Entropy **2009**, *11*, 867-887; doi:10.3390/e11040867

OPEN ACCESS

entropy

ISSN 1099-4300

www.mdpi.com/journal/entropy

Article

Maximum Entropy Estimation of Transition Probabilities of Reversible Markov Chains

Erik Van der Straeten

Queen Mary University of London, School of Mathematical Sciences, Mile End Road, London E1 4NS, UK; E-Mail: e.straeten@qmul.ac.uk

Received: 21 September 2009 / Accepted: 10 November 2009 / Published: 17 November 2009

Abstract: In this paper, we develop a general theory for the estimation of the transition probabilities of reversible Markov chains using the maximum entropy principle. A broad range of physical models can be studied within this approach. We use one-dimensional classical spin systems to illustrate the theoretical ideas. The examples studied in this paper are: the Ising model, the Potts model and the Blume-Emery-Griffiths model.

Keywords: maximum entropy principle; Markov chain; parameter estimation; statistical mechanics; spin chain models; thermodynamics

1. Introduction

Usually, the only knowledge about a physical system is the measurement of the average values of a few relevant observables. The maximum entropy principle [1–3] is a tool to obtain the least biased distribution for the equilibrium distribution of the system that reproduces these measurements. The rough line of reasoning of this approach is as follows [4]. The Boltzmann-Gibbs entropy functional of a distribution $P(j)$ is defined by

$$S = -\sum_{j\in\Gamma} P(j) \ln P(j) \tag{1}$$

where Γ is the (discrete) phase space. This entropy functional is most often used in statistical mechanics. The equilibrium distribution is obtained by maximising the entropy functional under the constraints that the average values of the relevant observables $H_1(j), \ldots, H_t(j)$ take on certain values. To solve this optimisation problem, usually the method of Lagrange multipliers is used. For every constraint, a

Lagrange multiplier θ_i is introduced and the following function is maximised

$$\mathcal{L} = S - \sum_{i=1}^{t} \theta_i \sum_{j \in \Gamma} P(j) H_i(j) \tag{2}$$

to obtain the equilibrium distribution. After variation with respect to $P(j)$ one obtains

$$P(j) \quad \sim \quad \exp\left(- \sum_{i=1}^{t} \theta_i H_i(j) \right) \tag{3}$$

With this distribution, one can calculate expressions for the equilibrium values $\langle H_i(j) \rangle$ of the observables as a function of the Lagrange multipliers θ_i. These relations can then be used to estimate the values of θ_i after measurement of $\langle H_i(j) \rangle$. In this way, one obtains the least biased estimate of the parameters θ_i, because $P(j)$ satisfies the maximum entropy principle. The standard example of this procedure is the estimation of the temperature by the measurement of the energy.

The maximum entropy principle can be used to introduce thermodynamic parameters in simple theoretical models [5–7]. One starts from a mathematical model and calculates the entropy and the average of the relevant observables as a function of the model parameters. Then the maximisation procedure is carried out over the model parameters only, instead of over all the possible probability distributions. The usefulness of this approach is already shown for specific models containing 2 and 5 parameters, see [5, 6] and [7] respectively. In this paper we show that the maximisation procedure can be carried out for reversible N-state Markov chains. This problem of conditional optimisation is more general and is also studied in the information theory framework [8, 9]. However, we focus on the applications of this technique in the context of statistical mechanics and use the maximum entropy principle to relate microscopic and macroscopic quantities of physical models. To illustrate the theoretical ideas, one-dimensional classical spin systems are studied, the Ising model [10], the Potts model [11, 12] and the Blume-Emery-Griffiths model [13, 14]. These examples serve to show that our theoretical procedure is very general and that a broad range of physical models can be studied within this approach. It is not the aim of the present paper to make progress in the understanding of the aforementioned models.

The outline of the paper is as follows. In the next section we fix our notation and repeat briefly some results obtained in [15] that will be used throughout the paper. The basic idea of this work is introduced in Section 3. with the use of a simple example, the 2-state Markov chain. The main result of this paper is obtained in Section 4. in which we apply the maximum entropy principle to the reversible N-state Markov chain. In Section 5. we make the connection between our theory and thermodynamics and examine under which conditions our technique coincides with the standard approach to introduce thermodynamic parameters in theoretical models. In Section 6., the example of the 3-state Markov chain is thoroughly studied. The final section contains a summary of our results and a brief discussion of the different assumptions we made throughout the paper.

2. Notation

We consider a finite state space Γ with N states. A Markov chain is defined by initial probabilities $p(z)$ and transition probabilities $w(y,z)$, with $y,z \in \Gamma$. The equation of motion is simply

$$p_{t+1}(z) = \sum_{y\in\Gamma} p_t(y)w(y,z), \quad \text{with} \quad p_0(y) = p(y) \tag{4}$$

Throughout the paper, we will assume that $w(z,y) \neq 0$ for all $z,y \in \Gamma$. This means that we study irreducible Markov chains [16]. The N initial probabilities $p(z)$ and the N^2 transition probabilities $w(z,y)$ can be interpreted as the parameters of the Markovian model. However, they are not independent because of the normalisation conditions

$$1 = \sum_{y\in\Gamma} p(y), \tag{5}$$

$$1 = \sum_{y\in\Gamma} w(z,y), \quad \forall z \in \Gamma \tag{6}$$

As a consequence, the Markovian model contains only $(N-1)+(N^2-N) = N^2-1$ independent parameters. Usually, extra restrictions on the parameters are assumed. A Markov chain is called stationary [16], when the following condition holds

$$p(z) = \sum_{y\in\Gamma} p(y)w(y,z), \qquad \forall z \in \Gamma \tag{7}$$

This is a set of $N-1$ extra equations (the normalisation is already taken into account). A stronger constraint is detailed balance

$$p(z)w(z,y) = p(y)w(y,z), \qquad \forall z,y \in \Gamma \tag{8}$$

This is a set of $N(N-1)/2$ extra equations. Throughout the paper, we will assume that this condition is satisfied. This means that we study reversible Markov chains [16].

A path $\gamma = (x_0, x_1, x_2 \ldots, x_n)$ of the Markov chain with length $n+1$ has probability

$$p(x_0)w(x_0,x_1)\ldots w(x_{n-1},x_n) \tag{9}$$

In [15] the record of transitions k is introduced. This is a sequence of numbers $k(z,y)$, one for each pair of states $z,y \in \Gamma$, counting how many times the transition from z to y is contained in a given path of the Markov chain. The ensemble average of the elements of the transition records and the entropy S [17, 18] of the Markov chain can be calculated as follows:

$$\begin{aligned}
\langle k(z,y)\rangle &= \sum_{x_0\in\Gamma}\ldots\sum_{x_n\in\Gamma} p(x_0)w(x_0,x_1)\ldots w(x_{n-1},x_n)\left[\sum_{i=0}^{n-1}\delta_{x_i,z}\delta_{x_{i+1},y}\right] \\
S &= -\sum_{x_0\in\Gamma}\ldots\sum_{x_n\in\Gamma} p(x_0)w(x_0,x_1)\ldots w(x_{n-1},x_n)\ln\left[p(x_0)w(x_0,x_1)\ldots w(x_{n-1},x_n)\right]
\end{aligned}$$

with $\delta_{i,j}$ the Kronecker delta. For stationary Markov chains, these expressions simplify to [15]

$$\frac{1}{n}\langle k(z,y)\rangle = p(z)w(z,y) \tag{10}$$

$$\frac{1}{n}S = -\sum_{z\in\Gamma}\sum_{y\in\Gamma} p(z)w(z,y)\ln w(z,y) - \frac{1}{n}\sum_{z\in\Gamma} p(z)\ln p(z) \tag{11}$$

The second term in the expression of the entropy is unimportant for large chains and will be ignored in the remaining part of this paper. The technical consequences of taking these finite size effects into account are already thoroughly studied for the 2-state Markov chain [5, 19].

The conditional probability $P(k; x_0)$ to observe a Markov chain with certain transition record k given the initial condition $x_0 \in \Gamma$ is

$$P(k; x_0) \quad = \quad c(k; x_0) \prod_{x\in\Gamma} \prod_{y\in\Gamma} w(x, y)^{k(x,y)} = c(k; x_0) \exp\left(\sum_{x\in\Gamma} \sum_{y\in\Gamma} k(x, y) \ln w(x, y) \right) \tag{12}$$

where the prior probability $c(k; x_0)$ counts the number of paths that have the same transition record k. Notice that the value of $c(k; x_0)$ can vanish. An obvious example is a combination of an initial condition x_0 with a transition record in which no transition $x_0 \to x$ with $x \in \Gamma$ occurs. However, this is not the only possibility to obtain $c(k; x_0) = 0$. To see this, observe that there are two ways to count the number of occurrences of a state $x \in \Gamma$ given k and $x_0 \in \Gamma$

$$\delta_{x,x_0} + \sum_{y\in\Gamma} k_{y,x} \quad \text{and} \quad \delta_{x,x_n} + \sum_{y\in\Gamma} k_{x,y} \tag{13}$$

where the Kronecker deltas δ_{x,x_0} and δ_{x,x_n} take into account the first and last state of the path respectively. Given k and $x_0 \in \Gamma$, only when following equality

$$\delta_{x,x_0} + \sum_{y\in\Gamma} k_{y,x} = \delta_{x,x_n} + \sum_{y\in\Gamma} k_{x,y}, \qquad \forall x \in \Gamma \tag{14}$$

is fulfilled, one ends up with a unique value for the number of occurrences of every state $x \in \Gamma$. Therefore, expression (14), is a necessary condition in order to obtain a non-vanishing value for $c(k; x_0)$. This shows that the elements of the transition record are not independent. In Section 5., the importance of this observation will become clear.

A Markov chain can be interpreted as a sequence of letters where the transition record k counts the number of occurrences of two-letter words. Markov chains with a finite memory and generalisations of k are examined in the information theory framework [20, 21] and find applications in, e.g., the computational biology [22, 23]. In the present paper we study Markov chains in the context of statistical mechanics and apply our results to physical models with only nearest neighbor interactions. The notion of two-letter words is sufficient for these applications. The extension of our theoretical results to systems with next (or higher order) nearest neighbor interactions is merely technical and can be obtained by increasing the number of states of the chain which allows to maintain the Markov property.

3. Example: The 2-state Markov Chain

In this section we study a simple example, the 2-state Markov chain. The two states are denoted $+$ and $-$. The different parameters of the Markovian model are

$$p(+), p(-) \quad \text{and} \quad w(+, +), w(+, -), w(-, -), w(-, +) \tag{15}$$

However, the number of independent parameters is reduced by 3 because of the normalisation conditions (5), (6). The detailed balance condition (8) further reduces this number by 1. We conclude

that this simple microscopic model contains only 2 independent parameters. We chose $w(+,-)$ and $w(-,+)$ to be these parameters and use (5), (6) and (8) to obtain the following relations

$$\begin{aligned} p(+) &= \frac{w(-,+)}{w(-,+)+w(+,-)}, \qquad p(-) = \frac{w(+,-)}{w(-,+)+w(+,-)} \\ w(+,+) &= 1-w(+,-), \qquad w(-,-) = 1-w(-,+) \end{aligned} \tag{16}$$

In Section 2. we introduced the transition record k. The matrix k contains only 4 elements for this example

$$k = \begin{bmatrix} k(+,+) & k(+,-) \\ k(-,+) & k(-,-) \end{bmatrix} \tag{17}$$

The 2-state Markov chain can be interpreted as a one-dimensional Ising chain [10] with constant length $n+1$ and two different spin-values ± 1. Two relevant observables are

$$H_1(\sigma) = -J\sum_{i=0}^{n-1}\sigma_i\sigma_{i+1} \quad \text{and} \quad H_2(\sigma) = \sum_{i=0}^{n}\sigma_i \tag{18}$$

with J a constant. The spin variables σ_i are scalars that can take on the values ± 1. The two states of the Markov chain $+,-$ correspond to the spin values $+1,-1$ respectively. Therefore, one can express $H_1(\sigma)$ and $H_2(\sigma)$ as a function of the elements of the transition record k as follows:

$$H_1(k) = -J[k(+,+)+k(-,-)] + J[k(+,-)+k(-,+)] \tag{19}$$

$$H_2(k) = k(+,+)+k(-,+)-k(-,-)-k(+,-) \tag{20}$$

The correspondence between $H_1(\sigma)$ and $H_1(k)$ is exact, while we ignored the contribution of the initial spin σ_0 to obtain $H_2(k)$ from $H_2(\sigma)$. This is only a finite size effect that is unimportant for large chains. This means that the correspondence between $H_2(\sigma)$ and $H_2(k)$ is also exact for infinite chains. With (10) and (16), one can immediately write out the ensemble averages of these variables as a function of the independent parameters $w(+,-)$ and $w(-,+)$

$$\begin{aligned} \frac{\langle H_1(k)\rangle}{Jn} &= \frac{4w(-,+)w(+,-)-w(-,+)-w(+,-)}{w(-,+)+w(+,-)} \\ \frac{\langle H_2(k)\rangle}{n} &= \frac{w(-,+)-w(+,-)}{w(-,+)+w(+,-)} \end{aligned} \tag{21}$$

Also the entropy (11) of the Markov chain can be expressed as a function of $w(+,-)$ and $w(-,+)$ only

$$\begin{aligned} \frac{S}{n} = & -\frac{w(-,+)}{w(-,+)+w(+,-)}\big([1-w(+,-)]\ln[1-w(+,-)]+w(+,-)\ln w(+,-)\big) \\ & -\frac{w(+,-)}{w(-,+)+w(+,-)}\big(w(-,+)\ln w(-,+)+[1-w(-,+)]\ln[1-w(-,+)]\big) \end{aligned} \tag{22}$$

We use $\langle H_1(k)\rangle$ and $\langle H_2(k)\rangle$ as constraints in the maximisation procedure (2)

$$\mathcal{L} = S - \theta_1\langle H_1(k)\rangle - \theta_2\langle H_2(k)\rangle \tag{23}$$

By solving the following set of equations

$$\frac{\partial \mathcal{L}}{\partial w(+,-)} = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial w(-,+)} = 0 \tag{24}$$

one can express θ_1 and θ_2 as a function of the microscopic parameters

$$4J\theta_1 = \ln \frac{1-w(+,-)}{w(+,-)} \frac{1-w(-,+)}{w(-,+)} \quad \text{and} \quad 2\theta_2 = \ln \frac{1-w(-,+)}{1-w(+,-)} \tag{25}$$

By inverting (25), one gets expressions for $w(+,-)$ and $w(-,+)$ as a function of θ_1 and θ_2

$$\begin{aligned} 1-w(-,+) &= e^{2\theta_2}[1-w(+,-)] \\ 1-w(+,-) &= \left(\cosh(\theta_2) - \sqrt{\sinh^2(\theta_2) + e^{-4J\theta_1}}\right)\left(1-e^{-4J\theta_1}\right)^{-1} e^{-\theta_2} \end{aligned} \tag{26}$$

In combination with (21), one finally obtains formulas for $\langle H_1(k)\rangle$ and $\langle H_2(k)\rangle$ as a function of θ_1 and θ_2. These relations can then be used to estimate the values of θ_1 and θ_2 after measurement of $\langle H_1(k)\rangle$ and $\langle H_2(k)\rangle$.

4. General Theory

Our microscopic model is the N-state Markov chain with parameters $p(z)$ and $w(z,y)$. The only constraints on these microscopic parameters are the normalisation conditions (5), (6) and the detailed balance conditions (8). To proceed from this mathematical model to a physical model one has to make a choice for the relevant observables $H_i(k)$. Then one can introduce Lagrange multipliers $\{\theta_i, \alpha, \zeta(z), \eta(z,y)\}$ and maximise the following function

$$\begin{aligned} \frac{1}{n}\mathcal{L} &= \frac{1}{n}S - \frac{1}{n}\sum_{i=1}^{t}\theta_i\langle H_i(k)\rangle - \alpha\sum_{z\in\Gamma}p(z) - \sum_{z\in\Gamma}\zeta(z)\sum_{y\in\Gamma}w(z,y) \\ &\quad - \sum_{z\in\Gamma}\sum_{y\in\Gamma, y>z}\eta(z,y)\left[p(z)w(z,y) - p(y)w(y,z)\right] \end{aligned} \tag{27}$$

over the parameters $p(z)$ and $w(z,y)$. Notice the fundamental difference between the constraints that are taken into account by the Lagrange multipliers $\{\theta_i\}$ and $\{\alpha, \zeta(z), \eta(z,y)\}$. The former should be determined as a function of the model parameters $p(z)$ and $w(z,y)$. The latter are mathematical tools to take into account some basic microscopic constraints. These multipliers should be eliminated out of the theory since they are not connected to macroscopic observables.

Before (27) can be maximised over $p(z)$ and $w(z,y)$, the parameter dependence of S and $\langle H_i(k)\rangle$ must be know. We already obtained a formula for the entropy as a function of $p(z)$ and $w(z,y)$ only, see expression (11). In this paper, we assume that the observables $H_i(k)$ are linear combinations of the elements of the transition record of the Markov chain

$$\sum_{i=1}^{t}\theta_i H_i(k) = \sum_{z\in\Gamma}\sum_{y\in\Gamma}\Theta(z,y)k(z,y) \tag{28}$$

where the elements of the matrix Θ are some linear combination of the Lagrange multipliers θ_i. Taking the ensemble average of (28) and using expression (10) results in

$$\frac{1}{n}\sum_{i=1}^{t}\theta_i\langle H_i(k)\rangle = \sum_{z\in\Gamma}\sum_{y\in\Gamma}\Theta(z,y)p(z)w(z,y) \tag{29}$$

In Appendix 1, the optimisation of the function (27) is carried out analytically. One ends up with the following set of equations

$$\begin{aligned} \ln \frac{w(x,y)}{w(x,x)} \frac{w(y,x)}{w(y,y)} &= \Theta(x,x) + \Theta(y,y) - \Theta(x,y) - \Theta(y,x) \\ \ln \frac{w(y,y)}{w(x,x)} &= \Theta(x,x) - \Theta(y,y) \end{aligned} \tag{30}$$

for all $x, y \in \Gamma$. These $N(N-1)/2 + N - 1 = (N-1)(N+2)/2$ equations together with the $N(N-1)/2$ detailed balance conditions (8) and the $1+N$ normalisation conditions (5), (6) are a closed set of equations for the $N + N^2$ microscopic parameters $p(x)$ and $w(x,y)$. To obtain relations for $p(x)$ and $w(x,y)$ as a function of the parameters θ_i (contained in the elements of the matrix Θ), one has to invert this set of equations. A part of this inversion can be performed generally. Start by choosing an arbitrary state r and rewrite the relations (8) and (5) as follows:

$$p(r) = \left(1 + \sum_{y \in \Gamma'} \frac{w(r,y)}{w(y,r)}\right)^{-1}, \qquad p(x) = p(r)\frac{w(r,x)}{w(x,r)}, \quad \forall x \in \Gamma' \tag{31}$$

with $\Gamma' = \Gamma \backslash \{r\}$. Then, the remaining detailed balance conditions (8) can be rewritten as follows:

$$w(r,x)w(x,y)w(y,r) = w(x,r)w(y,x)w(r,y), \qquad \forall x, y \in \Gamma' \tag{32}$$

Notice that (31) expresses the probabilities $p(x)$ as a function of the transition probabilities $w(x,y)$ only. Therefore to obtain relations for $p(x)$ and $w(x,y)$ as a function of the parameters θ_i one only has to invert the relations (30), (32) together with the normalisation conditions (6). This part of the inversion will depend on the particular form of the matrix Θ and has to be performed for every physical model individually.

We want to emphasise that our procedure fits in the estimation theory [24, 25]. In that approach, the average values of some observables are used to estimate the values of the model parameters. In the present paper, we make a separation between the physical model of a theory and the underlying mathematical model. The latter is the N-state Markov chain while the former model is introduced by identifying some physically relevant observables. Usually, the number of microscopic parameters $p(x)$ and $w(x,y)$ of the mathematical model is larger than the number of relevant observables $H_i(k)$ of the physical model. By measuring $\langle H_i(k) \rangle$, *only* the values of the corresponding parameters θ_i can be estimated. Then, one can calculate an estimation of *all* the values of $p(x)$ and $w(x,y)$ with the formulas obtained in this section. As such, no a priori choice for these parameters is necessary and one obtains the least biased values for $p(x)$ and $w(x,y)$ given only the measured information and some basic microscopic constraints.

5. Thermodynamics

In statistical mechanics, the starting point to describe a given model is usually the Hamiltonian $H(j)$ with $j \in \Gamma$ and Γ the phase space. Then, the standard way of introducing the temperature T is by postulating the Boltzmann-Gibbs form $\exp[-H(j)/T]$ for the equilibrium distribution. This approach is motivated by the maximum entropy principle that we already outlined in the introduction.

Indeed, when the Hamiltonian is identified as the only relevant observable, expression (3) for the equilibrium distribution simplifies to $\exp\left[-\theta_1 H(j)\right]$. Notice that this corresponds to the choice $t = 1$ and $H_1(j) = H(j)$. Using the laws of thermodynamics, one can show that θ_1 is indeed the inverse temperature.

In previous sections, we used the maximum entropy principle to obtain the least biased values of the microscopic parameters of a mathematical model gives some macroscopic constraints. Since this problem fits in the estimation theory, we did not give a thermodynamic interpretation of the Lagrange multipliers θ_i. However, such a deeper understanding is highly desirable for the application of our theory to physical models like, e.g., the Ising chain. Therefore, in this section we study this thermodynamic interpretation in more detail. We first outline briefly the concept of exponential families which is very important in this context. Then we illustrate the relation between the Lagrange multipliers θ_i and the temperature for the 2-state Markov chain. Finally, we show under which conditions our technique coincides with the standard approach to introduce thermodynamic parameters into theoretical models.

5.1. Curved exponential family

A distribution with parameters $w = [w_1, \ldots, w_s]$ belongs to the t-parameter exponential family when it can be written as follows:

$$P(j) = c(j) \exp\left(G(w) - \sum_{i=1}^{t} \theta_i(w) g_i(j)\right) \tag{33}$$

where t is the smallest integer for which the exponential form can be obtained, $c(j)$ is a prior probability and $G(w)$ is determined by the normalisation condition

$$G(w) = -\ln\left(\sum_j c(j) \exp\left(-\sum_{i=1}^{t} \theta_i(w) g_i(j)\right)\right) \tag{34}$$

The family is said to be curved when $t > s$ [26] (s is the dimension of the parameter vector w, see above). The special case for which $s = t$ is called a full exponential family. Then, one can interpret the functions $\theta_i(w)$ as the new parameters of the distribution $\theta = [\theta_1, \ldots, \theta_t]$. The Boltzmann-Gibbs form is obtained when also the functions $g_i(j)$ coincide with the relevant observables $H_i(j)$

$$P(j) = c(j) \exp\left(G(\theta) - \sum_{i=1}^{t} \theta_i H_i(j)\right) \tag{35}$$

In the next section, we will show that the subtle differences between the curved and the full exponential family are very important for the thermodynamic interpretation of the parameters θ. A generalisation of the concept of exponential families with applications in the context of nonextensive statistical mechanics is proposed by Naudts [25, 27, 28].

5.2. Example: the 2-state Markov chain

We studied the 2-state Markov chain already in Section 3. and interpreted this model as a one-dimensional Ising chain. We identified two relevant observables $H_1(k)$ and $H_2(k)$, see

expressions (19) and (20) respectively, and used $\langle H_1(k)\rangle$ and $\langle H_2(k)\rangle$ as constraints in the maximisation procedure. As a consequence, the matrix Θ (28) becomes

$$\Theta = \begin{bmatrix} \Theta(1,1) & \Theta(1,2) \\ \Theta(2,1) & \Theta(2,2) \end{bmatrix} = \begin{bmatrix} -J\theta_1 + \theta_2 & J\theta_1 - \theta_2 \\ J\theta_1 + \theta_2 & -J\theta_1 - \theta_2 \end{bmatrix} \tag{36}$$

Using (30), the parameters θ_1 and θ_2 can then be expressed as a function of the microscopic parameters. It is easy to check that this procedure results in the same formulas for θ_1 and θ_2 that we obtained before (25), as it should be. Clearly, $\langle H_2(k)\rangle$ is just the magnetisation M of the chain, while $\langle H_1(k)\rangle$ is usually interpreted as the internal energy U of the one-dimensional Ising model. Within this interpretation, the parameters θ_1 and θ_2 can be related to the temperature T and an external applied field F as follows: $\theta_1 = 1/T$ and $\theta_2 = -F/T$. One can check this, e.g., by showing that the following thermodynamic relations [4] hold

$$\frac{\partial \beta G}{\partial \beta} = U - FM \quad \text{and} \quad \frac{\partial G}{\partial F} = -M \tag{37}$$

with G the free energy $G = U - FM - TS$. The final expression for the magnetisation as a function of T and F is

$$\frac{M}{n} \quad = \quad \sinh(F/T)\left(\sinh^2(F/T) + e^{-4J/T}\right)^{-1/2} \tag{38}$$

This is the well-known result for the one-dimensional Ising model [10]. We proceed by studying the equilibrium distribution that is obtained by our optimisation procedure in order to understand why our final formula for the magnetisation (38) coincides with the standard result. The probability $P(k)$ to observe a Markov chain with certain transition record k is proportional to

$$P(k) \quad \sim \quad w(+,+)^{k(+,+)} w(+,-)^{k(+,-)} w(-,-)^{k(-,-)} w(-,+)^{k(-,+)} \tag{39}$$

Now we want to express this probability as a function of the relevant variables $H_1(k)$, $H_2(k)$ and n. Expressions for $H_1(k)$ and $H_2(k)$ as a function of the elements of k are already given in (19) and (20). The length of the Ising chain $n+1$ is just the sum of all the elements of k plus 1, i.e.,

$$n = k(+,+) + k(+,-) + k(-,-) + k(-,+) \tag{40}$$

In this way, we obtain only 3 equations for 4 variables, the 4 elements of the transition record. However notice that the difference between the values of $k(-,+)$ and $k(+,-)$ can only be 0 or 1. Therefore in the thermodynamic limit we have an extra constraint for the elements of k

$$k(+,-) = k(-,+) \tag{41}$$

In this way we end up with a closed set of equations (19,20,40,41) for the 4 elements of k. By solving this set of equations for $H_1(k)$, $H_2(k)$ and n, we can rewrite expression (39) as follows:

$$P(k) \quad \sim \quad \exp[-\theta_1 H_1(k) - \theta_2 H_2(k)] \tag{42}$$

see (25) for the definitions of θ_1 and θ_2. We omitted the dependence of n to obtain (42), because the length of the Ising chain is assumed to be constant. Rewriting $P(k)$ in this form makes the

thermodynamic interpretation of the parameters θ_1 and θ_2 as $\theta_1 = 1/T$ and $\theta_2 = -F/T$ immediately clear because expression (42) is just the Boltzmann-Gibbs distribution $\exp[-\theta_1 H(k)]$, with $H(k) = H_1(k) + H_2(k)\theta_2/\theta_1$. This is indeed the Hamiltonian of the Ising chain. We conclude that for this simple example, our technique to introduce the thermodynamic temperature in a statistical model coincides with the standard approach. As a consequence, it is no surprise that our expression for the magnetisation (38) is equal to the well-known result for the one-dimensional Ising model.

Notice that ignoring equation (41) in this procedure results in an extra contribution to expression (42)

$$P(k) \quad \sim \quad \exp[-\theta_1 H_1(k) - \theta_2 H_2(k) - \theta_3 H_3(k)] \tag{43}$$

with

$$H_3(k) = k(+,-) - k(-,+) \quad \text{and} \quad \theta_3 = \frac{1}{2} \ln \frac{w(-,-)}{w(+,+)} \frac{w(-,+)}{w(+,-)} \tag{44}$$

Observe that the distribution (43) is a member of the curved exponential family. Indeed, the 2 independent parameters are $w = [w(+,-), w(-,+)]$. However, in order to rewrite $P(k)$ in an exponential form, one needs 3 functions $\theta = [\theta_1, \theta_2, \theta_3]$ of these parameters instead of 2. Therefore, the distribution (43) belongs to the curved exponential family while the distribution (42) is a member of the full exponential family. As a consequence, the interpretation of the parameters of the distribution (43) is not immediately clear. As we mentioned before, the difference between the values of $k(-,+)$ and $k(+,-)$ can only be 0 or 1. So for this particular example, the difference between the full and curved exponential family only occurs for finite systems. However, the problem is more general. In this paper, we study mathematical models with an arbitrary number of microscopic parameters. The number of physically relevant observables $H_i(k)$ is usually a lot smaller. Therefore, it is not obvious whether the distribution $P(k)$ belongs to the full or curved exponential family in the variables $H_i(k)$. Or equivalently, it is not obvious whether it is possible to rewrite the probability $P(k)$ in the Boltzmann-Gibbs form. We examine this question for the N-state Markov chain in the next section.

5.3. Boltzmann-Gibbs distribution

The conditional probability $P(k; x_0)$ to observe a Markov chain with certain transition record k given the initial condition $x_0 \in \Gamma$ can be written as (12). We also derived relations between the elements of the transition record (14) that must be satisfied in order to obtain a non-vanishing value for $P(k; x_0)$. The dependence of $P(k; x_0)$ on the initial condition x_0 is unimportant for large Markov chains. Therefore, we ignore the dependence of x_0 in the remaining part of this section and write the probability $P(k)$ to observe a Markov chain with certain transition record k as follows:

$$P(k) \quad = \quad c(k) \exp\left(\sum_{x \in \Gamma} \sum_{y \in \Gamma} k(x,y) \ln w(x,y) \right) := c(k) \exp\left(\Upsilon(k)\right) \tag{45}$$

When finite size effects are ignored, expression (14) simplifies to

$$\sum_{y \in \Gamma} k(r,y) = \sum_{y \in \Gamma} k(y,r), \qquad \forall r \in \Gamma \tag{46}$$

Notice that these relations are the generalisation of expression (41). In this paper we have assumed that the detailed balance conditions (8) hold and derived relations (30) between $w(x,y)$ using the maximum entropy principle. The aim of this section is to examine whether the conditions (46) together with (8) and (30) are sufficient to rewrite $P(k)$ as the Boltzmann-Gibbs distribution (35)

$$P(k) \sim c(k)\exp\left(-\sum_{i=1}^{t}\theta_i H_i(k)\right) = c(k)\exp\left(-\sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\Theta(x,y)\right). \tag{47}$$

See expression (28) for the definition of the matrix Θ. Notice that we omitted the dependence of the normalisation $G(\theta)$. By using the conditions (8), (46) and (30), we rewrite $\Upsilon(k)$ (45) as follows:

$$\begin{aligned}
\Upsilon(k) &= \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\ln w(x,y) + \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\ln w(x,y)\\
&= \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\ln w(x,y) + \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\ln w(y,x)\frac{p(y)}{p(x)}\\
&= \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\ln w(x,y)w(y,x)\\
&= \frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\left[\Theta(x,x)+\Theta(y,y)-\Theta(x,y)-\Theta(y,x)+\ln w(x,x)w(y,y)\right]
\end{aligned} \tag{48}$$

Then, we use (46) and (30) again together with $\sum_{x,y}k(x,y)=n$ to prove the following equality

$$\frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\left[\Theta(x,x)+\Theta(y,y)+\ln w(x,x)w(y,y)\right] = n\left[\Theta(r,r)+\ln w(r,r)\right] \tag{49}$$

with r an arbitrary state. Since we assumed n to be constant, this term can be absorbed in the normalisation of the distribution $P(k)$. Therefore, we end up with the following expression

$$P(k) \sim c(k)\exp\left(-\frac{1}{2}\sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\left[\Theta(x,y)+\Theta(y,x)\right]\right) \tag{50}$$

The distribution $P(k)$ is only of the Boltzmann-Gibbs form when the following condition holds (compare (50) with (47))

$$0 = \sum_{x\in\Gamma}\sum_{y\in\Gamma}k(x,y)\left[\Theta(x,y)-\Theta(y,x)\right] \tag{51}$$

or equivalently

$$0 = \sum_{x\in\Gamma}\sum_{y\in\Gamma}[k(x,y)-k(y,x)]\Theta(x,y) \tag{52}$$

A sufficient condition for the equality (51) to hold is obviously $\Theta(x,y)=\Theta(y,x)$ for all $x,y\in\Gamma$. However, the results of previous section show that this constraint is to restrictive. Indeed, we showed for a simple example that one can rewrite the distribution $P(k)$ in the Boltzmann-Gibbs form, without the matrix Θ (36) being symmetric. The crucial observation to obtain this result was that following constraint, $k(+,-)=k(-,+)$, is fulfilled for the 2-state Markov chain. For N-state Markov chains,

the latter equality can be generalised to (46). We proceed by eliminating some of the elements of the transition record out of expression (51) by using the conditions (46). Chose an arbitrary state r and replace in expression (51), $k(x,r)$ by

$$k(x,r) \quad = \quad k(r,x) + \sum_{y\in\Gamma'} [k(y,x) - k(x,y)] \tag{53}$$

with $\Gamma' = \Gamma\backslash\{r\}$. This results in the following condition

$$0 \quad = \quad \sum_{x\in\Gamma'}\sum_{y\in\Gamma'} k(x,y)\,[\Theta(x,y) - \Theta(y,x) - \Theta(x,r) + \Theta(r,x) + \Theta(y,r) - \Theta(r,y)] \tag{54}$$

A sufficient condition for this equality to hold is

$$\Theta(r,x) + \Theta(x,y) + \Theta(y,r) \quad = \quad \Theta(x,r) + \Theta(y,x) + \Theta(r,y) \tag{55}$$

Notice that this is a similar constraint to the detailed balance condition for the transition probabilities (32). This derivation does not depend on the arbitrary chosen state r. As such, expression (55) should hold for all $r, x, y \in \Gamma$. For the examples studied in this paper, condition (55) will always be satisfied. As a consequence, the relation between the Lagrange multipliers θ_i (contained in the matrix Θ) and the thermodynamic parameters is immediately clear. Therefore, in the remaining part of this paper, we will omit the substep of explicitly checking thermodynamic relations like (37) during our analysis.

6. Example: The 3-state Markov Chain

In this section we study the 3-state Markov chain. The three states are denoted 1, 2 and 3. The different parameters of the Markovian model are

$$p(1), p(2), p(3) \quad \text{and} \quad w(1,1), w(1,2), w(1,3), w(2,1), w(2,2), w(2,3), w(3,1), w(3,2), w(3,3)$$

However, the number of independent parameters is reduced by 4 because of the normalisation conditions (5), (6). The detailed balance conditions (8) further reduce the number of independent parameters by 3. We conclude that this microscopic model contains 5 independent parameters. In Section 2. we introduced the transition record k. For this example, the matrix k contains 9 elements

$$k = \begin{bmatrix} k(1,1) & k(1,2) & k(1,3) \\ k(2,1) & k(2,2) & k(2,3) \\ k(3,1) & k(3,2) & k(3,3) \end{bmatrix} \tag{56}$$

In the next two sections we study two physical models that are contained in this 3-state Markov chain. The relevant observables of these two models are different and, as such, the elements of the matrix Θ are not equal. As a consequence, the constraints (30) that relate the thermodynamic parameters (contained in the matrix Θ) to the microscopic parameters $w(x,y)$ will be different for the two physical models. However, the relations (31), (32) and (6) are the same because they only depend on the mathematical 3-state model and not on the particular choice of the relevant observables. Therefore we write out the formulas (31), (32) and (6) here, before we proceed with studying the expressions (30) in the next two sections. We choose $r = 1$ in expressions (31) and (32)

$$p(1) = \frac{w(2,1)w(3,1)}{w(2,1)w(3,1) + w(1,2)w(3,1) + w(2,1)w(1,3)}$$

$$p(2) = \frac{w(1,2)}{w(2,1)}p(1), \;\; p(3) = \frac{w(1,3)}{w(3,1)}p(1) \tag{57}$$

$$w(3,1)w(1,2)w(2,3) = w(1,3)w(2,1)w(3,2) \tag{58}$$

and write out the normalisation conditions (5) for the transition probabilities

$$\begin{aligned} &w(1,1) = 1 - w(1,2) - w(1,3), \;\; w(2,2) = 1 - w(2,1) - w(2,3) \\ &w(3,3) = 1 - w(3,1) - w(3,2) \end{aligned} \tag{59}$$

6.1. *Potts model*

The 3-state Markov chain can be interpreted as a one-dimensional Potts model [11, 12]. This system corresponds to a chain of $n+1$ spins. Contrary to the Ising model, the spin variables σ_i are vectors with unit length that can point in 3 directions specified by the angles $2q\pi/3$ with $q = 0, 1, 2$. Two relevant observables are

$$H_1(\sigma) = -J\sum_{i=0}^{n-1}\sigma_i \cdot \sigma_{i+1} \quad \text{and} \quad H_2(\sigma) = \sum_{i=0}^{n}\mathbb{1}\cdot\sigma_i \tag{60}$$

where J is a constant and $\mathbb{1}$ is a unit vector that points in one of the spin directions. Clearly, $\langle H_2(\sigma)\rangle$ is just the magnetisation M of the chain along the direction of $\mathbb{1}$, while $\langle H_1(\sigma)\rangle$ is usually interpreted as the internal energy U of the one-dimensional Potts model. The three states of the Markov chain $1, 2, 3$ correspond to the three different spin directions. The contribution to $H_1(\sigma)$ is $-J$ or $J/2$ depending on whether $\sigma_i = \sigma_{i+1}$ or $\sigma_i \neq \sigma_{i+1}$ respectively. Therefore, one can express $H_1(\sigma)$ as a function of the elements of the transition record k as follows:

$$\begin{aligned} H_1(k) = &-J[k(1,1) + k(2,2) + k(3,3)] \\ &+J\frac{1}{2}[k(1,2) + k(1,3) + k(2,1) + k(2,3) + k(3,1) + k(3,2)] \end{aligned} \tag{61}$$

In order to obtain a similar expression for $H_2(\sigma)$, we chose arbitrarily the direction of $\mathbb{1}$ along the state with label 1

$$H_2(k) = k(1,1) + k(2,1) + k(3,1) - \frac{1}{2}[k(1,2) + k(2,2) + k(3,2) + k(1,3) + k(2,3) + k(3,3)] \tag{62}$$

Analogous to the example of the Ising model, see Section 3., we ignored the unimportant finite size contribution of the initial spin σ_0 to obtain $H_2(k)$ from $H_2(\sigma)$. We use $\langle H_1(k)\rangle$ and $\langle H_2(k)\rangle$ as constraints in the maximisation procedure. As a consequence, the matrix Θ (28) becomes

$$\Theta = \begin{bmatrix} \Theta(1,1) & \Theta(1,2) & \Theta(1,3) \\ \Theta(2,1) & \Theta(2,2) & \Theta(2,3) \\ \Theta(3,1) & \Theta(3,2) & \Theta(3,3) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} -2J\theta_1 + 2\theta_2 & J\theta_1 - \theta_2 & J\theta_1 - \theta_2 \\ J\theta_1 + 2\theta_2 & -2J\theta_1 - \theta_2 & J\theta_1 - \theta_2 \\ J\theta_1 + 2\theta_2 & J\theta_1 - \theta_2 & -2J\theta_1 - \theta_2 \end{bmatrix} \tag{63}$$

Using (30), the parameters θ_1 and θ_2 can then be expressed as a function of the microscopic parameters as follows:

$$3J\theta_1 \;\; = \;\; \ln\frac{w(1,1)}{w(1,2)}\frac{w(2,2)}{w(2,1)} = \ln\frac{w(1,1)}{w(1,3)}\frac{w(3,3)}{w(3,1)} = \ln\frac{w(2,2)}{w(2,3)}\frac{w(3,3)}{w(3,2)}$$

$$\frac{3}{2}\theta_2 = \ln\frac{w(2,2)}{w(1,1)} = \ln\frac{w(3,3)}{w(1,1)} \tag{64}$$

Together with (58) and (59) these expressions form a closed set of equations that relate the parameters θ_1 and θ_2 to the microscopic parameters $w(x,y)$. In Appendix 2, this set is inverted analytically. We proceed by writing out the ensemble average of $H_2(k)$ with (10), (57) and (59)

$$\frac{\langle H_2(k)\rangle}{n} = -\frac{1}{2} + \frac{3}{2}\frac{w(2,1)}{w(2,1)+2w(1,2)} \tag{65}$$

In combination with the results of Appendix 2, one then obtains a formula for $\langle H_2(k)\rangle$ as a function of θ_1 and θ_2. These parameters can be related to the temperature T and an external applied field F as follows: $\theta_1 = 1/T$ and $\theta_2 = -F/T$. The final expression for the magnetisation $M = \langle H_2(k)\rangle$ as a function of T and F is

$$\frac{M}{n} = \frac{1}{4} + \frac{3}{4}\frac{e^{3J/2T} - \left(1+e^{3J/2T}\right)e^{-3F/2T}}{\sqrt{\left[e^{3J/2T} - \left(1+e^{3J/2T}\right)e^{-3F/2T}\right]^2 + 8e^{-3F/2T}}} \tag{66}$$

An identical expression is obtained in [12]. The authors assume that the equilibrium distribution is of the Boltzmann-Gibbs form and solve the one-dimensional Potts model with the technique of the transfer-matrix. Since the Hamiltonian of this model satisfies (55), the resulting equilibrium distribution of our approach is also of the Boltzmann-Gibbs form. That's the reason why the final expressions for M as a function of T and F of the two different approaches coincide.

6.2. *Blume-Emery-Griffiths model*

The 3-state Markov chain can be interpreted as a one-dimensional Blume-Emery-Griffiths model [13, 14]. This system corresponds to a chain of $n+1$ spins. The spin variables σ_i are scalars that can take on three values $+1, 0, -1$. Two relevant observables are

$$H_1(\sigma) = -J\sum_{i=0}^{n-1}\sigma_i\sigma_{i+1} - K\sum_{i=0}^{n-1}\sigma_i^2\sigma_{i+1}^2 + \Delta\sum_{i=0}^{n}\sigma_i^2 \quad \text{and} \quad H_2(\sigma) = \sum_{i=0}^{n}\sigma_i \tag{67}$$

where J, K, Δ are constants. Clearly, $\langle H_2(\sigma)\rangle$ is just the magnetisation M of the chain, while $\langle H_1(\sigma)\rangle$ is usually interpreted as the internal energy U of the one-dimensional Blume-Emery-Griffiths model. The three states of the Markov chain $1, 2, 3$ correspond to the spin values $+1, 0, -1$ respectively. Within this interpretation, one can express $H_1(\sigma)$ and $H_2(\sigma)$ as a function of the elements of the transition record k as follows:

$$\begin{aligned} H_1(k) &= -J[k(1,1)+k(3,3)-k(1,3)-k(3,1)] - K[k(1,1)+k(3,3)+k(1,3)+k(3,1)] \\ &\quad +\Delta[k(1,1)+k(2,1)+k(3,1)+k(1,3)+k(2,3)+k(3,3)] \end{aligned} \tag{68}$$

$$H_2(k) = k(1,1)+k(2,1)+k(3,1)-[k(1,3)+k(2,3)+k(3,3)] \tag{69}$$

Analogous to previous examples, we ignored the contribution of $\Delta\sigma_0^2$ to obtain $H_1(k)$ from $H_1(\sigma)$ and the contribution of σ_0 to obtain $H_2(k)$ from $H_2(\sigma)$. We use $\langle H_1(k)\rangle$ and $\langle H_2(k)\rangle$ as constraints in the maximisation procedure. As a consequence, the matrix Θ (28) becomes

$$\Theta = \begin{bmatrix} \Theta(1,1) & \Theta(1,2) & \Theta(1,3) \\ \Theta(2,1) & \Theta(2,2) & \Theta(2,3) \\ \Theta(3,1) & \Theta(3,2) & \Theta(3,3) \end{bmatrix} = \begin{bmatrix} \theta_1(-J-K+\Delta)+\theta_2 & 0 & \theta_1(J-K+\Delta)-\theta_2 \\ \theta_1\Delta+\theta_2 & 0 & \theta_1\Delta-\theta_2 \\ \theta_1(J-K+\Delta)+\theta_2 & 0 & \theta_1(-J-K+\Delta)-\theta_2 \end{bmatrix} \tag{70}$$

Using (30), the parameters θ_1 and θ_2 can then be expressed as a function of the microscopic parameters as follows:

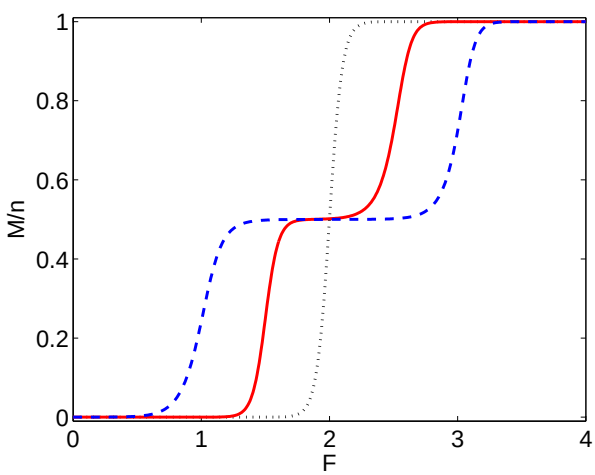
$$\begin{aligned} -\theta_1(J+K) &= \ln\frac{w(1,2)}{w(1,1)}\frac{w(2,1)}{w(2,2)} = \ln\frac{w(2,3)}{w(2,2)}\frac{w(3,2)}{w(3,3)} \\ -4\theta_1 J &= \ln\frac{w(1,3)}{w(1,1)}\frac{w(3,1)}{w(3,3)} \\ -\theta_1(J+K-\Delta)+\theta_2 &= \ln\frac{w(2,2)}{w(1,1)} \\ -\theta_1(J+K-\Delta)-\theta_2 &= \ln\frac{w(2,2)}{w(3,3)} \end{aligned} \tag{71}$$

Together with (58) and (59) these expressions form a closed set of equations that relate the parameters θ_1 and θ_2 to the microscopic parameters $w(x,y)$. In Appendix 3, this set is inverted analytically. As in previous example, the parameters θ_1 and θ_2 are related to the temperature T and an external applied field F as follows: $\theta_1 = 1/T$ and $\theta_2 = -F/T$. We proceed by writing out the ensemble average of $H_2(k)$ (69) with (10), (57) and (59)

$$\frac{\langle H_2(k)\rangle}{n} = \frac{w(3,1)w(2,1)-w(2,1)w(1,3)}{w(3,1)w(2,1)+w(1,3)w(2,1)+w(1,2)w(3,1)} \tag{72}$$

In combination with the results of Appendix 3, one finally obtains an expression for the magnetisation $M = \langle H_2(k)\rangle$ as a function of T and F. A plot of the magnetisation as a function of the external applied field at constant temperature $\theta_1 = 1/T = 20$ is shown in Figure 1 for the following values of the constants of $H_1(k)$ (68) $K = 0$, $J = -1$ and $\Delta = 0; 0.5; 1$. It is known that multiple plateaus show up is this curve depending on the value of Δ [14, 29]. This interesting behaviour can also be observed in Figure 1.

Figure 1. Plot of the magnetisation of the one-dimensional Blume-Emery-Griffiths model as a function of the external applied field at constant temperature $1/T = 20$. The values of the constants of H_1 (68) are $K = 0$, $J = -1$ and $\Delta = 0; 0.5; 1$ for the dotted, the solid, the dashed line respectively.



7. Discussion

In this paper, we present a general procedure to estimate parameters in Markovian models. The Markov chain is a mathematical model that is defined by initial probabilities $p(z)$ and transition probabilities $w(z, y)$. We interpret $p(z)$ and $w(z, y)$ as the microscopic parameters of the Markovian model. Then, relations between $p(z), w(z, y)$ and some relevant control parameters θ_i are determined with the maximum entropy principle. Finally, one ends up with formulas that express the average values of the relevant observables $H_i(k)$ as a function of the corresponding control parameters θ_i only. These expressions can be used to estimate the values of θ_i after the measurement of $\langle H_i(k)\rangle$. We want to stress that the dependence on the microscopic parameters is completely eliminated out of the theory. This means that no a priori choice for the values of $p(x)$ or $w(x, y)$ is necessary. This is important because the values of these parameters are not measurable.

We made a clear separation between the physical model of a theory and the underlying mathematical model. The latter is the N-state Markov chain while the former model is introduced by identifying some relevant observables. As such, different physical models can be contained in one type of Markov chain. This is illustrated in Section 6. where we examined two different physical models that are contained in the 3-state Markov chain. We showed that is possible to perform the aforementioned optimisation procedure in full generality for the N-state Markov chain. This results in relations (30) between the microscopic parameters of the mathematical model and some relevant control parameters. These formulas are the main result of this paper.

In Section 5.3. we studied under which conditions the equilibrium distribution of our approach is of the Boltzmann-Gibbs form. Obtaining this type of equilibrium distribution is advantageous because a thermodynamic interpretation of the control parameters is obvious in that case. We derived a sufficient condition that is satisfied for all the examples studied in this paper. As such, the final formulas for the average values of the relevant observables as a function of the thermodynamic parameters that are obtained in this paper have been studied before. The general procedure to obtain these formulas is the novel contribution of this paper. Notice that further generalisations of our technique are still possible. We assumed that the relevant observables are linear combinations of the elements of the transition record of the Markov chain. In [30], this condition is lifted. In that paper, the specific example of the 2-state Markov chain with a mean-field Hamiltonian is studied with the same technique as described in the present paper. It is an interesting topic for further research to examine the effect of allowing mean-field Hamiltonians in the theory for the general N-state Markov chain. We also assumed that all the transitions are allowed ($w(x, y) \neq 0$ for all $x, y \in \Gamma$). Lifting this assumption will usually cause the violation of the detailed balance condition. The generalisation of the results reported in the present paper to non-equilibrium steady states is currently under study. Notice that allowing for a vanishing transition probability $w(x, y)$ will not cause the violation of the detailed balance condition when the transition $y \to x$ is also not allowed. The specific example of a 6-state Markov chain with that property is studied in [7]. Throughout the present paper, we ignored finite size effects. The technical consequences of taking these effects into account are already thoroughly examined for the 2-state Markov chain in [5, 19].

Acknowledgements

The author wishes to thank Prof. Jan Naudts for helpful comments.

Appendix

Appendix 1

In this appendix we maximise the function

$$\begin{aligned}\frac{1}{n}\mathcal{L} &= -\sum_{z\in\Gamma}\sum_{y\in\Gamma} p(z)w(z,y)\ln w(z,y) - \sum_{z\in\Gamma}\sum_{y\in\Gamma}\Theta(z,y)p(z)w(z,y) - \alpha\sum_{z\in\Gamma}p(z)\\ &\quad -\sum_{z\in\Gamma}\zeta(z)\sum_{y\in\Gamma}w(z,y) - \sum_{z\in\Gamma}\sum_{y\in\Gamma,y>z}\eta(z,y)\left[p(z)w(z,y)-p(y)w(y,z)\right],\end{aligned} \tag{73}$$

over the parameters $p(z)$ and $w(z,y)$. Therefore, we set the first derivative of $\mathcal{L}$ with respect to these parameters equal to zero. The resulting equations for differentiating with respect to $w(u,u)$ (74), $p(u)$ (75), $w(u,v)$ (76), $w(v,u)$ (77) with $v>u$ are

$$\zeta(u) = -p(u)\left[1+\ln w(u,u)+\Theta(u,u)\right], \tag{74}$$

$$\begin{aligned}0 &= \sum_{y\in\Gamma} w(u,y)\ln w(u,y) + \sum_{y\in\Gamma}\Theta(u,y)w(u,y)+\alpha\\ &\quad + \sum_{y\in\Gamma,y>u}\eta(u,y)w(u,y) - \sum_{y\in\Gamma,y<u}\eta(y,u)w(u,y),\end{aligned} \tag{75}$$

$$0 = p(u)\left[1+\ln w(u,v)\right]+\Theta(u,v)p(u)+\zeta(u)+\eta(u,v)p(u), \tag{76}$$

$$0 = p(v)\left[1+\ln w(v,u)\right]+\Theta(v,u)p(v)+\zeta(v)-\eta(u,v)p(v). \tag{77}$$

One can simplify the expressions (76) and (77) with the use of the formula for $\zeta(u)$ (74)

$$0 = \ln\frac{w(u,v)}{w(u,u)}+\Theta(u,v)-\Theta(u,u)+\eta(u,v), \tag{78}$$

$$0 = \ln\frac{w(v,u)}{w(v,v)}+\Theta(v,u)-\Theta(v,v)-\eta(u,v). \tag{79}$$

Combining these two equations results in

$$\Theta(u,u)+\Theta(v,v)-\Theta(u,v)-\Theta(v,u) = \ln\frac{w(u,v)}{w(u,u)}\frac{w(v,u)}{w(v,v)}. \tag{80}$$

We proceed by rewriting expression (75) as follows:

$$\begin{aligned}0 &= w(u,u)\left[\ln w(u,u)+\Theta(u,u)\right]+\alpha\\ &\quad + \sum_{y\in\Gamma,y>u} w(u,y)\left[\ln w(u,y)+\Theta(u,y)+\eta(u,y)\right]\\ &\quad + \sum_{y\in\Gamma,y<u} w(u,y)\left[\ln w(u,y)+\Theta(u,y)-\eta(y,u)\right].\end{aligned} \tag{81}$$

Then, we use (78) and (79) to transform (81) further to

$$0 = w(u,u)\left[\ln w(u,u)+\Theta(u,u)\right]+\alpha$$

$$
\begin{aligned}
&+ \sum_{y\in\Gamma, y>u} w(u,y)\left[\ln w(u,u) + \Theta(u,u)\right] \\
&+ \sum_{y\in\Gamma, y<u} w(u,y)\left[\ln w(u,u) + \Theta(u,u)\right] \\
-\alpha &= \ln w(u,u) + \Theta(u,u). \qquad (82)
\end{aligned}
$$

The latter equation is valid for all $u \in \Gamma$. The parameter α can then be eliminated, by combining these equations two by two

$$
\Theta(u,u) - \Theta(v,v) = \ln \frac{w(v,v)}{w(u,u)}. \qquad (83)
$$

Appendix 2

In this appendix we invert the set of equations (64), (58), (59)

$$
\begin{aligned}
e^{3J\theta_1} &= \frac{1-w(1,2)-w(1,3)}{w(1,2)} \frac{1-w(2,1)-w(2,3)}{w(2,1)}, && (84)\\
e^{\frac{3}{2}\theta_2} &= \frac{1-w(2,1)-w(2,3)}{1-w(1,2)-w(1,3)}, && (85)\\
\frac{1}{w(1,2)} \frac{1-w(2,1)-w(2,3)}{w(2,1)} &= \frac{1}{w(1,3)} \frac{1-w(3,1)-w(3,2)}{w(3,1)}, && (86)\\
\frac{1-w(1,2)-w(1,3)}{w(1,2)} \frac{1}{w(2,1)} &= \frac{1}{w(2,3)} \frac{1-w(3,1)-w(3,2)}{w(3,2)}, && (87)\\
1-w(2,1)-w(2,3) &= 1-w(3,1)-w(3,2), && (88)\\
w(3,1)w(1,2)w(2,3) &= w(1,3)w(2,1)w(3,2), && (89)
\end{aligned}
$$

to obtain formulas in closed form for the microscopic parameters $w(x,y)$ as a function of θ_1 and θ_2 only. Equations (86), (88) and (89) can be simplified to

$$
w(1,3) = w(1,2), \quad w(3,1) = w(2,1) \quad \text{and} \quad w(3,2) = w(2,3). \qquad (90)
$$

Notice that this restricts the values of $w(1,2)$ and $w(1,3)$ to the interval $[0..1/2]$, because of the normalisation condition. Using (90), equation (87) can be rewritten as follows:

$$
w(1,2) = \frac{w(2,3)^2}{2w(2,3)^2 + w(2,1)[1-w(2,1)-w(2,3)]}. \qquad (91)
$$

Inserting (90) and (91) into equations (84) and (85) results in two equations in the variables $w(2,3)$ and $w(2,1)$. By inverting these two equations

$$
\begin{aligned}
w(2,1) &= 1 - w(2,3)\left(1 + e^{\frac{3}{2}J\theta_1}\right), \\
w(2,3) &= -\frac{1}{2} \frac{e^{\frac{3}{2}J\theta_1} + \left(1+e^{\frac{3}{2}J\theta_1}\right) e^{\frac{3}{2}\theta_2} - \sqrt{\left[e^{\frac{3}{2}J\theta_1} - \left(1+e^{\frac{3}{2}J\theta_1}\right) e^{\frac{3}{2}\theta_2}\right]^2 + 8e^{\frac{3}{2}\theta_2}}}{2 - \left(1+e^{\frac{3}{2}J\theta_1}\right) e^{\frac{3}{2}J\theta_1}}, \qquad (92)
\end{aligned}
$$

one finally obtains a closed chain of equations for the transition probabilities.

Appendix 3

In this appendix we invert the set of equations (71), (58), (59). We first introduce a shorthand notation

$$\eta_1 = e^{-\theta_1(J+K)}, \;\; \eta_2 = e^{-2\theta_1 J}, \;\; \eta_3 = e^{-\theta_1(J+K-\Delta)+\theta_2}, \;\; \eta_4 = e^{-\theta_1(J+K-\Delta)-\theta_2}, \tag{93}$$

and the substitution

$$\begin{aligned} X_1 &= \frac{w(1,2)}{w(1,1)}, \;\; X_2 = \frac{w(2,1)}{w(2,2)}, \;\; X_3 = \frac{w(2,3)}{w(2,2)}, \\ X_4 &= \frac{w(3,2)}{w(3,3)}, \;\; X_5 = \frac{w(1,3)}{w(1,1)}, \;\; X_6 = \frac{w(3,1)}{w(3,3)}, \end{aligned} \tag{94}$$

to obtain the following expressions for the normalisation conditions (59)

$$w(1,1) = (1+X_1+X_5)^{-1}, \;\; w(2,2) = (1+X_2+X_3)^{-1}, \;\; w(3,3) = (1+X_4+X_6)^{-1}, \tag{95}$$

and the equations (71), (58)

$$\eta_1 = X_1X_2, \; \eta_1 = X_3X_4, \qquad \eta_2^2 = X_5X_6, \;\; X_1X_3X_6 = X_2X_4X_5, \tag{96}$$

$$X_5 = \eta_3(1+X_2+X_3) - 1 - X_1, \qquad X_4 = \eta_4(1+X_2+X_3) - 1 - X_6. \tag{97}$$

We proceed by rewriting (96) as follows:

$$\eta_1 = X_1X_2, \;\; \eta_1 = X_3X_4, \;\; \eta_2X_3 = X_2X_5, \;\; \eta_2X_2 = X_3X_6. \tag{98}$$

Then, we use the expressions (97) for X_4 and X_5 to transform (98) further to

$$X_1 = \frac{\eta_1}{X_2}, \qquad X_6 = \eta_4(1+X_2+X_3) - 1 - \frac{\eta_1}{X_3}, \tag{99}$$

$$\eta_2X_3 = X_2\eta_3(1+X_2+X_3) - X_2 - \eta_1, \qquad \eta_2X_2 = X_3\eta_4(1+X_2+X_3) - X_3 - \eta_1. \tag{100}$$

Finally, we rewrite (100) as follows:

$$X_3 = \frac{X_2\eta_3(1+X_2) - X_2 - \eta_1}{\eta_2 - X_2\eta_3}, \tag{101}$$

$$\begin{aligned} (X_2\eta_2+\eta_1)(X_2\eta_3-\eta_2)^2 &= \big[X_2\eta_3(1+X_2) - (X_2+\eta_1)\big]\times \\ &\quad \big[\eta_4\eta_2(1+X_2) - \eta_4(X_2+\eta_1) + X_2\eta_3 - \eta_2\big], \end{aligned} \tag{102}$$

to obtain a closed chain of equations for the variables X_i with $i = 1\ldots6$. Expression (102) is a cubic equation in the variable X_2 which can have 3 real solutions. However, it is well known that one-dimensional systems with short range interactions do not exhibit phase transitions when the equilibrium distribution is of the Boltzmann-Gibbs form. Therefore, only one of the solutions of the cubic equation is physically meaningful. The other solutions are complex or result in values for some of the transition probabilities outside the interval $[0,1]$.

References

1. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620–630.
2. Jaynes, E.T. Information theory and statistical mechanics II. *Phys. Rev.* **1957**, *108*, 171–190.
3. Rosenkrantz, R.D. *E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics*; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1989.
4. Callen, H.B. *Thermodynamics and an Introduction to Thermostatistics*; Wiley: New York, NY, USA, 1985.
5. Van der Straeten, E.; Naudts, J. A one-dimensional model for theoretical analysis of single molecule experiments. *J. Phys. A-Math. Theor.* **2006**, *39*, 5715–5726.
6. Van der Straeten, E.; Naudts, J. Residual entropy in a model for the unfolding of single polymer chains. *Europhys. Lett.* **2008**, *81*, 28007:1–28007:5.
7. Van der Straeten, E.; Naudts, J. The 3-dimensional random walk with applications to overstretched DNA and the protein titin. *Physica A* **2008**, *387*, 6790–6800.
8. Van Campenhout, J.M.; Cover, T.M. Maximum entropy and conditional probability. *IEEE Trans. Inf. Theor.* **1981**, *IT-27*, 483–489.
9. Csiszár, I.; Cover, T.M.; Choi, B.S. Conditional limit theorems under Markov conditioning. *IEEE Trans. Inf. Theor.* **1987**, *IT-33*, 788–801.
10. Toda, M.; Kubo, R.; Saitô, N. *Statistical Physics I: Equilibrium Statistical Mechanics*; Springer-Verlag: Berlin, Germany, 1983.
11. Wu, F.Y. The potts model. *Rev. Mod. Phys.* **1982**, *54*, 235–268.
12. Kassan-Ogly, F.A. One-dimensional 3-state and 4-state standard potts models in magnetic field. *Phase Transit.* **2000**, *71*, 39–55.
13. Blume, M.; Emery, V.J.; Griffiths, R.B. Ising model for the λ transition and phase separation in He^3-He^4 mixtures. *Phys. Rev. A* **1971**, *4*, 1071–1077.
14. Mancini, F.; Mancini, F.P. Magnetic and thermal properties of a one-dimensional spin-1 model. *Condens. Matter Phys.* **2008**, *11*, 543–553.
15. Naudts, J.; Van der Straeten, E. Transition records of stationary Markov chains. *Phys. Rev. E* **2006**, *74*, 040103:1–040103:4.
16. Kelly, F.P. *Reversibility and Stochastic Networks*; Wiley: Chichester, UK, 1979.
17. Cornfeld, I.P.; Fomin, S.V.; Sinai, Y.G. *Ergodic Theory*; Springer-Verlag: New York, NY, USA, 1982.
18. Gaspard, P. Time-reversed dynamical entropy and irreversibility in Markovian random processes. *J. Stat. Phys.* **2004**, *117*, 599–615.
19. Van der Straeten, E.; Naudts, J. A two-parameter random walk with approximate exponential probability distribution. *J. Phys. A-Math. Theor.* **2006**, *39*, 7245–7256.
20. Csiszár, I.; Shields, P.C. The consistency of the BIC Markov order estimator. *Ann. Stat.* **2000**, *28*, 1601–1619.
21. Csiszár, I. Large-scale typicality of Markov sample paths and consistency of MDL order estimators. *IEEE Trans. Inf. Theor.* **2002**, *48*, 1616–1628.

22. Waterman, M.S. *Introduction to Computational Biology*; Chapman & Hall/CRC: Boca Raton, FL, USA, 1995.
23. Schbath, S. An overview on the distribution of word counts in Markov chains. *J. Comput. Biol.* **2000**, *7*, 193–201.
24. Sage, A.P.; Melsa, J.L. *Estimation Theory with Applications to Communications and Control*; Mc-Graw-Hill: New York, NY, USA, 1971.
25. Naudts, J. Parameter estimation in nonextensive thermostatistics. *Physica A* **2006**, *365*, 42–49.
26. Murray, M.K.; Rice, J.W. *Differential Geometry and Statistics*; Chapman & Hall: London, UK, 1993.
27. Naudts, J. Generalised exponential families and associated entropy functions. *Entropy* **2008**, *10*, 131–149.
28. Naudts, J. The q-exponential family in statistical physics. *Cent. Eur. J. Phys.* **2009**, *7*, 405–413.
29. Chen, X.Y.; Jiang, Q.; Shen, W.Z.; Zhong, C.G. The properties of one-dimensional spin-S (S$\geq$1) antiferromagnetic Ising chain with single-ion anisotropy. *J. Magn. Magn. Mater.* **2003**, *262*, 258–263.
30. Van der Straeten, E.; Naudts, J. The globule-coil transition in a mean field approach. *arXiv:cond-mat/0612256* **2006**.

© 2009 by the author; licensee Molecular Diversity Preservation International, Basel, Switzerland. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license http://creativecommons.org/licenses/by/3.0/.