Quasi-stationary distributions for queueing and other models

Phil Pollett

Received: 25 January 2022 / Accepted: 28 February 2022 / Published online: 22 March 2022
© Crown 2022

1 Background

We will examine questions concerning quasi-stationary behaviour in evanescent processes. The idea has its origins in biological modelling, where typically we are interested in limiting behaviour conditional on non-extinction. For queueing processes, we are typically interested in the behaviour within a busy period in stable or near stable queues [8,10], or, for unstable queues, prior to last exit from the empty state [4].

Let \((X(t), t \geq 0)\) be a Markov chain in continuous time whose state space \(S = \{0\} \cup C\) consists of transient states \(C = \{1, 2, \ldots\}\) and an absorbing state \(0\). Let \(Q = (q_{ij}, i, j \in S)\) be the \(q\)-matrix of transition rates, assumed to be stable, conservative and regular, so that there is a unique transition function \(P(t) = (p_{ij}(t), i, j \in S)\) associated with \(Q\), and \(p_{ij}(t) = \Pr(X(t) = j | X(0) = i)\). We assume that \(C\) is irreducible, and that \(0\) is reached with probability 1 from any state in \(C\). Thus, in particular, for \(i, j \in C\), \(p_{ij}(t) \to 0\) as \(t \to \infty\), and, for \(i \in C\), \(p_{i0}(t) \to 1\). We are interested in the limit of the ratio

\[
\frac{p_{ij}(t)}{1 - p_{i0}(t)} = \Pr(X(t) = j | X(t) \neq 0, X(0) = i), \quad i, j \in C,
\]

called a limiting conditional distribution (LCD). The first general results on LCDs [16] were facilitated by a finer classification of transient states [9] and a common rate at which the transition probabilities decay: there is a \(\lambda \geq 0\), called the decay parameter (of \(C\)), such that \(t^{-1} \log p_{ij}(t) \to -\lambda\) as \(t \to \infty\), for all \(i, j \in C\). \(C\) is then \(\lambda\)-recurrent or \(\lambda\)-transient according to whether \(\int_{0}^{\infty} e^{\lambda t} p_{ij}(t) \, dt\) diverges or converges (for some, and then all, \(i, j \in C\)), and, when \(C\) is \(\lambda\)-recurrent, it is \(\lambda\)-positive or \(\lambda\)-null according to whether the limit \(\lim_{t \to \infty} e^{\lambda t} p_{ij}(t)\) is positive or zero (for some, and then all, \(i, j \in C\)).

When positive, its value is determined by \((m_i, i \in C)\) and \((x_i, i \in C)\) satisfying

\[
\sum_{i \in C} m_i p_{ij}(t) = e^{-\lambda t} m_j \quad \text{and} \quad \sum_{j \in C} p_{ij}(t) x_j = e^{-\lambda t} x_i.
\]

Unique positive solutions to (2) (called the \(\lambda\)-invariant measure and vector, respectively) are guaranteed when \(C\) is \(\lambda\)-recurrent. \(C\) is then \(\lambda\)-positive if and only if

\[\text{Phil Pollett}\]

pkp@maths.uq.edu.au

1 School of Mathematics and Physics, The University of Queensland, Brisbane, Australia
The condition that $\kappa$ convergence for ergodic discrete-time Markov chains [5] (which has become important chains, which highlight the many complications in the theory of $R$[6,7], and more in line with work on algebraic transience [11] and sub-geometric convergence for ergodic discrete-time Markov chains [5] (which has become important in the analysis of Markov chain Monte Carlo methods). The condition that $\kappa$ exists is

$$A^{-1} := \sum_{k \in C} m_k x_k < \infty,$$

whence $\lim_{t \to \infty} e^{\lambda t} p_{ij}(t) = Ax_i m_j$. So, one can see, at least formally from (1), that, since $p_{i0}(t) = 1 - \sum_{j \in C} p_{ij}(t)$,

$$\Pr(X(t) = j | X(t) \neq 0, X(0) = i) = \frac{e^{\lambda t} p_{ij}(t)}{e^{\lambda t} \sum_{k \in C} p_{ik}(t)} \to \frac{m_j}{\sum_{k \in C} m_k}.$$  

$\lambda$-positivity is indeed sufficient [17], the limit taken to be 0 when $\sum_{k \in C} m_k = \infty$.

One might think this completes the picture. However, $\lambda$-positivity is not necessary for the existence of an LCD (see the example below). Further, the decay parameter cannot usually be determined from $Q$, and $\lambda$-positivity cannot usually be checked from $Q$. Of course $p_{ij}(t)$ is seldom available explicitly, but there are “$q$-matrix versions” of (2),

$$\sum_{i \in C} m_i q_{ij} = -\lambda m_j \quad \text{and} \quad \sum_{j \in C} q_{ij} x_j = -\lambda x_i,$$

and positive solutions to (3) satisfy (2) under conditions that are easy to check [12].

**Example** Consider the M/M/1 queue with arrival rate $p$ and departure rate $q (> p)$ modified so that it is killed when the queue size first reaches 0. Set $a = p + q$, $b = \sqrt{pq}$, and $\theta = 2\sqrt{pq}$. Seneta [14] showed that as $t \to \infty$,

$$p_{ij}(t) = i j b^{j-i} \frac{2 e^{-(a-\theta) t}}{\theta \sqrt{2\pi \theta}} \left( \frac{1}{t^{3/2}} + O \left( \frac{1}{t^{5/2}} \right) \right), \quad i, j \in C,$$

which implies that $\lambda = a - \theta = p + q - 2\sqrt{pq}$ is the decay parameter, and

$$\lim_{t \to \infty} t^{3/2} e^{\lambda t} p_{ij}(t) = i j b^{j-i} \frac{2}{\theta \sqrt{2\pi \theta}}, \quad i, j \in C.$$  

Notice that this limit is of the form $Ax_i m_j$, where $m_j = j b^{j-i}$ and $x_i = i b^{-i}$ specify the unique positive solutions to (3), and $A > 0$. Seneta also showed that

$$\lim_{t \to \infty} t^{3/2} e^{\lambda t} (1 - p_{i0}(t)) = \frac{ib^{-i}}{\lambda \sqrt{2\pi \theta}}, \quad i \in C.$$  

Notice also that this limit is of the form $Bx_i$, where $B > 0$. So, the LCD exists:

$$\lim_{t \to \infty} \frac{p_{ij}(t)}{1 - p_{i0}(t)} = \lim_{t \to \infty} \frac{t^{3/2} e^{\lambda t} p_{ij}(t)}{t^{3/2} e^{\lambda t} (1 - p_{i0}(t))} = (1 - b)^2 j b^{j-1}, \quad i, j \in C.$$  

Yet, $C$ is $\lambda$-transient: $\int_0^\infty e^{\lambda t} p_{ij}(t) \, dt = 2i/\theta < \infty$.

**2 Speculation** One way to approach the question of whether LCDs exist for $\lambda$-transient chains is to characterise the smallest $\kappa > 1$ such that $t^\kappa e^{\lambda t} p_{ij}(t)$ has a strictly positive limit for all $i, j \in C$. Such a characterisation is presently unavailable. I conjecture (for $\lambda$-null and $\lambda$-transient chains) that (i) when such a $\kappa$ exists, it is the same for all $i, j \in C$, that (ii) the limit is always of the form $Ax_i m_j$, where $(m_i, i \in C)$ and $(x_i, i \in C)$ satisfy (2) perhaps with an inequality (\leq), and (iii) there is a $\kappa_0 \leq \kappa$ such that $t^{\kappa_0} e^{\lambda t} (1 - p_{i0}(t)) \to Bx_i$.

**3 Discussion** The approach proposed here contrasts with work on discrete-time chains, which highlight the many complications in the theory of $R$-transient chains [6,7], and more in line with work on algebraic transience [11] and sub-geometric convergence for ergodic discrete-time Markov chains [5] (which has become important in the analysis of Markov chain Monte Carlo methods). The condition that $\kappa$ exists is
equivalent to requiring that the function $g_{ij}(t) := e^{\lambda t} p_{ij}(t)$ is regularly varying with index $\alpha = \kappa^{-1}$ (the same index of regular variation for all $i$ and $j$), or, equivalently, that there is a $\kappa > 1$, the same for all $i, j \in C$, such that $t^\kappa g_{ij}(t)$ is slowly varying [15]. Conjectures (i) and (ii) are true for $\lambda$-null recurrent chains (Lemma 1 of [13]), and indeed (iii) with $\kappa_0 = \kappa$ under an additional condition. Critical to the argument is that the $\lambda$-subinvariant measures and vectors ((2) with the inequality) are unique and $\lambda$-invariant. This is not true in the $\lambda$-transient case. One might hope to adapt Kingman’s arguments based on inequalities derived from the Chapman–Kolmogorov equations [9], but I cannot see how. In addition to the example detailed above, the conjectures are supported by several other contrasting models: the M/M/1 queue with $p > q$, the random walk on $\mathbb{Z}$ in continuous time [9], the birth–death immigration process [1], quasi-birth–death processes [3], and various branching models [2]. Interestingly, the critical Markov branching process provides an example for which $\kappa = 2$ and $\kappa_0 = 1$.

References

1. Anderson, W.: Continuous-Time Markov Chains: An Applications-Oriented Approach. Springer-Verlag, New York (1991)
2. Asmussen, S., Hering, H.: Branching Processes. Birkhauser, Boston (1983)
3. Bean, N., Bright, L., Latouche, G., Pearce, C., Pollett, P., Taylor, P.: The quasistationary behaviour of quasi-birth-and-death processes. Ann. Appl. Probab. 7, 134–155 (1997)
4. Coolen-Schrijner, P., Hart, A., Pollett, P.: Quasistationarity of continuous-time Markov chains with positive drift. J. Austral. Math. Soc. 41, 423–441 (2000)
5. Douc, R., Moulines, E., Soulier, P.: Computable convergence rates for sub-geometric ergodic Markov chains. Bernoulli 13, 831–848 (2007)
6. Foley, R.D., McDonald, D.R.: Yaglom limits can depend on the starting state. Adv. Appl. Probab. 50, 1–34 (2017)
7. Kesten, H.: A ratio limit theorem for (sub) Markov chains on $\{1, 2, \ldots\}$ with bounded jumps. Adv. Appl. Probab. 27, 652–691 (1995)
8. Kijima, M.: Quasi-stationary distributions of single-server phase-type queues. Math. Operat. Res. 18, 423–437 (1993)
9. Kingman, J.: The exponential decay of Markov transition probabilities. Proc. Lond. Math. Soc. 13, 337–358 (1963)
10. Kyprianou, E.: The quasi-stationary distributions of queues in heavy traffic. J. Appl. Probab. 9, 821–831 (1972)
11. Mao, Y.-H., Song, Y.-H.: On geometric and algebraic transience for discrete-time Markov chains. Stochastic Process. Appl. 124, 1648–1678 (2014)
12. Pollett, P.: On the equivalence of $\mu$-invariant measures for the minimal process and its $q$-matrix. Stochastic Process. Appl. 22, 203–221 (1986)
13. Pollett, P.: Similar Markov chains. J. Appl. Probab. 38A, 53–65 (2001)
14. Seneta, E.: Quasi-stationary behaviour in the random walk with continuous time. Austral. J. Statist. 8, 92–98 (1966)
15. Seneta, E.: Regularly Varying Functions . Lecture Notes in Mathematics. Springer-Verlag, New York (1976)
16. Seneta, E., Vere-Jones, D.: On quasi-stationary distributions in discrete-time Markov chains with a denumerable infinity of states. J. Appl. Probab. 3, 403–434 (1966)
17. Vere-Jones, D.: Some limit theorems for evanescent processes. Austral. J. Statist. 11, 67–78 (1969)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.