Computation of Mean First Passage Times in a Markov Chain

Kenneth Lange

Departments of Computational Medicine,
Human Genetics, and Statistics
University of California
Los Angeles, CA 90095
Phone: 310-206-8076
E-mail klange@ucla.edu

Research supported in part by USPHS grants GM53275 and HG006139.

March 10, 2020

Introduction

Let $\boldsymbol{P} = (p_{ij})$ be the transition probability matrix of a finite state Markov chain. The first passage time T_{ij} is the number number of steps it takes to reach j starting from i. We will assume all $p_{ii} = 0$ and set $T_{ii} = 0$ in contrast to the standard first passage time formulation. Let $m_{ij} = \mathrm{E}(T_{ij})$ be the expected value of T_{ij} and collect the m_{ij} into a matrix $\boldsymbol{M}$. One can calculate the m_{ij} recursively via the system of equations

$$m_{ij} = 1 + \sum_k p_{ik} m_{kj}$$

for all $j \neq i$. It is natural to solve this equation iteratively via the recurrence

$$\begin{aligned} \boldsymbol{M}^{(n+1)} &= \boldsymbol{P}\boldsymbol{M}^{(n)} + \mathbf{1}\mathbf{1}^t \qquad (1) \\ \operatorname{diag}(\boldsymbol{M}^{(n+1)}) &= \mathbf{0} \end{aligned}$$

starting from $\boldsymbol{M}^{(0)} = \mathbf{0}$. This recurrence has several advantages: (a) it exploits fast matrix times matrix multiplication, (b) it correctly maintains the diagonal entries, (c) it is monotonic in the sense that $\boldsymbol{M}^{(n+1)} \geq \boldsymbol{M}^{(n)}$ for all n, and (d) it converges to the minimal solution of the equations. Monotonicity follows by induction from the form of the updates and the obvious condition $\boldsymbol{M}^{(1)} \geq \boldsymbol{M}^{(0)}$. It is also clear by induction that if $\boldsymbol{M}$ solves the system of equations, then $\boldsymbol{M} \geq \boldsymbol{M}^{(n)}$ for all n. In view of this bound, the monotonic sequence $\boldsymbol{M}^{(n)}$ converges to a limit $\boldsymbol{M}^{(\infty)} \leq \boldsymbol{M}$.

Proposition 0.1. *The iterate* $m_{ij}^{(n)}$ *equals* $\mathrm{E}(T_{ij} 1_{\{T_{ij} \leq n\}})$ *and converges to* $\mathrm{E}(T_{ij})$ *whenever the latter is finite.*

Proof. This identification is clearly true for $n = 0$. Suppose $n > 0$ and X_1 denotes the state of the chain after one step starting at $X_0 = i$. Then

$$\begin{aligned}
m_{ij}^{(n+1)} &= \mathrm{E}(T_{ij}1_{\{T_{ij}\le n+1\}}) \\
&= \sum_k p_{ik}\,\mathrm{E}(T_{ij}1_{\{T_{ij}\le n+1\}} \mid X_1 = k) \\
&= \sum_k p_{ik}\,\mathrm{E}\left[(T_{kj}+1)1_{\{T_{kj}+1\le n+1\}}\right] \\
&= 1 + \sum_k p_{ik}\,\mathrm{E}(T_{kj}1_{\{T_{kj}\le n\}}) \\
&= 1 + \sum_k p_{ik} m_{kj}^{(n)}
\end{aligned}$$

proves our first claim. The monotone convergence theorem implies that $m_{ij}^{(n)}$ converges to $\mathrm{E}(T_{ij})$ whenever the later is finite. □

To understand the rate of convergence of the iteration scheme (1), we first observe that it operates column by column on $\boldsymbol{M}$. Thus, we can isolate each column and analyze its convergence rate separately. With out loss of generality, we choose column one. For the sake of convenience, we now decompose $\boldsymbol{P}$ into the block matrix

$$\boldsymbol{P} = \begin{pmatrix} 0 & \boldsymbol{a} \\ \boldsymbol{b} & \boldsymbol{Q} \end{pmatrix}.$$

If we let $\boldsymbol{m}_1$ denote the first column of $\boldsymbol{M}$ minus its top entry, then our recurrence for $\boldsymbol{m}_1$ takes the form

$$\boldsymbol{m}_1^{(n+1)} = \boldsymbol{1} + \boldsymbol{Q}\boldsymbol{m}_1^{(n)} \tag{2}$$

with $\boldsymbol{m}_1 = \boldsymbol{0}$.

Proposition 0.2. *Suppose $\boldsymbol{P}$ is irreducible. In the notation just established, the map $f(\boldsymbol{m}) = \boldsymbol{1} + \boldsymbol{Q}\boldsymbol{m}$ is a contraction for some norm with contraction* *constant $\rho < 1$. Hence, the iterates (2) converge at a linear rate to the corresponding vector of mean first passage times.*

Proof. The matrix $\boldsymbol{I} - \boldsymbol{Q}$ is diagonally dominant because the row sums of $\boldsymbol{P}$ equal 1. At least one row sum of $\boldsymbol{Q}$ is strictly less than 1 since otherwise $\boldsymbol{P}$ is reducible. Hence, $\boldsymbol{I} - \boldsymbol{Q}$ is irreducibly diagonally dominant, and Theorem 6.2.6 of [2] implies that $\boldsymbol{I} - \boldsymbol{Q}$ is nonsingular. Theorem 6.2.9 of [2] in turn implies that $\boldsymbol{Q}$ has dominant eigenvalue $\omega < 1$. For given $\epsilon > 0$, Theorem 1.3.6 of [2] implies that there exists a vector norm $\|\boldsymbol{x}\|$ with induced matrix norm satisfying $\|\boldsymbol{Q}\| \leq \omega + \epsilon$. Taking $\rho = \omega + \epsilon < 1$ gives

$$\|f(\boldsymbol{m}) - f(\boldsymbol{n})\| \quad \leq \quad \|\boldsymbol{Q}\| \cdot \|\boldsymbol{m} - \boldsymbol{n}\| \quad = \quad \rho\|\boldsymbol{m} - \boldsymbol{n}\|.$$

It follows that $f(\boldsymbol{m})$ is a contraction. Finally, an appeal to the classical contraction theorem mapping theorem (Theorem 8.2.2 of [2]) yields convergence at the linear rate ρ. □

References

[1] Lange, K (2010) *Applied Probability*, 2nd ed. Springer

[2] Ortega JM (1990) *Numerical Analysis: A Second Course*. SIAM