THIELE CENTRE
FOR APPLIED MATHEMATICS IN NATURAL SCIENCE

Summary statistics for end-point conditioned continuous-time Markov chains

Asger Hobolth and Jens Ledet Jensen

Research Report

No. 07 | August 2010

Summary statistics for end-point conditioned continuous-time Markov chains

Asger Hobolth[1] and Jens Ledet Jensen[1]

[1]*Department of Mathematical Sciences, Aarhus University, Ny Munkegade 118, Building 1530, DK-8000 Aarhus C, Denmark. E-mail: {asger,jlj}@imf.au.dk*

Abstract

Continuous-time Markov chains are a widely used modelling tool. Applications include DNA sequence evolution, ion channel gating behavior and mathematical finance. We consider the problem of calculating properties of summary statistics (e.g. mean time spent in a state, mean number of jumps between two states and the distribution of the total number of jumps) for discretely observed continuous time Markov chains. Three alternative methods for calculating properties of summary statistics are described and the pros and cons of the methods are discussed. The methods are based on (i) an eigenvalue decomposition of the rate matrix, (ii) the uniformization method, and (iii) integrals of matrix exponentials. In particular we develop a framework that allows for analyses of rather general summary statistics using the uniformization method.

Key words: Continuous-time Markov chain, dwelling time, EM algorithm, transition number, uniformization.

Mathematics subject classification 2000: 60-08 (computational methods); 60J22 (computational methods in Markov chains); 60J27 (Markov chains with continuous parameter).

1 Motivation and background

We review and extend three ways of calculating conditional properties of summary statistics of a continuous time Markov chain (CTMC). The conditioning is with respect to the starting point $x(0)$ and the ending point $x(T)$ of a process $x(t)$ considered in the interval $0 \leq t \leq T$. Typical summary statistics are either the time T_α spent in a state α or the number of transitions $N_{\alpha\beta}$ from state α to state β. Since, generally, for a statistic H, we have that

$$\mathrm{E}[H|x(0)=a,x(T)=b] = \frac{\mathrm{E}[H \cdot 1(x(T)=b))|x(0)=a]}{P(x(T)=b|x(0)=a)},$$

we formulate our results through terms of the form $\mathrm{E}[H \cdot 1(x(T)=b))|x(0)=a]$. The first approach to calculating the conditional properties is through an eigenvalue

decomposition of the rate matrix of the process, and will only be mentioned briefly (Section 2). The second approach (Section 3), based on the uniformization method, will be dealt with in more detail, and in particular we derive new formulas for various covariance terms. The uniformization method is the most general of the three methods as it allows for calculation of moments and the distribution itself. The third approach (Section 4) seems to be less known and is based on Van Loan's (1978) method of calculating integrals involving the matrix exponential.

In this introductory section we describe a few applications where conditional properties of summary statistics are needed. First, we review the very important problem of estimating the transition rates of a CTMC from observations at a finite set of time points using the expectation-maximization (EM) algorithm (Subsection 1.1). We identify five end-point conditioned mean values needed in the EM algorithm. Second, we consider summary statistics arising from the use of CTMCs in molecular biology (Subsection 1.2).

1.1 Summary statistics needed in the EM algorithm

We start by considering a CTMC $x(t)$ with states $\{1, 2, \ldots, m\}$ defined by a rate matrix $Q = (q_{ij})$, where $q_{ii} = -\sum_{j \neq i} q_{ij}$. When the process is observed at discrete time points $s_1 < s_2 < \cdots < s_k = \tau$ only, maximization of the likelihood function is often done via the EM-algorithm. Let $y = (x(s_1), \ldots, x(s_k))$ be the actual observations. As we describe below the expectation step of the algorithm gives rise to end-point conditioned mean values. Furthermore, for finding the asymptotic variance of the maximum likelihood estimates we need end-point conditioned second order moments. The log likelihood function $\ell(Q)$ based on continuous observations in the interval $0 \leq t \leq \tau$ is given by

$$\ell(Q) = \sum_{\alpha=1}^{m} q_{\alpha\alpha} T_\alpha(0, \tau) + \sum_{\alpha \neq \beta} q_{\alpha\beta} N_{\alpha\beta}(0, \tau),$$

where $T_\alpha(0, \tau)$ is the time spent in state α and $N_{\alpha\beta}(0, \tau)$ is the number of jumps from state α to state β. In the expectation step of the EM-algorithm we calculate $\mathrm{E}[\ell(Q)|y]$. Let $A(0, \tau)$ be either $T_\alpha(0, \tau)$ or $N_{\alpha\beta}(0, \tau)$. We must then calculate $\mathrm{E}[A(0, \tau)|y]$. Since A is an additive statistics $A(0, \tau) = \sum_{i=1}^{k} A(s_{i-1}, s_i)$ and because of the Markov property, we have that $\mathrm{E}[A(0, \tau)|y] = \sum_{i=1}^{k} \mathrm{E}[A(s_{i-1}, s_i)|x(s_{i-1}), x(s_i)]$. Thus, the EM-algorithm requires the calculation of end-point conditioned mean values.

For calculating the asymptotic variance of the maximum likelihood estimates we also need second order moments of the form $\mathrm{E}[A(0, \tau)\tilde{A}(0, \tau)|y]$, where again A and $\tilde{A}$ are additive statistics of the form T_α or $N_{\alpha\beta}$ (see e.g. Hobolth and Jensen, 2005). Once more because of the additive structure and the Markov property we end up with end-point conditioned second order moments of the form

$$\mathrm{E}[A(s_{i-1}, s_i)\tilde{A}(s_{i-1}, s_i)|x(s_{i-1}), x(s_i)].$$

Summarizing, inference in a CTMC via the EM-algorithm requires the calcula-

tion of the following end-point conditioned mean values:

$$\begin{array}{ll} \text{(i)} & \mathrm{E}[T_\alpha 1(x(T)=b)|x(0)=a] \\ \text{(ii)} & \mathrm{E}[N_{\alpha\beta} 1(x(T)=b)|x(0)=a] \\ \text{(iii)} & \mathrm{E}[T_\alpha T_\beta 1(x(T)=b)|x(0)=a] \\ \text{(iv)} & \mathrm{E}[N_{\alpha\beta} N_{\gamma\delta} 1(x(T)=b)|x(0)=a] \\ \text{(v)} & \mathrm{E}[N_{\alpha\beta} T_\gamma 1(x(T)=b)|x(0)=a] \end{array} \tag{1}$$

All the above mentioned mean values have integral representations which are linear combinations of the following terms (see e.g. Hobolth and Jensen, 2005):

$$I(a,b,\alpha,\beta) = \int_0^T P_{a\alpha}(t) P_{\beta b}(T-t) dt, \tag{2}$$

$$I(a,b,\alpha,\beta,\gamma,\delta) = \int_0^T \int_0^t P_{a\alpha}(u) P_{\beta\gamma}(t-u) P_{\delta b}(T-t) du dt, \tag{3}$$

where $P_{\alpha\beta}(t) = (\exp(Qt))_{\alpha\beta}$ is the transition probability. In particular we note that the mean value (i) in (1) of the time spent in state α is given by $I(a,b,\alpha,\alpha)$ and the mean value (ii) in (1) of the number of jumps from state α to state β is given by $q_{\alpha\beta} I(a,b,\alpha,\beta)$.

Further discussion of transition rate estimation based on incomplete observations can be found in Metzner, Horenko and Schütte (2007) and the references therein. Bladt and Sørensen (2009) describe an application in mathematical finance. They consider the problem of estimating transition rates between credit ratings from observations at discrete points (e.g. weekly observations). Holmes and Rubin (2002), Hobolth and Jensen (2005) and Kosiol, Holmes and Goldman (2007) use the EM algorithm for analysis of DNA sequence data. The DNA sequences are observed at present day and related by a phylogenetic tree. Continuous-time Markov chains are not only an important modelling tool in mathematical finance and molecular evolution. Ball and Milne (2005) describe how insights into the gating mechanism of a single ion channel can be obtained by modelling the system using a continuous-time Markov chain on a finite state space.

1.2 Further summary statistics used in applications of CTMCs

Let R denote a set of transitions of interest and let $N_R = \sum_{(\alpha,\beta)\in R} N_{\alpha\beta}$ be the number of such transitions. Minin and Suchard (2008) describe how the distribution of N_R is of interest in evolutionary developmental biology. They consider in particular a CTMC along a small evolutionary tree with observations at the tips of the tree. For the case of a two state Markov chain they derive closed form expressions for $P(N_R = k, x(T) = b|x(0) = a)$. For the general case they consider moments of N_R. As an example they consider the mean number of transitions and transversions within a small multiple alignment of DNA sequences. Note that the moments of N_R can be found through the moments of $N_{\alpha\beta}$, $(\alpha,\beta) \in R$. As a step in their analysis Minin and Suchard (2008) write the moment generating function of N_R through a

matrix exponential. More generally, the joint Laplace transform and moment generating function of all the T_α's and $N_{\alpha\beta}$'s is given as a matrix exponential in Bladt *et al.* (2002).

In an evolutionary setting Siepel, Pollard and Haussler (2006) also consider the distribution $P(N_R = k, x(T) = b)|x(0) = a)$ for the case where R is the set of all transitions. These authors use the statistic N_R to detect lineages in a phylogenetic tree that are under selection in a specific genomic region.

2 The eigenvalue decomposition method

The transition matrix $P(t)$ is given by the matrix exponential $P(t) = \exp(Qt)$. If Q is diagonalizable with real eigenvalues, $Q = U\Lambda U^{-1}$ where Λ is a diagonal matrix, the integrals in (2) and (3) can be evaluated easily. These formulas were given in Holmes and Rubin (2002) for (2) and Hobolth and Jensen (2005) for (3) and correspond to the formulas in Minin and Suchard (2008) for the statistic N_R using that $N_R = \sum_{(\alpha,\beta)\in R} N_{\alpha\beta}$. A reversible process always admits an eigenvalue decomposition with real eigenvalues.

Hobolth and Jensen (2005) also consider the case where some of the eigenvalues are complex numbers and illustrate the results for the case of four states, corresponding to the four possible nucleotides {`A,G,C,T`}. When the rate matrix Q is no longer diagonalizable a possibility is to use a Jordan decomposition. However, in this case the calculations for evaluating the integrals (2) and (3) become more involved.

3 The uniformization method

The uniformization method was originally introduced as a way of calculating $P(t) = \exp(Qt)$ (Jensen, 1953). Let $\mu = \max_i(-q_{ii})$, and define the transition matrix $R = \frac{1}{\mu}Q + I$. Then

$$\exp(Qt) = \exp\left(\mu(R - I)t\right) = \sum_{n=0}^{\infty} R^n \frac{(\mu t)^n}{n!} e^{-\mu t} = \sum_{n=0}^{\infty} R^n \,\mathrm{Pois}(n; \mu t), \qquad (4)$$

where $\mathrm{Pois}(n; \lambda)$ is the probability from a Poisson distribution with mean λ. More fundamentally, the uniformization method gives rise to an alternative description of the process itself. Let $z_0, z_1, \ldots$ be a Markov chain with transition matrix R. Independent of the chain let $0 = T_0 < T_1 < T_2 < \cdots$ be the times of a Posison process with rate μ. Next, we define $\{x(t), t \geq 0\}$, by setting $x(t) = z_k$ in the time interval $T_k \leq t < T_{k+1}$. It is clear that this is a continuous time Markov chain with rates $\mu R_{ij} = q_{ij}$, $j \neq i$.

Consider now a statistic H defined on $\{x(t), 0 \leq t \leq T\}$, which can be written as a function of the number of Poisson events J, the times $0 = T_0 < T_1 < \cdots < T_J < T$, and the states $z_0, z_1, \ldots, z_J$. To study the properties of H we condition on the value of J, and use that the times are the ordered values of uniformly distributed times and that the states are from a Markov chain with transition matrix R. Also, conditioned

on $J = n$, we generally use a recursion in n to evaluate the properties of H. As an illustration, consider the original use of the uniformization method

$$P(x(T) = b|x(0) = a)$$
$$= \sum_{n=0}^{\infty} P(z_n = b|z_0 = a)\,\mathrm{Pois}(n;\mu T) = \sum_{n=0}^{\infty} (R^n)_{ab}\,\mathrm{Pois}(n;\mu T), \tag{5}$$

where R^n can be calculated by the recursion $R^{n+1} = (R^n)R$.

What kind of general statistic H do we want to consider? The class of statistics must contain the time spent in a state $T_\alpha = \sum_{i=0}^{J} 1(z_i = \alpha)(T_{i+1} - T_i)$, where $T_{J+1} = T$, as well as the number of transitions between two states $N_{\alpha\beta} = \sum_{i=1}^{J} 1((z_{i-1}, z_i) = (\alpha, \beta))$. These two cases are covered by a general statistic of the form

$$H = \psi(z_0)f(T_1) + \sum_{i=1}^{J} \phi(z_{i-1}, z_i)f(T_{i+1} - T_i), \tag{6}$$

where $f(t) = t$ and $\phi(z_1, z_2) = \psi(z_2) = 1(z_2 = \alpha)$ for T_α, and $f \equiv 1$, $\psi(z) \equiv 0$ and $\phi(z_1, z_2) = 1((z_1, z_2) = (\alpha, \beta))$ for $N_{\alpha\beta}$. Note also, that the general form of H in (6) need not be an additive statistic.

3.1 Computing mean values using uniformization

We want to find the mean of a general statistic $\mathrm{E}[H \cdot 1(x(T) = b)|x(0) = a]$ with H defined in (6). We are particularly interested in the two special cases (i) where $H = T_\alpha$ and (ii) where $H = N_{\alpha\beta}$. In Theorem 1 we treat the general case and in Corollary 2 we consider the two special cases. Theorem 1 shows how the uniformization method allows a division of the conditional mean problem into three components that are each easy to handle. The components are concerned with properties of the number of jumps n, the inter arrival times $T_{i+1} - T_i$, and the discrete Markov chain $(z_0, \ldots, z_n)$.

First we introduce some notation. Let $M(n)$ be the $m \times m$ matrix with entries

$$M(n, a, b) = \mathrm{E}\Big[1(z_n = b)\big(\psi(z_0) + \sum_{i=1}^{n} \phi(z_{i-1}, z_i)\big)|z_0 = a\Big], \tag{7}$$

and let $\phi = (\phi(a, b))$. Also let $\mathrm{diag}(\psi)$ be a diagonal matrix with entries $\psi(j)$, $j = 1, \ldots, m$. For two matrices A and B we let $A * B$ be the matrix with entries given by the product of the corresponding entries of A and B. Finally, we define $\delta(n, T, f) = \mathrm{E}[f(TW_n)]$, where W_n is a Beta distributed random variable with parameters $(1, n)$.

Theorem 1 (General statistics)**.** *Let H be the general statistic defined in* (6)*. Then we have*

$$\mathrm{E}[H \cdot 1(x(T) = b)|x(0) = a] = \sum_{n=0}^{\infty} \delta(n, T, f)M(n, a, b)\,\mathrm{Pois}(n;\mu T). \tag{8}$$

The matrix $M(n)$ is determined by the recursion

$$M(n) = M(n-1)R + R^{n-1}(\phi * R), \quad n \geq 1 \tag{9}$$

with starting value $M(0) = \mathrm{diag}(\psi)$. *The solution of the recursion is*

$$M(n) = \mathrm{diag}(\psi)R^n + \sum_{\ell=0}^{n-1} R^\ell(\phi * R)R^{n-1-\ell}. \tag{10}$$

Proof. The properties of the inter arrival times $W_i = T_i - T_{i-1}$, conditioned on $J = n$ and with $T_0 = 0$ and $T_{n+1} = T$, can be studied as follows. Let $W_1, \ldots, W_{n+1}$ be independent exponentially distributed variables with mean $1/\mu$, and let $S_{n+1} = \sum_{i=1}^{n+1} W_i$ be the sum of these variables. Then the conditional distribution of the vector $T_1, T_2 - T_1, \ldots, T_{n+1} - T_n$ given $J = n$ is the same as the conditional distribution of $W_1, W_2, \ldots, W_{n+1}$ given that $S_{n+1} = T$. In particular we note that $(W_1, \ldots, W_{n+1})/T$ conditional on $S_{n+1} = T$ follows a Dirichlet distribution with parameter $(1, \ldots, 1)$. Consequently the marginal distribution of W_i/T conditional on $S_{n+1} = T$ is a Beta distribution with parameter $(1, n)$.

We now use the uniformization method

$$\begin{aligned}
&\mathrm{E}[H \cdot 1(x(T) = b)|x(0) = a, J = n] \\
&\quad = \mathrm{E}\Big[\big(\psi(z_0)f(W_1) + \sum_{i=1}^{n} \phi(z_{i-1}, z_i)f(W_{i+1})\big)1(z_n = b)|z_0 = a, S_{n+1} = T\Big] \\
&\quad = \delta(n, T, f)\mathrm{E}\Big[1(z_n = b)\big(\psi(z_0) + \sum_{i=1}^{n} \phi(z_{i-1}, z_i)\big)|z_0 = a\Big] \\
&\quad = \delta(n, T, f)M(n, a, b).
\end{aligned}$$

Here

$$M(0, a, b) = \mathrm{E}[\psi(z_0)1(z_0 = b)|z_0 = a] = \psi(a)1(a = b),$$

and for $n \geq 1$ we obtain the following recursion upon dividing the mean value according to the value of z_{n-1}

$$\begin{aligned}
M(n, a, b) &= \mathrm{E}\Big[1(z_n = b)\big(\psi(z_0) + \sum_{i=1}^{n} \phi(z_{i-1}, z_i)\big)|z_0 = a\Big] \\
&= \sum_{c=1}^{m} \mathrm{E}\Big[1(z_n = b)1(z_{n-1} = c)\big(\psi(z_0) \\
&\qquad + \sum_{i=1}^{n-1} \phi(z_{i-1}, z_i) + \phi(z_{n-1}, z_n)\big)|z_0 = a\Big] \\
&= \sum_{c=1}^{m} M(n-1, a, c)R_{cb} + \sum_{c=1}^{m} (R^{n-1})_{ac}\phi(c, b)R_{cb}
\end{aligned}$$

In matrix form this recursion has the form (9), and it is easy to see that the solution is given by (10). □

To use the solution (8) for the cases $H = T_\alpha$ or $H = N_{\alpha\beta}$ we must specify $\delta(n, T, f)$ and the matrix $M(n)$. Let the matrix which is one at entry (α, β) and zero elsewhere be denoted $U(\alpha, \beta)$.

Corollary 2. *In the case* $H = T_\alpha$ *we have* $\phi = U(\alpha, \alpha)$, $\mathrm{diag}(\psi) = U(\alpha, \alpha)$ *and* $\delta(n, T, f) = T/(n+1)$. *The solution* (10) *for* $M(n)$ *becomes*

$$M(n, a, b) = \sum_{\ell=0}^{n} (R^{n-\ell})_{a\alpha} R^{\ell}_{\alpha b}, \tag{11}$$

and the mean value (8) *is*

$$\mathrm{E}[T_\alpha \cdot 1(x(T) = b)|x(0) = a] = \sum_{n=0}^{\infty} \frac{T}{n+1} \Big[\sum_{\ell=0}^{n} (R^{\ell})_{a\alpha} (R^{n-\ell})_{\alpha b} \Big] \mathrm{Pois}(n; \mu T).$$

Corollary 3. *In the case* $H = N_{\alpha\beta}$ *we have* $\phi = U(\alpha, \beta)$, $\mathrm{diag}(\psi) = 0$ *and* $\delta(n, T, f) = 1$. *The solution* (10) *for* $M(n)$ *becomes*

$$M(n, a, b) = \sum_{\ell=1}^{n-1} (R^{n-\ell})_{a\alpha} R_{\alpha\beta} R^{\ell-1}_{\beta b}, \tag{12}$$

and the mean value (8) *is*

$$\mathrm{E}[N_{\alpha\beta} \cdot 1(x(T) = b)|x(0) = a] = \sum_{n=0}^{\infty} \Big[\sum_{\ell=1}^{n-1} (R^{n-\ell})_{a\alpha} R_{\alpha\beta} R^{\ell-1}_{\beta b} \Big] \mathrm{Pois}(n; \mu T).$$

We note in passing that the mean value in (8) for the cases $H = T_\alpha$ or $H = N_{\alpha\beta}$ can be obtained from the integral representations (2) and (3) on inserting the original uniformization result (5) for $P(t)$. As an example we have

$$\begin{aligned}
&\mathrm{E}[T_\alpha \cdot 1(x(T) = b)|x(0) = a] \\
&\quad = I(a, b, \alpha, \alpha) \\
&\quad = \int_0^T \Big[\sum_{i=0}^{\infty} (R^i)_{a\alpha} \frac{(\mu t)^i}{i!} e^{-\mu t} \Big] \Big[\sum_{j=0}^{\infty} (R^j)_{\alpha b} \frac{(\mu(T-t))^j}{j!} e^{-\mu(T-t)} \Big] dt \\
&\quad = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} (R^i)_{a\alpha} (R^j)_{\alpha b} \frac{T}{(i+j+1)} e^{-\mu T} \frac{(\mu T)^{i+j}}{(i+j)!} \\
&\quad = \sum_{n=0}^{\infty} \frac{T}{n+1} \Big[\sum_{\ell=0}^{n} (R^{\ell})_{a\alpha} (R^{n-\ell})_{\alpha b} \Big] e^{-\mu T} \frac{(\mu T)^n}{n!}.
\end{aligned}$$

This derivation, however, cannot be generalized to the most general form of H in (6) as we do not have an integral representation in the general case.

For the case where $\phi(z_1, z_2) = \psi(z_2)$ and $f \equiv 1$, a recursion equivalent to (9) can be found in Narayana and Neuts (1992). The explicit forms (11) and (12) for $M(n)$ can be found in Bladt and Sørensen (2005).

3.2 Computing covariances using uniformization

In this subsection we consider two statistics H and $\tilde{H}$ of the form (6) and want to calculate $\mathrm{E}[H \cdot \tilde{H} \cdot 1(x(T) = b)|x(0) = a]$. We are particularly interested in the cases

(iii)–(v) in (1) where H and $\tilde{H}$ are the time spent in a state and/or the number of jumps from one state to another. First, we introduce some more notation. Define

$$\delta_1(n,T,f,\tilde{f}) = \mathrm{E}[f(TX)\tilde{f}(TX)] \quad \text{and} \quad \delta_2(n,T,f,\tilde{f}) = \mathrm{E}[f(TX)\tilde{f}(TY)],$$

where $(X, Y, 1-X-Y)$ follows a Dirichlet distribution with parameter $(1,1,n-1)$. We note that δ_1 depends on X only and that the marginal distribution of X is a Beta-distribution with parameters $(1,n)$. Define the two matrices $M_1(n)$ and $L(n)$ as

$$M_1(n,a,b) = \mathrm{E}\Big[1(z_n = b)\big(\psi(z_0)\tilde{\psi}(z_0) + \sum_{i=1}^{n} \phi(z_{i-1},z_i)\tilde{\phi}(z_{i-1},z_i)\big)\Big| z_0 = a\Big]$$

and

$$L(n,a,b) = \mathrm{E}\Big[1(z_n = b)\sum_{i=1}^{n} \phi(z_{i-1},z_i)\Big| z_0 = a\Big],$$

and define $\tilde{L}(n)$ similarly to $L(n)$ with ϕ replaced by $\tilde{\phi}$. Note that both $M_1(n)$, $L(n)$ and $\tilde{L}(n)$ has the same structure as $M(n)$ defined in equation 7, and therefore satisfy similar recursions and admit the same form of explicit solutions as in Theorem 1. Finally, define the matrix $M_2(n)$ as

$$\begin{aligned} M_2(n,a,b) = \mathrm{E}\Big[1(z_n = b)\Big(&\psi(z_0)\sum_{i=1}^{n}\tilde{\phi}(z_{i-1},z_i) + \tilde{\psi}(z_0)\sum_{i=1}^{n}\phi(z_{i-1},z_i) \\ &+ \sum_{i,j:i\neq j}\phi(z_{i-1},z_i)\tilde{\phi}(z_{j-1},z_j)\Big)\Big| z_0 = a\Big]. \end{aligned}$$

Theorem 4 (Products of general statistics)**.** *Let H and $\tilde{H}$ be two general statistics of the form* (6)*. Then we have*

$$\begin{aligned} &\mathrm{E}[H\cdot\tilde{H}\cdot 1(x(T)=b)|x(0)=a] \\ &\quad = \sum_{n=0}^{\infty}\Big[\delta_1(n,T,f,\tilde{f})M_1(n,a,b) + \delta_2(n,T,f,\tilde{f})M_2(n,a,b)\Big]\,\mathrm{Pois}(n;\mu T). \end{aligned} \quad (13)$$

The matrix $M_1(n)$ satisfies the recursion

$$M_1(n) = M_1(n-1)R + R^{n-1}(\phi * \tilde{\phi} * R)$$

with starting values $M_1(0) = \mathrm{diag}(\psi\tilde{\psi})$. The solution is

$$M_1(n) = \mathrm{diag}(\psi\tilde{\psi})R^n + \sum_{\ell=0}^{n-1} R^{\ell}(\phi * \tilde{\phi} * R)R^{n-1-\ell}.$$

The matrices $L(n)$ and $\tilde{L}(n)$ satisfy similar recursions and solutions with $\psi\tilde{\psi}$ replaced by 0 and $\phi\tilde{\phi}$ replaced by ϕ or $\tilde{\phi}$. The matrix $M_2(n)$ satisfies the recursion

$$\begin{aligned} M_2(n) = M_2(n-1)R &+ \mathrm{diag}(\psi)R^{n-1}(\tilde{\phi} * R) + \mathrm{diag}(\tilde{\psi})R^{n-1}(\phi * R) \\ &+ \tilde{L}(n-1)(\phi * R) + L(n-1)(\tilde{\phi} * R). \end{aligned}$$

The solution to the latter recursion can be written as

$$M_2(n) = \operatorname{diag}(\psi)\sum_{\ell=0}^{n-1} R^\ell(\tilde{\phi} * R)R^{n-1-\ell} + \operatorname{diag}(\tilde{\psi})\sum_{\ell=0}^{n-1} R^\ell(\phi * R)R^{n-1-\ell} + \sum_{\ell=0}^{n-2}\sum_{s=0}^{n-2-\ell} R^\ell\left[(\tilde{\phi} * R)R^s(\phi * R) + (\phi * R)R^s(\tilde{\phi} * R)\right]R^{n-2-\ell-s}.$$

Proof. Conditional on the number of jumps n we can divide the product of H and $\tilde{H}$ according to terms where the inter arrival times are the same or different

$$\begin{aligned} H\tilde{H} = {} & \psi(z_0)\tilde{\psi}(z_0)f(W_1)\tilde{f}(W_1) + \sum_{i=1}^{n}\psi(z_0)\tilde{\phi}(z_{i-1},z_i)f(W_1)\tilde{f}(W_{i+1}) \\ & + \sum_{i=1}^{n}\tilde{\psi}(z_0)\phi(z_{i-1},z_i)\tilde{f}(W_1)f(W_{i+1}) \\ & + \sum_{i=1}^{n}\phi(z_{i-1},z_i)\tilde{\phi}(z_{i-1},z_i)f(W_{i+1})\tilde{f}(W_{i+1}) \\ & + \sum_{i,j:i\neq j}\phi(z_{i-1},z_i)\tilde{\phi}(z_{j-1},z_j)f(W_{i+1})\tilde{f}(W_{j+1}). \end{aligned}$$

Taking means with respect to $(W_1,\ldots,W_{n+1})$ conditioned on $W_1+\cdots+W_{n+1}=T$ and conditioning on $z_0=a$ we therefore get

$$\begin{aligned} & \mathrm{E}[H\tilde{H}1(x(T)=b)|x(0)=a, J=n] \\ & \quad = \delta_1(n,T,f,\tilde{f})M_1(n,a,b) + \delta_2(n,T,f,\tilde{f})M_2(n,a,b). \end{aligned}$$

The derivation of the recursion for $M_1(n)$ is as in the proof of Theorem 1 for $M(n)$ with ψ replaced by $\psi\tilde{\psi}$ and with ϕ replaced by $\phi * \tilde{\phi}$.

To establish the recursion for $M_2(n)$ we proceed as in the proof of Theorem 1 dividing the mean value according to the value of z_{n-1}. This gives after some manipulations

$$\begin{aligned} M_2(n,a,b) = \sum_{c=1}^{m}\Big[& M_2(n-1,a,c)R_{cb} + (R^{n-1})_{ac}[\psi(a)\tilde{\phi}(c,b) + \tilde{\psi}(a)\phi(c,b)]R_{cb} \\ & + \tilde{L}(n-1,a,c)\phi(c,b)R_{cb} + L(n-1,a,c)\tilde{\phi}(c,b)R_{cb}\Big]. \end{aligned}$$

The starting value of the recursion is $M_2(0)=0$. The recursions for $L(n)$ and $\tilde{L}(n)$ are derived as for $M(n)$ in the proof of Theorem 1. The solutions of the latter recursions correspond to the solution of the recursion for $M(n)$. Using these it can be shown that the solution of the recursion for $M_2(n)$ is as stated in the theorem. □

From the properties of the Beta- and Dirichlet distributions we obtain the following corollary.

Corollary 5. *In the case where both H and $\tilde{H}$ correspond to either the time spent in a state or the number of jumps between two states the covariance term can be found from Theorem 4 on using that*

$$\delta_1(n,T,f,\tilde{f}) = \begin{cases} 1 & \text{if } f = \tilde{f} \equiv 1 \\ \frac{T}{(n+1)} & \text{if } f(t) = t \text{ and } \tilde{f} \equiv 1 \text{ or } f \equiv 1 \text{ and } \tilde{f}(t) = t \\ \frac{2T^2}{(n+1)(n+2)} & \text{if } f(t) = \tilde{f}(t) = t \end{cases}$$

and

$$\delta_2(n,T,f,\tilde{f}) = \begin{cases} 1 & \text{if } f = \tilde{f} \equiv 1 \\ \frac{T}{(n+1)} & \text{if } f(t) = t \text{ and } \tilde{f} \equiv 1 \text{ or } f \equiv 1 \text{ and } \tilde{f}(t) = t \\ \frac{T^2}{(n+1)(n+2)} & \text{if } f(t) = \tilde{f}(t) = t. \end{cases}$$

For the case where $\phi(z_1, z_2) = \psi(z_2)$, $f \equiv 1$, $\tilde{\phi} = \phi$ and $\tilde{f} = f$ a recursion equivalent to the recursion for $M_2(n)$ can be found in Narayana and Neuts (1992). Generally, the formulas of this section appear to be new. Bladt and Sørensen (2009) use numerical differentiation (with respect to the entries of the rate matrix Q) to find the covariance terms of this section.

3.3 Computing distributions using uniformization

In this subsection we use uniformization to derive the distribution of the number of state changes and the time spent in states using uniformization.

Siepel, Pollard and Haussler (2006) use uniformization to derive the distribution of the total number of substitutions $N = \sum_{i=1}^{J} 1(z_{i-1} \neq z_i)$. Other statistics of interest are the number of substitutions $N_{\alpha\beta} = \sum_{i=1}^{J} 1((z_{i-1}, z_i) = (\alpha, \beta))$ between two different states α and β and the number of visits $N_\alpha = \sum_{i=1}^{J} 1(z_{i-1} = \alpha)$ to a state α. In this section we consider a general count statistic

$$N_H = \psi(z_0) + \sum_{i=1}^{J} \phi(z_{i-1}, z_i), \tag{14}$$

where entries in both ϕ and ψ can be zero or one only. We let $P(k,n)$ be the $m \times m$ matrix with entries

$$P(k,n,a,b) = P\Big(\psi(z_0) + \sum_{i=1}^{n} \phi(z_{i-1}, z_i) = k, z_n = b | z_0 = a\Big). \tag{15}$$

Thus $P(k,n,a,b)$ is the probability of recording k counts of interest when the Markov chain starts in state a and must be in state b at time n. Note, that

$$P(0,0) = \text{diag}(1-\psi) \quad \text{and} \quad P(1,0) = \text{diag}(\psi),$$

and that $P(k,n) = 0$ for $k > n+1$.

Theorem 6. *Let N_H be given as in (14) where entries in both ϕ and ψ can be zero or one only. We have*

$$P(N_H = k, x(T) = b | x(0) = a) = \sum_{n=0}^{\infty} P(k, n, a, b) \operatorname{Pois}(n; \mu t)$$

where for $n \geq 1$ and $1 \leq k \leq n+1$ the matrix $P(k,n)$ is given by the recursion

$$P(k,n) = P(k-1, n-1)(R * \phi) + P(k, n-1)(R * (1-\phi)).$$

Proof. The uniformization method gives

$$P(N_H = k, x(T) = b | x(0) = a) = \sum_{n=0}^{\infty} P(k, n, a, b) \operatorname{Pois}(n; \mu t)$$

where $P(k,n)$ is defined in (15).

The recursion for $P(k,n,a,b)$ is derived by dividing according to the value of z_{n-1}

$$\begin{aligned}
&P(k,n,a,b) \\
&= \sum_{c=1}^{m} P\Big(\psi(z_0) + \sum_{i=1}^{n} \phi(z_{i-1}, z_i) = k, z_{n-1} = c, z_n = b | z_0 = a\Big) \\
&= \sum_{c=1}^{m} P\Big(\psi(z_0) + \sum_{i=1}^{n-1} \phi(z_{i-1}, z_i) = k-1, z_{n-1} = c | z_0 = a\Big) \phi(c,b) R_{cb} \\
&\quad + \sum_{c=1}^{m} P\Big(\psi(z_0) + \sum_{i=1}^{n-1} \phi(z_{i-1}, z_i) = k, z_{n-1} = c | z_0 = a\Big) (1 - \phi(c,b)) R_{cb} \\
&= \sum_{c=1}^{m} P(k-1, n-1, a, c) \phi(c,b) R_{cb} + \sum_{c=1}^{m} P(k, n-1, a, c)(1 - \phi(c,b)) R_{cb},
\end{aligned}$$

on using that $\phi(c,b)$ is either one or zero. □

Let us illustrate the result of the theorem for the case considered in Siepel, Pollard and Haussler (2006). Thus, we consider the total number of substitutions, $N = \sum_{i=1}^{J} 1(z_{i-1} \neq z_i)$. For this case $P(0,0) = \operatorname{diag}(1, 1, \ldots, 1)$, $P(1,0) = 0$, and the recursion is given by

$$P(k,n,a,b) = \sum_{c: c \neq b} P(k-1, n-1, a, c) R_{cb} + P(k, n-1, a, b) R_{bb}.$$

The interpretation of the recursion is quite clear. We are dividing the probability according to the last jump being from a state c to the state b, and the two terms in the recursion correspond to the last jump being a real substitution ($c \neq b$) or a virtual substitution ($c = b$).

In the case of the number of transitions from state α to state β given by the statistic $N_{\alpha\beta}$ we obtain the starting values $P(k,0,a,b) = 0$ for $k \geq 0$. For $n \geq 1$ and $1 \leq k \leq n+1$ we have the recursion

$$P(k,n,a,b) = \begin{cases} P(k-1, n-1, a, \alpha) R_{\alpha b} + \sum_{c: c \neq \alpha} P(k, n-1, a, c) R_{cb} & \text{if } b = \beta \\ \sum_{c=1}^{m} P(k, n-1, a, c) R_{cb} & \text{if } b \neq \beta. \end{cases}$$

Again the interpretation is quite clear. The probability is calculated according to the last jump being from a state c. If the ending state b is different from β then the jump is never from α to β; this is the last case. If the ending state is $b = \beta$ then the jump is from α to β when $c = \beta$. These considerations justify the first case.

As a final application of the uniformization method we consider a sum of inter arrival times like for example the time spent in a state α, $T_\alpha = \sum_{i=0}^{J} 1(z_i = \alpha)(T_{i+1} - T_i)$. Such a statistic has a mixed distribution with point probabilities at zero and T, and has a continuous distribution between these two points. The trick to handle this statistic is that conditional on $J = n$ the distribution depends on the number of terms in the sum only. Furthermore, the distribution of the number of terms is given through $P(k, n)$ from Theorem 6 above. In the theorem below we treat a general statistic of the form

$$T_H = \psi(z_0)T_1 + \sum_{i=1}^{J} \phi(z_{i-1}, z_i)(T_{i+1} - T_i), \tag{16}$$

where both ϕ and ψ can be zero and one only. We let $f(t; a, b)$ be the conditional density of T_H given $x(0) = a$ on the set $x(T) = b$,

$$P(0 < T_H < t, x(T) = b | x(0) = a) = \int_0^t f(y; a, b)dy, \quad t < T.$$

In the theorem we let $f_B(u; \lambda_1, \lambda_2)$ be the Beta density given by

$$f_B(u; \lambda_1, \lambda_2) = \frac{\Gamma(\lambda_1 + \lambda_2)}{\Gamma(\lambda_1)\Gamma(\lambda_2)} u^{\lambda_1 - 1}(1 - u)^{\lambda_2 - 1}. \tag{17}$$

Theorem 7. *The distribution of T_H in (16) is given by*

$$P(T_H = 0, x(T) = b | x(0) = a) = \sum_{n=0}^{\infty} P(0, n, a, b) \operatorname{Pois}(n; \mu T),$$

$$P(T_H = T, x(T) = b | x(0) = a) = \sum_{n=0}^{\infty} P(n + 1, n, a, b) \operatorname{Pois}(n; \mu T),$$

$$f(t; a, b) = \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{1}{T} f_B(t/T; k, n - k + 1) P(k, n, a, b) \operatorname{Pois}(n; \mu T)$$

where $P(k, n, a, b)$ is given in Theorem 6.

Proof. Let $N_H = \psi(z_0) + \sum_{i=1}^{J} \phi(z_{i-1}, z_i)$. Using the uniformization method we condition on $J = n$. For T_H to be zero we must have $N_H = 0$, and the probability of this is given by $P(0, n, a, b)$. Similarly, for T_H to be T we must have $N_H = n + 1$, and the probability of this is given by $P(n+1, n, a, b)$. For $0 < N_H = k < (n+1)$ the density of T_H is the density of $\sum_{j=1}^{k} W_i$ given that $\sum_{j=1}^{n+1} W_i = T$. In this conditional distribution $\frac{1}{T}\sum_{j=1}^{k} W_i$ has a Beta distribution with parameters $(k, n - k + 1)$. This then gives the formula for $f(t; a, b)$ on using that the conditional probability of $N_H = k$ is $P(k, n; a, b)$. □

4 Integrals of matrix exponentials

For the time spent in a state T_α or the number of jumps between two states $N_{\alpha\beta}$ the first two moments can be calculated from the integral representations (2) and (3). These representation and an eigenvalue decomposition of the rate matrix are used in Section 2. When an eigenvalue decomposition is not available we derived in Section 3 alternative expressions based on the uniformization method. There is, however, a third approach to the calculation of integrals of matrix exponentials as those in (2) and (3). The purpose of this section is to draw attention to the theory developed by Van Loan (1978) for calculating such integrals. We describe the method of Van Loan (1978) in its most simple form.

Consider the problem of evaluating the integral

$$\int_0^T e^{Q(T-t)} B e^{Qt} dt \tag{18}$$

where B is a matrix of the same dimension as Q. The special case with $B = U(\alpha, \beta)$ gives the integral in (2). For evaluating this integral Van Loan (1978) introduces a matrix A, with a dimension twice that of Q,

$$A = \begin{bmatrix} Q & B \\ 0 & Q \end{bmatrix}. \tag{19}$$

The structure of A implies that the matrix exponential must be of the form

$$e^{At} = \begin{bmatrix} F(t) & G(t) \\ 0 & F(t) \end{bmatrix} \quad \text{with } F(0) = I \text{ and } G(0) = 0.$$

Using $\frac{d}{dt} e^{At} = A e^{At}$ we obtain the equation

$$\begin{bmatrix} F'(t) & G'(t) \\ 0 & F'(t) \end{bmatrix} = \begin{bmatrix} Q & B \\ 0 & Q \end{bmatrix} \begin{bmatrix} F(t) & G(t) \\ 0 & F(t) \end{bmatrix}.$$

We thus have two differential equations to solve. One homogeneous linear differential equation

$$F'(t) = QF(t) \quad \text{with} \quad F(0) = I,$$

(the solution is $F(t) = e^{Qt}$), and an inhomogeneous linear differential equation

$$G'(t) = QG(t) + BF(t) \quad \text{with} \quad G(0) = 0.$$

The inhomogeneous linear differential equation has solution

$$G(t) = \int_0^t e^{(t-x)Q} B F(x) dx = \int_0^t e^{(t-x)Q} B e^{Qx} dx.$$

Thus, the integral (18) is the upper right corner in e^{At}.

A number of approaches exist for evaluating e^{At} and are implemented in software packages; see Moler and Van Loan (2003) for a review. Calculating the matrix exponential e^{At} therefore provides a very direct way of calculating the desired integral

representation. The only caveat is the accuracy of the matrix exponential utility provided by the software.

To handle the integral (3) we can use the part of Van Loan's methodology concerning integrals of the type

$$\int_0^T \int_0^t e^{Q(T-u)} B_1 e^{Q(t-u)} B_2 e^{Qu} du dt.$$

This integral is given by the upper right corner of e^{At} with

$$A = \begin{bmatrix} Q & B_1 & 0 \\ 0 & Q & B_2 \\ 0 & 0 & Q \end{bmatrix}.$$

5 Discussion

We have presented three approaches for calculating expectations of summary statistics for end-point conditioned continuous time Markov chains. The eigenvalue decomposition method and Van Loan's method both take the integral representations (2) and (3) as a starting point. The integral representations (2) and (3) cover the most important summary statistics (time spent in a state and number of jumps from one state to another). However, in general a summary statistic H of the form (6) does not necessarily admit an integral representation. In Section 3 we provide a framework for calculating properties of general summary statistics. The framework is based on the uniformization method, and as shown in Theorem 1 the calculation is decomposed in three simple parts that are each easy to handle.

All three methods of calculating summary statistic are in principle easy to implement but they all have their limitations. The method based on an eigenvalue decomposition becomes rather involved if the eigenvalues are not real. (However, we would like to emphasize that a reversible substitution process always admits real eigenvalues.) Van Loan's method seems very attractable but requires access to a reliable matrix exponentiation software package. Uniformization is the most general method but requires that an infinite sum is truncated. A discussion of truncation error can be found in Narayana and Neuts (1992) and Grassmann (1993).

References

Ball, F. and Milne, R.K. (2005). Simple derivations of properties of counting processes associated with Markov renewal processes. *J. Appl. Probab.*, **42**, 1031–1043.

Bladt, M., Neuts, M. F., Meini, B. and Sericola, B. (2002). Distributions of reward functions on continuous-time, Markov chains. In Matrix-analytic Methods: Theory and Applications (ed. G. Latouche). Singapore: World Scientific Publishing Company.

Bladt, M. and Sørensen, M. (2005). Statistical inference for discretely observed Markov jump processes. *J. R. Statist. Soc B.*, **67**, 395–410.

Bladt, M. and Sørensen, M. (2009). Efficient estimation of transition rates between credit ratings from observations at discrete time points. *Quantitative Finance*, **9**, 147–160.

Grassmann, W.K. (1993). Rounding errors in certain algorithms involving Markov chains. ACM Transactions on Mathematical Software, **19**, 496–508.

Hobolth, A. and Jensen, J.L. (2005). Statistical inference in evolutionary models of DNA sequences via the EM algorithm. *Statistical Applications in Genetics and Molecular Biology*, **4**, article 18.

Holmes, I. and Rubin, G.M. (2002). An expectation maximization algorithm for training hidden substitution models. *J. Mol. Biol.* **317**, 753–764.

Jensen, A. (1953). Markoff chains as an aid in the study of Markoff processes. *Skand. Aktuarietidskr.* **36**, 87–91.

Kosiol, C., Holmes, I. and Goldman, N. (2007). An empirical codon model for protein sequence evolution. *Mol. Biol. Evol.*, **24**, 1464–1479.

Metzner, P., Horenko, I. and Schütte, C. (2007). Generator estimation of Markov jump processes based on incomplete observations nonequidistant in time. *Physical Review E*, **76**, 066702.

Minin, V.N. and Suchard, M.A. (2008a). Counting labeled transitions in continuous-time Markov models of evolution. *J. Math. Biol.*, 56, 391–412.

Moler, C. and Van Loan, C. (2003). Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. *SIAM Review*, **45**, 3–49.

Narayana, S. and Neuts, M.F. (1992). The first two moment matrices of the counts for the markovian arrival process. *Stochastic Models*, **8**, 459–477.

Siepel, A., Pollard, K.S. and Haussler, D. (2006). New methods for detecting lineage-specific selection. *Proceedings of the 10th International Conference on Research in Computational Molecular Biology (RECOMB)*, 190–205.

Van Loan, C.F. (1978). Computing integrals involving the matrix exponential. *IEEE Transactions on Automatic Control*, **23**, 395–404.