A Method for Evaluating the Distribution of the Total Cost of a Random Process over its Lifetime

P.K. Pollett[a] **and V.T. Stefanov**[b]

[a]Department of Mathematics, The University of Queensland
Queensland 4072, Australia (pkp@maths.uq.edu.au)

[b]Department of Mathematics and Statistics, The University of Western Australia, Crawley
Western Australia 6009, Australia (stefanov@maths.uwa.edu.au)

Abstract: For a random process $(X(t),\ t \geq 0)$, suppose that there is a cost f_x associated with being in state x. This paper is concerned with evaluating the distribution and the expected value of the total cost Γ over the life of the process. The existing literature contains results for particular classes of process and particular choices of f, usually linear functions of the state. We will describe a method which assumes only that f is non-negative. We characterize both the distribution and the expected value of Γ as extremal solutions of systems of linear equations. Of particular interest in biological applications is the case when there is a single absorbing state, corresponding to population extinction, where we are usually interested in evaluating the cost of the process up to the time of extinction. We will illustrate our results with reference to three important Markovian models: the pure-birth process, the birth-death process, and the linear birth-death and catastrophe process.

Keywords: *Hitting times; Extinction times; Population processes*

1. INTRODUCTION

Let $(X(t),\ t \geq 0)$ be a continuous-time Markov chain taking values in the non-negative integers $S = \{0, 1, \dots\}$ and let A be a fixed subset of S. Consider the path integral

$$\Gamma = \int_0^\tau f_{X(t)}\, dt, \tag{1}$$

where $f : A \to [0, \infty)$ and $\tau = \inf\{t > 0 : X(t) \notin A\}$ is the first exit time of A. Here f_j may be regarded as the cost per unit time of staying in state j. For example, f_j might represent the cost of an epidemic when there are j individuals infected. It could be the amount of nutrient consumed when a population is in state j, or, it could be the storage cost associated with an inventory consisting of j items. Γ is then the total cost over the period that the chain spends in A.

Path integrals of this kind has been used in a variety of applications: for example, in modelling epidemics (Puri, 1967, Jerwood, 1970, Downtown, 1972, Gani and Jerwood, 1972), in the assessment of plant stress (Billard et al., 1998), in modelling queueing, storage and traffic flow (Moran, 1959, Gaver, 1969, Gani, 1970) and in assessing the risk of computer virus attacks (Soh et al., 1995). However, in each case this work is limited by an assumption that f_x is a linear function of x. We will describe a method which assumes that f is an arbitrary non-negative function. We characterize the Laplace transform of both the distribution and the expected value of Γ as extremal solutions of systems of linear equations. Our results are closely related to the corresponding results (f identically 1) on the distribution of first passage times (Syski, 1992). Explicit formulae are available for special Markov chains. For example, in the case of birth-death processes, Flajolet and Guillemin (2000) and Ball and Stefanov (2001) have obtained results on transforms of the distributions of first passage times, and other characteristics, in terms of continued fractions.

2. THE DISTRIBUTION OF Γ

Let $Q = (q_{ij},\ i, j \in S)$ be the q-matrix of transition

*This work was funded by the ARC (Grant No. A00104575).

rates of the chain (assumed to be stable and conservative), so that q_{ij} represents the rate of transition from state i to state j, for $j \neq i$, and $q_{ii} = -q_i$, where $q_i := \sum_{j \neq i} q_{ij}$ $(< \infty)$ represents the total rate out of state i. It will not be necessary to assume that Q is regular, so that there may actually be many processes with the given set of rates. However, we will take $(X(t),\ t \geq 0)$ to be the *minimal chain* associated with Q.

We shall evaluate the Laplace transform of the distribution of path integral (1), conditional on the chain starting in state $i \in A$, making the harmless assumption that $q_j > 0$, for all $j \in A$, so that A contains no absorbing states. Let $y_i(\theta) = \mathrm{E}_i(e^{-\theta\Gamma})$, with the understanding that $y_i(\theta) = 1$ when $i \notin A$. (Here and henceforth we will use the notation $\mathrm{E}_i(\cdot) = \mathrm{E}(\cdot\,|X(0) = i)$ and $\mathrm{P}_i(\cdot) = \mathrm{Pr}(\cdot\,|X(0) = i)$.) The following result is a simple extension of the standard characterization of hitting times (for example, Theorem 9 on Page 86 of Syski (1992)). Its proof can be found in Pollett and Stefanov (2002).

Proposition 1 *For each $\theta > 0$, $y(\theta) = (y_i(\theta),\ i \in S)$ is the maximal solution to the system*

$$\sum_{j \in S} q_{ij} z_j = \theta f_i z_i\,, \quad i \in A, \tag{2}$$

with $0 \leq z_j \leq 1$ for $j \in A$ and $z_j = 1$ for $j \notin A$, in the sense that $y(\theta)$ satisfies these equations, and, if $z = (z_i,\ i \in S)$ is any such solution, then $y_i(\theta) \geq z_i$ for all $i \in S$.

Formal differentiation of (2) suggests a corresponding result on the expected value of the path integral, conditional on the chain starting in state $i \in A$. In fact, using similar arguments, we can arrive at the following result (Pollett, 2003), which is an extension of Theorem 10 on Page 86 of Syski (1992).

Proposition 2 *$e = (e_i,\ i \in A)$, where $e_i = \mathrm{E}_i(\Gamma)$, is the minimal non-negative solution to the system*

$$\sum_{j \in A} q_{ij} z_j + f_i = 0\,, \quad i \in A. \tag{3}$$

Remarks. If we set $f_i = 1$ for all $i \in A$, then $\Gamma = \tau$, and so the above results can be used to determine the distribution and the expectation of τ (the time to first exit from A). In the case when Q is regular, these reduce to well known and widely used results on hitting times; see, for example, Section 9.2 of Anderson (1991).

Of particular interest are the cases (i) $A = S$ with S irreducible, and (ii) $S = A \cup \{0\}$, with A irreducible and 0 being an absorbing state that is accessible from A. In both cases, Γ counts the cost over the lifetime of the chain. Note that if Q is not regular, then, in case (i), τ is the explosion time of the chain (which is almost surely finite for all starting states). The above results might therefore be useful in biological applications, where we may wish to account for explosive behaviour by allowing the chain to perform infinitely-many transitions in a finite time. Case (ii) was considered by Stefanov and Wang (2000) for birth-death processes. They derived an explicit expression for the expectation $\mathrm{E}_i(\Gamma)$, building on earlier work of Hernández-Suárez and Castillo-Chavez (1999), who studied the case $i = 1$ and the linear function $f_j = j$.

On dividing equation (2) by f_i, we see that, conditional on $X(0) = i$, Γ has the same distribution as τ for the Markov chain with transition rates $Q^* = (q^*_{ij},\ i, j \in S)$ given by $q^*_{ij} = q_{ij}/f_i$ for all $i \in A$ such that $f_i > 0$, and $q^*_{ij} = q_{ij}$ otherwise. This fact was first noticed by McNeil (1970) in the context of birth-death processes. It is intuitively reasonable, for if T_j is the total time that the process spends in state j during the period up to time τ, and N_j is the number of visits to j during that period, then

$$\Gamma = \sum_{j \in A} f_j T_j \quad \text{and} \quad T_j = \sum_{n=1}^{N_j} X_{jn},$$

where $\{X_{jn}, n = 1, 2, \dots\}$ are independent and identically distributed exponential random variables with parameter q_j. Since the distribution of N_j does not depend on the holding times, but rather on the transition probabilities $p_{ij} = q_{ij}/q_i$ of the jump chain, then, for states j with $f_j > 0$, $f_j T_j$ has the same distribution as the sum of N_j independent and identically distributed exponential random variables with parameter q_j/f_j. Therefore, since (in an obvious notation) $p^*_{ij} = p_{ij}$, and $q^*_i = q_i/f_i$ for all $i \in A$ such that $f_i > 0$, we would expect Γ to have the same distribution as τ for the modified chain. This observation will be useful in studying specific models for which the distribution and the expectation of τ are known in sufficient generality to accommodate state-dependent transition rates. For example, in the case of birth-death processes, there are explicit expressions for the expected value of various hitting times, and, expressions for transforms of their distributions are available in terms of continued fractions (Flajolet and Guillemin, 2000, Ball and Stefanov, 2001), while in several special cases the hitting time densities are known explicitly (Di Crescenzo, 1998).

3. SOME APPLICATIONS

In this section we give several applications to specific models. We start with the simplest Markov chain that can exhibit explosive behaviour.

The pure-birth process. This process has $q_{i,i+1} = q_i > 0$, $i \geq 0$, with all other transition rates equal to 0. The minimal chain has a lifetime τ, which is almost surely finite for all starting states if and only if the series $\sum_{j=0}^{\infty} 1/q_j$ converges. Equations (2) and (3) have explicit solutions. From (3) we get $\mathrm{E}_i(\Gamma) = \sum_{j=i}^{\infty} f_j/q_j$. From (2) we get $\mathrm{P}_i(\Gamma < \infty) = 1$ for all $i \in S$ if and only if $\mathrm{E}_0(\Gamma) < \infty$, in which case

$$\mathrm{E}_i(e^{-\theta\Gamma}) = \prod_{j=i}^{\infty} \frac{q_j}{q_j + \theta f_j}, \quad \theta > 0,\ i \in S.$$

Birth-death processes. These have $q_{i,i+1} = a_i$, $i \geq 0$, $q_{i,i-1} = b_i$, $i \geq 1$, $q_0 = a_0$ and $q_i = a_i + b_i$, $i \geq 1$, with all other transition rates equal to 0. We will assume that the birth rates $(a_i,\ i \geq 0)$ and the death rates $(b_i,\ i \geq 1)$ are all strictly positive, except perhaps a_0, which might be 0. Thus, we can accommodate the two cases referred to above: if $a_0 > 0$, then S is irreducible; otherwise, $S = A \cup \{0\}$, with $A = \{1, 2, \dots\}$ irreducible and 0 an absorbing state that is accessible from A.

Proposition 2 can be used to obtain explicit formulae for the expected value of the path integral. It is easy to show that, when S is irreducible,

$$\mathrm{E}_i(\Gamma) = \sum_{j=i}^{\infty} \frac{1}{\lambda_j \pi_j} \sum_{k=0}^{j} f_k \pi_k\,, \quad i \geq 0,$$

where the *potential coefficients* $(\pi_j,\ j \geq 0)$ are given by $\pi_0 = 1$ and

$$\pi_i = \prod_{j=1}^{i} \frac{\lambda_{j-1}}{\mu_j}, \quad i \geq 1.$$

When 0 is an absorbing state, we need to distinguish two cases depending on whether or not the series $A = \sum_{i=1}^{\infty} 1/(\mu_i \pi_i)$ converges, where now $(\pi_j,\ j \geq 1)$ are defined by $\pi_1 = 1$ and

$$\pi_i = \prod_{j=2}^{i} \frac{\lambda_{j-1}}{\mu_j}, \quad i \geq 2.$$

When $A = \infty$, a condition that corresponds to the process being non-explosive with absorption probability 1, we have

$$\mathrm{E}_i(\Gamma) = \sum_{j=1}^{i} \frac{1}{\mu_j \pi_j} \sum_{k=j}^{\infty} f_k \pi_k,$$

for all $i \geq 1$, this being finite if and only if $\sum_{k=1}^{\infty} f_k \pi_k < \infty$. When $A < \infty$, we have

$$\mathrm{E}_i(\Gamma) = \sum_{j=1}^{i} \frac{1}{\mu_j \pi_j} \left(\frac{C(f)}{A} - \sum_{k=1}^{j-1} f_k \pi_k \right),$$

for $i \geq 1$, where

$$C(f) = \sum_{j=2}^{\infty} \frac{1}{\mu_j \pi_j} \sum_{k=1}^{j-1} f_k \pi_k = \sum_{j=1}^{\infty} \frac{1}{\lambda_j \pi_j} \sum_{k=1}^{j} f_k \pi_k,$$

with $\mathrm{E}_i(\Gamma)$ being finite if and only if $C(f) < \infty$.

Proposition 1 can be used to identify the distribution of Γ for specific cases. To illustrate this, consider the *linear birth-death process*, which has $a_i = ai$ and $b_i = bi$ for $i \geq 0$, where a and b are positive constants. Suppose that $a < b$, so that the time τ to absorption is finite with probability 1 for all starting states, and consider the path integral $\Gamma = \int_0^{\tau} X(t)\, dt$. With this specification, equations (2) can be solved explicitly, and it is a simple matter to identify the minimal solution. We find that $\mathrm{E}_i(e^{-\theta\Gamma}) = (\gamma(\theta))^i$, where $\gamma(\theta)$ is the smaller of the two zeros of $as^2 - (a + b + \theta)s + b$ (which are both real and positive). Indeed, the Laplace transform can be inverted to give the probability density

$$d\mathrm{P}_i(\Gamma \leq t) = \frac{i}{t} e^{-(a+b)t} \left(\frac{b}{a}\right)^{i/2} I_i(2t\sqrt{ab})\, dt,$$

where $I_i(z)$ is the usual modified Bessel function of the first kind. This accords with the first-passage time density of state 0 for the $M/M/1$ queue (Abate, Kijima and Whitt (1991)); see the remark at the end of the previous section.

The birth-death and catastrophe process. This process, first studied by Brockwell (1985), extends the linear birth-death process by allowing for downward jumps of arbitrary size (catastrophes). It has transition rates given by

$$q_{ij} = \begin{cases} i\rho a, & i \geq 0,\ j = i+1, \\ -i\rho, & i \geq 0,\ j = i, \\ i\rho d_{i-j}, & i \geq 2,\ 1 \leq j < i, \\ i\rho \sum_{k \geq i} d_k, & i \geq 1,\ j = 0, \end{cases}$$

with all other transition rates equal to 0. Here ρ and a are positive, d_i is positive for at least one value of i in $A = \{1, 2, \dots\}$ and $a + \sum_{i \geq 1} d_i = 1$. Clearly 0 is an absorbing state for the process and A is an irreducible class. It is easy to establish that the chain is non-explosive (Corollary 1 of Pollett and Taylor (1993)). Brockwell (1985) showed that the probability of extinction starting with i individuals is 1 for all $i \in A$ if and only if D (the expected increment size), given by

$$D = a - \sum_{i=1}^{\infty} i d_i = 1 - \sum_{i=1}^{\infty} (i+1) d_i,$$

is less than or equal to 0; the process is said to be *subcritical*, *critical* or *supercritical* according as D is negative, zero or positive.

We will consider only the subcritical case. Let

$$d(s) = a + \sum_{i=1}^{\infty} d_i s^{i+1}, \quad |s| < 1,$$

and $b(s) = d(s) - s$, so that, for example,

$$D = -b'(1-) = 1 - d'(1-) \; (< 0).$$

Observe that $b(0) = a \; (> 0)$, $b(1) = 1$, that b is strictly convex on $[0, \infty)$ and that b has a unique zero σ on $(0, 1)$. We will study the path integral (1) with τ being the time to absorption. First we will evaluate $\mathrm{E}_i(\Gamma)$. We seek the minimal non-negative solution to the system

$$i\rho a z_{i+1} - i\rho z_i + i\rho \sum_{j=1}^{i-1} d_{i-j} z_j + f_i = 0, \qquad (4)$$

where $i \geq 1$. On multiplying by s^{i-1} and then summing over i, we find that a non-negative solution exists whenever $g(\sigma) < \infty$, where g is the generating function of the sequence $(g_i, \; i \geq 1)$ given by $g_i = f_i/i$: $g(s) = \sum_{i=1}^{\infty} g_i s^i$. The generating function of the minimal solution is then easily evaluated in terms of g. We find that

$$\sum_{i=1}^{\infty} \mathrm{E}_i(\Gamma) s^{i-1} = \frac{g(\sigma) - g(s)}{\rho b(s)}, \quad |s| < \sigma.$$

Since $1/b(s)$ has a power series expansion near $s = 0$ with positive coefficients (Lemma V.12.1 of Harris (1963)), we may write

$$\frac{1}{\rho b(s)} = \sum_{j=0}^{\infty} e_j s^j, \quad |s| < \sigma,$$

where $e_j > 0$; note that $e_0 = 1/(a\rho)$. Thus, $\mathrm{E}_i(\Gamma) < \infty$ if and only if $g(\sigma) < \infty$, in which case

$$\mathrm{E}_i(\Gamma) = g(\sigma) e_{i-1} - \sum_{j=0}^{i-2} g_{i-1-j} e_j, \;\; i \geq 1. \qquad (5)$$

(Empty sums are taken to be 0.) To illustrate this, take $f_i = \alpha^{i-1}$, where $\alpha > 0$, so that

$$g(s) = -\frac{1}{\alpha} \log(1 - \alpha s), \quad |s| < \frac{1}{\alpha},$$

and hence $g(\sigma) < \infty$ provided $\alpha < 1/\sigma$. For example, if $\alpha = 1$, then $f_i = 1$ and $g(\sigma) < \infty$ (since $\sigma < 1$). We deduce that $\mathrm{E}_i(\tau) < \infty$ and

$$\mathrm{E}_i(\tau) = -\log(1 - \sigma) e_{i-1} - \sum_{j=0}^{i-2} \frac{e_j}{i - 1 - j}, \; i \geq 1,$$

or equivalently,

$$\sum_{i=1}^{\infty} \mathrm{E}_i(\tau) s^{i-1} = \frac{1}{\rho b(s)} \log\left(\frac{1 - s}{1 - \sigma}\right), \; |s| < \sigma.$$

This is equation (3.1) of Brockwell (1985). As a further illustration, take $f_i = i$. In this case we have $g_i = 1$. Hence, $g(s) = s/(1 - s)$, and so

$$\mathrm{E}_i(\Gamma) = e_{i-1}\left(\frac{\sigma}{1 - \sigma}\right) - \sum_{j=0}^{i-2} e_j, \quad i \geq 1,$$

or equivalently,

$$\sum_{i=1}^{\infty} \mathrm{E}_i(\tau) s^{i-1} = \frac{(\sigma - s)}{\rho(1 - \sigma)(1 - s) b(s)}, \; |s| < \sigma.$$

More explicit results can be obtained in the case where the catastrophe size follows a geometric law. Suppose that $d_i = d(1 - q) q^{i-1}$, $i \geq 1$, where $d (> 0)$ satisfies $a + d = 1$, and $0 \leq q < 1$. (The linear birth-death process is recovered on setting $q = 0$.) It is easy to see that $D = a - d/(1 - q)$. In order that $D < 0$ we require $c > 1$, where $c = q + d/a$. We also have

$$\begin{aligned} b(s) &= \frac{(d + qa)s^2 - (1 + qa)s + a}{1 - qs} \\ &= \frac{a(1 - s)(1 - cs)}{1 - qs}, \end{aligned}$$

and hence $\sigma = 1/c \; (< 1)$. The coefficients of the power series for $1/(\rho b(s))$ are easily evaluated using partial fractions. We find that

$$e_j = \frac{dc^j - (1 - q)a}{\rho a(d - (1 - q)a)}, \quad j \geq 0,$$

Thus, $\mathrm{E}_i(\Gamma)$ can be evaluated by substituting these expressions into (5), remembering that the expectation is finite whenever $g(\sigma) < \infty$. For example, if $f_i = \alpha^{i-1}$, where $\alpha > 0$, the expectation will be finite when $\alpha < q + d/a$, while if $f_i = i$, we get

$$\mathrm{E}_i(\Gamma) = \frac{q + (1 - q)i}{\rho(d - (1 - q)a)}.$$

Proposition 1 can be used to identify the distribution of Γ for specific choices of f. To illustrate this, suppose that $f_i = i$. For each $\theta > 0$, we seek the maximal solution to

$$\begin{aligned} &\rho a z_{i+1} - \rho z_i + \rho \sum_{j=1}^{i-1} d_{i-j} z_j \\ &\qquad + \rho z_0 \sum_{j=i}^{\infty} d_j = \theta z_i, \quad i \geq 1, \qquad (6) \end{aligned}$$

satisfying $z_0 = 1$ and $0 \leq z_j \leq 1$ for $j \geq 1$. On multiplying by s^{i-1} and then summing over i, we find that

$$\mathrm{E}_i(e^{-\theta\Gamma}) = \frac{1}{1 - s} - \frac{\theta(\gamma - s)}{(1 - \gamma)(1 - s)(\rho b(s) - \theta s)},$$

where $\gamma = \gamma(\theta)$ is the unique solution to $\rho b(s) = \theta s$ on the interval $0 < s < \sigma$. In the case of geometric catastrophes, we get

$$\mathrm{E}_i(e^{-\theta\Gamma}) = \frac{\beta(\theta) - q}{1 - q} \left(\beta(\theta)\right)^{i-1}, \quad i \geq 1,$$

where $\beta(\theta)$ is the smaller of the two zeros of $a\rho s^2 - (\rho(1 + qa) + \theta)s + \rho(d + qa) + q\theta$.

4. ACKNOWLEDGEMENTS

I would like to thank the referee for helpful comments on an earlier draft of the manuscript.

5. REFERENCES

Abate, J., M. Kijima and W. Whitt, Decompositions of the $M/M/1$ transition function, *Queuing Systems Theory Appl.*, 9, 323–336, 1991.

Anderson, W., *Continuous-Time Markov Chains: An Applications-Oriented Approach*, Springer-Verlag, New York, 1991.

Ball, F. and V. Stefanov, Further approaches to computing fundamental characteristics of birth-death processes, *J. Appl. Probab.*, 38, 995–1005, 2001.

Billard, L., P.W.A. Dayananda and S. Elmes, Assessment of plant stress due to the presence of disease, *Biometrics*, 54, 877–887, 1998.

Brockwell, P., The extinction time of a birth, death and catastrophe process and of a related diffusion model, *Adv. Appl. Probab.*, 17, 42–52, 1985.

Di Crescenzo, A., First-passage-time densities and avoiding probabilities for birth-and-death processes with symmetric sample paths, *J. Appl. Probab.*, 35, 383–394, 1998.

Downtown, F., The area under the infectives trajectory of the general stochastic epidemic, *J. Appl. Prob.*, 9, 414–417, 1972.

Flajolet, P. and F. Guillemin, The formal theory of birth-and-death processes, lattice path combinatorics and continued fractions, *Adv. Appl. Probab.*, 32, 750–778, 2000.

Gani, J., First emptiness problems in queuing, storage and traffic theory, *Proc. 6th Berkeley Symp. Math. Stat. Prob.*, 3, 515–532, 1970.

Gani, J. and D. Jerwood, The cost of a general stochastic epidemic, *J. Appl. Probab.*, 9, 257–269, 1972.

Gaver, D.P., Highway delays resulting from flow-stopping incidents, *J. Appl. Prob.*, 6, 137–153, 1969.

Harris, T., *The Theory of Branching Processes*, Springer-Verlag, Berlin, 1963.

Hernández-Suárez, C. and C. Castillo-Chavez, A basic result on the integral for birth-death Markov processes, *Math. Biosci.*, 161, 95–104, 1999.

Jerwood, D., A note on the cost of the simple epidemic, *J. Appl. Prob.*, 3, 339–352, 1970.

Moran, P.A.P., *The Theory of Storage*, Methuen, London, 1959.

McNeil, D. Integral functionals of birth and death processes and related limiting distributions, *Ann. Math. Statist.*, 41, 480–485, 1970.

Pollett, P.K., Integrals for continuous-time Markov chains, *Math. Biosci.*, 182, 113–225, 2003.

Pollett, P.K. and V.T. Stefanov, Path integrals for continuous-time Markov chains, *J. Appl.Probab.*, 39, 901–904, 2002.

Pollett, P. and P. Taylor, On the problem of establishing the existence of stationary distributions for continuous-time Markov chains, *Probab. Eng. Inf. Sci.*, 7, 529–543, 1993.

Puri, P.S., A class of stochastic models of response after infection in the absence of defense mechanism, *Proc. 5th Berkeley Symp. Math. Stat. Prob.*, 4, 537–547, 1967.

Soh, B.C., T.S. Dillon and P. County, Quantitative risk assessment of computer virus attacks on computer networks, *Computer Networks*, 27, 1447–1456, 1995.

Stefanov, V. and S. Wang, A note on integrals for birth-death processes, *Math. Biosci.*, 168, 161–165, 2000.

Syski, R., *Passage Times for Markov Chains*, IOS Press, Amsterdam, 1992.