Total Reward Variance in Discrete and Continuous Time Markov Chains

Karel Sladký[1] and Nico M. van Dijk[2]

[1] Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod vodárenskou věží 4, 182 08 Praha 8, Czech Republic
e-mail: sladky@utia.cas.cz
[2] University of Amsterdam, Department of Economic Sciences and Econometrics,
Roetersstrat 11, 1018 WB Amsterdam, The Netherlands
e-mail: nivd@fee.uva.nl

Abstract. This note studies the variance of total cumulative rewards for Markov reward chains in both discrete and continuous time. It is shown that parallel results can be obtained for both cases.

First, explicit formulae are presented for the variance within finite time. Next, the infinite time horizon is considered. Most notably, it is concluded that the variance has a linear growth rate. Explicit expressions are provided, related to the standard average reward case, to compute this growth rate.

1 Introduction

The usual optimization criteria examined in the literature on optimization of Markov reward processes, e.g. total discounted or mean reward, may be quite insufficient to characterize the problem from the point of the decision maker. To this end it is necessary to select more sophisticated criteria that also reflect variability-risk features of the problem, most notably by taking into account the variance of the cumulative rewards. For a detailed discussion of such approaches see the review paper by White [8].

In this paper therefore we aim to establish results for the variance of the cumulative rewards. More precisely, it will be shown that similar results can be obtained for the discrete time case (as partly already obtained in the literature [1], [2], [5], [6]) and the continuous time case, which seems to be new. In addition to reward rates, also transition rewards are herein included.

Particularly, for the average (or mean) reward case, it can be concluded that the variance of the total reward has an asymptotic linear growth rate in time. Relations for this growth rate to be computed are provided.

2 Formulation

In both the discrete time and continuous time case consider a Markov reward chain with finite state space $\mathcal{S} = \{1, 2, \ldots, N\}$, denoted by the Markov chains:

$$\{X^{\mathrm{d}}(n) | n = 0, 1, 2, \ldots\} \quad \text{in discrete time}$$
$$\{X^{\mathrm{c}}(t) | t \geq 0\} \quad \text{in continuous time}$$

Throughout we use the superindex d and c to indicate whether the discrete (d) and continuous (c) time case is in order.

Discrete time case

The discrete time Markov reward chain is characterized by

p_{ij} : the one step transition probabilities for a transition from $i \to j$,

r_{ij} : the one-step reward accrued to a transition from $i \to j$,

$\tilde{r}_i = \sum_j p_{ij} r_{ij}$: the expected one-step reward in state i.

Let

$P = [p_{ij}]$ be the transition probability matrix, and

$\Pi^{\mathrm{d}} = \lim_{m \to \infty} \frac{1}{m} \sum_{n=0}^{m-1} P^n$ the limiting matrix.

(Note that $\Pi^{\mathrm{d}} = \lim_{n \to \infty} P^n$ if P is aperiodic.)

Continuous time case

The continuous time Markov reward chain is characterized by

q_{ij} : transition rate for a transition from $i \to j$ $(j \neq i)$ with $q_{ii} = -\sum_{j, j \neq i} q_{ij}$,

r_{ij} : instantaneous transition reward when a transition from $i \to j$ takes place,

r_i : reward rate per unit of time incurred in state i,

$\tilde{r}_i = r_i + \sum_{j, j \neq i} q_{ij} r_{ij}$: total reward rate per unit of time incurred in state i.

Let

$Q = [q_{ij}]$ be the generator matrix. Then

$P^{\mathrm{c}}(t) = e^{tQ}$ is the transition matrix over time t, and

$\Pi^{\mathrm{c}} = \lim_{t \to \infty} P^{\mathrm{c}}(t)$ $(\Leftrightarrow \Pi^{\mathrm{c}} Q = Q \Pi^{\mathrm{c}} = 0)$ the limiting matrix.