BAYESIAN ANALYSIS FOR REVERSIBLE MARKOV CHAINS

by

Persi Diaconis
Silke W. W. Rolles

April 2004

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065
BAYESIAN ANALYSIS REVERSIBLE MARKOV CHAINS

by

Persi Diaconis
Department of Statistics
Stanford University

Silke W. W. Rolles
Department of Mathematics
University of California, Los Angeles

April 2004

This research was supported in part by
National Science Foundation grant DMS-0072360

Department of Statistics
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
Bayesian analysis for reversible Markov chains

Persi Diaconis\(^1\) Silke W.W. Rolles\(^2\)

April 2, 2004

Abstract

We introduce a natural conjugate prior for the transition matrix of a reversible Markov chain. This allows estimation and testing. The prior arises from random walk with reinforcement in the same way the Dirichlet prior arises from Polya’s urn. We give closed form normalizing constants, a simple method of simulation from the posterior and a characterization along the lines of W.E. Johnson’s characterization of the Dirichlet prior.

Contents

1 Introduction

2 A class of prior distributions
  2.1 A minimal sufficient statistics ........................................ 2
  2.2 Definition of the densities \( \phi_{v_0,a} \) ................................. 3
  2.3 Random walk with reinforcement ....................................... 5

3 The density \( \phi_{v_0,a} \) for special graphs
  3.1 The line graph (Birth and death chains) ............................. 7
  3.2 Trees with loops ...................................................... 8
  3.3 The triangle .......................................................... 8
  3.4 The complete graph .................................................. 9

4 Properties of the family of priors
  4.1 Closure under sampling ................................................. 10
  4.2 Uniqueness ............................................................ 10
  4.3 The priors are dense .................................................. 12
  4.4 Computing some moments .............................................. 13
  4.5 Simulating from the posterior ....................................... 14

5 Applications ................................. 15

1 Introduction

Modelling with Markov chains is an important part of time series analysis, genomics and many other applications. Reversible Markov chains are a mainstay of computational statistics through the Gibbs sampler, Metropolis algorithm and their many variants. Reversible chains are widely used natural models in physics and chemistry where reversibility (often called detailed balance) is a stochastic analog of the time reversibility of Newtonian mechanics.

\(^1\)Department of Statistics, Stanford University, Stanford, CA 94305-4065, USA.
\(^2\)Department of Mathematics, University of California, Los Angeles, Box 951555, Los Angeles, CA 90095-1555, USA. e-mail: srolles@math.ucla.edu
This paper develops tools for a Bayesian analysis of the transition probabilities, stationary distribution and future prediction of a reversible Markov chain. We observe \( X_0 = v_0, X_1 = v_1, \ldots, X_n = v_n \) from a reversible Markov chain with a finite state space \( V \). Neither the stationary distribution \( \pi(v) \) nor the transition kernel \( k(v,v') \) are assumed known. Reversibility entails \( \pi(v)k(v,v') = \pi(v')k(v',v) \) for all \( v, v' \in V \). We also assume we know which transitions are possible (for which \( v, v' \in V \) is \( k(v,v') > 0 \)).

In Section 2, we introduce a family of natural conjugate priors. These are defined via closed form densities and by a generalization of Polya's urn to random walk with reinforcement on a graph. The density gives normalizing constants needed for testing independence versus reversibility or reversibility versus a full Markovian specification. The random walk gives a simple method of simulating from the posterior (Section 4.5).

Properties of the prior are developed in Section 4. The family is closed under sampling (Proposition 4.1). Mixtures of our conjugates are shown to be dense (Proposition 4.5). A characterization of the priors via predictive properties of the posterior is given (Section 4.2).

Two practical examples are given in Section 5. Several simple hypotheses are tested for a data set arising from the DNA of the human HLA-B gene. Section 5 also contains remarks about statistical analysis for reversible chains.

2 A class of prior distributions

We observe \( X_0 = v_0, X_1 = v_1, \ldots, X_n = v_n \) from a reversible Markov chain with a finite state space \( V \) and unknown transition kernel \( k(\cdot, \cdot) \).

Let \( G = (V, E) \) be the finite graph with vertex set \( V \) and edge set \( E \) defined as follows: \( e = \{v, v'\} \in E \) (i.e. there is an edge between \( v \) and \( v' \)) if and only if \( k(v,v') > 0 \). We assume that \( k(v,v') > 0 \) iff \( k(v',v) > 0 \). In particular, all edges of \( G \) are undirected and an edge is denoted by the set of its endpoints. For some vertices \( v \), we may have \( k(v,v) > 0 \). Define the simplex

\[
\Delta := \left\{ x = (x_e)_{e \in E} \in (0,1]^E : \sum_{e \in E} x_e = 1 \right\}. \tag{2.1}
\]

Recall that the distribution of a reversible Markov chain can be described by putting on the edge between \( v \) and \( v' \) the weight \( x_{(v,v')} := \pi(v)k(v,v') = \pi(v')k(v',v) \). If the weights are normalized so that \( \sum_{e \in E} x_e = 1 \), this is a unique way to describe the distribution of the Markov chain. A transition from \( v \) to \( v' \) is made with probability proportional to the weight \( x_{(v,v')} \). Denote by \( Q_{v_0,x} \) the distribution of the Markov chain induced by the weights \( x = (x_e)_{e \in E} \in \Delta \) which starts with probability 1 in \( v_0 \). Using this notation, our assumption says that the observed data comes from a distribution in the class

\[
Q := \{ Q_{v_0,x} : v_0 \in V, x \in \Delta \}. \tag{2.2}
\]

2.1 A minimal sufficient statistics

If the endpoints of an edge \( e \) agree, we call \( e \) a loop. Let

\[
E_{\text{loop}} := \{ e \in E : e \text{ is a loop} \}. \tag{2.3}
\]
Let \( \pi := (\pi_0, \pi_1, \ldots, \pi_n) \) be an admissible path in \( G \). For \( v \in V \) and \( e \in E \), define
\[
k_v(\pi) := |\{i \in \{1, 2, \ldots, n\} : (v, \pi_i) = (\pi_{i-1}, \pi_i)\}|, \text{ for } v \in V,
\]
\[
k_e(\pi) := \begin{cases} 
|\{i \in \{1, 2, \ldots, n\} : (\pi_{i-1}, \pi_i) = e\}| & \text{if } e \in E \setminus E_{\text{loop}}, \\
2 \cdot |\{i \in \{1, 2, \ldots, n\} : (\pi_{i-1}, \pi_i) = e\}| & \text{if } e \in E_{\text{loop}}.
\end{cases}
\] (2.4) (2.5)

I.e. \( k_v(\pi) \) equals the number of times the path \( \pi \) leaves vertex \( v \); for an edge \( e \) which is not a loop, \( k_e(\pi) \) is the number of traversals of \( e \) by \( \pi \), and for a loop \( e \), \( k_e(\pi) \) is twice the number of traversals of \( e \). Recall that the edges are undirected; hence \( k_e(\pi) \) counts the traversals of \( e \) in both directions. Set
\[
Z_n := (X_0, X_1, \ldots, X_n).
\] (2.6)

**Proposition 2.1** The vector of transition counts \( (k_e(Z_n))_{e \in E} \) is a minimal sufficient statistic for the model \( Q_{v_0} := \{Q_{v_0, x} : x \in \Delta\} \).

**Proof.** Let \( \pi \) be an admissible path in \( G \). In order to prove that \( (k_e(Z_n))_{e \in E} \) is a sufficient statistics, we need to show that
\[
Q_{v_0, \pi}(Z_n = \pi | (k_e(Z_n))_{e \in E})
\] (2.7)
does not depend on \( x \). If \( \pi \) does not start in \( v_0 \), (2.7) equals zero. Otherwise, we have
\[
Q_{v_0, \pi}(Z_n = \pi) = \frac{\prod_{e \in E} x_e^{k_e(\pi)}}{\prod_{v \in V} x_v^{k_v(\pi)}}.
\] (2.8)

It is not hard to see that \( k_v(\pi) \) can be expressed in terms of the \( k_e(\pi) \) and the first observation \( v_0 \). Hence, \( Q_{v_0, \pi} \)-probability of \( \pi \) depends only on \( k_e(\pi), e \in E, \) and \( v_0 \). Thus, (2.7) equals one divided by the number of admissible paths \( \pi' \) with starting point \( v_0 \) and \( k_e(\pi') = k_e(Z_n) \) for all \( e \in E \), which is independent of \( x \).

Suppose \( K := (k_e)_{e \in E} \) is not minimal. Then there exists a sufficient statistics \( K' \) which needs less information than \( K \). Consequently, there exist two admissible paths \( \pi \) and \( \pi' \) starting in \( v_0 \) such that \( K(\pi) \neq K(\pi') \) and \( K'(\pi) = K'(\pi') \). Then
\[
\frac{Q_{v_0, \pi}(Z_n = \pi | K'(Z_n) = K'(\pi))}{Q_{v_0, \pi}(Z_n = \pi' | K'(Z_n) = K'(\pi'))} = \frac{Q_{v_0, \pi}(Z_n = \pi)}{Q_{v_0, \pi}(Z_n = \pi')} = \prod_{e \in E} x_e^{k_e(\pi) - k_e(\pi')} \prod_{v \in V} x_v^{k_v(\pi') - k_v(\pi)}.
\] (2.9)

Since by assumption \( (k_e(\pi))_{e \in E} \neq (k_e(\pi'))_{e \in E} \), the last quantity depends on \( x \). This contradicts the fact that \( K' \) is a sufficient statistics. \( \blacksquare \)

### 2.2 Definition of the densities \( \phi_{v_0, a} \)

Our aim is to define a class of prior distributions in terms of measures on \( \Delta \). We prepare the definition with some notation.

For an edge \( e \), denote the set of its endpoints by \( \bar{e} \). Denote the cardinality of a set \( S \) by \( |S| \). Recall the definition (2.3) of the set \( E_{\text{loop}} \). Set
\[
l := |V| + |E_{\text{loop}}| \quad \text{and} \quad m := |E|.
\] (2.10)
For $x = (x_e)_{e \in E} \in (0, \infty)^E$ and a vertex $v$, define $x_v$ to be the sum of all components $x_e$ with $e$ incident to $v$:

$$x_v := \sum_{\{e \in E \}} x_e. \quad (2.11)$$

In sums such as this the sum is over edges including loops. Similarly, define $a_v$ for $a := (a_e)_{e \in E} \in (0, \infty)^E$.

There is a simple way to delineate a generating set of cycles of $G$. We call a maximal subgraph of $G$ which contains all loops but no cycle a spanning tree of $G$. Choose a spanning tree $T$. Each edge $e \in E \setminus E_{\text{loop}}$ which is not in $T$ forms a cycle $c_e$ when added to $T$. (By definition, a loop is never a cycle and never contained in a cycle.) There are $m-l+1$ such cycles and we enumerate them arbitrarily: $c_1, \ldots, c_{m-l+1}$. This set of cycles forms an additive basis for the homology $H_1$ and also serves for our purposes. In general, the first Betti number $\beta_1$ is the dimension of $H_1$. For the complete graph, $\beta_1(K_n) = \binom{n-1}{2}$. Further details can be found in Giblin ([Gib81], Section 1.16). In Section 3.4, we show how such a basis of cycles can be obtained for the complete graph.

Orient the cycles $c_1, \ldots, c_{m-l+1}$ and all edges $e \in E$ in an arbitrary way. For every $x \in \Delta$, define a matrix $A(x) = (A_{i,j}(x))_{1 \leq i, j \leq m-l+1}$ by

$$A_{i,i}(x) = \sum_{e \in c_i} \frac{1}{x_e}, \quad A_{i,j}(x) = \sum_{e \in c_i \cap c_j} \pm \frac{1}{x_e} \text{ for } i \neq j, \quad (2.12)$$

where the signs in the last sum are chosen to be $+1$ or $-1$ depending on whether the edge $e$ has in $c_i$ and $c_j$ the same orientation or not.

**Definition 2.2** For all $v_0 \in V$ and $a := (a_e)_{e \in E} \in (0, \infty)^E$, define

$$\phi_{v_0,a}(x) := Z_{v_0,a}^{-1} \prod_{e \in E \setminus E_{\text{loop}}} \frac{a_e - \frac{1}{2}}{x_e} \prod_{e \in E_{\text{loop}}} \frac{a_e}{x_e} \Gamma \left( \frac{a_e}{2} \right) \sqrt{\det(A(x))} \quad (2.13)$$

$$\langle x := (x_e)_{e \in E} \in \Delta \rangle \text{ with}$$

$$Z_{v_0,a} := \frac{\prod_{e \in E} \Gamma(a_e)}{\Gamma \left( \frac{a_{v_0}}{2} \right) \prod_{v \in V \setminus \{v_0\}} \Gamma \left( \frac{a_v+1}{2} \right) \prod_{e \in E_{\text{loop}}} \Gamma \left( \frac{a_e+1}{2} \right)} \cdot \frac{(m-1)! \pi^{l-1}}{2^{l-1} + \sum_{e \in E} a_e}. \quad (2.14)$$

The definition of $\phi_{v_0,a}$ does not depend on the choice of the cycles $c_i$ used in the definition of $A(x)$. Let us explain why: Let $T$ denote the set of all spanning trees of $G$. Then

$$\det A(x) = \sum_{T \in T} \prod_{e \in E(T)} \frac{1}{x_e}. \quad (2.15)$$

Clearly, the right-hand side of (2.15) does not depend on the choice of the cycles $c_i$.

The identity (2.15) is proved for graphs without loops in [Mau76] (page 145, theorem 3'). By definition, $A(x)$ does not depend on $x_e$, $e \in E_{\text{loop}}$. Furthermore, since every spanning tree contains all loops, the right-hand side of (2.15) does not depend on $x_e$, $e \in E_{\text{loop}}$ either. In particular, both sides of (2.15) are the same for $G$ and the graph obtained from $G$ by removing all loops; hence they are equal.
2.3 Random walk with reinforcement

Let \( \sigma \) denote the Lebesgue measure on \( \Delta \), normalized such that \( \sigma(\Delta) = 1 \). The measures \( \phi_{v_0,a}d\sigma \) on \( \Delta \) arise in the study of edge-reinforced random walk, as was observed by Coppersmith and Diaconis (see [Dia88]). Let us explain this connection: The edges of \( G \) are given strictly positive weights; at time 0 edge \( e \) has weight \( a_e > 0 \). Edge-reinforced random walk on \( G \) with starting point \( v_0 \) is defined as follows: The process starts at \( v_0 \) at time 0. In each step, the random walker traverses an edge with probability proportional to its weight. Each time an edge \( e \in E \setminus E_{\text{loop}} \) is traversed, its weight is increased by 1. Each time a loop \( e \in E_{\text{loop}} \) is traversed, its weight is increased by 2.

Denote the set of non-negative integers by \( \mathbb{N}_0 \). Let \( \Omega \) be the set of all \( (v_i)_{i \in \mathbb{N}_0} \in V^{\mathbb{N}_0} \) such that \( \{v_i, v_{i+1}\} \in E \) for all \( i \in \mathbb{N}_0 \). Let \( X_n : V^{\mathbb{N}_0} \rightarrow V \) denote the projection onto the \( n \)th coordinate. Recall that \( Z_n = (X_0, X_1, \ldots, X_n) \). Denote by \( P_{v_0,a} \) the distribution on \( \Omega \) of an edge-reinforced random walk with starting point \( v_0 \) and initial edge weights \( a = (a_e)_{e \in E} \).

Let \( \alpha_e(Z_n) := k_e(Z_n)/n \) be the proportion of traversals of edge \( e \) up to time \( n \). It was observed by Coppersmith and Diaconis that \( \alpha(Z_n) := (\alpha_e(Z_n))_{e \in E} \) converges almost surely to a random variable with distribution \( \phi_{v_0,a}d\alpha \) (see [Dia88] and also [KR00]). In particular, \( \phi_{v_0,a}d\sigma \) is a probability measure on \( \Delta \). This fact is not at all obvious from the definition of \( \phi_{v_0,a} \).

It turns out that edge-reinforced random walk on \( G \) is a mixture of reversible Markov chains, where the mixing measure described as a measure on edge weights \( (x_e)_{e \in E} \) is given by \( \phi_{v_0,a}d\sigma \). This is made precise by the following theorem:

**Theorem 2.3** Let \( (X_n)_{n \in \mathbb{N}_0} \) be edge-reinforced random walk with initial weights \( a = (a_e)_{e \in E} \) starting at \( v_0 \), and let \( Z_n = (X_0, X_1, \ldots, X_n) \). For any admissible path \( \pi = (v_0, \ldots, v_n) \), the following holds:

\[
P_{v_0,a}(Z_n = \pi) = \int_{\Delta} \prod_{i=1}^{n} \frac{x_{\{v_{i-1},v_i\}}}{x_{v_{i-1}}} \phi_{v_0,a}(x) \, d\sigma(x); \tag{2.16}
\]

here \( x := (x_e)_{e \in E} \). Hence, if \( P_{v_0,a} \) is the mixture of Markov chains where the mixing measure, described as a measure on edge weights \( (x_e)_{e \in E} \), is given by \( \phi_{v_0,a}d\sigma \), then

\[
P_{v_0,a} = P_{v_0,a}. \tag{2.17}
\]

**Proof.** If \( G \) has no loops, then the claim is true by Theorem 3.1 of [Ro03].

Let \( G \) be a graph with loops. Define a graph \( G' := (V', E') \) as follows: Replace every loop of \( G \) by an edge of degree 1 incident to the same vertex (see Figure 1). More precisely, for all \( e \in E_{\text{loop}} \), let \( v(e) \) be the vertex \( e \) is incident to and let \( v'(e) \) be an additional vertex, different from all the others. Then, set \( g(e) := \{v(e), v'(e)\} \) and

\[
V' := V \cup \{v'(e) : e \in E_{\text{loop}}\}, \tag{2.18}
\]

\[
E' := [E \setminus E_{\text{loop}}] \cup \{g(e) : e \in E_{\text{loop}}\}. \tag{2.19}
\]

The graph \( G' \) has no loops and the claim of the theorem is true for \( G' \).

Let \( P'_{v_0,b} \) be the distribution of a reinforced random walk on \( G' \) starting at \( v_0 \) with initial weights \( b = (b_{e'})_{e' \in E'} \) defined by

\[
b_{e'} := \begin{cases} 
0 & \text{if } e' \in E \setminus E_{\text{loop}} \\
a_e & \text{if } e' = g(e) \text{ for some } e \in E_{\text{loop}}.
\end{cases} \tag{2.20}
\]
Any finite admissible path \( \pi = (\pi_0 = v_0, \pi_1, \ldots, \pi_n) \) in \( G \) can be mapped to an admissible path \( \pi' = (\pi_0' = v_0, \pi_1', \ldots, \pi_n') \) in \( G' \) by mapping every traversal of a loop \( e \in E_{\text{loop}} \) in \( \pi \) to a traversal of \((v(e), v'(e), v(e))\) in \( \pi' \) (i.e. a traversal of the edge \( g(e) \) back and forth in \( \pi' \)). The probability that the reinforced random walk on \( G \) traverses \( \pi \) agrees with the probability that the reinforced random walk on \( G' \) traverses \( \pi' \). (Note that for \( G \) and \( G' \) the following is true: Between any two successive visits to \( v(e) \), the sum of the weights of all edges incident to \( v(e) \) increases by 2.) Since the claim of the theorem is true for \( G' \), it follows that

\[
P_{v_0,a}(Z_n = \pi) = P_{v_0,b}(Z_n = \pi') = \int_{\Delta} \prod_{i=1}^{n'} \frac{x_{\{\pi_{i-1},\pi_i\}}'}{x_{\pi_i'}} \phi'_{v_0,b}(x) \, d\sigma(x), \tag{2.21}
\]

where \( \phi'_{v_0,b} \) denotes the density corresponding to \( G' \), starting point \( v_0 \), and initial weights \( b \). We claim that the right-hand side of (2.21) equals

\[
\int_{\Delta} \prod_{i=1}^{n} \frac{x_{\{\pi_{i-1},\pi_i\}}}{x_{\pi_i}} \phi_{v_0,a}(x) \, d\sigma(x). \tag{2.22}
\]

Note that a traversal of \( e \in E_{\text{loop}} \) contributes \( x_e/x_{v(e)} \) to the integrand in (2.22), whereas a traversal of \((v(e), v'(e), v(e))\) contributes \( x_{g(e)}/x_{v(e)} \) to the integrand in (2.21). Furthermore, \( e \in E_{\text{loop}} \) contributes

\[
\frac{\Gamma \left( \frac{a_e+1}{2} \right)}{\Gamma(a_e)} \cdot 2^{a_e} \cdot (x_e)^{\frac{a_e-1}{2}} \tag{2.23}
\]

to \( \phi_{v_0,a} \), whereas the contribution of the edge \( g(e) \) and the vertex \( v'(e) \) to the density \( \phi'_{v_0,b} \) equals

\[
\frac{\Gamma \left( \frac{a_{g(e)}+1}{2} \right)}{\Gamma(a_e)} \cdot 2^{a_e} \cdot \frac{x_{g(e)}^{a_e-\frac{1}{2}}}{(x_{v'(e)})^{a_{v'(e)}+\frac{1}{2}}} = \frac{\Gamma \left( \frac{a_e+1}{2} \right)}{\Gamma(a_e)} \cdot 2^{a_e} \cdot \frac{(x_{g(e)})^{a_e-\frac{1}{2}}}{(x_{v'(e)})^{a_{v'(e)}+\frac{1}{2}}}
\]

\[
= \frac{\Gamma \left( \frac{a_e+1}{2} \right)}{\Gamma(a_e)} \cdot 2^{a_e} \cdot (x_{g(e)})^{\frac{a_e-1}{2}}. \tag{2.24}
\]

Finally, \(|V| + |E_{\text{loop}}| = |V'| \) and \(|E| = |E'| \). Consequently, the expression in (2.22) agrees with the right-hand side of (2.21) and the claim follows. \( \blacksquare \)
Definition 2.4 Let $\mathcal{D} := \{P_{v_0,a} : v_0 \in V, a = (a_e)_{e \in E} \in (0, \infty)^E\}$ with $P_{v_0,a}$ as in Theorem 2.3.

We will prove in Section 4.2 that $\mathcal{D}$ is a natural family of prior distributions for reversible Markov chains.

3 The density $\phi_{v_0,a}$ for special graphs

In this section, we write down the densities $\phi_{v_0,a}$ for some special graphs.

3.1 The line graph (Birth and death chains)

Consider the line graph with vertex set $V = \{i : 0 \leq i \leq n\}$ and edge set $E = \{(i, i+1) : 0 \leq i \leq n - 1\}$. Given $a = (a_{i-1,i})_{1 \leq i \leq n}$, let $b_i := a_{i-1,i}$. The variables in the simplex $\Delta$ are denoted $z_i := x_{i-1,i}$.

Figure 2: The line graph

Recall that the density of the beta distribution with parameters $b_1, b_2 > 0$ is given by

$$\beta[b_1, b_2](p) := \frac{\Gamma(b_1 + b_2)}{\Gamma(b_1)\Gamma(b_2)} p^{b_1-1} (1-p)^{b_2-1} \quad (0 < p < 1).$$

Set

$$p_i = \frac{z_i}{z_i + z_{i+1}}, \quad 1 \leq i \leq n - 1,$$

and $p := (p_i)_{1 \leq i \leq n-1}$. Clearly $p_i$ is the probability that the Markov chain with edge weights $z_i$ makes a transition to $i-1$ given it is at $i$. If we make the change of variables (3.2) in the density $\phi_{v_0,a}$, then we obtain the transformed density $\tilde{\phi}_{v_0,a}$ given by

$$\tilde{\phi}_{v_0,a}(p) = \begin{cases} \prod_{i=1}^{n-1} \beta \left[ \frac{b_i + 1}{2}, \frac{b_{i+1}}{2} \right] (p_i) & \text{if } v_0 = 0, \\
\prod_{i=1}^{v_0-1} \beta \left[ \frac{b_i}{2}, \frac{b_{i+1} + 1}{2} \right] (p_i) \beta \left[ \frac{b_{v_0}}{2}, \frac{b_{v_0}}{2} \right] (p_{v_0}) \prod_{i=v_0+1}^{n-1} \beta \left[ \frac{b_i + 1}{2}, \frac{b_{i+1}}{2} \right] (p_i) & \text{if } v_0 \in \{1, 2, \ldots, n - 1\}, \\
\prod_{i=1}^{n-1} \beta \left[ \frac{b_i}{2}, \frac{b_{i+1} + 1}{2} \right] (p_i) & \text{if } v_0 = n; \end{cases}$$

here the empty product is defined to be 1.

With the change of variables (3.2), the conjugate prior can be described as a product of independent beta variables with carefully linked parameters. If loops are allowed, the edge weights are independent Dirichlet by a similar argument (see Section 3.2). The next example contains a generalization.
3.2 Trees with loops

Recall that the density of the Dirichlet distribution with parameters $b_i > 0$, $1 \leq i \leq d$ is given by

$$ D[b_i; 1 \leq i \leq d](p_i; 1 \leq i \leq d) := \frac{\Gamma\left(\sum_{i=1}^{d} b_i\right)}{\prod_{i=1}^{d} \Gamma(b_i)} \prod_{i=1}^{d} p_i^{b_i-1}, \quad \left( p_i \geq 0, \sum_{i=1}^{d} p_i = 1 \right). \quad (3.4) $$

Let $T = (V, E)$ be a tree. Suppose that there is a loop attached to every vertex, i.e. $\{v\} \in E$ for all $v \in V$. Let $v_0 \in V$. For every $v \in V \setminus \{v_0\}$ there exists a unique shortest path from $v_0$ to $v$. Let $e(v)$ be the unique edge incident to $v$ which is traversed by the shortest path from $v_0$ to $v$. Let $E_v := \{ e \in E : v \in e \}$ be the set of all edges incident to $v$. Set

$$ p_e := \frac{x_e}{x_v} \quad \text{for} \quad v \in V, e \in E_v, \quad (3.5) $$

$p := (p_e)_{e \in E}$, and $\tilde{p}_v := (p_e)_{e \in E_v}$. If we make the change of variables (3.5) in the density $\phi_{v_0,a}$, the transformed density $\tilde{\phi}_{v_0,a}$ is given by

$$ \tilde{\phi}_{v_0,a}(p) = D\left[\frac{a_e}{2}, e \in E_{v_0}\right](\tilde{p}_{v_0}) \prod_{v \in V \setminus \{v_0\}} D\left[\frac{a_e(v) + 1}{2}, a_e, e \in E_v \setminus \{e(v)\}\right](\tilde{p}_v). \quad (3.6) $$

Thus again, in the reparametrization (3.5), the conjugate prior is seen as a product of independent choices of edge weights. This is not true in the following example.

The fact that the density $\phi_{v_0,a}$ for a tree has this particular form, was first observed by Pemantle [Pem88].

3.3 The triangle

Consider the triangle with loops attached to all vertices. Let the vertex set be $V = \{0, 1, 2\}$ and the edge set $E = \{\{0\}, \{1\}, \{2\}, \{0, 1\}, \{0, 2\}, \{1, 2\}\}$ (see Figure 3). Let $b_i$ be the initial weight of the loop at vertex $i$ and let $c_i$ be the initial weight of the edge opposite of vertex $i$. Similarly, let $y_i := x_{i(1)}$ and let $z_0 := x_{\{1, 2\}}, z_1 := x_{\{0, 2\}}, z_2 := x_{\{0, 1\}}$.

Figure 3: The triangle with loops
The density $\phi_{0,a}$ for $a = (b_0, b_1, b_2, c_0, c_1, c_2)$ is given by

$$\phi_{0,a}(y_0, y_1, y_2, z_0, z_1, z_2) = Z_{0,a}^{-1} \cdot \frac{b_0^{y_0-1} b_1^{y_1-1} b_2^{y_2-1} z_0^{c_0-1} z_1^{c_1-1} z_2^{c_2-1}}{(z_1 + z_2)^{b_0+b_1+b_2} (z_0 + z_2)^{b_1+b_0+c_0+1} (z_0 + z_1)^{b_2+b_0+c_1+1}} \sqrt{z_0 z_1 + z_0 z_2 + z_1 z_2}$$  \hspace{1cm} (3.7)

with

$$Z_{0,a} = \frac{\Gamma(c_0) \Gamma(c_1) \Gamma(c_2) \Gamma\left(\frac{b_0}{2}\right) \Gamma\left(\frac{b_1}{2}\right) \Gamma\left(\frac{b_2}{2}\right)}{\Gamma\left(\frac{b_0+c_1+c_2}{2}\right) \Gamma\left(\frac{b_1+c_0+c_2+1}{2}\right) \Gamma\left(\frac{b_2+c_0+c_1+1}{2}\right)} \cdot \frac{480\pi}{2^{b_0+c_1+c_2}}.$$  \hspace{1cm} (3.8)

To calculate $Z_{0,a}$ from (2.14), use the identity

$$\frac{\Gamma\left(\frac{b_i}{2}\right)}{2^{b_i} \Gamma\left(\frac{b_i+1}{2}\right)} = \frac{\Gamma\left(\frac{b_i}{2}\right)}{2^{\sqrt{\pi}}} \quad (i = 0, 1, 2).$$  \hspace{1cm} (3.9)

For the triangle without loops, a derivation of the formula for the density $\phi_{0,a}$ can be found e.g. in [Kea90].

### 3.4 The complete graph

Perhaps the most important example is where all transitions are possible. This involves the complete graph $K_n$ on $n$ vertices with loops attached to all vertices. Let $V = \{1, 2, 3, \ldots, n\}$. Let $T_n$ be the spanning tree with edges $\{1, i\}$ and loops $\{i\}$, $1 \leq i \leq n$. This spanning tree induces the basis of cycles given by all triangles $(1, i, j)$, $2 \leq i < j \leq n$. Figure 4 shows $K_3, K_4$, and $K_5$ together with $T_3, T_4$, and $T_5$.

Figure 4: The complete graphs $K_3, K_4, K_5$ with loops together with a spanning tree

Cycles: $\{1, 2, 3\}$ $\{1, 2, 3, 4, 1, 3\}$ $\{1, 2, 3, 4, 1, 3, 5\}$ $\{1, 2, 3, 4, 1, 2, 5\}$ $\{1, 2, 3, 4, 1, 3, 5\}$ $\{1, 2, 3, 4, 1, 4, 5\}$

We remark that a different basis of cycles is given by $(i, i+1, j)$ for $1 \leq i < j + 1 \leq n$. This may be proved by induction using the Mayer-Vietoris decomposition theorem based on $K_{n-1}$ and a point.
Let \( a = (a_{i,j})_{1 \leq i,j \leq n} \) be given. For \( K_n \), set \( b_i := a_{i,i} \), \( a_i = \sum_{j=1}^{n} a_{i,j} \), and \( b := \sum_{1 \leq i,j \leq n} a_{i,j} \). The variables of the simplex are \( x = (x_{i,j})_{1 \leq i,j \leq n} \). Abbreviating \( y := x_{i,i} \) and \( x_i = \sum_{j=1}^{n} x_{i,j} \), the density \( \phi_{1,a} \) is given by

\[
\phi_{1,a}(x) = \frac{1}{Z_{1,a}} \cdot \prod_{1 \leq i < j \leq n} \frac{x_{i,j}^{a_{i,j}} - \frac{1}{2}}{x_i^{a_{i,i}}} \prod_{i=1}^{n} \frac{y_i^b - 1}{y_i} \sqrt{\det(A_n(x))} \tag{3.10}
\]

with \( A_n(x) \) defined in (2.12) and

\[
Z_{1,a} = \frac{\prod_{1 \leq i,j \leq n} \Gamma(a_{i,j})}{\Gamma(\frac{a_1}{2}) \prod_{i=2}^{n} \Gamma(\frac{a_i + 1}{2}) \prod_{i=1}^{n} \Gamma\left(\frac{a_i + 1}{2}\right)} \cdot \frac{(n(n+1) - 1)! \pi^{n - \frac{1}{2}}}{2^{1 - 2n + b}}. \tag{3.11}
\]

4 Properties of the family of priors

4.1 Closure under sampling

Recall the definition (2.5) of \( k_e(\pi) \) and recall that \( Z_n = (X_0, \ldots, X_n) \).

**Proposition 4.1** Under the prior distribution \( \mathbb{P}_{v_0,a} \) with observations \( X_0 = v_0, X_1 = v_1, \ldots, X_n = v_n \), the posterior is given by \( \mathbb{P}_{v_n, (a_e + k_e(Z_n))_{e \in E}} \). In particular, the family \( \mathcal{D} \) is closed under sampling.

**Proof.** Suppose we are given \( n + 1 \) observations \( \pi = (\pi_0, \pi_1, \ldots, \pi_n) \) sampled from \( \mathbb{P}_{v_0,a} \). Then \( \pi_0 = v_0 \). We claim that the posterior is given by \( \mathbb{P}_{\pi_n, (a_e + k_e(\pi))_{e \in E}} \). Recall that the posterior distribution is the \( \mathbb{P}_{v_0,a} \)-distribution of \( \{X_{n+k}\}_{k \geq 0} \) given \( Z_n = \pi \). By Theorem 2.3, \( \mathbb{P}_{v_0,a} = \mathbb{P}_{v_0,a} \). The \( \mathbb{P}_{v_0,a} \)-distribution of \( \{X_{n+k}\}_{k \geq 0} \) given \( Z_n = \pi \) is the distribution of an edge-reinforced random walk starting at the vertex \( \pi_n \) with initial values \( a_e + k_e(\pi) \). Using the identity (2.17) from Theorem 2.3 again, it follows that the posterior equals \( \mathbb{P}_{\pi_n, (a_e + k_e(\pi))_{e \in E}} \), which is an element of \( \mathcal{D} \). \( \blacksquare \)

4.2 Uniqueness

In this section, we give a characterization of our priors along the lines of W.E. Johnsons characterization of the Dirichlet prior. See Zabell [Zab82] for history and Zabell [Zab95] for a version for non-reversible chains. The closely related topic of de Finetti's theorem for Markov chains is developed by Diaconis and Freedman ([Fre62], [DF80]). See also Regazzini et al. [FLPR02].

**Definition 4.2** Two finite admissible paths \( \pi \) and \( \pi' \) are called equivalent if they have the same starting point and satisfy \( k_e(\pi) = k_e(\pi') \) for all \( e \in E \). We define \( P \) to be partially exchangeable if \( P(Z_n = \pi) = P(Z_n = \pi') \) for any equivalent paths \( \pi \) and \( \pi' \) of length \( n \).

For \( n \in \mathbb{N}_0 \) and \( v \in V \), define

\[
k_n(v) := |\{i \in \{0,1,\ldots,n\} : X_i = v\}|. \tag{4.1}
\]

It seems natural to take a class \( \mathcal{P} \) of distributions for \( Z_n \) with the following properties:
P 1 For all $P \in \mathcal{P}$ there exists $v_0 \in V$ such that $P(X_0 = v_0) = 1$.

P 2 For all $P \in \mathcal{P}$, $v_0$ as in P 1, and any admissible path $\pi$ of length $n \geq 1$ starting at $v_0$, we have $P(Z_n = \pi) > 0$.

P 3 Every $P \in \mathcal{P}$ is partially exchangeable.

P 4 For all $P \in \mathcal{P}$, $v \in V$, and $e \in E$ there exists a function $f_{P,v,e}$ taking values in $[0,1]$ such that for all $n \geq 0$

$$P(X_{n+1} = v|Z_n) = f_{P,v,e}(X_n, k_n, Z_n).$$

The condition P 4 says that given $X_0, X_1, \ldots, X_n$, the probability that $X_{n+1} = v$ depends only on the following quantities: the observation $X_n$, the number of times $X_n$ has been observed so far, the edge $\{X_n, v\}$, and the number of times transitions between $X_n$ and $v$ (and between $v$ and $X_n$) have been observed so far.

We make the following assumptions on the graph $G$:

G 1 For all $v \in V$ degree$(v) \neq 2$.

G 2 The graph $G$ is 2-edge-connected, i.e. removing an edge does not make $G$ disconnected.

For example, a triangle with loops or the complete graph $K_n$, $n \geq 4$, with or without loops satisfies G 1 and G 2 while a path fails both G 1 and G 2.

Recall that $Q_{v_0,a}$ is the distribution of the reversible Markov chain starting in $v_0$, making a transition from $v$ to $v'$ with probability proportional to $x_{v,v'}$ whenever $\{v, v'\} \in E$.

Theorem 4.3 Suppose the graph $G$ satisfies G 1 and G 2.

1. The set $\mathcal{D}$ defined in Definition 2.4 satisfies P 1 – P 4.

2. On the other hand, if P 1 – P 4 are satisfied for a set $\mathcal{P}$ of probability distributions, then for all $P \in \mathcal{P}$ there exist $v_0 \in V$ and $a \in (0,\infty)^E$ such that either

$$P(X_{n+1} = v|Z_n, k_n(X_n) \geq 3) = P_{v_0,a}(X_{n+1} = v|Z_n, k_n(X_n) \geq 3) \quad \text{for all } n \geq 0 \quad \text{or}$$

$$P(X_{n+1} = v|Z_n, k_n(X_n) \geq 3) = Q_{v_0,a}(X_{n+1} = v|Z_n, k_n(X_n) \geq 3) \quad \text{for all } n \geq 0. \quad (4.2)$$

The second part of the theorem states that either $P$ and $P_{v_0,a}$ or $P$ and $Q_{v_0,a}$ essentially agree; only the conditional probabilities to leave from a state which has been visited at most twice could be different.

Proof of Theorem 4.3. It is straightforward to check that $\mathcal{D}$ has the properties P 1 – P 4. For the converse, let $P \in \mathcal{P}$. By P 3, $P$ is partially exchangeable. Hence, if $G$ has no loops, then Theorem 1.2 of [Rol03] implies that there exist $v_0 \in V$ and $a \in (0,\infty)^E$ such that either (4.2) holds or $P(X_{n+1} = v|Z_n, k_n(X_n) \geq 3) = P_{v_0,a}(X_{n+1} = v|Z_n, k_n(X_n) \geq 3)$ for all $n$. In this case, the claim follows from (2.17).

If $G$ has loops, consider the graph $G'$ defined in the proof of Theorem 2.3 and the induced process $X' := (X'_n)_{n \in \mathbb{N}_0}$ on $G'$ with reflection at the vertices $v'(e)$, $e \in E_{\text{loop}}$. The process $X'$ satisfies P 1 – P 4. Hence, the claim holds for $X'$ and consequently for $(X_n)_{n \in \mathbb{N}_0}$. ■
Remark 4.4 The preceding theorem holds under the assumption that the graph $G$ is 2-edge-connected ($G \geq 2$). If $G$ is not 2-edge-connected, a similar statement can be proved for a different class of priors: One replaces the class $D$ defined in Definition 2.4 by the mixing measures of so-called modified edge-reinforced random walk; for the definition of this process see Definition 2.1 of [Rol03]. A uniqueness statement similar to Theorem 4.3 follows from Theorem 2.1 of [Rol03].

4.3 The priors are dense

As shown by Dalal and Hall [DH83] and Diaconis and Ylvisaker [DY85] for classical exponential families, mixtures of conjugate priors are dense in the space of all priors. This holds for reversible Markov chains.

Proposition 4.5 The set of convex combinations of priors in $D$ is weak-star dense in the set of all prior distributions on reversible Markov chains on $G$.

Proof. For an infinite admissible path $\pi = (\pi_0, \pi_1, \pi_2, \ldots)$ in $G$, define $\alpha(\pi) := (\alpha_e(\pi))_{e \in E}$ by $\alpha_e(\pi) := \lim_{n \to \infty} k_e(\pi_0, \pi_1, \ldots, \pi_n)/n$ to be the limiting fraction of crossings of the edge $e$ by the path $\pi$. Let $Z_\infty := (Z_0, Z_1, Z_2, \ldots)$. Note that $\alpha(Z_\infty)$ is defined $P_{v_0, a}$-a.s. Define $\tau_n$ to be the $n$th return time to $v_0$. Since $G$ is finite, $\tau_n < \infty P_{v_0, a}$-a.s. for all $n \in \mathbb{N}$ and all $a \in (0, \infty)^E$.

Let $f : \Delta \to \mathbb{R}$ be bounded and continuous. Denote the expectation with respect to $P_{v_0, a}$ by $E_{v_0, a}$. Since $X_{\tau_n} = v_0$, Proposition 4.1 implies that $P_{v_0, a}[\alpha_e + k_e(\tau_n)] = P_{v_0, a}[\alpha(\tau_n)] = M_n$. Hence,

$$E_{v_0, a} [f(\alpha(Z_\infty))] = E_{v_0, a} [f(\alpha(Z_\infty)) | Z_{\tau_n}] = M_n.$$  \hspace{1cm} (4.3)

Clearly, $(M_n)_{n \geq 1}$ is a bounded martingale. Hence, by the martingale convergence theorem,

$$\lim_{n \to \infty} E_{v_0, a} [f(\alpha(Z_\infty))] = E_{v_0, a} [f(\alpha(Z_\infty)) | Z_{\infty}] = f(\alpha(Z_\infty)) = \int f(\alpha(Z_\infty)) d\delta(\alpha(Z_\infty)).$$  \hspace{1cm} (4.4)

$P_{v_0, a}$-a.s.; here $\delta_a$ denotes the point mass in $a$. Since $\Delta$ is compact, there is a countable dense subset of the set of bounded continuous functions on $\Delta$. Hence, the above shows that for $P_{v_0, a}$-almost all $Z_\infty$,

$$P_{v_0, a}[\alpha(Z_\infty) \in - \delta_a(\alpha(Z_\infty))] \Rightarrow \delta_a(\alpha(Z_\infty)) \text{ weakly as } n \to \infty.$$  \hspace{1cm} (4.5)

The $P_{v_0, a}$-distribution of $\alpha(Z_\infty)$ is absolutely continuous with respect to Lebesgue measure on $\Delta$ with the density $\phi_{v_0, a}$ which is strictly positive in the interior of $\Delta$. Hence, for Lebesgue-almost all $a \in \Delta$ there is a sequence $a_n \in \Delta$ such that $P_{v_0, a_n}(\alpha(Z_\infty) \in \cdot) \Rightarrow \delta_a(\cdot)$ weakly. By the Krein-Milman theorem, convex combinations of point masses are weak-star dense in the set of all measures on $\Delta$. Using a standard argument it follows that the set of convex combinations of distributions of the form $P_{v_0, a}(\alpha(Z_\infty) \in \cdot)$ is dense in the set of all probability measures on $\Delta$. Consequently, the set of convex combinations of distributions of the form $P_{v_0, a}$ is dense in the set of all prior distributions.
4.4 Computing some moments

For any edge $e_0 \in E$, we can calculate the probability that the mixture of Markov chains with mixing measure $\phi_{v_0,a} \, d\sigma$ traverses $e_0$ back and forth starting at an endpoint of $e_0$. This gives a closed form for certain moments of the prior $P_{v_0,a}$.

**Proposition 4.6** For $e_0 \in E \setminus E_{loop}$ with endpoints $v$ and $v'$ we have

$$
\int_{\Delta} \frac{(x_{e_0})^2}{x_v x_{v'}} \phi_{v_0,a}(x) \, d\sigma(x) = \begin{cases} 
\frac{a_{e_0}(a_{e_0} + 1)}{(a_v + 1)(a_{v'} + 1)} & \text{if } v_0 \notin \{v, v'\}, \\
\frac{a_{e_0}(a_{e_0} + 1)}{a_v(a_{v'} + 1)} & \text{if } v = v_0.
\end{cases}
$$

(4.6)

For a loop $e_0 \in E_{loop}$ incident to $v$ we have

$$
\int_{\Delta} \frac{x_{e_0}}{x_v} \phi_{v_0,a}(x) \, d\sigma(x) = \begin{cases} 
\frac{a_{e_0}}{a_v + 1} & \text{if } v \neq v_0, \\
\frac{a_{e_0}}{a_v} & \text{if } v = v_0.
\end{cases}
$$

(4.7)

**Proof.** Case $e_0 \in E \setminus E_{loop}$: Suppose $v_0$ is an endpoint of $e_0$, say $v = v_0$. Then,

$$
\int_{\Delta} \frac{(x_{e_0})^2}{x_v x_{v'}} \phi_{v_0,a}(x) \, d\sigma(x) = P_{v_0,a}(X_0 = v_0, X_1 = v', X_2 = v_0);
$$

(4.8)

this is the probability that the mixture of Markov chains traverses the edge $e_0$ back and forth starting at $v_0$. By (2.17), $P_{v_0,a} = P_{v_0,a}$. Hence (4.8) equals the probability that an edge-reinforced random walk traverses $e_0$ back and forth, namely

$$
P_{v_0,a}(X_0 = v_0, X_1 = v', X_2 = v_0) = \frac{a_{e_0}(a_{e_0} + 1)}{a_v(a_{v'} + 1)}.
$$

(4.9)

Here we used that the sum of the weights of all edges incident to $v'$ equals $a_{v'} + 1$ after $e_0$ has been traversed once. This proves the claim in the case $v_0 \notin e_0$.

Suppose $v_0 \notin e_0$. Define $b := (b_e)_{e \in E}$ by $b_{e_0} := a_{e_0} + 2$ and $b_e := a_e$ for $e \in E \setminus \{e_0\}$. Then, using the definition of $\phi_{v_0,a}$, we obtain

$$
\frac{(x_{e_0})^2}{x_v x_{v'}} \phi_{v_0,a}(x) = \frac{Z_{v_0,b}}{Z_{v_0,a}} \phi_{v_0,b}(x) \quad \text{for all } x \in \Delta.
$$

(4.10)

Using the definition of the normalizing constants $Z_{v_0,a}$ and $Z_{v_0,b}$ and the identity $\Gamma(x + 1) = x \Gamma(x)$, it follows that

$$
\frac{Z_{v_0,b}}{Z_{v_0,a}} = \frac{\Gamma(\frac{a_{e_0} + 1}{2}) \Gamma(\frac{a_{v'} + 1}{2}) \Gamma(a_{e_0} + 2)}{4 \Gamma(\frac{a_{e_0} + 3}{2}) \Gamma(\frac{a_{v'} + 3}{2}) \Gamma(a_{e_0})} = \frac{a_{e_0}(a_{e_0} + 1)}{(a_v + 1)(a_{v'} + 1)}.
$$

(4.11)

Since $\int_{\Delta} \phi_{v_0,b}(x) \, d\sigma(x) = 1$, the claim follows by integrating both sides of (4.10) over $\Delta$.

**Case $e_0 \in E_{loop}$:** The proof follows the same line as in the case $e_0 \notin E_{loop}$. Let $e_0 = \{v\}$ be incident to $v$. We prove only the case $v \neq v_0$. Defining $b$ as above, (4.10) is valid with

$$
\frac{Z_{v_0,b}}{Z_{v_0,a}} = \frac{\Gamma(\frac{a_{e_0} + 1}{2}) \Gamma(\frac{a_{v'} + 1}{2}) \Gamma(a_{e_0} + 2)}{4 \Gamma(\frac{a_{e_0} + 3}{2}) \Gamma(\frac{a_{v'} + 3}{2}) \Gamma(a_{e_0})} = \frac{a_{e_0}(a_{e_0} + 1)}{(a_v + 1)(a_{v'} + 1)} = \frac{a_{e_0}}{(a_v + 1)};
$$

(4.12)
here we used again the identity $\Gamma(z + 1) = z\Gamma(z)$. The claim follows.

Recall the definitions (2.4) and (2.5) of $k_v(\pi)$ and $k_e(\pi)$ for a finite admissible path $\pi$ in $G$. Abbreviate $k_v := k_v(\pi)$, $k_e := k_e(\pi)$. For $x = (x_e)_{e \in E} \in \Delta$, denote by $Q_x(\pi)$ the probability that the reversible Markov chain with transition probabilities induced by the weights $(x_e)_{e \in E}$ on the edges traverses the path $\pi$. Note that if $\pi$ is a closed path, i.e. if the starting and endpoint of $\pi$ agree, then $Q_x(\pi)$ is independent of the starting point of $\pi$. The argument as in the proof of Proposition 4.6 gives

**Proposition 4.7** For any finite admissible path $\pi$ starting at $v_0$, we have

\[
\int_{\Delta} Q_x(\pi) \phi_{v_0,a}(x) \, d\sigma(x) = \prod_{e \in E \setminus E_{\text{loop}}} \frac{k_{e-1}}{k_{v_0-1}} \left( a_e + i \right)^{k_{e-1}} \prod_{e \in E_{\text{loop}}} \frac{k_{e/2-1}}{k_{u_0-1}} \left( a_e + 2i \right)^{k_{e/2}}.
\] (4.13)

For any finite admissible path $\pi$ with the same starting and endpoint which avoids $v_0$, we have

\[
\int_{\Delta} Q_x(\pi) \phi_{v_0,a}(x) \, d\sigma(x) = \prod_{v \in V \setminus \{v_0\}} \frac{k_{v-1}}{k_{u_0-1}} \left( a_v + 1 + 2i \right)^{k_{v-1}}.
\] (4.14)

Here the empty product is defined to be 1.

If $\pi$ is a closed path, we call $Q_x(\pi)$ a cycle probability. The transition probabilities of a Markov chain with finite state space $V$ that visits every state with probability 1 are completely determined by all its cycle probabilities (see e.g. [Fre62], Corollary on page 116).

**4.5 Simulating from the posterior**

In this subsection, we show how the posterior distribution of the unknown stationary distribution for the underlying Markov chain can be simulated using reinforced random walks.

Suppose our posterior distribution is $P_{v_0,a} = \phi_{v_0,a} \, d\sigma$. Let $X^{(i)} := (X^{(i)}_n)_{n \geq 0}$, $i \geq 1$, be independent reinforced random walks with the same initial edge weights $a = (a_e)_{e \in E}$. Let $Z^{(i)}_n := (X^{(i)}_0, X^{(i)}_1, \ldots, X^{(i)}_n)$ and recall that $k_e(Z^{(i)}_n)$ equals the number of traversals of edge $e$ by the process $X^{(i)}$ up to time $n$.

**Proposition 4.8** For any interval $I \subseteq \mathbb{R}$ and all $e \in E$, we have

\[
\lim_{n \to \infty} \lim_{m \to \infty} \frac{1}{m} \left\{ i \leq m : \frac{k_e(Z^{(i)}_n)}{n} \in I \right\} = P_{v_0,a}(x_e \in I).
\] (4.15)

**Proof.** For every $n$, the random variables $k_e(Z^{(i)}_n)/n$, $i \geq 1$, are i.i.d. Hence, by the Glivenko-Cantelli theorem, for all $x \in \mathbb{R}$,

\[
\lim_{m \to \infty} \frac{1}{m} \left\{ i \leq m : \frac{k_e(Z^{(i)}_n)}{n} \leq x \right\} = P_{v_0,a} \left( \frac{k_e(Z_n)}{n} \leq x \right) = P_{v_0,a} \left( \frac{k_e(Z_n)}{n} \leq x \right).
\] (4.16)
For the last equality, we used (2.17). Since $P_{v_0,a}$ is a mixture of Markov chains, $k_e(Z_n)/n$ converges to the normalized weight of the edge $e$ $P_{v_0,a}$-a.s. and hence weakly. Since the $P_{v_0,a}$-distribution of $x_e$ is continuous,

$$\lim_{n \to \infty} P_{v_0,a}\left(\frac{k_e(Z_n)}{n} \leq x\right) = P_{v_0,a}(x_e \leq x),$$

and the claim follows.

**Proposition 4.9** For all $e \in E$,

$$\lim_{n \to \infty} \int \frac{k_e(Z_n)}{n} dP_{v_0,a} = \int x_e dP_{v_0,a}.$$  \hspace{1cm} (4.18)

**Proof.** By (2.17), $P_{v_0,a} = P_{v_0,a}$. Since the proportion $k_n(e)/n$ converges $P_{v_0,a}$-a.s. to the normalized weight of the edge $e$, the claim follows from the dominated convergence theorem.

**Remark 4.10** The Markov chain with distribution induced by the edge weights $(x_e)_{e \in E} \in \Delta$ has the stationary distribution $\pi(v) = \frac{x_e}{2} = \frac{1}{2} \sum_{e \in E} x_e$. Thus, Propositions 4.8 and 4.9 allow simulation of the $P_{v_0,a}$-distribution and the mean of $\pi(v)$.

## 5 Applications

Reversibility can serve as a natural intermediate between independence and fully non-parametric Markovian dependence. On $|V|$ states, with no restrictions the number of free parameters is $|V| - 1$ with independence, $|V|(|V| - 1)$ for full Markov and $\frac{|V|(|V| - 1)}{2} - 1$ for reversibility. As Table 1 below indicates, these numbers vary widely for $|V|$ large.

**Table 1:** Degrees of freedom for independent, reversible, and full Markov specification

| $|V|$              | 3  | 4  | 5  | 10 | 20  | 50  | 100 | 1000 |
|-------------------|----|----|----|----|-----|-----|-----|------|
| independent $|V| - 1$ | 2  | 3  | 4  | 9  | 19  | 49  | 99  | 999  |
| reversible $|V|(|V| - 1)/2 - 1$ | 2  | 5  | 9  | 44 | 189 | 1224| 4949| 49949 |
| full Markov $|V|(|V| - 1)$ | 6  | 12 | 20 | 90 | 380 | 2450| 9900| 99900 |

In this section, we illustrate the use of our priors for testing a variety of simple hypotheses. As a first numerical illustration, we use a small data set (Table 2) generated by asking a student to generate a random sequence of length 100 with the symbols “red (R), white (W), blue (B).” The data is listed in order left to right and from top to bottom. It is well known that human generated sequences of this type fail simple tests of randomness (Bar-Hillel and Wagenaar [BHW91]). This example is simply illustrative.

Table 4 shows a genetic data set from the DNA sequence of the humane HLA-B gene. This gene plays a central role in the immune system. The data displayed in Table 4 is downloaded from the webpage of the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genome/guide/human/).

In Example a, we test i.i.d. $\frac{1}{3}$ versus i.i.d. for the data from Table 2 and we test i.i.d. $\frac{1}{3}$ versus i.i.d. for the DNA-data. In Example b, we test i.i.d. versus reversible. In Example c, we test reversible versus full Markov. In Example d, we compare i.i.d. with full Markov.
Table 2: Sequence of length 100 consisting of “R”, “W”, “B”

RRWBRWWBRR
WBRWWBBRBWR
BBRBRWBRWR
BRRWBWRWR
BRBWBWRWR

Let $n_R$, $n_W$, and $n_B$ denote the number of occurrences of $R$, $W$, and $B$, respectively. Then

$$n_R = 34, \quad n_W = 34, \quad n_B = 32.$$

Table 3: Occurrences $N_{ij}$ of the string $ij$ for $i, j \in \{R, W, B\}$

<table>
<thead>
<tr>
<th></th>
<th>R</th>
<th>W</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>5</td>
<td>24</td>
<td>4</td>
</tr>
<tr>
<td>W</td>
<td>7</td>
<td>7</td>
<td>20</td>
</tr>
<tr>
<td>B</td>
<td>21</td>
<td>3</td>
<td>8</td>
</tr>
</tbody>
</table>

Table 4: The human HLA-B gene. Part of the DNA sequence of length 3370.

```
1  ratagttaga  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc
61  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
121  ctgggaatgg  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc
181  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
241  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct  tccggcttc
301  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct  tccggcttc
361  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
421  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
481  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
541  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
601  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
661  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
721  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
781  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
841  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
901  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
961  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1021  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1081  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1141  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1201  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1261  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1321  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1381  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1441  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1501  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1561  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1621  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1681  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1741  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1801  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1861  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1921  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
1981  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2041  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2101  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2161  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2221  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2281  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2341  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2401  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2461  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2521  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2581  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2641  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2701  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2761  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2821  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2881  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
2941  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
3001  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
3061  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
3121  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
3181  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
3241  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
3301  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
3361  gataggtatg  cagcgctagc  tccggcttc  ggggaagtct  tccggcttc  ggggaagtct
```
Let \( n_a, n_c, n_g, \) and \( n_t \) denote the number of occurrences of \( a, c, g, \) and \( t, \) respectively. Then
\[
n_a = 621, \quad n_c = 974, \quad n_g = 1064, \quad n_t = 711.
\]

Table 5: Occurrences \( N_{ij} \) of the string \( ij \) for \( i, j \in \{ a, c, g, t \} \)

<table>
<thead>
<tr>
<th></th>
<th>( a )</th>
<th>( c )</th>
<th>( g )</th>
<th>( t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a )</td>
<td>91</td>
<td>160</td>
<td>261</td>
<td>108</td>
</tr>
<tr>
<td>( c )</td>
<td>213</td>
<td>351</td>
<td>161</td>
<td>249</td>
</tr>
<tr>
<td>( g )</td>
<td>251</td>
<td>224</td>
<td>388</td>
<td>201</td>
</tr>
<tr>
<td>( t )</td>
<td>66</td>
<td>239</td>
<td>254</td>
<td>152</td>
</tr>
</tbody>
</table>

**Example 1a.** A Bayes test of \( H_0: \text{i.i.d.}(\frac{1}{3}) \) versus \( H_1: \text{i.i.d.}(\text{unknown}) \) for the data from Table 2.

A 'standard' test can be based on the Bayes factor
\[
\frac{P(\text{data}|H_0)}{P(\text{data}|H_1)}. 
\]

See Good [Goo68] for an extensive discussion. For \( H_1, \) we use a Dirichlet(1,1,1) prior. This gives
\[
P(\text{data}|H_0) = \left( \frac{1}{3} \right)^{100} \approx 1.94033 \cdot 10^{-48},
\]
\[
P(\text{data}|H_1) = \frac{\Gamma(3)\Gamma(n_R + 1)\Gamma(n_W + 1)\Gamma(n_B + 1)}{\Gamma(n_R + n_W + n_B + 3)} = \frac{\Gamma(3)\Gamma(35)\Gamma(35)\Gamma(33)}{\Gamma(103)} \approx 4.77096 \cdot 10^{-50},
\]
and the bayes factor equals
\[
\frac{P(\text{data}|H_0)}{P(\text{data}|H_1)} \approx 41.
\]

This is not surprising since the observed number of \( R, W, B \) are 34, 34, 32, respectively.

**Example 2a.** A Bayes test of \( H_0: \text{i.i.d.}(\frac{1}{4}) \) versus \( H_1: \text{i.i.d.}(\text{unknown}) \) for the DNA-data.

In this case, we use a Dirichlet(1,1,1,1) prior for \( H_1. \) This yields
\[
P(\text{data}|H_0) = \left( \frac{1}{4} \right)^{3370} \approx 1.142429015368253 \cdot 10^{-2029},
\]
\[
P(\text{data}|H_1) = \frac{\Gamma(4)\Gamma(n_a + 1)\Gamma(n_c + 1)\Gamma(n_g + 1)\Gamma(n_t + 1)}{\Gamma(n_a + n_c + n_g + n_t + 4)} 
= \frac{\Gamma(4)\Gamma(622)\Gamma(975)\Gamma(1065)\Gamma(712)}{\Gamma(3374)} \approx 1.140417804695619 \cdot 10^{-1999},
\]
and hence the Bayes factor equals
\[
\frac{P(\text{data}|H_0)}{P(\text{data}|H_1)} \approx 1.00176 \cdot 10^{-30}.
\]
Thus, $H_0$ is strongly rejected. Again, this is not surprising since the observed number of $a$, $c$, $g$, $t$ are $n_a = 621$, $n_c = 974$, $n_g = 1064$, $n_t = 711$, respectively.

**Example 1b.** A Bayes test of $H_0$: i.i.d.(unknown) versus $H_1$: reversible for the data from Table 2.

Here we use a Dirichlet$(1,1,1)$ prior for the null hypothesis and the prior based on the complete graph $K_3$ with loops (see Figure 3) and all edge weights equal to 1. Then, $P(data|H_0)$ is as in Example 1a. In order to calculate $P(data|H_1)$, we first determine the transition counts $k_e$ for our data (see Table 3):

$k_{\{R,W\}} = 31$, $k_{\{R,B\}} = 25$, $k_{\{B,W\}} = 23$, $k_{\{R\}} = 10$, $k_{\{W\}} = 14$, $k_{\{B\}} = 16$,

and also $k_v = n_v - \delta_R(v)$: $k_R = 33$, $k_W = 34$, $k_B = 32$. Using the first part of Proposition 4.7, we obtain

$$P(data|H_1) = \frac{\prod_{e \in \{(R,W),(R,B),(B,W)\}} k_e^{(1+\delta(e))} \prod_{j \in \{W,B\}} k_j^{(4+2i)}}{\prod_{i=0}^{k_R-1} (3+2i) \prod_{j \in \{W,B\}} \prod_{i=0}^{k_j-1} (4+2i)}$$

$$= \frac{(31)! (25)! (23)! 4^{6} (1+2i)^{7} \prod_{i=0}^{6} (1+2i) \prod_{i=0}^{7} (1+2i)}{32 \prod_{i=0}^{32} (3+2i) \prod_{i=0}^{33} (4+2i) \prod_{i=0}^{31} (4+2i)}$$

$$\approx 2.63663 \cdot 10^{-49}$$

So the Bayes factor is

$$\frac{P(data|H_0)}{P(data|H_1)} \approx 0.180949$$

and the null hypothesis is rejected.

**Example 2b.** A Bayes test of $H_0$: i.i.d.(unknown) versus $H_1$: reversible for the DNA-data.

Here we use a Dirichlet$(1,1,1,1)$ prior for the null hypothesis and the prior based on the complete graph $K_4$ with loops (see Figure 4) and all edge weights equal to 1. The probability $P(data|H_0)$ is calculated in Example 2a. In order to calculate $P(data|H_1)$, we first determine the transition counts $k_e$ for our data (see Table 6) and also $k_v = n_v - \delta_a(v)$:

$k_a = 620$, $k_c = 974$, $k_g = 1064$, $k_t = 711$. \hspace{1cm} (5.1)

<table>
<thead>
<tr>
<th>a</th>
<th>c</th>
<th>g</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>182</td>
<td>373</td>
<td>512</td>
<td>174</td>
</tr>
<tr>
<td>373</td>
<td>702</td>
<td>385</td>
<td>488</td>
</tr>
<tr>
<td>512</td>
<td>385</td>
<td>776</td>
<td>455</td>
</tr>
<tr>
<td>174</td>
<td>488</td>
<td>455</td>
<td>304</td>
</tr>
</tbody>
</table>

Table 6: The undirected transition counts $k_{\{i,j\}}$, $i,j \in \{a,c,g,t\}$
We abbreviate \( E' = \{\{a, c\}, \{a, g\}, \{a, t\}, \{c, g\}, \{c, t\}, \{g, t\}\} \). By the first part of Proposition 4.7,

\[
P(data|H_1) = \frac{\prod_{e \in E'} \prod_{i=0}^{k_e-1} (1 + i) \prod_{j \in \{a, c, g, t\}} \prod_{i=0}^{k_{ij}/2-1} (1 + 2i)}{\prod_{i=0}^{k_1-1} (4 + 2i) \prod_{j \in \{a, c, g\}} \prod_{i=0}^{k_{ij}-1} (5 + 2i)} = \frac{(373)!(512)!(174)!(385)!(488)!(455)! \prod_{i=0}^{90} (1 + 2i) \prod_{i=0}^{350} (1 + 2i) \prod_{i=0}^{387} (1 + 2i) \prod_{i=0}^{151} (1 + 2i)}{\prod_{i=0}^{710} (4 + 2i) \prod_{i=0}^{619} (5 + 2i) \prod_{i=0}^{973} (5 + 2i) \prod_{i=0}^{1063} (5 + 2i)} \approx 2.166939224648291 \cdot 10^{-1961}.\]

So the Bayes factor is

\[
\frac{P(data|H_0)}{P(data|H_1)} \approx 5.2628 \cdot 10^{-39}
\]

and the null hypothesis is strongly rejected.

**Example 1c.** A Bayes test of \( H_0 \): reversible versus \( H_1 \): full Markov for the data from Table 2.

Here we use our conjugate prior on reversible chains with all constants chosen as one. We use product Dirichlet measure for the rows in the full Markov case. Now,

\[
P(data|H_1) = \frac{\prod_{i} \Gamma(3)\Gamma(6)\Gamma(25)\Gamma(5)}{\Gamma(36)} \cdot \frac{\prod_{i} \Gamma(3)\Gamma(8)\Gamma(8)\Gamma(21)}{\Gamma(37)} \cdot \frac{\prod_{i} \Gamma(3)\Gamma(22)\Gamma(4)\Gamma(9)}{\Gamma(35)}
\approx 9.62182 \cdot 10^{-42}.
\]

\( P(data|H_0) \) was calculated in Example 1b. Hence,

\[
\frac{P(data|H_0)}{P(data|H_1)} \approx 2.74026 \cdot 10^{-8}.
\]

We see that a straight-forward Bayes test strongly rejects reversibility. This is not surprising since the transition counts are far from symmetric.

**Example 2c.** A Bayes test of \( H_0 \): reversible versus \( H_1 \): full Markov for the DNA-data.

Again, we use our conjugate prior on reversible chains with all constants chosen as one. We use product Dirichlet measure for the rows in the full Markov case. This yields

\[
P(data|H_1) = \prod_{i \in \{a, c, g, t\}} \frac{\Gamma(4)\prod_{j \in \{a, c, g, t\}} \Gamma(N_{ij} + 1)}{\Gamma(k_i + 4)} \approx \frac{\Gamma(4)^4\Gamma(92)\Gamma(161)\Gamma(262)\Gamma(109)\Gamma(321)\Gamma(352)\Gamma(162)\Gamma(250)}{\Gamma(624)\Gamma(798)}
\cdot \frac{\Gamma(252)\Gamma(225)\Gamma(389)\Gamma(202)\Gamma(67)\Gamma(240)\Gamma(255)\Gamma(153)}{\Gamma(1068)\Gamma(715)}
\approx 4.16382063735625 \cdot 10^{-1956}.
\]
The probability \( P(data|H_0) \) was calculated in Example 2b. Hence,
\[
\frac{P(data|H_0)}{P(data|H_1)} \approx 5.20421 \cdot 10^{-6}.
\]
We see that a straight-forward Bayes test rejects reversibility.

**Example 1d.** A Bayes test of \( H_0: \text{i.i.d. (unknown)} \) versus \( H_1: \text{full Markov for the data from Table 2} \).

Using the Bayes factors computed above, we see strong rejection of i.i.d. versus Markov:
\[
\frac{P(data|H_0)}{P(data|H_1)} \approx 4.95848 \cdot 10^{-9}.
\]

Of course, an i.i.d. process is a reversible Markov chain.

**Example 2d.** A Bayes test of \( H_0: \text{i.i.d. (unknown)} \) versus \( H_1: \text{full Markov for the DNA-data} \).

Again, with the Bayes factors as above, the null hypothesis is strongly rejected:
\[
\frac{P(data|H_0)}{P(data|H_1)} \approx 2.73887 \cdot 10^{-44}.
\]

In using the Dirichlet prior for testing uniformity with multinomial data and for testing independence in contingency tables I.J. Good found the symmetric Dirichlet prior with density proportional to \( \prod_{i=1}^{d} 2^{-1} \) an important tool. Good’s many insights into these testing problems may be accessed through his book [Goo68] and the survey article [GC87].

We have used the analog of the symmetric Dirichlet for the reversible Markov chain context with all edge weights \( a_e \) equal to a constant \( c \) say. As \( c \) tends to infinity, this prior tends to a point mass supported on the simple random walk on the graph. As \( c \) tends to zero this prior tends to an improper prior which gives the maximum likelihood as its posterior.

Good also worked with \( c \)-mixtures of symmetric Dirichlet priors. We suspect that parallel, useful things can be done in our case as well.

We have not found any literature about statistical analysis of reversible Markov chains with unknown transitions and append two data analytic remarks here. First, under reversibility, the count \( N(v, v') \) of \( v \) to \( v' \) transitions has the same expectation as the count \( N(v', v) \) of \( v' \) to \( v \) transitions, namely \( \pi(v)k(v, v') \). This suggests looking at ratios \( N(v, v')/N(v', v) \) or differences \( N(v, v') - N(v', v) \). For example, from Table 3, \( \frac{N_{RW}}{N_{WR}} = \frac{24}{7}, \frac{N_{RA}}{N_{AR}} = \frac{4}{21}, \frac{N_{RW}}{N_{BW}} = \frac{20}{3} \); these are way off.

In large samples, these counts have limiting normal distributions by results of Höglund [Hög74]. A second data analytic tool would be to estimate the stationary distribution (perhaps by the method of moments estimator \( \hat{k}(v) = \frac{1}{n}\sum_{i \leq n} X_i = v \)) and also estimate the transition matrix, and then compare \( \hat{k}(v)\hat{k}(v, v') \) with \( \hat{k}(v')\hat{k}(v, v') \).

An interesting problem not tackled here is finding natural priors on the set of reversible Markov chains with a fixed stationary distribution. For definiteness, consider the uniform stationary distribution. Then the problem is to put a prior on \( S(n) \), the symmetric doubly stochastic \( n \times n \) matrices. We make two remarks. First, determining the Euclidean volume of \( S(n) \) is a
long-standing open problem, see Clara Chan and al. [CRY00] for recent results. Second, \( S(n) \) is a compact, convex subset of \( \mathbb{R}^{\frac{n^2}{2}} \). Its extreme points are well known to be the symmetrized permutation matrices (see Stanley [Sta78]). Thus, if \( \pi \) is a permutation matrix on \( n \) letters with \( e(\pi) \) the usual \( n \times n \) permutation matrix, let \( \tilde{e}(\pi) = \frac{1}{2}[e(\pi) + e(\pi^{-1})] \). The extreme points of \( S(n) \) are \( (\tilde{e}(\pi)) \) as \( \pi \) ranges over permutations in \( S_n \). We may put a prior on \( S(n) \) by taking a random convex combination of the \( \tilde{e}(\pi) \). Alas, \( S(n) \) is not a simplex, so symmetric weights on the extreme points may not lead to symmetric measures on \( S(n) \).

Acknowledgement: We would like to thank Franz Merkl for some interesting discussions.

References


