An efficient approximation method for phase-type distributions

by

Jung Hee Kim

and

Bok Sik Yoon

Hong Ik University
Department of mathematics

Abstract

The Phase-type(PH) distribution, defined as a distribution of the time until the absorption in a finite continuous-time Markov chain state with one absorbing state, has been widely used for various stochastic modelling. But great computational burdens often make us hesitate to apply PH methods. In this paper, we propose a seemingly efficient approximation method for phase type distributions. We first describe methods to bound the first passage time distribution in continuous-time Markov chains. Next, we adapt these bounding methods to approximate phase-type distributions. Numerical computation results are given to verify their efficiency.

1. Introduction

A nonnegative random variable T is said to be of phase type(PH) if T is the time until absorption in a finite-state continuous-time Markov chain having an absorbing state. PH distributions, which generalize traditional Erlang distributions, have been widely used in probabilistic modelling because of their versatility and the relative easiness of numerical implementations[cf. Neuts(1981), Shaked and Shanthikumar(1984)]. But as we can see from equation (2) in the next section, the actual computations of PH distributions is not a simple matter. In this paper, we propose a seemingly efficient approximation method for phase type distributions. We first describe methods to bound the first passage time distribution in continuous-time Markov chains. Next, we adapt these bounding methods to approximate phase-type distributions. Numerical computation results are given to verify their efficiency.

2. Computation of PH distributions

To represent a PH distribution formally, consider an absorbing continuous-time Markov chain $\{X(t), t \geq 0\}$ with state space $\{1, 2, \ldots, m, \Delta\}$, where states $1, 2, \ldots, m$ are transient and state Δ is absorbing. The infinitesimal generator of the Markov chain will be of the form

$$R = \begin{bmatrix} A & -Ae \\ 0 & 0 \end{bmatrix} \qquad (1)$$

where $A = \{a_{ij}\}$ is an $m \times m$ matrix with negative diagonal elements, ν_i, $i = 1, \ldots, m$ and nonnegative off-diagonal elements and $Ae \leq 0$. Here e denotes the m-dimensional column vector $(1, 1, \ldots, 1)^T$ and 0 denotes the m-dimensional column vector of zeros. Let (α, α_Δ) be an initial probability vector, that is, $\alpha_i = P(X(0) = i)$, $i = 1, \ldots, m$, and $\alpha_\Delta = P(X(0) = \Delta)$. Letting $T = \inf\{t : X(t) = \Delta\}$ denotes the time until absorption, the distribution function of T is

$$F(t) = 1 - \alpha \exp\{At\} e, \qquad t \geq 0. \qquad (2)$$

Note that if $\alpha_\Delta > 0$, then F has an atom at 0 [i.e. $P(t = 0) = \alpha_\Delta$] and is absolutely continuous on $(0, \infty)$.

There may be several ways to compute (2). One direct method is to compute the infinite series

$$\exp(At) = \sum_{n=0}^{\infty} \frac{(At)^n}{n!} \qquad (3)$$

with appropriate truncation. But this method is very time-consuming and numerically unstable. Alternatively we may use uniformization technique[Shaked and Shanthikumar(1984), Yoon(1988)], in which we choose a finite $\lambda \geq \max_{1 \leq i \leq m}\{\nu_i\}$ and compute F by

$$F(t) = \sum_{n=0}^{\infty}\left(\sum_{r=0}^{n} g(r)\right)\frac{e^{-\lambda t}(\lambda t)^n}{n!}, \quad t \geq 0, \qquad (4)$$

where $g(0) = \alpha_{\triangle}$, $g(n) = \alpha U^{n-1} r_0$, $r_0 = -Ae/\lambda$ and the elements of the $m \times m$ matrix U are

$$u_{ij} = \begin{cases} a_{ij}/\lambda & \text{if } i \neq j \\ 1-\nu_i/\lambda & \text{if } i = j \end{cases} \qquad (5)$$

But still the computational burden is huge.

3. Bouding first passage time distributions in Markov Chains

Consider the continuous-time Markov chain X={X(t), t≥0} with the transition probability matrix Q = (q_{ij}) and state finite space S. Let P(t) denote the matrix of transition functions, that is, $P(t)_{ij}=P\{X(t)=j|X(0)=i\}$. Define a monotone sequence $\{T_n, n = 0,1,\ldots\}$of stopping times for X as follows :

$$\begin{aligned} T_0 &= 0 \quad \text{w.p.1 and} \\ T_n &= \min\{\ \inf\{t: X(t) \neq X(T_{n-1}),\ t > T_{n-1}\},\ T_{n-1} + \Delta\ \},\quad n = 1,2,\ldots . \end{aligned} \qquad (6)$$

Clearly for each n, $T_n \leq n\Delta$ w.p. 1

Now, letting $\delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{otherwise} \end{cases}$ and $\nu = Qe$,

$$P_{2:ij} = \begin{cases} (1-e^{-\nu_i\Delta})\frac{q_{ij}}{\nu_i} + e^{-\nu_i\Delta}\delta_{ij},\ \text{for } i,j \in S & \text{if } \nu_i \neq 0 \\ \delta_{ij} & \text{if } \nu = 0 \end{cases} \qquad (7)$$

, from the observation that T_n, $n = 1,2,\ldots,$ are stopping times of X, we can easily see that $E[\ P(T_1)] = P_2$ and

$$P_2(t) = E[\ P(T_{\lfloor t/\Delta \rfloor})] = P_2^{\lfloor t/\Delta \rfloor}, \quad t \geq 0 \qquad (8)$$

, where $\lfloor x \rfloor$ implies the largest integer less than or equal to x.

Proposition 1 :

For any fixed t, $P_2(t) \rightarrow P(t)$ as $\Delta \rightarrow 0$.

proof :

First we show that

$$\lim_{\Delta \to \infty} T_{\lfloor t/\Delta \rfloor} = t \quad a.s. \qquad (9)$$

Then, the proposition is immediate from (8). To show (10) note from the definition T_n that $T_n - T_{n-1} < \Delta$ iff a transition of X occurs during $(T_{n-1}, T_n + \Delta)$. Let Z(t) be the number of transition of X during (0,t). Since $T_n \leq n\Delta$ a.s., it is then obvious that

$$T_n \geq (n - Z(T_n))\Delta \geq n\Delta - Z(n\Delta)\Delta \qquad a.s.$$

Hence,

$$\lfloor t/\Delta \rfloor \Delta - Z(t)\Delta \leq T_{\lfloor t/\Delta \rfloor} \leq \lfloor t/\Delta \rfloor \Delta \qquad a.s.$$

Thus, as $\Delta \rightarrow 0$ one sees that $T_{\lfloor t/\Delta \rfloor} \rightarrow t$ a.s. (QED)

Now let's consider the following monotone sequence of stopping times of the uniformized continuous time Markov chain $\{ \overline{X}(t),\ t \geq 0 \}$.

$$\begin{aligned} &T_0 = 0 \quad w.p.1, \\ &T_n = \min\{ \inf\{ t : N(t) \neq N(T_{n-1}) \},\ T_{n-1} + \Delta \}, \end{aligned} \qquad (9)$$

Let P5 denote the transition matrix of the discrete time Markov chain obtained by observing $\overline{X}(t)$ at Δ T_n's that is, $P_5 \equiv E[\ P(T_1)]$. Then it is easily verified that

$$\begin{aligned} P_5 &= (1 - e^{-\lambda\Delta})(I + \frac{1}{\lambda}R) + e^{-\lambda\Delta}I \\ &= I + \frac{1 - e^{-\lambda\Delta}}{\lambda}R \ . \end{aligned} \qquad (10)$$

We may now approximate P(t) as above by

$$P_5(t) = P_5^{\lfloor t/\Delta \rfloor} \ . \qquad (11)$$

For this approximation we have the following proposition in which (i)follows in a manner similar to that in Proposition 1 and (ii) is verified by the observation: the limiting distribution of the discrete time Markov chain characterized by P_5 is the same as that of the original continuous time Markov chain characterized by R.

Proposition 2:

(i)For any fixed λ and t, $P_5 \rightarrow P(t)$ as $\Delta \rightarrow 0$.

(ii)For any fixed λ and Δ, $P_5(t) \rightarrow P(\infty)$ as $t \rightarrow \infty$, whenever $P(\infty)$ is well defined.

Note that $T_{\lfloor t/\Delta \rfloor} \leq t$ with probability 1.

Let

$$T_0 = 0 \quad \text{w.p. } 1$$
$$T_n = \inf\{ t : N(t+\Delta) = N(t),\ t \geq T_{n-1} \} + \Delta \qquad (12)$$

where $\{ N(t), t \geq 0 \}$ is the uniformizing Poisson process. Clearly $\{ T_n, n=1,2,\ldots \}$ is a sequence of renewal epochs of a renewal process and T_1 is the time to a gap of size Δ in the Poisson process $\{ N(t), t \geq 0 \}$. Suppose the first arrival takes place after time Δ. Then $T_1 = \Delta$ (this happens with probability $e^{-\lambda\Delta}$). Otherwise, the remaining time to T after the first arrival has the same distribution as T_1 (this happens with probability $1 - e^{-\lambda\Delta}$). Based on this observation one sees that

$$P_6 \equiv E[\ P(T_1)\ ] = (1-e^{-\lambda\Delta})(I + \frac{1}{\lambda}R)P_6 + e^{-\lambda\Delta} I\ , \qquad (13)$$

since when a transition occurs before time Δ, the state will be changed according to $(I + \frac{1}{\lambda}R)$ and the Markov chain starts over again with the current state as an initial state.

Solving (13), we have

$$P_6 = \left(I - \frac{1-e^{-\lambda\Delta}}{\lambda e^{-\lambda\Delta}} R \right)^{-1}. \qquad (14)$$

Then we approximate P(t) by

$$P_6(t) = P_6^{\lceil t/\Delta \rceil}, \qquad t \geq 0. \qquad (15)$$

Note that $T_{\lceil t/\Delta \rceil} \geq t$ with probability 1.

The following properties of $P_6(t)$ can be established without difficulty[Yoon and Shanthikumar(1988)].

Proposition 3 :

(i)For any fixed λ and t, $P_6 \rightarrow P(t)$ as $\Delta \rightarrow 0$.

(ii)For any fixed λ and Δ, $P_6(t) \rightarrow P(\infty)$ as $t \rightarrow \infty$, , whenever $P(\infty)$ is well defined.

In P_5 and P_6 we need to set the uniformization rate as small as possible to gain better accuracy for the same Δ. In $P_2(t)$, $P_5(t)$ and $P_6(t)$, if we choose Δ such that $t/\Delta = 2^k$, then to obtain $P_2^{\lfloor t/\Delta \rfloor}$ we need only k matrix multiplications.

4. Bounding PH Time Distributions

In this section we discuss the bounds for PH distributions using the mathods discussed in preceeding sections. Consider a Markov chain with generator R and let $T_{\cdot\Delta} = \inf\{t : X(t) = \Delta,\ t \geq 0\}$ be the first passage time to Δ from some initial state X(0) =

Then $F(t) \equiv P\{T_{\cdot\Delta} \leq t\} = P_{\cdot\Delta}(t)$, so that all the methods approximating P(t) can be directly used to approximate F(t). Moreover, we can establish bounds on F(t) without any restriction on X as in

Theorem 1

$$\max\{P_{2:.\Delta}(t), P_{5:.\Delta}(t)\} \le F(t) \le P_{6:.\Delta}(t). \qquad (16)$$

Proof :

Since state Δ is absorbing, it is obvious that $P\{X(s)=\Delta\} \le P\{X(t)=\Delta\}$ for all $s \le t$.
Now observing that for $P_2(t)$

$$P_{2:.\Delta}(t) = E[\, P\{X(T_{\lfloor t/\Delta \rfloor})=\Delta\}\,]$$

and $T_{\lfloor t/\Delta \rfloor} \le \lfloor t/\Delta \rfloor \Delta \le t$ a.s., one immediately sees that

$$P_{2:.\Delta} \le P\{X(t)=\Delta\} = F(t).$$

Similarily we get $P_{5:.\Delta} \le F(t)$. The upper bound is also true since $T_{\lceil t/\Delta \rceil}$ defined for $P_6(t)$ is $\ge \lceil t/\Delta \rceil \Delta \ge t$ a.s. $\diamond$

The above result seems to be very useful and illustrates the value of the approximations $P_2(t)$, $P_5(t)$ and $P_6(t)$.

5. Numerical Example :

(1) Generalized Erlang-2 with rates 1 and 2

$$F(t) = 1-2e^{-t}+e^{-2t}$$

$$R = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

Table 1. Bounding values for F(t)

Time	From $P_2(t)$	True Value	From $P_6(t)$
1/8	0.0137057	0.013807	0.0139275
1/2	0.154585	0.154818	0.155231
1	0.399349	0.399576	0.400163
2	0.747531	0.747645	0.748138
3	0.902858	0.902905	0.903189
4	0.963687	0.963704	0.963846
5	0.986563	0.98657	0.986635
6	0.995046	0.995049	0.995078
7	0.998176	0.998177	0.998189
8	0.999329	0.999329	0.999334
오차 최대값	0.000233		0.000587

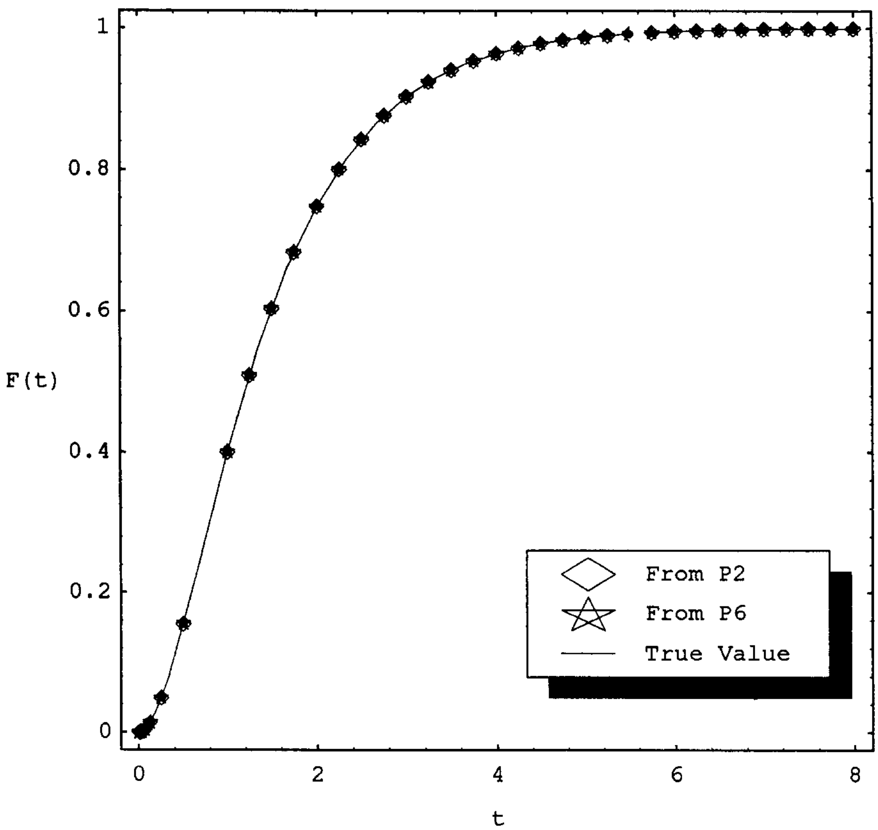

Figure 1. comparison of P2, P6, true values

References

Yoon, B.S. and J.G. Shanthikumar, "Bounds and Approximations for the Transient Behavior of Continuous-time Markov Chains," *Probability in the Engineering and Informational Sciences*, v3, 175-198, 1989.

Neuts, M.F. (1981c). *Matrix-Geometric Solutions in Stochastic Models: Algorithmic Approach.* Johns Hopkins University Press, Baltimore, MD. (Chapter 2 is excellent review on PH distributions)

Shaked, M. and J.G.Shanthikumar, "Phase type distributions" *in Encyclopedia of Statistical Sciences,* N.L.Johnonson and S.Kotz, Eds, pp709-715, Wiley, New York, 1984.