in\{2,4,6,8,12\}}i\left(e^{-\beta i}-e^{\beta i}\right)\\
&=\frac{1}{\sum_{y\in E}e^{\beta E_y}}\ln\left(\sum\nolimits_{y\in E}e^{\beta E_y}\right)-\frac{2}{\sum_{y\in E}e^{-\beta E_y}}\sum\nolimits_{i\in\{2,4,6,8,12\}}i\left(\beta\sinh(\beta i)\right)\\
&=\frac{1}{\sum_{y\in E}e^{\beta E_y}}\ln\left(\sum\nolimits_{y\in E}e^{\beta E_y}\right)-\frac{2}{\sum_{y\in E}e^{-\beta E_y}}\sum\nolimits_{i\in\{2,4,6,8,12\}}i\left(-\beta\sinh(-\beta i)\right)\\
&=H_{-\beta}(\boldsymbol{\tau}_1)=h_{\tau_1}(-\beta).
\end{aligned}$$

Hence $h_{\tau_1}(\beta)=H_\beta(\boldsymbol{\tau}_1)$ (and similarly $h_{\tau_2}(\beta)=H_\beta(\boldsymbol{\tau}_2)$) is an even function in the 3×3 grid (and similarly the horizontal) model (also see Figure 36). Note that in the proof above, the evenness of the function $z(\beta)=\beta\sinh(\beta i)$ was used.

Now, speaking of entropy, it is not bad to have a comparison of energy entropy in the 3×3 grid, horizontal and circular models. First, we present a diagram for the second derivative of entropy as follows:

By Proposition 3.6 and Corollary 3.7, we have

$$\begin{aligned}
s\left(\beta\right)&=H''_\beta(\boldsymbol{\xi})=\beta\,\boldsymbol{Cov}_\beta\left(E_x,E_x^2\right)-\boldsymbol{Var}_\beta\left(E_x\right)\left(2\beta\boldsymbol{E}_\beta\left(E_x\right)+1\right)\\
&=\beta\boldsymbol{E}_\beta\left(\left(E_x-\boldsymbol{E}_\beta\left(E_x\right)\right)^3\right)-\boldsymbol{Var}_\beta\left(E_x\right).
\end{aligned}$$

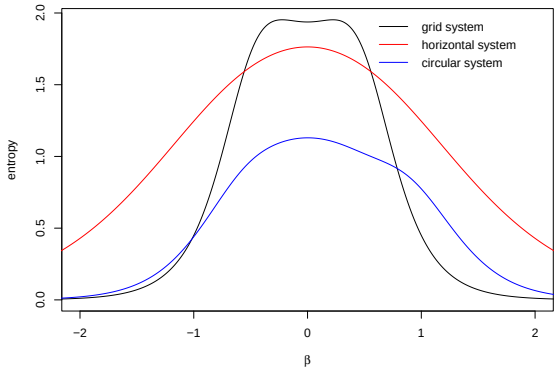


FIGURE 35. Comparing $H_\beta(\boldsymbol{\tau}_i)$, $i = 1, 2, 3$ for the distribution given by Figures 29 to 33.

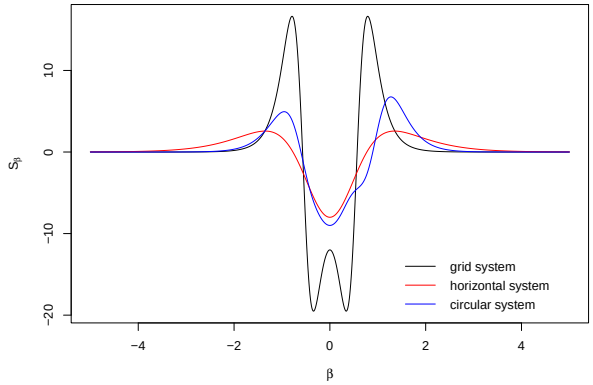


FIGURE 36. The second drivative of entropy.

In Figure 37, for values of β where $s(\beta) = H''_\beta(\xi)$ is positive (negative), $H_\beta(\boldsymbol{\xi})$ is convex (concave). This fact can be seen in the Figure 37.

We observe that, the entropy values for the horizontal and circular systems are smaller than corresponding values in the $n \times n$ grid. It can be mentioned that the convexity or concavity of $H_\beta(\boldsymbol{\xi})$ can be proved theoretically, and the corresponding inequality is given in Proposition 3.6 and Corollary 3.7. Also. we can see, in all models, $H_\beta(\boldsymbol{\xi})$ decrease as $|\beta|$ become greater. The maximum entropy occurs at $\beta = 0$ i.e., $\arg\max_\beta H_\beta(\boldsymbol{\xi}) = 0$, $H_0(\boldsymbol{\xi}) = 6.238325$. Also, the entropy function in the 3×3 grid and horizontal models is symmetric about $\beta = 0$, because by (Eq. (5.1)), we have

$$\begin{aligned} h(-\beta) = H_{-\beta}(\boldsymbol{\xi}) &= -\beta E_{-\beta}(E_x) + \ln\left(\sum_{y\in E} e^{\beta E_y}\right) = -\beta e(-\beta) + \ln\left(\sum_{y\in E} e^{-\beta E_y}\right) \\ &= \beta e(-\beta) + \ln\left(\sum_{y\in E} e^{-\beta E_y}\right) = \beta E_\beta(E_x) + \ln\left(\sum_{y\in E} e^{\beta E_y}\right) \\ &= H_\beta(\boldsymbol{\xi}) = h(\beta). \end{aligned}$$

Hence $h(\beta) = H_\beta(\boldsymbol{\xi})$ is an even function in the 3×3 grid and horizontal model (also see Figure 38).

In Theorem 3.9, we observed that solving equation $r(\beta) = E_\beta(E_x) - \bar{E}_x = 0$ leads to finding MME and MLE for β. Now we want to solve this equation, which cannot be solved analytically, with numerical methods. For example, in Figure 38, the root of the equation $r(\beta) = E_\beta(E_x) - \bar{E}_x = 0$ for state case $\bar{E}_x = 0$ is specified in three models.

One of the ways to calculate the root of equation $r(\beta) = \boldsymbol{E}_\beta(E_x) - \bar{E}_x = 0$ is to use regression models. We want to find this root for in the 3×3 grid, horizontal and circular models.

The working method is that first some random samples are taken from the distribution of Eq. (3.3), then the average of these values is calculated as $\bar{E}_x$. After that, by putting these values in equation $r(\beta) = \boldsymbol{E}_\beta(E_x) - \bar{E}_x = 0$, the root of this equation is found. In fact, with this method, we find several pairs, where the first element is the average energy and the second element is the corresponding root value in equation $r(\beta) = \boldsymbol{E}_\beta(E_x) - \bar{E}_x = 0$. With the help of regression models, we consider the average energy as an independent variable and the root value as a dependent variable and observe that these two variables fit a line. Therefore, using the regression model, the root of equation $r(\beta) = \boldsymbol{E}_\beta(E_x) - \bar{E}_x = 0$, which is the same as MME and MLE for β is found. Let the sample mean be $\bar{E}_x = 5$. Now, we want to compare the function $r(\beta) = \boldsymbol{E}_\beta(E_x) - 5$ presented in Examples 3.5, 3.10, and 3.12. In Figure 39 three functions $r(\beta)$ are plotted for the 3×3 grid, horizontal and circular models. As we expect when the sample mean is 5, we expect the root for $r(\beta)$. The roots of $r(\beta)$ are $-0.0739403 \times 5 = -0.3697015$, $-0.146003 \times 5 = -0.730015$, $-0.125151 \times 5 = -0.625755$, respectively.

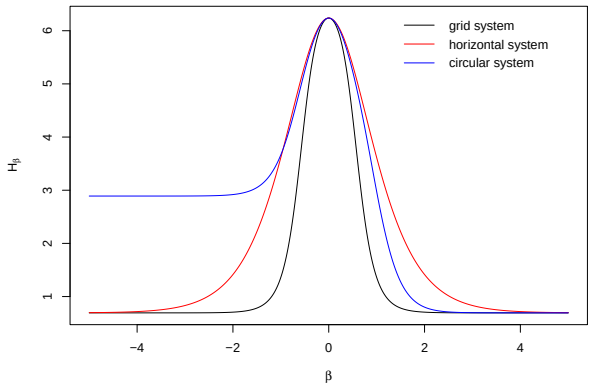


FIGURE 37. The entropy w.r.t β.

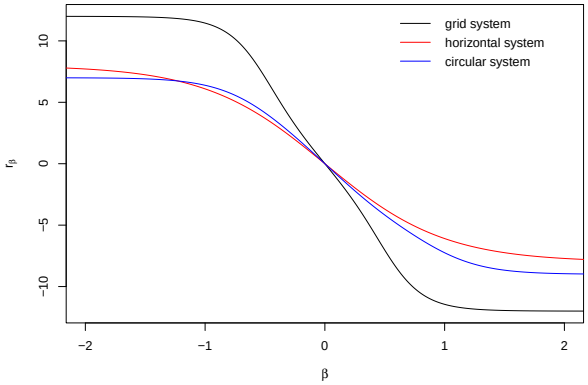


FIGURE 38. Comparison $r(\beta)$ in the 3×3 grid, horizontal and circular models for $\bar{E}_x = 0$.

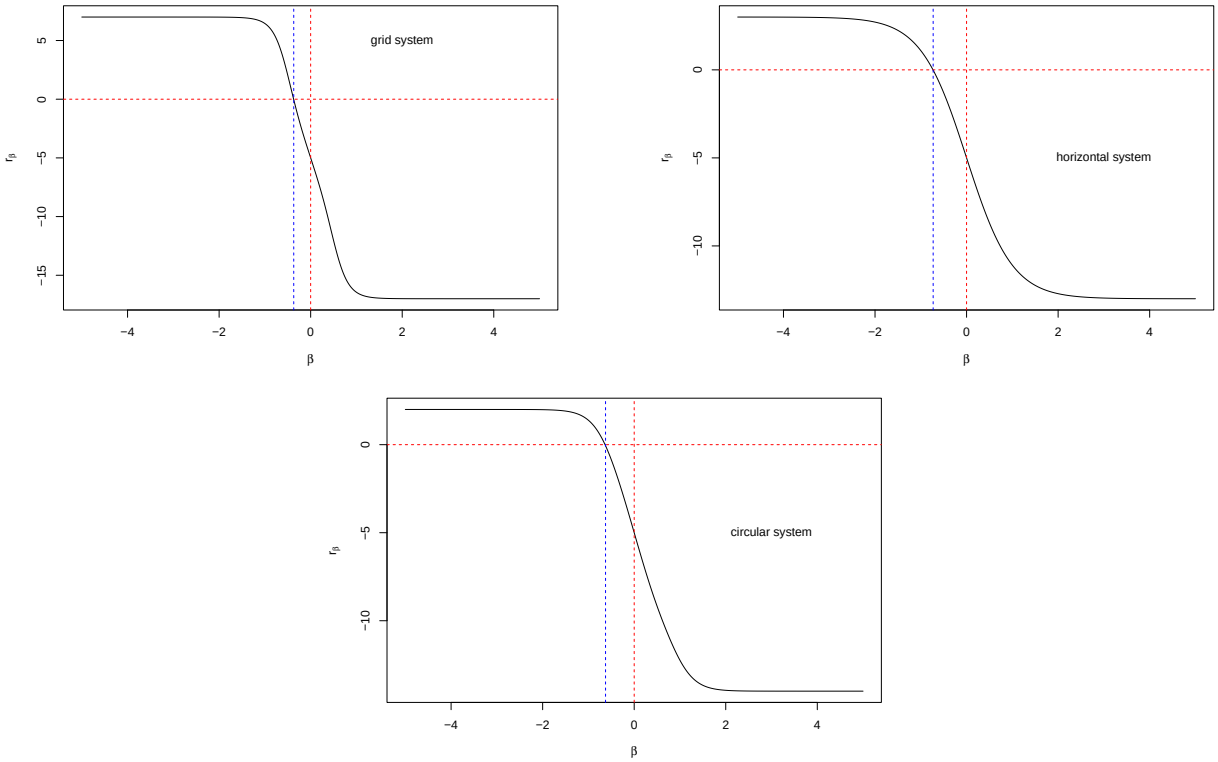


FIGURE 39. Finding the root for MLE in the 3×3 grid, horizontal and circular models.

TABLE 8. Summary of statistical analysis to estimate linear regression coefficients.

	grid	**horizontal**	**circular**
estimate	-0.0739403	-0.146003	-0.125151
std.error	0.0006454	0.002061	0.001534
t_{value}	-114.6	-70.84	-81.61
p_{value}	<0.0001	<0.0001	<0.0001

From Figure 39, we understand when we collect i.i.d. sample from Boltzmann distribution and calculate the average of sample, we can find the root of $r(\beta)$ according to regression equation. You can see for other values of β, but don't forget by according the energy of systems, average of sample can't be arbitrary value. The statistical criteria are summarized in Table 8.

As can be seen from Figure 40, and Table 8, all models are good to explain the relation between root of $r(\beta)$ and the sample mean $\bar{E}_x$. The regression models are calculated as follows:

$$3\times 3 \text{ grid model}: \quad \hat{\beta}=\tilde{\beta}=-0.0739403\bar{E}_x,$$

$$\text{horizontal model}: \quad \hat{\beta}=\tilde{\beta}=-0.146003\bar{E}_x,$$

$$\text{circular model}: \quad \hat{\beta}=\tilde{\beta}=-0.125151\bar{E}_x.$$

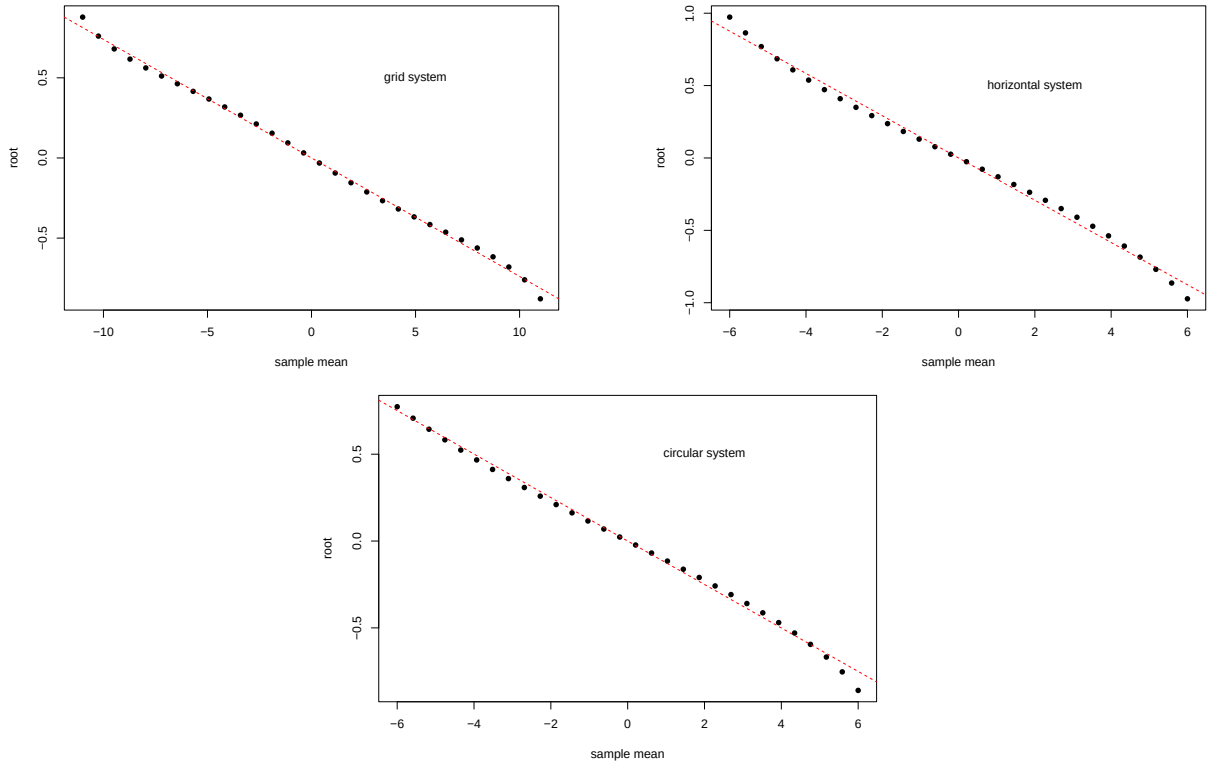


FIGURE 40. Linear regression models in the 3×3 grid, horizontal and circular models that shows the linear relationship between MLE (MME) and $\bar{E}_x$.

Criterion R^2 for the 3×3 grid, horizontal and circular models respectively are 0.9978, 0.9943, 0.9957. These quantities indicate the high degree of linearity of the models.

Finally, we want to calculate the Kullback-Leibler divergence between the two Boltzmann distributions given in Eq. (3.3) as a function of the parameter β. The first distribution $\boldsymbol{\xi}$ is the Boltzmann distribution for the 3×3 grid model and the second distribution $\boldsymbol{\xi'}$ is the Boltzmann distribution for the horizontal model. In other words $\xi_x = \frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}$ and $\xi'_x = \frac{e^{-\beta E'_x}}{\sum_{y\in E} e^{-\beta E'_y}}$ where E_x energies correspond to the 3×3 grid system and E'_x energies correspond to the horizontal model. Hence Kullback-Leibler divergence between $\boldsymbol{\xi}$ and $\boldsymbol{\xi'}$ is calculated as follows:

$$\begin{aligned} D_\beta(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}) &= \sum\nolimits_{x\in E} \frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}} \ln\left(\frac{\frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}}{\frac{e^{-\beta E'_x}}{\sum_{y\in E} e^{-\beta E'_y}}}\right) \\ &= -H_\beta(\boldsymbol{\xi}) - \sum\nolimits_{x\in E} \frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}} \ln\left(\frac{e^{-\beta E'_x}}{\sum_{y\in E} e^{-\beta E'_y}}.\right). \end{aligned}$$

Similarly

$$D_\beta(\boldsymbol{\xi'} \parallel \boldsymbol{\xi}) = -H_\beta(\boldsymbol{\xi'}) - \sum\nolimits_{x\in E} \frac{e^{-\beta E'_x}}{\sum_{y\in E} e^{-\beta E'_y}} \ln\left(\frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}\right).$$

The symmetric Kullback-Leibler divergence is defined as follows $\bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'}) = \frac{D_\beta(\boldsymbol{\xi}\|\boldsymbol{\xi'}) + D_\beta(\boldsymbol{\xi'}\|\boldsymbol{\xi})}{2}$. Therefore

$$\begin{aligned} \bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'}) = &-\frac{H_\beta(\boldsymbol{\xi}) + H_\beta(\boldsymbol{\xi'})}{2} - \frac{1}{2}\Big[\sum\nolimits_{x\in E} \frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}} \ln\Big(\frac{e^{-\beta E'_x}}{\sum_{y\in E} e^{-\beta E'_y}}\Big) \\ &+ \sum\nolimits_{x\in E} \frac{e^{-\beta E'_x}}{\sum_{y\in E} e^{-\beta E'_y}} \ln\Big(\frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}}\Big)\Big]. \end{aligned}$$

By the above formulas or using package 'seewave' with function kl.dist, we call plot Kullback-Leibler functions based on β as follows. The output of the **R** codes for Kullback-Leibler divergence between the two Boltzmann distributions in the 3×3 grid and horizontal systems is shown in Figure 41(a).

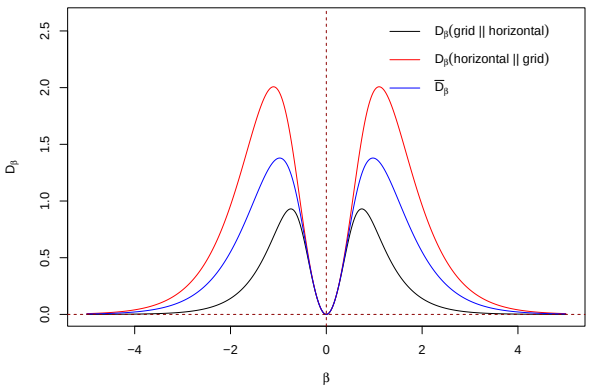
In Figure 41(a), we obtain minimum values of Kullback-Leibler function in $\beta = 0$, that means

$$\begin{aligned} &\arg\min_\beta D_\beta(\boldsymbol{\xi'} \parallel \boldsymbol{\xi}) = \arg\min_\beta D_\beta(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}) = \arg\min_\beta \bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'}) = 0, \\ &D_0(\boldsymbol{\xi'} \parallel \boldsymbol{\xi}) = D_0(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}) = \bar{D}_0(\boldsymbol{\xi}, \boldsymbol{\xi'}) = 0. \end{aligned}$$

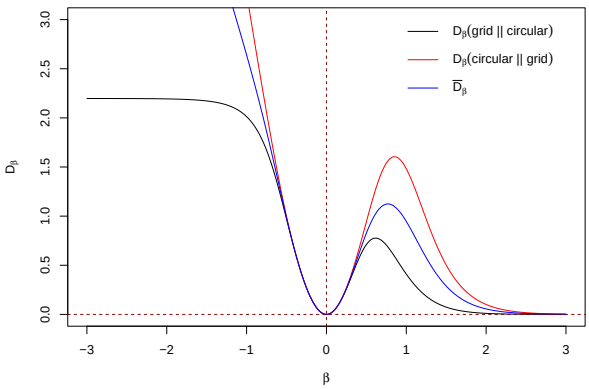
It means in $\beta = 0$, the 3×3 grid and horizontal systems are identically distributed. All functions $D_\beta(\boldsymbol{\xi'} \parallel \boldsymbol{\xi})$, $D_\beta(\boldsymbol{\xi} \parallel \boldsymbol{\xi'})$ and $\bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'})$ are symmetric and bimodal. For example. $\arg\max_\beta \bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'}) = \{-0.973389734,\ 0.973389734\}$, Hence $\bar{D}_{-0.973389734}(\boldsymbol{\xi}, \boldsymbol{\xi'}) = \bar{D}_{0.973389734}(\boldsymbol{\xi}, \boldsymbol{\xi'}) = 1.37904349$.

Also, $Kl_1(\beta) = D_\beta(\boldsymbol{\xi'} \parallel \boldsymbol{\xi})$, $Kl_2(\beta) = D_\beta(\boldsymbol{\xi} \parallel \boldsymbol{\xi'})$ and $Kl_3(\beta) = \bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'})$ are symmetric about $\beta = 0$ (also see Figure 41(a)), because it was shown in this section, the entropy function $h(\beta) = H_\beta(\boldsymbol{\xi})$ in the 3×3 grid and horizontal models is symmetric about $\beta = 0$, and so

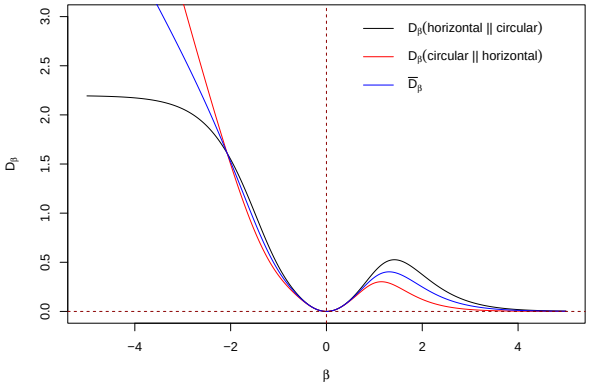
$$\begin{aligned} Kl_1(-\beta) = D_{-\beta}(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}) &= -H_{-\beta}(\boldsymbol{\xi}) - \sum\nolimits_{x\in E} \frac{e^{\beta E_x}}{\sum_{y\in E} e^{\beta E_y}} \ln\left(\frac{e^{\beta E'_x}}{\sum_{y\in E} e^{\beta E'_y}}\right) \\ &= -h(-\beta) - \sum\nolimits_{x\in E} \frac{e^{-\beta(-E_x)}}{\sum_{y\in E} e^{-\beta E_y}} \ln\left(\frac{e^{-\beta(-E'_x)}}{\sum_{y\in E} e^{-\beta E'_y}}\right) \\ &= -h(\beta) - \sum\nolimits_{x\in E} \frac{e^{-\beta E_x}}{\sum_{y\in E} e^{-\beta E_y}} \ln\left(\frac{e^{-\beta E'_x}}{\sum_{y\in E} e^{-\beta E'_y}}\right) \end{aligned}$$



(a) $\boldsymbol{\xi}$ and $\boldsymbol{\xi'}$ are the Boltzmann distribution corresponding to the 3×3 grid and horizontal models, respectively



(b) $\boldsymbol{\xi}$ and $\boldsymbol{\xi'}$ are the Boltzmann distribution corresponding to the 3×3 grid and circular models, respectively.



(c) $\boldsymbol{\xi}$ and $\boldsymbol{\xi'}$ are the Boltzmann distribution corresponding to the horizontal and circular models, respectively.

FIGURE 41. Comparison $D_\beta(\boldsymbol{\xi'} \parallel \boldsymbol{\xi})$, $D_\beta(\boldsymbol{\xi} \parallel \boldsymbol{\xi'})$ and $\bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'})$ based on different β values.

$$= D_\beta(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}) = Kl_1(\beta).$$

Hence $Kl_1(\beta) = D_\beta(\boldsymbol{\xi'} \parallel \boldsymbol{\xi})$ is an even function. Similarly, $Kl_2(\beta) = D_\beta(\boldsymbol{\xi} \parallel \boldsymbol{\xi'})$ is even. Since $Kl_3(\beta) = \frac{Kl_1(\beta)+Kl_2(\beta)}{2}$ is a convex linear combination of even functions $Kl_1(\beta)$ and $Kl_2(\beta)$, it is easy to conclude that $Kl_3(\beta) = \bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'})$ is also even.

The Kullback-Leibler divergence between the two Boltzmann distributions in the 3×3 grid and circular systems are shown in Figure 41(b).

In Figure 41(b), We get the same result as the previous case, i.e.,

$$\arg\min_\beta D_\beta(\boldsymbol{\xi'} \parallel \boldsymbol{\xi}) = \arg\min_\beta D_\beta(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}) = \arg\min_\beta \bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'}) = 0,$$
$$D_0(\boldsymbol{\xi'} \parallel \boldsymbol{\xi}) = D_0(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}) = \bar{D}_0(\boldsymbol{\xi}, \boldsymbol{\xi'}) = 0.$$

It means in $\beta = 0$, the 3×3 grid and circular systems are identically distributed. All functions $D_\beta\left(\boldsymbol{\xi'} \parallel \boldsymbol{\xi}\right)$, $D_\beta\left(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}\right)$ and $\bar{D}_\beta\left(\boldsymbol{\xi}, \boldsymbol{\xi'}\right)$ are asymmetric and we have $\arg\max_{\beta>0}\bar{D}_\beta\left(\boldsymbol{\xi}, \boldsymbol{\xi'}\right) = 0.769676968$ and $\bar{D}_{0.769676968}\left(\boldsymbol{\xi}, \boldsymbol{\xi'}\right) = 1.1239533$.

In Figure 41(c), We get the same result as the previous case, i.e.,

$$\arg\min_\beta D_\beta(\boldsymbol{\xi'} \parallel \boldsymbol{\xi}) = \arg\min_\beta D_\beta(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}) = \arg\min_\beta \bar{D}_\beta(\boldsymbol{\xi}, \boldsymbol{\xi'}) = 0,$$
$$D_0(\boldsymbol{\xi'} \parallel \boldsymbol{\xi}) = D_0(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}) = \bar{D}_0(\boldsymbol{\xi}, \boldsymbol{\xi'}) = 0.$$

It means in $\beta = 0$, the horizontal and circular systems are identically distributed. All functions $D_\beta\left(\boldsymbol{\xi'} \parallel \boldsymbol{\xi}\right)$, $D_\beta\left(\boldsymbol{\xi} \parallel \boldsymbol{\xi'}\right)$ and $\bar{D}_\beta\left(\boldsymbol{\xi}, \boldsymbol{\xi'}\right)$ are asymmetric and we have $\arg\max_{\beta>0}\bar{D}_\beta\left(\boldsymbol{\xi}, \boldsymbol{\xi'}\right) = 1.30813081$ and $\bar{D}_{1.308130818}\left(\boldsymbol{\xi}, \boldsymbol{\xi'}\right) = 0.402840491$.

Figure 41 shows that the necessary and sufficient condition for two distributions $\boldsymbol{\xi}$ and $\boldsymbol{\xi'}$ to be the same distribution is that $\beta = 0$.

6. Conclusions and Recommendations

Probabilistic models occur in many natural phenomena. One of the ways to model these phenomena is to use a stochastic process, especially the time-continuous jump process. Meanwhile, the Poisson process, the birth and death process, and the Markov process are mostly used to model natural phenomena. A method to simulate MCMC is Metropolis–Hastings and Gibbs sampling, in order to simulate a stochastic process, especially in the Ising model, can be a solution. In this paper, at first introduce the time-continuous stochastic process, the Poisson process, and the Chapman-Kolmogorov, simulated the birth and death process by **R** software with some practical problems. After that, the master equations were presented and solved by the Kolmogorov forward equation. In the Ising model, the limiting and stationary distribution were found by using the birth and death process. Also, we presented Metropolis-Hastings algorithm, Gibbs sampler to simulate for the Ising model. Besides those, we see direct sampling in **R** software. At the end, we simulate the Potts model. We advise readers that to use Random graphs for simulating the Potts model, and use the Bayesian inference to simulate both of them.

The limitations that prevented some statistical inferences from being applied in this paper were experimental limitations. Therefore, if researchers can overcome these limitations, using appropriate data for the energy of states, statistical inferences of parameter β such as confidence intervals, statistical hypothesis tests, uniformly minimum variance unbiased estimator (UMVUE), minimum risk equivariant estimator (MRE) etc. can also be studied. Of course, in this paper, to find the MLE of β, we came to solve an equation that forced us to use statistical methods, regression models, and find the root of the equation that leads to finding the MLE. By removing the experimental limitations and having appropriate data on the energy of the states, this problem is solved and the MLE can be easily found.

The Boltzmann probability mass function is a stationary distribution that was studied for some concepts such as Kolmogorov forward equation and also the $n \times n$ grid, horizontal, and circular systems. Apart from this distribution, there are other probability mass functions that can be used to describe the states of physical particles such as atoms or molecules, or biological systems. Researchers can use these distributions for the purposes of this paper, and they can even benefit from other well-known nonstationary distribution in physics. Also, with appropriate data for the energy of these states, they can make statistical inferences about the parameters of the model. By considering the Ising model on n with different patterns where points take values independently and follow a Markov chain, researchers can calculate the stationary distribution as well as the Boltzmann probability function, its induced probability function, and the maximum likelihood estimator for the parametrs.

References

[1] N. Abbas, K. U. Rehman, W. Shatanawi, and A. A. Al-Eid, *Theoretical study of non-Newtonian micropolar nanofluid flow over an exponentially stretching surface with free stream velocity*, Advances in Mechanical Engineering, *14*(7) (2022), 1-9.

[2] N. Abbas, W. Shatanawi, K. U. Rehman, and T. A. M. Shatnawi, *Velocity and thermal slips impact on boundary layer flow of micropolar nanofluid over a vertical nonlinear stretched Riga sheet*, Nanoengineering and Nanosystems, *238*(3-4) (2024), 107-117.

[3] E. Aurell, G. Del Ferraro, E. Domínguez and R. Mulet, *Cavity master equation for the continuous time dynamics of discrete-spin models*, Physical Review E, *95*(5) (2017), 052119.

[4] D. Beysens, *Brownian motion in strongly fluctuating liquid*, Thermodynamics of Interfaces and Fluid mechanics, *3* (2019), 1-8.

[5] K. Binder, D. Stauffer and H. Müller-Krumbhaar, *Calculation of Dynamic Critical Properties from a Cluster Reaction Theory*, Physical Review B, *10*(9) (1974), 3853.

[6] M. J. Catanzaro, V. Y. Chernyak, and J. R. Klein, *A higher Boltzmann distribution*, Journal of Applied and Computational Topology, *1*(2) (2017), 215-240.

[7] E. Cinlar, *Introduction to stochastic processes*, Dover Publications, 2013.

[8] D. Dhar, *Stochastic evolution in Ising models*, Stochastic Processes Formalism and Applications: Proceedings of the Winter School Held at the University of Hyderabad, India, Springer, (1982), 300-313.

[9] R. P. Dobrow, *Introduction to stochastic processes with R*, John Wiley & Sons, 2016.

[10] W. Ebeling and I. S. Sokolov, *Statistical Thermodynamics and Stochastic: Theory of Nonequilibrium Systems*, World Scientific; New Jersey, *34* (2005).

[11] P. A. Flinn, *Monte Carlo calculation of phase separation in a two-dimensional Ising system*, Journal of Statistical Physics, *10* (1974), 89-97.

[12] R. Folk, *The Survival of Ernst Ising and the Struggle to Solve His Model*, Chapter submitted to the book: Order, Disorder and Criticality: Advanced Problems of Phase Transition Theory. Ed. by Yu. Holovatch, World Scientific, Singapore, *7* (2023), 1-77.

[13] Z. Gao and X. Wang, *The Application of Ising Model in Simulating Neural Networks*, 2nd International Conference on Mechatronics, IoT and Industrial Informatics, Melbourne, Australia, (2024), 836-846.

[14] H. Ge and H. Qian, *Physical origins of entropy production, free energy dissipation, and their mathematical representations*, Physical Review E, *81*(5) (2010), 051133.

[15] G. H. Gilmer, *Ising Model Simulations of Crystal Growth*, In: Vanselow, R., Howe, R. (eds), Chemistry and Physics of Solid Surfaces V, (1984), 297-316.

[16] R. J. Glauber, *Time-Dependent Statistics of the Ising Model*, Journal of Mathematical Physics, *4* (1963), 294-307.

[17] A. Gut, *Probability: A Graduate Course 2nd edition* , Springer New York, 2014.

[18] G. Haag, *Derivation of the Chapman–Kolmogorov Equation and the Master Equation*, Modelling with the Master Equation, Springer International Publishing AG, (2017), 39-61.

[19] P. G. Hoel, S. C. Port, and C. J. Stone, *Introduction to Stochastic Processes*, Waveland Pr Inc, 1986.

[20] J. Honerkamp, *Statistical Physics. An Advanced Approach with Applications*, Springer, Berlin, 1998.

[21] H. Hooyberghs, S. Van Lombeek, C. Giuraniuc, B. Van Schaeybroeck, and J. O. Indekeu, *Ising model for distribution networks*, Philosophical Magazine, *92*(1-3) (2012), 168-191.

[22] A. Hosseiny, M. Bahrami, A. Palestrini, and M. Gallegat, *Metastable Features of Economic Networks and Responses to Exogenous Shocks*, PLoS ONE, *11*(e016036) (2016), 1-22.

[23] M. N. Khan, S. Nadeem, N. Abbas, and A. M. Zidan, *Heat and mass transfer investigation of a chemically reactive Burgers nanofluid with an induced magnetic field over an exponentially stretching surface*, Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering, SAGE Publications *235*(6) (2021).

[24] S. Kiesewetter and P. D. Drummond, *Coherent Ising machine with quantum feedback: The total and conditional master equation methods*, Physical Review A, *106*(2) (2022), 022409.

[25] S. Kullback and R. A. Leibler, *On Information and Sufficiency*, Annals of Mathematical Statistics, *22* (1951), 79-86.

[26] Z. Lei, *Irreversible Markov Chains for Particle Systems and Spin Models: Mixing and Dynamical Scaling. PhD thesis*, Université Paris sciences et lettres, (2018).

[27] S. Madhira and S. Deshmukh, *Introduction to Stochastic Processes Using R*, Springer, 2023.

[28] E.V. Murashkin and Y. N. Radayev, *Generalization of the algebraic Hamilton–Cayley theory*, Mechanics of Solids., *56*(6) (2021), 996-1003.
[29] S. Nadeem, A. Amin, and N. Abbas, *On the stagnation point flow of nanomaterial with base viscoelastic micropolar fluid over a stretching surface*, Alexandria Engineering Journal, *59*(9) (2020), 1751-1760.
[30] M. Niss, *History of the Lenz-Ising model 1920-1950: from ferromagnetic to cooperative phenomena*, Archive for History of Exact Sciences, *59*(3) (2005), 267-318.
[31] J. A. Pachter, Y. Jen. Yang, and K. A. Dill, *Entropy, irreversibility and inference at the foundations of statistical physics*, Nature Reviews Physics, (2024), 1-12.
[32] A. Papale and A. Rosa, *The Ising model in swollen vs. compact polymers: Mean-field approach and computer simulations*, The European Physical Journal E, *41*(12) (2018), 141.
[33] R. B. Potts, *Some generalized order-disorder transformations*, Mathematical Proceedings of the Cambridge Philosophical Society, *48*(1) (1952), 106-109.
[34] S. M. Ross, *Stochastic Processes 2nd edition*, Wiley, 1995.
[35] S. M. Ross and E. A. Peköz, *A second course in probability*, Cambridge University Press, 2023.
[36] B. Senoglu and B. Surucu, *Goodness-of-fit tests based on Kullback-Leibler information*, IEEE Transactions on Reliability, *53*(3) (2004), 357-361.
[37] C. A. Shannon, *Mathematical Theory of Communication*, Bell System Technical Journal, *27* (1948), 379-423.
[38] T. Shoji, K. Aihara, and Y. Yamamoto, *Quantum model for coherent Ising machines: Stochastic differential equations with replicator dynamics*, Physical Review A, *96*(5) (2017), 053833.
[39] S. P. Singh, *The Ising Model: Brief Introduction and Its Application*, In Solid State Physics—Metastable, Spintronics Materials and Mechanics of Deformable Bodies—Recent Progress; Sivasankaran, S., Nayak, P.K., Günay, E., Eds.; IntechOpen: London, UK, (2020).
[40] G. Slade, *Probabilistic Models of Critical Phenomena*, The Princeton Companion to Mathematics, Princeton University Press, (2008).
[41] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry 3rd edition*, North Holland Publ, Amsterdam, 2007.
[42] L. Zdeborová and F. Krzakala, *Statistical physics of inference: Thresholds and algorithms*, Advances in Physics, *65*(5) (2016), 453-552.