NUMERICAL COMPUTATION OF LARGE DEVIATIONS EXPONENTS VIA SIMULATION

Peter W. Glynn*
Department of Engineering-Economic
Systems and Operations Research
Stanford University
Stanford, CA 94305-4023, USA

Abstract

Consider the problem of numerically computing the exponential rate at which the tail of the hitting time of a Markov chain to a given set decreases to zero. One approach involves computing the solution to an eigenvalue problem. In this paper, we use a regenerative representation for the exponential rate constant to construct an associated simulation-based estimator. A strong law and central limit theorem for the estimator are also presented.

Keywords

Large deviations, simulation, regeneration, eigenvalue problem, hitting time distribution, Markov chains.

1 Introduction

It has long been known that the hitting time distributions of Markov processes in both discrete and continuous time often exhibit exponential decay in their tail probabilities. However, the numerical computation of the corresponding decay rate constants has received less attention.

In this paper, we consider a simulation-based approach for computing such decay rate constants (referred to, in the sequel, as the "large deviations exponent"). Our approach depends upon a regenerative representation for such exponents that was implicit in the work of Ney and Nummelin (1987) on large deviations for additive functionals of Markov processes.

This paper is organized as follows. Section 2 sets up the basic mathematical framework in which our analysis will be conducted. In Section 3, we represent the large deviations exponent of interest in terms of a regenerative quantity, and provide some associated theory. Section 4 describes our estimator, and provides a law of large numbers and central limit theorem for our methodology.

*This research was supported by the Army Research Office under Contract No. DAAH04-94-G-0021

2 The Basic Framework

Let $X = (X_n : n \geq 0)$ be a (time-homogeneous) Markov chain living on a finite or countably infinite state space S, with one-step transition matrix $P = (P(x, y) : x, y \in S)$. For a given non-empty subset $A \subseteq S$ such that $A^c \neq \emptyset$, let

$$T = \inf\{n \geq 0 : X_n \in A\}$$

be the first hitting time of A. Our interest is in studying the rate at which the tail probabilities of T decrease to zero, given that the process is initially started in A^c.

Consider the assumption:

A1 The restriction of P to A^c, namely $B = (P(x, y) : x, y \in A^c)$, is irreducible.

For $x \in S$, let $P_x(\cdot) = P(\cdot | X_0 = x)$. Using elementary arguments, it is straightforward to establish the following result, due to Vere-Jones (1962) and Kingman (1963).

Theorem 1 *Under Assumption A1, there exists $\lambda \geq 0$ such that for each $x, y \in A^c$,*

$$\frac{1}{n} \log P_x(X_n = y, T > n) \to -\lambda \tag{1}$$

as $n \to \infty$.

Theorem 1 suggests the approximation

$$P_x(X_n = y, T > n) \approx \exp(-\lambda n) \tag{2}$$

for n large. (An important consequence of the theorem is that the exponential rate constant λ is independent of x and y.) This is, of course, a special type of "large deviations" asympototic. Our interest is in efficient numerical computation of the "large deviations exponent" λ.

Before proceeding, it should be noted that $A1$ does not generally imply either

$$\frac{1}{n} \log P_\mu(X_n = y, T > n) \to -\lambda$$

or that

$$\frac{1}{n} \log P_x(T > n) \to -\lambda$$

as $n \to \infty$. (Here, $P_\mu(\cdot)$ is the probability distribution of X, conditional on X_0 following distribution μ.) It is easy to construct non-pathological examples in which the large deviations exponents for $P_\mu(X_n = y, T > n)$ and $P_x(T > n)$ differ from λ. Even simple random walks provide such examples; see Kingman (1963) for details. (Of course, it is obvious that all such examples must involve chains in which $|A^c| = +\infty$.) Thus, it is evident that additional restrictions must be imposed on X in order to guarantee that these various large deviations exponents coincide.

3 The Regenerative Viewpoint

The regenerative structure of discrete state space Markov chains is commonly used to study the steady-state of such stochastic processes. Here, we consider the use of regeneration in studying the hitting time distribution introduced in Section 2.

For $x, y \in A^c$, let

$$a(n) = P_x(X_n = y, T > n).$$

Set $\tau = \inf\{n \geq 1 : X_n = x\}$ and note that the strong Markov property, applied at time τ, implies that

$$\begin{aligned} a(n) &= P_x(X_n = y, T \wedge \tau > n) + \sum_{j=1}^{n} P_x(X_n = y, T > n, \tau = j) && (3) \\ &= b(n) + \sum_{j=1}^{n} a(n-j)g(j), && (4) \end{aligned}$$

where

$$\begin{aligned} b(n) &= P_x(X_n = y, T \wedge \tau > n) \\ g(n) &= P_x(\tau = n, T > \tau) \end{aligned}$$

Consequently, $a = (a(n) : n \geq 0)$ satisfies a (defective) renewal equation. The standard approach to studying such equations is to invoke the following assumption:

A2 There exists $\theta \geq 0$ such that

$$\sum_{n=1}^{\infty} e^{\theta n} g(n) = 1$$

Assumption A2 may be rewritten in the form: There exists $\theta \geq 0$ such that

$$E_x[e^{\theta\tau} : T > \tau] = 1.$$

In any case, if we let $\tilde{a}(n) = e^{\theta n}a(n)$, $\tilde{b}(n) = e^{\theta n}b(n)$, $\tilde{g}(n) = e^{\theta n}g(n)$, we note that by multiplying through Equation 4 by $e^{\theta n}$ that we arrive at the following (proper) renewal equation:

$$\tilde{a}(n) = \tilde{b}(n) + \sum_{j=1}^{n} \tilde{a}(n-j)\tilde{g}(j) \qquad (5)$$

Since $(\tilde{g}(n) : n \geq 1)$ is a probability mass function, conventional renewal theory may be applied to Equation 5. To do so, we require that:

A3 i) $(\tilde{g}(n) : n \geq 1)$ is an aperiodic sequence;
ii) $\sum_{n=1}^{\infty} n\tilde{g}(n) < \infty$.

Observe that Assumption A3 ii) may be re-expressed as requiring that

$$E_x[\tau e^{\theta\tau} : T > \tau] < \infty,$$

where θ is defined as in Assumption A2. Suppose $y = x$. Then

$$\begin{aligned}
\sum_{n=0}^{\infty} \tilde{b}(n) &= \sum_{n=0}^{\infty} e^{\theta n} b(n) \\
&= \sum_{n=0}^{\infty} e^{\theta n} P_x(X_n = x, T \wedge \tau > n) \\
&= E_x \sum_{n=0}^{T \wedge \tau - 1} I(X_n = x) e^{\theta n} \\
&= 1,
\end{aligned}$$

since $X_n \neq x$ for $1 \leq n < \tau$. Under Assumptions A2 and A3, the renewal theorem therefore applies (see, for example, Feller (1968)), permitting us to conclude that if $y = x$, then

$$\tilde{a}(n) \to \sum_{j=0}^{\infty} \tilde{b}(j) / \sum_{j=0}^{\infty} j \tilde{g}(j)$$

as $n \to \infty$. We have therefore established the following result.

Proposition 1 *Under Assumptions A2 and A3,*

$$P_x(X_n = x, T > n) \sim e^{-\theta n} / E_x[\tau e^{\theta \tau}; T > \tau]$$

as $n \to \infty$.

Comparing with Equation 2, we conclude that $\lambda = \theta$. Consequently, the large deviations exponent λ may be characterized as the root λ of the equation

$$E_x[e^{\lambda \tau}; T > \tau] = 1$$

This characterization will form the basis of our numerical approach to computing λ.

Before proceeding to a discussion of the algorithm, we wish to argue that under Assumption A1, if Assumptions A2 and A3 holds for one particular "regeneration state" x, then Assumptions A2 and A3 holds for all $x \in A^c$. Without such a guarantee, the choice of regeneration state would be complicated greatly.

Observe that if Assumption A2 holds for some $x \in A^c$, then for $y \in A^c$,

$$\begin{aligned}
& E_y[e^{\theta \tau}; T > \tau] \\
=& \sum_{z \neq x, z \in A^c} e^{\theta} P(y, z) E_z[e^{\theta \tau}; T > \tau] + e^{\theta} P(y, x) \\
=& e^{\theta} \sum_{z \in A^c} P(y, z) E_z[e^{\theta \tau}; T > \tau]
\end{aligned}$$

Hence, if we set $h(y) = E_y[e^{\theta \tau}; T > \tau]$, we may conclude that for $y \in A^c$,

$$\sum_{z \in A^c} P(y, z) h(z) = e^{-\theta} h(y).$$

In other words, h is a (column) eigenvector of B corresponding to eigenvalue $e^{-\theta}$. Furthermore, under Assumption A1, it is trivially evident that h is strictly positive. Also, h is finite-valued under Assumption A1, because

$$\begin{aligned} 1 = h(x) &\geq E_x[e^{\theta\tau}; \tau_y < \tau < T] \\ &\geq P_x(\tau_y < \tau < T)h(y) \end{aligned}$$

where $\tau_y = \inf\{n \geq 1 : X_n = y\}$. Then we have established the following result.

Proposition 2 *Assume that Assumption A2 holds, for some $x \in A^c$. Then under Assumption A1, there exists a strictly positive finite-valued (column) eigenvector of B such that*

$$Bh = e^{-\theta}h.$$

Proposition 2 establishes that the large deviations exponent can be characterized in terms of the Perron-Frobenius eigenvalue of B. Thus, one potential means of computing the large deviations exponent would involve computing the solution of the above eigenvalue problem.

Let $R = (R(y, z) : y, z \in A^c)$ be the matrix with entries

$$R(y, z) = e^{\theta}B(y, z)h(z)/h(y).$$

It is easily verified that R is irreducible and stochastic, under our assumptions. Let $\tilde{P}_y(\cdot)$ and $\tilde{E}_y(\cdot)$ denote the probability distribution and expectation on the path-space of X under which $X_0 = y$ and X evolves according to one-step transition matrix R. Then, for $y \in A^c$, it is evident that

$$\tilde{E}_y Z = E_y[Z\frac{h(X_n)}{h(X_0)}; T > \tau] \cdot e^{\theta n}$$

for any non-negative $\mathcal{F}_n$-measurable r.v. Z ($\mathcal{F}_n \triangleq \sigma(X_0, \cdots, X_n)$). The above relation can be easily extended to $\mathcal{F}_{\tau_z}$-measurable Z:

$$\tilde{E}_y[Z; \tau_z < \infty] = E_y[Ze^{\theta\tau_z}\frac{h(X_{\tau_z})}{h(X_0)}; T > \tau_z, \tau_z < \infty] = \frac{h(z)}{h(y)}E_y[Ze^{\theta\tau_z}; T > \tau_z]. \tag{6}$$

(We used here the fact that $T > \tau_z$ is incompatible with $\tau_z = +\infty$.) Setting $y = z = x$ and $z \equiv 1$, we arrive at the identity

$$\tilde{P}_x(\tau_x < \infty) = E_x[e^{\theta\tau}; T > \tau] = 1.$$

Consequently, Assumption A2 guarantees that R is a recurrent matrix. Under the irreducibility of Assumption A1, we know that recurrence is a "solidarity" property, so that $\tilde{P}_y(\tau_y < \infty) = 1$ for $y \in A^c$. Relation 6 therefore implies that

$$E_y[e^{\theta\tau_y}; T > \tau_y] = 1$$

for $y \in A^c$.

Setting $y = z = x$ and $Z = \tau$, equation 6 yields

$$\tilde{E}_x(\tau; \tau < \infty) = E_x[\tau e^{\theta\tau}; T > \tau] < \infty$$

Thus, if Assumption A2 and A3 hold for some $x \in A^c$, then R is positive recurrent. Since positive recurrence is also a "solidarity" property, evidently $\tilde{E}_y(\tau < \infty)$ for $y \in A^c$. Relation 6 therefore implies that

$$\infty > \tilde{E}_y \tau_y = E_y[\tau_y e^{\theta\tau_y}; T > \tau_y],$$

verifying Assumption A3 for $y \in A^c$. We have therefore proved the following result.

Proposition 3 *Assume that Assumption A1 holds. Then,*

1. *if Assumption A2 is valid for one $x \in A^c$, Assumption A2 is valid for all $x \in A^c$.*
2. *if Assumption A3 is valid for one $x \in A^c$, Assumption A3 is valid for all $x \in A^c$.*

Before concluding this section, we note that we have essentially established that under Assumption A1, A2 is equivalent to requiring that B is R-recurrent (with $R = e^{\theta}$), whereas Assumptions A2 and A3 together imply that B is R-positive (again, with $R = e^{\theta}$). The concepts of R-recurrence and R-positivity are fundamental to the theory of quasi-stationary distributions; see for example, Seneta (1980).

In the next section, we describe our simulation-based estimator for computing the large deviations exponent λ (or equivalently, θ).

4 A Simulation-Based Estimator

If B is irreducible, then Proposition 3 establishes that if

$$E_y[e^{\theta\tau_y}; \tau_y < T] = 1 \tag{7}$$

holds for one $y \in A^c$, then Equation 7 holds for all $y \in A^c$. So, fix $x \in A^c$, and consider simulating n i.i.d. replicates of X up to time $T \wedge \tau_x$, conditioned on starting in state x. In particular, let

$$W_i(\gamma) \stackrel{\Delta}{=} e^{\gamma\tau_i} I(\tau_i < T_i)$$

be the value of the r.v. $e^{\gamma\tau_x} I(\tau_x < T)$ associated with i-th replicate. Set

$$\bar{W}_i(\gamma) = \frac{1}{n}\sum_{i=1}^{n} W_i(\gamma).$$

Clearly, $\bar{W}_n(\cdot)$, is strictly increasing and continuous, with $\bar{W}_n(-\infty) = 0$ and $\bar{W}_n(\infty) = +\infty$. Hence, for each n, there exists a unique root λ_n to the equation

$$\bar{W}_n(\lambda_n) = 1$$

It is easy to show that the empirical root λ_n is consistent for λ.

Proposition 4 *Under Assumptions A1 and A2.* $\lambda_n \to \lambda$ *almost surely as* $n \to \infty$. *where* λ *solves Equation 7.*

Proof: For $\varepsilon > 0$, the strong law of large numbers guarantees that

$$\begin{aligned} \bar{W}_n(\lambda - \varepsilon) &\to E_x[e^{(\lambda-\varepsilon)\tau_x}; \tau_x < T] < 1 \\ \bar{W}_n(\lambda + \varepsilon) &\to E_x[e^{(\lambda+\varepsilon)\tau_x}; \tau_x < T] > 1 \end{aligned}$$

as $n \to \infty$. Consequently, $\lambda_n \in (\lambda - \varepsilon, \lambda + \varepsilon)$ for n sufficiently large. Since ε was arbitrary, this proves the result. □

We turn next to the central limit theorem (CLT) for λ_n. Assume that:

A4 $E_x[e^{2\lambda\tau_x}; \tau_x < \tau] < \infty$.

Theorem 2 *Under Assumptions A1, A2 and A4,*

$$n^{1/2}(\lambda_n - \lambda) \Rightarrow \frac{\sqrt{Var_x W_1(\lambda)}}{w'_x(\lambda)} N(0,1)$$

as $n \to \infty$, *where* $w'_x(\lambda) = E_x[\tau_x e^{\lambda\tau_x}; \tau_x < T]$ *and* $Var_x(\cdot) \triangleq Var(\cdot|X_0 = x)$.

Proof: Note that

$$\bar{W}_n(\lambda_n) - \bar{W}_n(\lambda) = 1 - \bar{W}_n(\lambda),$$

and consequently

$$\bar{W}'_n(\xi_n)(\lambda_n - \lambda) = 1 - \bar{W}_n(\lambda), \tag{8}$$

where ξ_n lies between λ and λ_n. Now,

$$\bar{W}'_n(\xi_n) = \frac{1}{n}\sum_{i=1}^{n} \tau_i e^{\xi_n \tau_i} I(\tau_i < T_i).$$

Because $\bar{W}'_n(\cdot)$ is strictly increasing and $\lambda_n \to \lambda$ a.s. as $n \to \infty$. it is evident that for $\varepsilon > 0$ and n large enough,

$$\bar{W}'_n(\lambda - \varepsilon) \le \bar{W}'_n(\xi_n) \le \bar{W}'_n(\lambda + \varepsilon).$$

But the strong law of large numbers ensures that

$$\begin{aligned} \bar{W}'_n(\lambda - \varepsilon) &\to w'_x(\lambda - \varepsilon). \\ \bar{W}'_n(\lambda + \varepsilon) &\to w'_x(\lambda + \varepsilon) \end{aligned}$$

as $n \to \infty$. Since ε was arbitrary, we may conclude that $\bar{W}_n(\xi_n) \to w'_x(\lambda)$ a.s. as $n \to \infty$. On the other hand, Assumption A4 ensures that

$$n^{1/2}(1 - \bar{W}_n(\lambda)) \Rightarrow \sqrt{Var_x W_1(\lambda)} N(0,1) \tag{9}$$

as $n \to \infty$. Combining Equations 8 and 9 yields the desired conclusion. □

Given Proposition 3, it is perhaps surprising that the "second moment" hypothesis Assumption A4 is not a solidarity property in the presence of Assumption A1. In particular, it is easy to construct examples in which Assumption A4 holds for one $x \in A^c$ but not all $x \in A^c$, even in the presence of irreducibility. This raises the question of what happens when Assumption A4 fails to hold. We will study this issue further in future research.

Given the CLT provided by Theorem 2, we may construct confidence intervals for the large deviations exponent λ. In particular, let

$$V_n(\gamma) = \frac{1}{n-1}\sum_{i=1}^{n}(W_i(2\gamma) - \bar{W}_n(\gamma))^2.$$

It is easily verified that under the conditions of Theorem 2 (plus the additional proviso that $Var_x W_1(\lambda) > 0$),

$$[\lambda_n - z\frac{\sqrt{V_n(\lambda_n)}}{\bar{W}_n'(\lambda_n)}, \lambda_n + z\frac{\sqrt{V_n(\lambda_n)}}{\bar{W}_n'(\lambda_n)}]$$

is an asymptotic (as $n \to \infty$) $100(1-\delta)\%$ confidence interval for λ, where z solves the equation $P(N(0,1) \leq z) = 1 - \delta/2$.

References

[1] Feller, W. (1968). "An Introduction to Probability Theory and its Applications, Volumn 1" (3rd edition). John Wiley, New York.

[2] Kingman, J.F.C. (1963). The exponential decay of Markov transition probabilities. Proc. London Math. Soc. (3) 13, 337-338.

[3] Ney, P. and E. Nummelin (1987). Markov additive processes I: Eigenvalue properties and limit theorems. Annals of Probability 15, 561-592.

[4] Seneta, E. (1980) Non-negative Matrices and Markov Chains. Springer-Verlag, New York.

[5] Vere-Jones, D. (1962). Geometric ergodicity in denumerable Markov chains. Quart. J. Math. Oxford (2nd sec.) 13, 7-28.