Commentationes Mathematicae Universitatis Carolinae

Petr Kůrka

Merging of states of Markov chains with infinite probability rates

Commentationes Mathematicae Universitatis Carolinae, Vol. 20 (1979), No. 1, 173--182

Persistent URL: `http://dml.cz/dmlcz/105911`

COMMENTATIONES MATHEMATICAE UNIVERSITATIS CAROLINAE
20, 1 (1979)

MERGING OF STATES OF MARKOV CHAINS WITH INFINITE PROBABILITY

P. KŮRKA

Abstract: In the paper we investigate sequences of continuous time finite state Markov chains, some transition rates of which tend to infinity. We show that states which communicate infinitely fast with each other can be merged, thus obtaining Markov chains with fewer states and finite transition rates, which approximates the original one.

Key words: Markov chains, infinitely ergodic sets, transition rates.

AMS: 60J25
15A51

Finite filtered Markov chains were introduced by Richardson (1975) to construct probabilistic models of self-reproduction. They are continuous time finite state Markov chains, whose transition rates depend on parameters, and may tend to zero, infinity, or to a finite number.
Apart from self-reproduction, other applications of this concept are suggestive. Thus in the life of a population, a mutation happens considerably more rarely than normal reproduction, and soon after it happens, a stationary distribution of its occurence is attained. In chemistry, a reaction $A+B \rightleftarrows C+D$ may proceed by forming a complex E, which is highly unstable, and quickly decomposes to either A+B or C+D. This situation may be represented by a chain

$$A + B \underset{r_1}{\overset{s_1}{\rightleftarrows}} E \underset{r_2}{\overset{s_2}{\leftrightarrows}} C + D$$

where transition rates r_1, r_2 are infinitely larger than s_1 and s_2. If our unit of time is long enough, we can neglect the state E, because the process stays in it for an infinitely short period of time, and approximate the chain by

$$A + B \underset{s_2 r_1/(r_1+r_2)}{\overset{s_1 r_2/(r_1+r_2)}{\rightleftarrows}} C + D$$

We show in the present paper that such approximation is possible whenever all infinite transition rates are of the same order, i.e. if their ratio is finite. A set of states of such chain is infinitely ergodic, if between any two of its members there is a path of infinite transition rates, and no infinite transition rate leads out of it. A transition out of an infinitely ergodic set A occurs only after infinite transitions attain equilibrium on A, and soon after the process leaves A, it arrives to another infinitely ergodic set. In the limit, this new process over infinitely ergodic sets has Markovian character, so we obtain a finite filtered Markov chain with fewer states, which approximates the original one. The transition rates of this new chain are solutions of a system of linear equations, so the computation of the transition probability matrix is a bit simplified.
To simplify the notation, the finite filtered Markov chains are defined here as sequences of transition rate matrices.

Definition. A finite filtered Markov chain over a finite set of states $\mathcal{C}$ is a sequence of matrices

$(r_i(x,y))_{x,y\in\mathcal{C},i\in\omega}$ such that

1. $x\neq y \Longrightarrow r_i(x,y)\geq 0$, $x\in\mathcal{C}\Longrightarrow \sum_{y\in\mathcal{C}} r_i(x,y) = 0$ for any $i\in\omega$,
2. For any $x\neq y\in\mathcal{C}$ there exist $\lim_{i\to\infty} r_i(x,y)\in[0,\infty]$.

A finite filtered Markov chain (r_i) over $\mathcal{C}$ has one level, if whenever $\lim_{i\to\infty} r_i(x,y) = \infty$, $\lim_{i\to\infty} r_i(z,v) = \infty$ then $\lim_{i\to\infty} r_i(x,y)/r_i(z,v) < \infty$.

Definition. Let (r_i) be a finite filtered Markov chain over $\mathcal{C}$.

1. $(p_i(x,y)(t))_{x,y\in\mathcal{C},t\geq 0,i\in\omega}$ the transition probability matrix is given by $p_i(t) = \underline{\underline{\exp}}\,(r_i t) = \sum_{n=0} (r_i t)^n/n!$. $p_i(x,y)(t)$ is the probability that the chain r_i is in state y at time t, provided it started in x at time 0.

2. $\mathcal{F} = \{(x,y)\in\mathcal{C}\times\mathcal{C} \mid \lim_{i\to\infty} r_i(x,y) = \infty\}$, $\mathcal{F}^*$ is the reflexive and transitive closure of $\mathcal{F}$.

3. $\mathcal{D} = \{A\subseteq\mathcal{C} \mid A\neq 0,\ A\times A\subseteq\mathcal{F}^*,\ \mathcal{F}[A]\subseteq A\}$ is the set of infinitely ergodic sets of (r_i). (Here $\mathcal{F}[A] = \{y \mid (x,y)\in\mathcal{F}$ for some $x\in A\}$.) We have $A\in\mathcal{D}$ iff between any two of its members there is a path of infinite transition rates, and no infinite transition rate leads out of A.

4. $N = \mathcal{C}-\bigcup\mathcal{D}$ is the set of infinitely transient states (which may be empty). For any $x\in N$ there is a path of infinite transition rates leading from x to some $A\in\mathcal{D}$.

5. $(P_i(A,x))_{A\in\mathcal{D},x\in\mathcal{C},i\in\omega}$, the equilibrium matrix of (r_i) is given as follows:

$x\notin A\Longrightarrow P_i(A,x) = 0$,

$x\in A\Longrightarrow 0\leq P_i(A,x)\leq 1$, $\sum_{x\in A} P_i(A,x) = 1$ for any $A\in\mathcal{D}$,

$x\in A\Longrightarrow \sum_{y\in A} P_i(A,y)r_i(y,x) + P_i(A,x)\sum_{y\notin A} r_i(x,y) = 0$.

$(P_i(A,x))_{x\in A}$ is the equilibrium distribution of (r_i) on A.

6. $(Q_i(x,A))_{x\in\mathscr{C},A\in\mathscr{D},i\in\omega}$, the absorption matrix of (r_i) is given as follows:

$x\in A \Longrightarrow Q_i(x,A) = 1$, $x\in B\in\mathscr{D}$, $B\neq A \Longrightarrow Q_i(x,A) = 0$,

$x\in N$, $A\in\mathscr{D} \Longrightarrow \sum_{y\in N} r_i(x,y)Q_i(y,A) + \sum_{y\in A} r_i(x,y) = 0$

$Q_i(x,A)$ is the probability that the first set of $\mathscr{D}$ which the chain r_i visits is A, provided it started at x.

Observe that $P_iQ_i = I_{\mathscr{D}}$ (identity matrix), and that (P_ir_i), (r_iQ_i) are bounded sequences. Furthermore, if (r_i) has one level, then finite limits $\lim_{i\to\infty} P_i$, $\lim_{i\to\infty} Q_i$, $\lim_{i\to\infty} P_ir_i$, $\lim_{i\to\infty} r_iQ_i$ exist. To compute for a given $t>0$ $\lim_{i\to\infty} p_i(t) = \lim_{i\to\infty} \exp(r_it)$, it may be reasoned as follows:
Suppose that the process starts in some $x\in A$. Then before any transition with finite rate occurs, the infinite transitions attain equilibrium on A. The transition rate from A to say $y\notin A$ is then $\sum_{x\in A} P_i(A,x)r_i(x,y)$. (This was proved in Richardson (1975).) The process then jumps in negligible time from y to some B with probability $Q_i(y,B)$. In this way, the transition rate from A to B is $\sum_{x\in A}\sum_{y\notin A} P_i(A,x)r_i(x,y)Q_i(y,B)$, which is equal to (A,B)-entry of the matrix $P_ir_iQ_i$, and

$t>0$, $x\in A \Longrightarrow \lim_{i\to\infty}\sum_{y\in B} p_i(x,y)(t) = \lim_{i\to\infty}(\exp(P_ir_itQ_i))_{A,B}$

Now, starting from some $x\in\mathscr{C}$ the process first jumps to some A with probability $Q_i(x,A)$, then it behaves according to $P_ir_iQ_i$, and if it ends in B, it attains there equilibrium $(P_i(B,y))_{y\in B}$. This may be expressed by the equality

$$t>0 \Longrightarrow \lim_{i\to\infty} \exp(r_it) = \lim_{i\to\infty} Q_i \exp(P_ir_itQ_i)P_i.$$

Actually, for chains with one level this holds for any sequences (P_i), (Q_i), for which (P_i), (Q_i), (P_ir_i), (r_iQ_i) are convergent sequences, and $P_iQ_i = I$.

Theorem. Let (r_i) be a finite filtered Markov chain over $\mathcal{C}$ with one level, let $\mathcal{D}$ be its set of infinitely ergodic sets, let $(P_i),(Q_i)$, be sequences of $\mathcal{D}\times\mathcal{C}$ resp. $\mathcal{C}\times\mathcal{D}$ matrices such that there exist finite limits $\lim_{i\to\infty} P_i$, $\lim_{i\to\infty} Q_i$, $\lim_{i\to\infty} P_i r_i$, $\lim_{i\to\infty} r_i Q_i$ and $P_i Q_i = I_{\mathcal{D}}$. Then for any $t > 0$

$$\lim_{i\to\infty} \exp(r_i t) = \lim_{i\to\infty} Q_i \exp(P_i r_i t Q_i) P_i .$$

To prove the theorem, denote $c = \text{card}(\mathcal{C})$, $d = \text{card}(\mathcal{D})$, $\bar{d} = c - d$, and let $\bar{\mathcal{D}}$ be some index set with $\text{card}(\bar{\mathcal{D}}) = \bar{d}$, $\mathcal{D} \cap \bar{\mathcal{D}} = 0$.

If (r_i) is a bounded sequence, then $c = d$, P_i, Q_i are inverse to one another, and the theorem holds trivially. Suppose therefore that (r_i) is not bounded and denote $T_i = \max\{r_i(x,y) \mid x,y \in \mathcal{C}\}$, so $\lim_{i\to\infty} T_i = \infty$, and there exist finite limit $r = \lim_{i\to\infty} r_i/T_i$. We prove first two lemmas.

Lemma 1. The eigenvalues of r_i may be assigned to sets $\mathcal{D}, \bar{\mathcal{D}}$ in such way that $(\lambda_i(z))_{z\in\mathcal{D}}$ are eigenvalues of r_i for which $\lim_{i\to\infty} \lambda_i(z)/T_i = 0$

$(\lambda_i(z))_{z\in\bar{\mathcal{D}}}$ are eigenvalues of r_i for which $\lim_{i\to\infty} \text{Re}(\lambda_i(z)) = -\infty$.

Proof: The set of ergodic sets of $r = \lim_{i\to\infty} r_i/T_i$ is $\mathcal{D}$, so the multiplicity of the eigenvalue 0 of r is just d. By Gershgorin theorem (see Franklin (1968)), for any eigenvalue λ of r, $|\lambda - r(x,x)| \leq -r(x,x)$ for some $x \in \mathcal{C}$, so $\lambda \neq 0$ implies $\text{Re}(\lambda) < 0$. Since the eigenvalues depend on the matrix continuously, the lemma follows.

Lemma 2. Define a sequence of matrices $A_i = \prod_{z\in\bar{\mathcal{D}}} (r_i - I_c \cdot \lambda_i(z))$. Then there exist bounded sequences of matrices $(u_i(z,x))_{z\in\bar{\mathcal{D}}, x\in\mathcal{C}}$ $(v_i(x,z))_{x\in\mathcal{C}, z\in\bar{\mathcal{D}}}$ such that

$u_iA_i = 0$, $A_iv_i = 0$, $u_iv_i = I_{\bar{d}}$, $u_ir_iv_iu_i = u_ir_i$, $v_iu_ir_iv_i = r_iv_i$.

<u>Proof</u>: The proof is straightforward provided all non-zero eigenvalues of r are distinct. In this case, for sufficiently large i $(\lambda_i(z))_{z\in\bar{\mathcal{D}}}$ are distinct too, and we can define the z-th row of u_i as the left eigenvector of r_i corresponding to $\lambda_i(z)$, and the z-th column of v_i as the right eigenvector of r_i corresponding to $\lambda_i(z)$. We normalize these vectors so that their scalar product is 1, and both $(u_i(z,x))_{x\in\mathcal{C}}$, $(v_i(x,z))_{x\in\mathcal{C}}$ are bounded sequences. Then $u_iv_i = I_{\bar{d}}$, and since the factors of A_i commute with each other, $u_iA_i = A_iv_i = 0$. Furthermore $u_ir_i = \Lambda_iu_i$ $r_iv_i = v_i\Lambda_i$, where Λ_i is the diagonal matrix, whose diagonal is $(\lambda_i(z))_{z\in\bar{\mathcal{D}}}$. So $u_ir_iv_iu_i = \Lambda_iu_iv_iu_i = \Lambda_iu_i = u_ir_i$, $v_iu_ir_iv_i = v_iu_iv_i\Lambda_i = v_i\Lambda_i = r_iv_i$.

In the general case of multiple eigenvalues denote $B_i = \prod_{z\in\mathcal{D}} (r_i - I_c \cdot \lambda_i(z))$, $A = \lim_{i\to\infty} A_i/T_i^{\bar{d}}$, $B = \lim_{i\to\infty} B_i/T_i^{d}$.

It follows from the theorem on p.126 in Franklin (1968) that $X_c = \underline{\text{Ker}}\,(A) \oplus \underline{\text{Ker}}\,(B)$, where X_c is the complex vector space with dimension c, $\underline{\text{Ker}}\,(A) = \{x\in X_c \mid x\cdot A = 0\}$. By Cayley-Hamilton theorem, $AB = 0$, so $\underline{\text{Im}}\,(A) \subseteq \underline{\text{Ker}}\,(B)$, where $\underline{\text{Im}}\,(A) = \{x\cdot A \mid x\in X_c\}$. Since the dimension of both these spaces is d, we have $\underline{\text{Im}}\,(A) = \underline{\text{Ker}}\,(B)$.

Let u be any $\bar{\mathcal{D}}\times\mathcal{C}$ matrix whose rows form a basis for the space Ker (A), so $uA = 0$. Since $X_c = \underline{\text{Ker}}\,(A) \oplus \underline{\text{Im}}\,(A)$, there exists the unique $\mathcal{C}\times\bar{\mathcal{D}}$ matrix v with $uv = I_{\bar{d}}$, $Av = 0$. Since $X_c = \underline{\text{Ker}}\,(A_i) \oplus \underline{\text{Im}}\,(A_i)$ for any i, there exist matrices u_i, v_i with $u_iA_i = 0$, $A_iv_i = 0$, $u_iv_i = I_{\bar{d}}$, $\lim_{i\to\infty} u_i = u$, $\lim_{i\to\infty} v_i = v$, so (u_i), (v_i) are bounded.

Since $(u_ir_i)A_i = u_iA_ir_i = 0 = (u_ir_iv_iu_i)A_i$,

$(u_ir_i)v_i = (u_ir_iv_iu_i)v_i$, we have $u_ir_i = u_ir_iv_iu_i$.

Since $A_i(r_iv_i) = r_iA_iv_i = 0 = A_i\ (v_iu_ir_iv_i)$,

$u_i(r_iv_i) = u_i(v_iu_ir_iv_i)$, we have $r_iv_i = v_iu_ir_iv_i$.

Proof of the theorem: By lemma 2, we have $P_iA_iv_i/T_i^{\bar{d}-1} = 0$. If we carry out the multiplication in A_i, we get a polynomial in r_i, whose every term but absolute has the form

$(P_ir_i)(r_i/T_i)^{\bar{d}-k-1}\ v_i(\lambda_i(z_1)/T_i)\ldots(\lambda_i(z_k)/T_i)\quad 0 \le k < \bar{d}$

and so it is bounded. It follows that the absolute term $P_iv_iT_i \prod_{z\in\bar{\mathcal{D}}}(\lambda_i(z)/T_i)$ is bounded too, and if we multiply it by bounded sequence $\prod_{z\in\bar{\mathcal{D}}}(T_i/\lambda_i(z))$, we get that $P_iv_iT_i$ is bounded, so $\lim_{i\to\infty} P_iv_i = 0$.

Similarly we get that $u_iQ_iT_i$ is bounded, and $\lim_{i\to\infty} u_iQ_i = 0$. From this result and Lemma 2 it follows

$$\begin{bmatrix} P_i \\ u_i \end{bmatrix} \cdot [Q_i, v_i] = \begin{bmatrix} P_iQ_i, & P_iv_i \\ u_iQ_i, & u_iv_i \end{bmatrix} \xrightarrow[i]{} I_c$$

so

$$\lim_{i\to\infty} \det\begin{bmatrix} P_i \\ u_i \end{bmatrix} \cdot \det[Q_i, v_i] = 1 .$$

Therefore, for sufficiently large i $[Q_i, v_i]$ is regular, $\det[Q_i, v_i]$ is bounded away from zero, and $[Q_iv_i]^{-1}$ is bounded, too. Define a $\mathcal{D} \times \mathcal{C}$ matrix $\bar{P}_i = [I_d, 0]\,[Q_i, v_i]^{-1}$. There is $(P_i - \bar{P}_i)T_i\,[Q_i, v_i] = [0, P_iv_iT_i]$ which is bounded, so $(P_i - \bar{P}_i)T_i$ is bounded, and $\lim_{i\to\infty}(P_i - \bar{P}_i) = 0$. Again we have

$$\begin{bmatrix} \bar{P}_i \\ u_i \end{bmatrix} [Q_i, v_i] \xrightarrow[i]{} I_c$$

and we can define $\bar{Q}_i = \begin{bmatrix} \bar{P}_i \\ u_i \end{bmatrix}^{-1} \cdot \begin{bmatrix} I_d \\ 0 \end{bmatrix}$ so that $(Q_i - \bar{Q}_i)T_i$ is bounded, and $\lim_{i\to\infty}(Q_i - \bar{Q}_i) = 0$. There is

$\begin{bmatrix} \bar{P}_i \\ u_i \end{bmatrix} \cdot [\bar{Q}_i, v_i] = I_c$, so $[\bar{Q}_i, v_i] \begin{bmatrix} \bar{P}_i \\ u_i \end{bmatrix} = \bar{Q}_i\bar{P}_i + v_iu_i = I_c$, and

$r_i = (\bar{Q}_i\bar{P}_i + v_iu_i)r_i(\bar{Q}_i\bar{P}_i + v_iu_i) = \bar{Q}_i\bar{P}_ir_i\bar{Q}_i\bar{P}_i + \bar{Q}_i\bar{P}_ir_iv_iu_i + + v_iu_ir_i\bar{Q}_i\bar{P}_i + v_iu_ir_iv_iu_i$.

By Lemma 2 the middle two terms of this expression are zero since $\bar{P}_ir_iv_i = \bar{P}_iv_iu_ir_iv_i = 0$, $u_ir_i\bar{Q}_i = u_ir_iv_iu_i\bar{Q}_i = 0$

We have $r_i = [\bar{Q}_i, v_i] \begin{bmatrix} \bar{P}_ir_i\bar{Q}_i, & 0 \\ 0 & , u_ir_iv_i \end{bmatrix} \begin{bmatrix} \bar{P}_i \\ u_i \end{bmatrix}$

so the eigenvalues of r_i are divided between $\bar{P}_ir_i\bar{Q}_i$ and $u_ir_iv_i$. Since $\bar{P}_ir_i\bar{Q}_i$ is bounded, it has bounded eigenvalues, so by Lemma 1 the eigenvalues of $u_ir_iv_i$ are $(\lambda_i(z))_{z\in \mathcal{Z}}$. It follows that for any $t > 0$ $\lim_{i\to\infty} \exp(U_ir_itv_i) = 0$, and

$$\lim_{i\to\infty} [\exp(r_it) - \bar{Q}_i \exp(\bar{P}_ir_it\bar{Q}_i)\bar{P}_i] = 0 .$$

Furthermore,

$P_ir_iQ_i - \bar{P}_ir_i\bar{Q}_i = (P_i-\bar{P}_i)r_iQ_i + (\bar{P}_ir_i/T_i)(Q_i-\bar{Q}_i)T_i \xrightarrow[i]{} 0$

since r_iQ_i, $(Q_i-\bar{Q}_i)T_i$ are bounded, and $\lim_{i\to\infty}(P_i-\bar{P}_i) = 0$, $\lim_{i\to\infty} \bar{P}_ir_i/T_i = 0$. Since $P_ir_iQ_i$ is bounded,

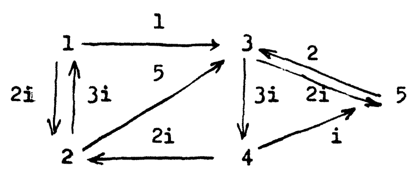
$$\lim_{i\to\infty} [\exp(P_ie_itQ_i) - \exp(\bar{P}_ir_it\bar{Q}_i)] = 0$$

and the theorem follows.

Besides some insight into the structure of finite filtered Markov chains, the above theorem yields also a reduction in computational complexity of the transition probability matrix $\exp(r_it)$. If we take for P_i, Q_i the equilibrium and absorption matrices of r_i, then the nontrivial values of P_i, i.e. $(P_i(A,x))_{x\in A}$ are obtained by solution of a system of linear equations with card (A) unknowns, and nontrivial values of Q_i, i.e, $(Q_i(x,A))_{x\in N}$ are solutions of a system of linear equations with card (N) unknowns. Since the computation of the exponential of a matrix is a rather complicated task, and the dimension of $P_i r_i Q_i$ may be substantially smaller than that of r_i, the whole procedure may be much simplified.

The theorem may be also used for Markov chains with finite (but sufficiently large) transifion rates. In this case the error of approximation is of the order exp (-st), where s is the value of some large transition rate.

Example. Consider a chain



with matrix

$$r_i = \begin{bmatrix} -2i-1, & 2i & , & 1 & , & 0 & , & 0 \\ 3i & , -3i-5 & , & 5 & , & 0 & , & 0 \\ 0 & , \ 0 & , & -5i & , & 3i & , & 2i \\ 0 & , \ 2i & , & 0 & , & -3i & , & i \\ 0 & , \ 0 & , & 2 & , & 0 & , & -2 \end{bmatrix}$$

Clearly $\mathcal{D} = \{\{1,2\},\{5\}\}$, $N = \{3,4\}$ and following (constant) matrices satisfy the assumptions of the theorem:

$$P_i = \begin{bmatrix} 3/5, & 2/5, & 0, & 0, & 0 \\ 0, & 0, & 0, & 0, & 1 \end{bmatrix} \qquad Q_i = \begin{bmatrix} 1 & , & 0 \\ 1 & , & 0 \\ 2/5 & , & 3/5 \\ 2/3 & , & 1/3 \\ 0 & , & 1 \end{bmatrix}$$

Then $P_i r_i Q_i$ is the matrix of the chain

$$\{1,2\} \underset{4/5}{\overset{39/25}{\rightleftarrows}} \{5\}$$

R e f e r e n c e s

D. RICHARDSON: Self-reproduction by template, Mathematical Biosciences 28(1975).

S. KARLIN: A first course in stochastic processes, Academic Press 1966.

J.N. FRANKLIN: Matrix theory, Prentice Hall, New Jersey 1968,

Czechoslovak Academy of Sciences

Center of Biomathematics

Budějovická 1083, 14220 Praha

Československo

(Oblatum 24.11. 1978)