Mathematical Methods in Operations Research 2005
Special issue in honor of *Arie Hordijk*

On Mean Reward Variance in Semi-Markov Processes

Karel Sladký

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, 182 08 Praha 8, Czech Republic

Received: date / Revised version: date

Abstract As an extension of the discrete-time case, this note investigates the variance of the total cumulative reward for the embedded Markov chain of semi-Markov processes. Under the assumption that the chain is aperiodic and contains a single class of recurrent states recursive formulae for the variance are obtained which show that the variance growth rate is asymptotically linear in time. Expressions are provided to compute this growth rate.

Keywords. Markov and semi-Markov processes with rewards, variance of cumulative reward, asymptotic behaviour

AMS Subject Classification. Primary 90C47
Secondary 60J27

1 Introduction

The usual average criteria examined in the literature on Markov reward processes can be insufficient to fully capture the various aspects for a decision maker. It may be preferable to also include more sophisticated criteria to reflect variability-risk features (for details see [17]). Most notably, the variance of cumulative rewards can be indicative and seems of interest.

The variance of cumulative reward structures has been studied extensively for discrete-time Markov reward chains. The majority of these results has been involved with the development of further optimization criterion (see [1], [2], [3], [4], [5], [7], [8], [10], [13], [15], [16]).

For the continuous-time case only structural results were reported in [5] while just recently explicit expressions were developed in [14]. (The present paper is related to the latter reference but focuses on the embedded Markov chain to also include non-exponential densities).

No results, however, seem to be reported for semi-Markov processes. Such an extension is not direct as higher moments and stochastic durations are to be taken into account.

In this paper therefore we aim to establish a first step in this direction. This step can be seen as essential for further optimization results in line with the discrete-time case.

2 Formulation and Notation

Consider a semi-Markov reward process $Y = \{Y(t), t \geq 0\}$ with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$ along with the embedded Markov chain $X = \{X_n, n = 0, 1, \ldots\}$. The transition and reward structure is characterized by

p_{ij} : transition probability from $i \to j$ $(i, j \in \mathcal{I}, j \neq i)$ of the embedded Markov chain X with generic stochastic matrix $\boldsymbol{P} = [p_{ij}]_{i,j=1}^{N}$,

η_{ij} : random time of the transition from $i \to j$, hence $\eta_i = \Sigma_{j \in \mathcal{I}}\, p_{ij}\eta_{ij}$ is the random time spent by the semi-Markov process Y in state i,

$F_{ij}(\tau)$: distribution function representing the conditional probability $\mathsf{P}(\eta_{ij} \leq \tau)$,

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

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

We will be interested in the variance of the total reward per unit of time and its behavior for the infinite time horizon.

To this end, let the vector $\mathcal{R}(t)$ denote the expected total reward of the semi-Markov process $Y(t)$ up to time t given its initial state at time $t = 0$. More precisely, we are interested in

$$g = \lim_{t \to \infty} \frac{1}{t} \cdot \mathcal{R}_i(t) \quad \text{with} \quad \mathcal{R}_i(t) = \mathsf{E}[\xi(t) | Y(0) = i] \tag{1}$$

where

$$\xi(t) = \left[\int_0^t r_{Y(s)} \mathrm{d}s \; + \sum_{k=0}^{N(t)} r_{Y(\tau_k^-), Y(\tau_k^+)} \right]$$

with $Y(s)$, denoting the state of the system at time s, $Y(\tau_k^-)$ and $Y(\tau_k^+)$ the state just prior and after the k-th jump, $N(t)$ the number of jumps up to time t and E and $\boldsymbol{\sigma}^2$ the standard symbols for expectation and variance.

Now consider the embedded Markov chain $X = \{X_n, n = 0, 1, 2, \ldots\}$ and let the vectors $\boldsymbol{R}(n)$, $\boldsymbol{S}(n)$, and $\boldsymbol{V}(n)$ denote the first moment, the second moment and the variance of the (random) total reward ξ_n received in the first n transitions of the embedded Markov chain X given its initial state at time $t = 0$ respectively.

That is,

$$R_i(n) = \mathsf{E}[\xi_n | X(0) = i],$$
$$S_i(n) = \mathsf{E}[\xi_n^2 | X(0) = i],$$
$$V_i(n) = \boldsymbol{\sigma}^2[\xi_n | X(0) = i]$$

where

$$\xi_n = \sum_{k=0}^{n-1} \left[r_{X_k} \cdot \eta_{X_k, X_{k+1}} + r_{X_k, X_{k+1}} \right].$$

Similarly, $\boldsymbol{D}(n)$ is the vector of expected times ζ_n spent in the first n transitions with elements

$$D_i(n) = \mathsf{E}[\zeta_n | X(0) = i] \quad \text{where} \quad \zeta_n = \sum_{k=0}^{n-1} \left[\eta_{X_k, X_{k+1}} \right].$$

From the theory of dynamic programming it is well known that under assumption AS1 below

$$g^{\mathrm{r}} := \lim_{n\to\infty} \frac{1}{n} \cdot R_i(n), \quad g^{\mathrm{t}} := \lim_{n\to\infty} \frac{1}{n} \cdot D_i(n) \quad \text{exist for all } i \in \mathcal{I}.$$

It can then also be shown (see [12], Chapt. 7.3) that under assumptions AS1–AS2 below then also:

$$g = \lim_{t\to\infty} \frac{1}{t} \cdot \mathcal{R}_i(t) =: \frac{g^{\mathrm{r}}}{g^{\mathrm{t}}} = \frac{\lim\limits_{n\to\infty} R_i(n)}{\lim\limits_{n\to\infty} D_i(n)}. \tag{2}$$

As mentioned in the introduction, such as for second order optimization purposes, and in analogy with (2) as our main interest we aim to study the "mean" variance defined by

$$G := \frac{\lim\limits_{n\to\infty} V_i(n)}{\lim\limits_{n\to\infty} D_i(n)}. \tag{3}$$

Notation. In what follows $\boldsymbol{I}$ denotes an identity matrix, and $\boldsymbol{e}$ is reserved for a unit column vector. By $\varepsilon(t)$ we denote a function in t such that $\varepsilon(t) \to 0$ exponentially fast as $t \to \infty$, i.e. for some α and β: $|\varepsilon(t)| \le \alpha \cdot e^{-\beta t}$. By $\boldsymbol{\varepsilon}(t)$ we denote a vector function such that each component $\boldsymbol{\varepsilon}_i(t) \to 0$ as $t \to \infty$ exponentially fast.

Assumptions. We make the following assumptions:

AS 1. The transition probability matrix $\boldsymbol{P}$ has a single class of recurrent states and is aperiodic.

AS 2. $F_{ij}(\tau)$ is a non-lattice distribution. The first two moments $d_i^{(1)}$, $d_i^{(2)}$ of the of (random) time η_i ($i \in \mathcal{I}$) spent by the process Y in any state during each visit are positive and finite, i.e. we assume that for $\ell = 1, 2$ and any $i, j = 1, \ldots, N$

$$0 < d_{ij}^{(\ell)} = \int_0^\infty \tau^\ell \, \mathrm{d}F_{ij}(\tau) < \infty \;\text{ hence also }\; 0 < d_i^{(\ell)} = \sum_{j=1}^{N} p_{ij} d_{ij}^{(\ell)} < \infty.$$

3 Preliminaries

As well-known, under AS 1 the rows of the limiting matrix $\boldsymbol{P}^* = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \boldsymbol{P}^k$ of the embedded Markov chain X are identical and equal to the (row) vector of steady state probabilities $\boldsymbol{\pi} = [\pi_1, \ldots, \pi_N]$ as determined by $\boldsymbol{\pi} = \boldsymbol{\pi} \cdot \boldsymbol{P}$ along with the normalizing condition $\boldsymbol{\pi} \cdot \boldsymbol{e} = 1$. Moreover, since $\boldsymbol{P}$ is aperiodic $\boldsymbol{P}^* = \lim_{k\to\infty} \boldsymbol{P}^k$ and the convergence is geometrically fast. Then

$$g = \frac{\sum_{j\in\mathcal{I}} \pi_j \cdot r_j^{(1)}}{\sum_{j\in\mathcal{I}} \pi_j \cdot d_j^{(1)}} \tag{4}$$

with $r_i^{(1)}$ defined by (16) and $d_i^{(1)}$ the first moment of the (random) time η_i. Also the average reward g^{r} and the average time g^{t} per transition of the embedded Markov chain are well defined by:

$$g^{\mathrm{r}} = \sum_{j\in\mathcal{I}} \pi_j \cdot r_j^{(1)}, \quad g^{\mathrm{t}} = \sum_{j\in\mathcal{I}} \pi_j \cdot {d_j}^{(1)}.$$

Obviously for the constant vectors $\boldsymbol{g}^{\mathrm{r}}$ and $\boldsymbol{g}^{\mathrm{t}}$ with elements g^{r} and g^{t} we have $\boldsymbol{g}^{\mathrm{r}} = \boldsymbol{P}^* \cdot \boldsymbol{r}^{(1)}$, $\boldsymbol{g}^{\mathrm{t}} = \boldsymbol{P}^* \cdot \boldsymbol{d}^{(1)}$, respectively.

From the theory of dynamic programming (e.g. [11], [12]) it is well-known that the vectors of expected reward and times for n transitions fulfil the recursive formulas

$$\boldsymbol{R}(n+1) = \boldsymbol{r}^{(1)} + \boldsymbol{P} \cdot \boldsymbol{R}(n) \;\text{ with }\; \boldsymbol{R}(0) = 0 \tag{5}$$

$$\boldsymbol{D}(n+1) = \boldsymbol{d}^{(1)} + \boldsymbol{P} \cdot \boldsymbol{D}(n) \;\text{ with }\; \boldsymbol{D}(0) = 0. \tag{6}$$

Furthermore, under assumption AS 1, there exists vectors $\boldsymbol{w}^{\mathrm{r}}$, $\boldsymbol{w}^{\mathrm{t}}$ such that

$$\boldsymbol{R}(n) = \boldsymbol{g}^{\mathrm{r}} \cdot n + \boldsymbol{w}^{\mathrm{r}} + \boldsymbol{\varepsilon}(n) \tag{7}$$

$$\boldsymbol{D}(n) = \boldsymbol{g}^{\mathrm{t}} \cdot n + \boldsymbol{w}^{\mathrm{t}} + \boldsymbol{\varepsilon}(n). \tag{8}$$

Hence, both $\boldsymbol{R}(n)$ and $\boldsymbol{D}(n)$ possess a linear growth rate g^{r} and g^{t} in n up to a geometric convergence to the null vector and vectors $\boldsymbol{w}^{\mathrm{r}}$ and $\boldsymbol{w}^{\mathrm{t}}$, respectively. The constant vectors $\boldsymbol{g}^{\mathrm{r}}$, $\boldsymbol{g}^{\mathrm{t}}$ along with vectors $\boldsymbol{w}^{\mathrm{r}}$, $\boldsymbol{w}^{\mathrm{t}}$ are uniquely determined by

$$\boldsymbol{w}^{\mathrm{r}} + \boldsymbol{g}^{\mathrm{r}} = \boldsymbol{r}^{(1)} + \boldsymbol{P} \cdot \boldsymbol{w}^{\mathrm{r}}, \quad \boldsymbol{P}^* \cdot \boldsymbol{w}^{\mathrm{r}} = 0 \tag{9}$$

$$\boldsymbol{w}^{\mathrm{t}} + \boldsymbol{g}^{\mathrm{t}} = \boldsymbol{d}^{(1)} + \boldsymbol{P} \cdot \boldsymbol{w}^{\mathrm{t}}, \quad \boldsymbol{P}^* \cdot \boldsymbol{w}^{\mathrm{t}} = 0. \tag{10}$$

From (9), (10) we can also immediately conclude the existence of a constant vector $\boldsymbol{g}$ along with vector $\boldsymbol{w}$ being the unique solution of the equation

$$\boldsymbol{w} + \boldsymbol{D}^{(1)} \cdot \boldsymbol{g} = \boldsymbol{r}^{(1)} + \boldsymbol{P} \cdot \boldsymbol{w}, \quad \boldsymbol{P}^* \cdot \boldsymbol{w} = 0, \tag{11}$$

where the ith element of the diagonal matrix $\boldsymbol{D}^{(1)}$ equals $d_i^{(1)}$, where the elements of the constant vector $\boldsymbol{g}$ are equal to $g = g^{\mathrm{r}}/g^{\mathrm{t}}$ and where for the ith element of $\boldsymbol{w}$, $w_i = w_i^{\mathrm{r}} - g \cdot w_i^{\mathrm{t}}$.

Conditional Expectations of Reward. Let E_i denote the conditional expectation given $X(0) = i$. Since the process X is time homogeneous, for $n > m$ we can conclude that

$$\mathsf{E}_i[\xi_n] = \mathsf{E}_i[\xi_m] + \mathsf{E}_i\{\sum_{j \in \mathcal{I}} \mathsf{P}(X_m = j) \cdot \mathsf{E}_j[\xi_{n-m}]\}. \tag{12}$$

Similarly we get

$$\begin{aligned} \mathsf{E}_i[\xi_n]^2 = \mathsf{E}_i[\xi_m]^2 &+ \mathsf{E}_i\{\sum_{j \in \mathcal{I}} \mathsf{P}(X_m = j) \cdot \mathsf{E}_j[\xi_{n-m}]^2\} \\ &+ 2 \cdot \mathsf{E}_i[\xi_m] \sum_{j \in \mathcal{I}} \mathsf{P}(X_m = j) \cdot \mathsf{E}_j[\xi_{n-m}]. \end{aligned} \tag{13}$$

4 Reward Variance of Embedded Markov Chains

In this section we focus our attention on the total reward variance generated by the embedded Markov chain $X = \{X_n, n = 0, 1, \ldots\}$ characterized by the transition probability matrix $\boldsymbol{P}$ and one-stage (random) rewards $\eta_{ij} \cdot r_i + r_{ij}$ $(i, j \in \mathcal{I}, j \neq i)$ accrued to the transition from state i into state j.

From (12) and (13) we directly conclude

$$\mathsf{E}_i[\xi_n] = r_i^{(1)} + \sum_{j \in \mathcal{I}} p_{ij} \cdot \mathsf{E}_j[\xi_{n-1}] \tag{14}$$

$$\begin{aligned} \mathsf{E}_i[\xi_n]^2 = r_i^{(2)} &+ \sum_{j \in \mathcal{I}} p_{ij} \cdot \mathsf{E}_j[\xi_{n-1}]^2 \\ &+ 2 \cdot \sum_{j \in \mathcal{I}} p_{ij} \cdot [r_i \cdot d_{ij}^{(1)} + r_{ij}] \cdot \mathsf{E}_j[\xi_{n-1}] \end{aligned} \tag{15}$$

where

$$r_i^{(1)} = \sum_{j\in\mathcal{I}} p_{ij}\cdot[r_i\cdot d_{ij}^{(1)} + r_{ij}] \tag{16}$$

$$r_i^{(2)} = \sum_{j\in\mathcal{I}} p_{ij}\cdot\{[r_i]^2\cdot d_{ij}^{(2)} + [r_{ij}]^2 + 2\cdot r_i\cdot d_{ij}^{(1)}\cdot r_{ij}\}. \tag{17}$$

By using the more appealing notation $R_i(n) = \mathsf{E}_i[\xi_n]$, $S_i(n) = \mathsf{E}_i[\xi_n]^2$, (14), (15) take on the forms:

$$R_i(n+1) = r_i^{(1)} + \sum_{j\in\mathcal{I}} p_{ij}\cdot R_j(n), \tag{18}$$

$$S_i(n+1) = r_i^{(2)} + 2\cdot\sum_{j\in\mathcal{I}} p_{ij}\cdot[r_i\cdot d_{ij}^{(1)} + r_{ij}]\cdot R_j(n) + \sum_{j\in\mathcal{I}} p_{ij}\cdot S_j(n). \tag{19}$$

For the variance $V_i(\cdot) = S_i(\cdot) - [R_i(\cdot)]^2$ we now arrive after some algebra by using (18), (19)

$$V_i(n+1) = r_i^{(2)} + 2\cdot\sum_{j\in\mathcal{I}} p_{ij}\cdot[r_i\cdot d_{ij}^{(1)} + r_{ij}]\cdot R_j(n) - \sum_{j\in\mathcal{I}} p_{ij}\cdot[R_i(n+1) + R_j(n)]\cdot[R_i(n+1) - R_j(n)] + \sum_{j\in\mathcal{I}} p_{ij}\cdot V_j(n). \tag{20}$$

Remark 1 (Transient case.) For the so-called transient case, i.e. when

$$\lim_{n\to\infty} R_i(n) = R_i, \quad \lim_{n\to\infty} S_i(n) = S_i, \quad \lim_{n\to\infty} V_i(n) = V_i,$$

(18), (19) and (20) obtain the closed form:

$$R_i = r_i^{(1)} + \sum_{j\in\mathcal{I}} p_{ij}\cdot R_j, \tag{21}$$

$$S_i = r_i^{(2)} + 2\cdot\sum_{j\in\mathcal{I}} p_{ij}\cdot(r_i\cdot d_{ij}^{(1)} + r_{ij})\cdot R_j + \sum_{j\in\mathcal{I}} p_{ij}\cdot S_j, \tag{22}$$

$$V_i = r_i^{(2)} - [R_i]^2 + \sum_{j\in\mathcal{I}} p_{ij}\cdot\{2\cdot(r_i\cdot d_{ij}^{(1)} + r_{ij})\cdot R_j + [R_j]^2\} + \sum_{j\in\mathcal{I}} p_{ij}\cdot V_j. \tag{23}$$

5 Mean Reward Variance in Markov Chains: Infinite Horizon

To investigate the asymptotic behaviour of the variance $V_i(n)$ we focus our attention on the recursive formula (20) and employ well-known facts on the asymptotic behaviour of $R_i(n)$ as based upon (7).

By (7) in the third term of (20), we can substitute:

$$R_i(n+1) + R_j(n) = 2 \cdot n \cdot g^{\mathrm{r}} + g^{\mathrm{r}} + w_i^{\mathrm{r}} + w_j^{\mathrm{r}} + \varepsilon(n), \tag{24}$$

$$R_i(n+1) - R_j(n) = g^{\mathrm{r}} + w_i^{\mathrm{r}} - w_j^{\mathrm{r}} + \varepsilon(n). \tag{25}$$

Hence

$$\begin{aligned}
&\sum_{j\in\mathcal{I}} p_{ij} \cdot [R_i(n+1) + R_j(n)] \cdot [R_i(n+1) - R_j(n)] \\
&= 2 \cdot n \cdot g^{\mathrm{r}} \cdot (g^{\mathrm{r}} + w_i^{\mathrm{r}} - \sum_{j\in\mathcal{I}} p_{ij} \cdot w_j^{\mathrm{r}}) + \sum_{j\in\mathcal{I}} p_{ij} \cdot \{[g^{\mathrm{r}} + w_i^{\mathrm{r}}]^2 - [w_j^{\mathrm{r}}]^2\} + \varepsilon(n) \\
&= 2 \cdot n \cdot g^{\mathrm{r}} \cdot r_i^{(1)} + \sum_{j\in\mathcal{I}} p_{ij} \cdot \{[g^{\mathrm{r}} + w_i^{\mathrm{r}}]^2 - [w_j^{\mathrm{r}}]^2\} + \varepsilon(n).
\end{aligned} \tag{26}$$

Similarly for the second term of (20) we obtain by (7)

$$\begin{aligned}
&\sum_{j\in\mathcal{I}} p_{ij} \cdot [r_i \cdot d_{ij}^{(1)} + r_{ij}] \cdot R_j(n) \\
&= \sum_{j\in\mathcal{I}} p_{ij} \cdot [r_i \cdot d_{ij}^{(1)} + r_{ij}] \cdot [n \cdot g^{\mathrm{r}} + w_j^{\mathrm{r}} + \varepsilon(n)] \\
&= n \cdot g^{\mathrm{r}} \cdot r_i^{(1)} + \sum_{j\in\mathcal{I}} p_{ij} \cdot [r_i \cdot d_{ij}^{(1)} + r_{ij}] \cdot w_j^{\mathrm{r}} + \varepsilon(n).
\end{aligned} \tag{27}$$

Substituting (26) and (27) in (20) now yields

$$\begin{aligned}
V_i(n+1) &= \sum_{j\in\mathcal{I}} p_{ij} \cdot V_j(n) + r_i^{(2)} + 2 \cdot \sum_{j\in\mathcal{I}} p_{ij} \cdot [r_i \cdot d_{ij}^{(1)} + r_{ij}] \cdot w_j^{\mathrm{r}} \\
&\quad - \sum_{j\in\mathcal{I}} p_{ij} \cdot \{[g^{\mathrm{r}} + w_i^{\mathrm{r}}]^2 - [w_j^{\mathrm{r}}]^2\} + \varepsilon(n) \\
&= \sum_{j\in\mathcal{I}} p_{ij} \cdot V_j(n) + s_i + \varepsilon(n),
\end{aligned} \tag{28}$$

where for the elements s_i of the vector $\boldsymbol{s}$ we obtain after some algebra:

$$\begin{aligned}
s_i &= r_i^{(2)} + \sum_{j\in\mathcal{I}} p_{ij} \cdot (2 \cdot [r_i \cdot d_{ij}^{(1)} + r_{ij}] \cdot w_j^{\mathrm{r}} + [w_j^{\mathrm{r}}]^2) - [g^{\mathrm{r}} + w_i^{\mathrm{r}}]^2 \\
&= \sum_{j\in\mathcal{I}} p_{ij} \cdot \{[r_i]^2 \cdot d_{ij}^{(2)} + 2 \cdot r_i \cdot d_{ij}^{(1)} \cdot [r_{ij} + w_j^{\mathrm{r}}] + [r_{ij} + w_j^{\mathrm{r}}]^2\} \\
&\quad - [g^{\mathrm{r}} + w_i^{\mathrm{r}}]^2.
\end{aligned} \tag{29}$$

Furthermore, by (9) for the last term of (29) we get

$$\begin{aligned}-[g^{\mathrm{r}}+w_i^{\mathrm{r}}]^2 &= [g^{\mathrm{r}}]^2-[w_i^{\mathrm{r}}]^2-2\cdot g^{\mathrm{r}}\cdot[g^{\mathrm{r}}+w_i^{\mathrm{r}}]\\ &= [g^{\mathrm{r}}]^2-[w_i^{\mathrm{r}}]^2-2\cdot g^{\mathrm{r}}\cdot\sum_{j\in\mathcal{I}}p_{ij}\cdot[r_i\cdot d_{ij}^{(1)}+r_{ij}+w_j^{\mathrm{r}}]\end{aligned}$$

which by substitution in (29) and some algebra yields

$$\begin{aligned}s_i &= \sum_{j\in\mathcal{I}}p_{ij}\cdot\{[r_i]^2\cdot d_{ij}^{(2)}+[r_{ij}]^2+2\cdot r_i\cdot d_{ij}^{(1)}\cdot r_{ij}\\ &\quad+2\cdot[r_i\cdot d_{ij}^{(1)}+r_{ij}]\cdot w_j^{\mathrm{r}}+[w_j^{\mathrm{r}}]^2\}+[g^{\mathrm{r}}]^2-[w_i^{\mathrm{r}}]^2\\ &\quad-2\cdot g^{\mathrm{r}}\cdot\sum_{j\in\mathcal{I}}p_{ij}\cdot[r_i\cdot d_{ij}^{(1)}+r_{ij}+w_j^{\mathrm{r}}]\\ &= \sum_{j\in\mathcal{I}}p_{ij}\cdot\{[r_i]^2\cdot d_{ij}^{(2)}+2\cdot r_i\cdot d_{ij}^{(1)}\cdot[r_{ij}-g^{\mathrm{r}}+w_j^{\mathrm{r}}]\\ &\quad+[r_{ij}-g^{\mathrm{r}}+w_j^{\mathrm{r}}]^2\}-[w_i^{\mathrm{r}}]^2. \qquad (30)\end{aligned}$$

Hence, in matrix form we have:

$$\boldsymbol{V}(n+1)=\boldsymbol{s}+\boldsymbol{P}\cdot\boldsymbol{V}(n)+\boldsymbol{\varepsilon}^{(1)}(n) \qquad (31)$$

where the elements of the vector $\boldsymbol{\varepsilon}^{(1)}(n)$ converge to zero geometrically, that is, for some numbers $c>0$ and $\delta\in(0,1)$: $\|\boldsymbol{\varepsilon}^{(1)}(n)\|\le c\cdot\delta^n$.

Growth rate. In order to investigate the behaviour of $\boldsymbol{V}(n)$ for n large, let

$$\boldsymbol{W}(n+1)=\boldsymbol{s}+\boldsymbol{P}\cdot\boldsymbol{W}(n) \qquad (32)$$

and

$$\boldsymbol{X}(n)=\boldsymbol{V}(n)-\boldsymbol{W}(n). \qquad (33)$$

Hence, in line with (5) for $\boldsymbol{R}(n)$, the vector $\boldsymbol{W}(n)$ can be regarded as the total reward vector over n steps for a Markov reward chain with one step reward vector $\boldsymbol{s}$. As a consequence, similarly to (7), we can conclude the existence of a vector $\boldsymbol{w}^{(2)}$ such that

$$\boldsymbol{W}(n)=\boldsymbol{g}^{(2)}\cdot n+\boldsymbol{w}^{(2)}+\boldsymbol{\varepsilon}(n). \qquad (34)$$

In words that is, $\boldsymbol{W}(n)$ possesses a linear growth rate $g^{(2)}$ in n up to a geometric convergence to the null vector and vector $\boldsymbol{w}^{(2)}$. The constant vector $\boldsymbol{g}^{(2)}$ along with vector $\boldsymbol{w}^{(2)}$ are uniquely determined by

$$\boldsymbol{w}^{(2)}+\boldsymbol{g}^{(2)}=\boldsymbol{s}+\boldsymbol{P}\cdot\boldsymbol{w}^{(2)},\quad \boldsymbol{P}^*\cdot\boldsymbol{w}^{(2)}=0 \qquad (35)$$

where elements of $\boldsymbol{s}$ are calculated by (29) or (30).

Iterating (33) we immediately conclude that

$$\|\boldsymbol{X}(n)\|\le\|\sum_{k=1}^{n}\boldsymbol{P}^k\cdot\boldsymbol{\varepsilon}^{(1)}(k)\|\le\sum_{k=1}^{n}\|\boldsymbol{\varepsilon}^{(1)}(k)\|<c\cdot\frac{1}{1-\delta}=:C. \qquad (36)$$

Theorem 1 *For the (constant) growth rate $g^{(2)}$ (vector $\boldsymbol{g}^{(2)}$) and vector $\boldsymbol{w}^{(2)}$ as determined by (35), some geometrically converging function $\boldsymbol{\varepsilon}(n)$, some constant C and some bounded vector $\boldsymbol{c}(n)$ with $\|\boldsymbol{c}(n)\| < C$*

$$\boldsymbol{V}(n) = n \cdot \boldsymbol{g}^{(2)} + \boldsymbol{w}^{(2)} + \boldsymbol{\varepsilon}(n) + \boldsymbol{c}(n) \text{ for all } n \tag{37}$$

and hence,

$$\boldsymbol{g}^{(2)} = \lim_{n\to\infty} \frac{\boldsymbol{V}(n)}{n} = \lim_{n\to\infty} \frac{\boldsymbol{W}(n)}{n} = \boldsymbol{P}^{\star} \cdot \boldsymbol{s}. \tag{38}$$

Proof The relations follow directly combining (31), (32), (33) and (36) with $\boldsymbol{c}(n)$ equal to $\boldsymbol{X}(n)$.

Remark 2 By (9) the coefficients w_i^{r}'s are normalized by $\sum_{i\in\mathcal{I}} \pi_i \cdot w_i^{\mathrm{r}} = 0$. Since by assumption AS 1 the Markov chain has a single class of recurrent states, the first equation in (9) still holds if we replace w_i^{r} by $\tilde{w}_i^{\mathrm{r}} = w_i^{\mathrm{r}} + c$ $(i = 1, 2, \ldots, N)$, where c is an arbitrary constant.

As a consequence, by examining the formulas (29), (30) we can conclude that (38) still holds if in (29), (30) we replace all w_i^{r}'s by $\tilde{w}_i^{\mathrm{r}}$'s.

Since $\sum_{i\in\mathcal{I}} \pi_i \cdot p_{ij} = \pi_j$ for arbitrary real c_i's it holds

$$\sum_{i\in\mathcal{I}} \sum_{j\in\mathcal{I}} \pi_i \cdot p_{ij} \cdot c_j = \sum_{i\in\mathcal{I}} \pi_i \cdot c_i.$$

Then, since by (9): $\sum_{i\in\mathcal{I}} \pi_i \cdot w_i^{\mathrm{r}} = 0$ and defining

$$\tilde{s}_i = r_i^{(2)} - [g^{\mathrm{r}}]^2 + 2 \cdot \sum_{j\in\mathcal{I}} p_{ij} \cdot [r_i \cdot d_{ij}^{(1)} + r_{ij}] \cdot w_j^{\mathrm{r}} \tag{39}$$

we can conclude that

$$g^{(2)} = \sum_{i\in\mathcal{I}} \pi_i \cdot s_i = \sum_{i\in\mathcal{I}} \pi_i \cdot \tilde{s}_i. \tag{40}$$

Remark 3 Assume that $r_i \equiv 0$ for all $i \in \mathcal{I}$. Then (cf. (29), (30)) s_i takes on the form:

$$\begin{aligned} s_i &= \sum_{j\in\mathcal{I}} p_{ij} \cdot \{[r_{ij} + w_j^{\mathrm{r}}]^2\} - [g^{\mathrm{r}} + w_i^{\mathrm{r}}]^2 \\ &= \sum_{j\in\mathcal{I}} p_{ij} \cdot \{[r_{ij} - g^{\mathrm{r}} + w_j^{\mathrm{r}}]^2\} - [w_i^{\mathrm{r}}]^2. \end{aligned} \tag{41}$$

Moreover, by (39) we have

$$\begin{aligned} \tilde{s}_i &= \sum_{j\in\mathcal{I}} p_{ij} \cdot \{[r_{ij}]^2 + 2 \cdot r_{ij} \cdot w_j^{\mathrm{r}}\} - [g^{\mathrm{r}}]^2 \\ &= \sum_{j\in\mathcal{I}} p_{ij} \cdot \{[r_{ij} - g^{\mathrm{r}}]^2 + 2 \cdot r_{ij} \cdot w_j^{\mathrm{r}}\}. \end{aligned} \tag{42}$$

6 Mean Reward Variance in Semi-Markov Processes

In analogy with (2) and (4) we are thus interested in the "mean" variance defined by

$$G := \frac{g^{(2)}}{g^{\mathrm{t}}} = \frac{\sum_{j\in\mathcal{I}} \pi_j \cdot s_j}{\sum_{j\in\mathcal{I}} \pi_j \cdot d_j^{(1)}}. \tag{43}$$

Below we present two approaches how to calculate this mean variance G.

Theorem 2 *The mean variance G given by (43) can be calculated using the limiting probabilities of the semi-Markov processes as*

$$G = \sum_{j\in\mathcal{I}} \overline{\pi}_j \cdot \overline{s}_j = \sum_{j\in\mathcal{I}} \overline{\pi}_j \cdot \hat{s}_j \qquad \text{independently of } \ i \in \mathcal{I} \tag{44}$$

where

$$\overline{s}_i = s_i/d_i^{(1)}, \quad \hat{s}_i = \tilde{s}_i/d_i^{(1)} \tag{45}$$

and

$$\overline{\pi}_i = \frac{\pi_i \cdot d_i^{(1)}}{\sum_{j=1}^{N} \pi_j \cdot d_j^{(1)}}. \tag{46}$$

Proof Since under assumptions AS 1 and AS 2, the rows of the limiting matrix $\overline{\boldsymbol{P}}^*$ of the considered semi-Markov process $Y(t)$ are identical and equal to the row vector $\overline{\boldsymbol{\pi}} = [\overline{\pi}_1, \ldots, \overline{\pi}_N]$ where elements $\overline{\pi}_i = \lim_{t\to\infty}\{Y(t)|Y(0) = i\}$ are given by (46) (see e.g. [11], [12]), the proof follows immediately by inserting (45) and (46) in (40).

In what follows we show how to calculate the mean variance using formulas similar to those for calculating the mean reward. To this end, let $\mathcal{G}$ be a diagonal matrix with all diagonal elements equal to $G = g^{(2)}/g^{\mathrm{t}}$. Premultiplying (10) by $\mathcal{G}$ and subtracting from (35) we immediately get

$$\boldsymbol{u} = \boldsymbol{s} - \mathcal{G} \cdot \boldsymbol{d}^{(1)} + \boldsymbol{P} \cdot \boldsymbol{u} \tag{47}$$

where

$$\boldsymbol{u} = \boldsymbol{w}^{(2)} - \mathcal{G} \cdot \boldsymbol{w}^{\mathrm{t}}. \tag{48}$$

So we have arrived at

Theorem 3 *The mean variance G can be calculated as a solution of (47). The solution $(G, \boldsymbol{u})$ of (47) is unique up to an additive constant to vectors $\boldsymbol{u}$, and unique under the additional normalizing condition $\boldsymbol{P}^\star \cdot \boldsymbol{u} = 0$.*

Acknowledgement

The author is grateful to Arie Hordijk for his stimulating work in the field of Markov Decision Theory.

The research was supported by the Grant Agency of the Czech Republic under Grants 402/05/0115 and 402/04/1294.

References

1. Benito F (1982) Calculating the variance in Markov processes with random reward. Trabajos de Estadistica y de Investigacion Operativa 33: 73–85
2. Filar J, Kallenberg LCM and Lee H-M (1989) Variance penalized Markov decision processes. Mathem. Oper. Research 14: 147–161
3. Huang Y and Kallenberg LCM (1994) On finding optimal policies for Markov decision chains: a unifying framework for mean-variance-tradeoffs. Mathem. Oper. Research 19: 434–448
4. Jaquette SC (1972) Markov decision processes with a new optimality criterion: Small interest rates. Ann. Math. Statist. 43: 1894–1901
5. Jaquette SC (1973) Markov decision processes with a new optimality criterion: Discrete time. Ann. Statist. 1: 496–505
6. Jaquette SC (1975) Markov decision processes with a new optimality criterion: Continuous time. Ann. Statist. 3: 547–553
7. Kadota Y (1997) A minimum average-variance in Markov decision processes. Bulletin of Informatics and Cybernetics 29: 83–89
8. Kawai H (1987) A variance minimization problem for a Markov decision process. European J. Oper. Research 31: 140–145
9. Kurano M (1987) Markov decision processes with a minimum-variance criterion. J. Mathem. Anal. Appl. 123: 572–583
10. Mandl P (1971) On the variance in controlled Markov chains. Kybernetika 7: 1–12
11. Puterman ML (1994) Markov decision processes – discrete stochastic dynamic programming. Wiley, New York
12. Ross SM (1970) Applied probability models with optimization applications. Holden–Day, San Francisco, Calif.
13. Sladký K and Sitař M (2004) Optimal solutions for undiscounted variance penalized Markov decision chains. In: Marti K, Ermoliev Y and Pflug G (eds.) Dynamic Stochastic Optimization. Springer–Verlag, Berlin, pp. 43–66
14. Sladký K and van Dijk NM (2004) On total reward variance for continuous-time Markov reward chains. Submitted for publication
15. Sobel MJ (1982) The variance of discounted Markov decision processes. J. Appl. Probab. 19: 794–802
16. Sobel MJ (1985) Maximal mean/standard deviation ratio in an undiscounted MDP. Oper. Research Lett. 4: 157–159
17. White DJ (1988) Mean variance and probability criteria in finite Markov decision processes: a review. J. Optim. Theory Appl. 56: 1–29