CENTRAL MOMENTS AND RISK-SENSITIVE OPTIMALITY IN CONTINUOUS-TIME MARKOV REWARD PROCESSES

Karel Sladký
Institute of Information Theory and Automation of the AS CR

Abstract In this note we consider continuous-time Markov decision processes with finite state space where the stream of rewards generated by the Markov processes is evaluated by an exponential utility function with a given risk sensitivity coefficient (so-called risk-sensitive models). For the risk-sensitive case, i.e. if the considered risk-sensitivity coefficient is non-zero, we establish explicit formulas for growth rate of expectation of the exponential utility function. Recall that in this case along with the total reward also its higher moments are taken into account. Using Taylor expansion of the utility function we present explicit formulae for calculating variance and higher central moments of the total reward generated by the Markov reward process along with its asymptotic behavior.

Keywords Continuous-time Markov reward chains, exponential utility, moment generating functions, formulae for central moments

JEL Classification C44, C61
AMS Classification 90C40

1 Introduction

The usual optimization criteria examined in the literature on stochastic dynamic programming, such as a total discounted or mean (average) reward structures, may be quite insufficient to characterize robustness of the problem from the point of a decision maker. To this end it may be preferable if not necessary to select more sophisticated criteria that also reflect stability and variability-risk features of the problem. Hence robustness and risk control are also important issues in practical applications. Perhaps the best known approaches stem from the classical work of Markowitz (1952) on mean variance selection rules, i.e. we optimize the weighted sum of average or total reward and its variance, and from the seminal paper titled "Risk-sensitive Markov decision processes" of Howard and Matheson (1972), based on evaluating generated reward by exponential utility functions. Higher moments and variance of cumulative rewards in Markov reward chains have been original studied only for discrete time models. Research in this direction has been initiated in early papers Mandl (1971), Jaquette (1975), Benito (1982) and Sobel (1982). For connections with risk sensitive models see e.g. Cavazos-Cadena and Fernandez-Gaucherand (1999), Cavazos-Cadena and Hernández-Hernández (2005) and Sladký (2008).

To the best of our knowledge higher moments of cumulative rewards for continuous-time Markov control processes were originally studied by Jaquette (1975). In the paper Van Dijk and Sladky (2006) results for the discrete-time case are extended to continuous-time Markov reward chains. As the essential step is an expression for the variance of the undiscounted cumulative reward and its asymptotic behavior. Limiting average variance for continuous-time models are also studied in Guo and Song (2009) and in Prieto-Rumeau and Hernández-Lerma (2009)

(see also the monograph by Guo and Hernández-Lerma (2009)), Wei and Chen (2016) and for discounted models in Guo and Ying (2012).

The present article is a continuous-time version of the author's paper Sladky (2018). The present paper is structured as follows. Section 2 contains notations and summary of basic facts on continuous-time Markov reward processes. Markov models with exponential utility function (called risk-sensitive Markov chains) are studied in section 3 along with the corresponding moment generating functions. Sections 4 and 5 are devoted to explicit formulas of higher moments and higher central models of total rewards generated in continuous-time Markov decision chains.

2 Notations and Preliminaries

In this note we consider Markov decision processes with finite state space $\mathcal{I} = \{1, 2, \ldots, N\}$ evolving in continuous-time. In particular, the development of the considered Markov decision process $X = \{X(t), t \geq 0\}$ (with finite state space $\mathcal{I}$) over time is governed by the transition rates $q(j|i,a)$, for $i, j \in \mathcal{I}$, depending on the selected action $a \in \mathcal{A}_i$. For $j \neq i$ $q(j|i,a)$ is the transition rate from state i to state j, $q(i|i,a) = \sum_{j \in \mathcal{I}, j \neq i} q(j|i,a)$ is the transition rate out of state i. Recall that for sufficiently small δ it holds for transition probabilities $P_{ij}(\cdot)$'s and transition rates $q_{ij}(\cdot)$'s (with $q_{ii}(\cdot) = -q_i(\cdot)$) that

$$P_{ij}(\cdot) = q_{ij}(\cdot) \cdot \delta + o(\delta^2) \quad \text{for} \quad i \neq j, \qquad P_{ii}(\cdot) = (1 - q_i(\cdot) \cdot \delta) + o(\delta^2)$$

and similarly for the corresponding one stage rewards we can conclude that

$$\begin{aligned} r_{ij} &:= r(i,j) \quad \text{for} \quad i \neq j \quad \text{is the transition reward from state } i \text{ to state } \quad j \\ r_i &:= r(i) \quad \text{is the reward rate earned in state } i \end{aligned}$$

Let $\xi(t) := \int_0^t r(X(\tau)\mathrm{d}\tau + \sum_{k=0}^{N(t)} r(X(\tau^-), X(\tau^+))$, obviously $\xi(t)$ is the (random) reward obtained up to time t, where $X(t)$ denotes the state at time t, $X(\tau^-)$, $X(\tau^+)$ is the state just prior and after the kth jump, and $N(t)$ is the number of jumps up to time t. Similarly $\xi(t', t) := \int_{t'}^t r(X(\tau))\mathrm{d}\tau + \sum_{k=N(t')}^{N(t)} r(X(\tau^-), X(\tau^+))$ is the total (random) reward obtained in the time interval $[t', t)$; hence $\xi(t+\Delta) = \xi(\Delta) + \xi(\Delta, t+\Delta)$ or $\xi(t+\Delta) = \xi(t) + \xi(t, t+\Delta)$.

In this note we shall suppose that the obtained random reward, say ξ, is evaluated by an exponential utility function, say $u^\gamma(\cdot)$, i.e. utility functions with constant risk sensitivity depending on the value of the risk sensitivity coefficient γ.

In case that $\gamma > 0$ (the risk seeking case) the utility assigned to the (random) reward ξ is given by $u^\gamma(\xi) := \exp(\gamma\xi)$, if $\gamma < 0$ (the risk averse case) the utility assigned to the (random) reward ξ is given by $u^\gamma(\xi) := -\exp(\gamma\xi)$, for $\gamma = 0$ it holds $u^\gamma(\xi) = \xi$ (risk neutral case). Hence we can write

$$u^\gamma(\xi) = \mathrm{sign}(\gamma) \exp(\gamma\xi) \tag{1}$$

and for the expected utility we have (E is reserved for expectation)

$$\bar{U}^{(\gamma)}(\xi) := \mathsf{E}u^\gamma(\xi) = \mathrm{sign}(\gamma)\mathsf{E}[\exp(\gamma\xi)], \quad \text{where} \quad \mathsf{E}\,[\exp(\gamma\xi)] = \sum_{k=0}^{\infty} \mathsf{E}\,\frac{1}{k!}(\gamma\xi)^k. \tag{2}$$

Then for the corresponding certainty equivalent $Z^\gamma(\xi)$ we have

$$u^\gamma(Z^\gamma(\xi)) = \mathrm{sign}(\gamma)\mathsf{E}[\exp(\gamma\xi)] \Longleftrightarrow Z^\gamma(\xi) = \gamma^{-1} \ln\{\mathsf{E}\,[\exp(\gamma\xi)]\}. \tag{3}$$

From (2),(3) we immediately conclude that

$$Z^{\gamma}(\xi) \approx \mathsf{E}\xi + \frac{\gamma}{2}\mathsf{Var}\,\xi. \tag{4}$$

A (Markovian) policy controlling the decision process is given as a piecewise constant right continuous function of time. In particular, $\pi = f(t)$, is a piecewise constant, right continuous vector function where $f(t) \in \mathcal{F} \equiv \mathcal{A}_1 \times \ldots \times \mathcal{A}_N$, and $f_i(t) \in \mathcal{A}_i$ is the decision (or action) taken at time t if the process $X(t)$ is in state i. Since π is piecewise constant, for each π we can identify the time points $0 < t_1 < t_2 \ldots < t_i < \ldots$ at which the policy switches; we denote by $f^i \in \mathcal{F}$ the decision rule taken in the time interval $(t_{i-1}, t_i]$. Policy which takes at all times the same decision rule, i.e. $\pi \sim f$, is called stationary; $Q(f)$ is the transition rate matrix with elements $q(j|i, f_i)$.

Let for $f \in \mathcal{F}$ $Q(f) = [q_{ij}(f_i)]$ be an $N \times N$ matrix whose ijth element $q_{ij}(f_i) = q(j|i, f_i)$ for $i \neq j$ and for the iith element we set $q_{ii}(f_i) = -q(i|i, f_i)$. The sojourn time of the considered process X in state $i \in \mathcal{I}$ is exponentially distributed with parameter $[q(i|i, f_i)]$. Hence the expected value of the reward obtained in state $i \in \mathcal{I}$ equals $\tilde{r}_i(f_i) = [q(i|i, f_i)]^{-1}\, r(i) + \sum_{j\in\mathcal{I}, j\neq i} q(j|i, f_i)\, r(i, j)$ and $\tilde{r}(f)$ is the (column) vector of reward rates at time t.

Using policy $\pi = f(t)$ means that if the Markov chain X was found to be in state i at time t, the action chosen at this time is $f_i(t)$, i.e. the ith coordinate of $f(t) \in \mathcal{F}$. For any policy $\pi = f(t)$ the accompanying set of transition rate matrices $\{Q(f(t)), t \geq 0\}$ determines a continuous-time (in general, nonstationary) Markov process.

3 Formulas for Higher Moments of Random Reward

Supposing that the obtained random reward up to time t, say $\xi(t)$, is evaluated by an exponential utility function, say $u^{\gamma}(\cdot)$, with the risk sensitivity coefficient γ, let for $\pi \sim (f)$, $U_i^{(\gamma)}(t, f) := \mathsf{E}_i^{\pi}[\exp(\gamma\xi(t))]$ considered as the moment generating function of $\xi(t)$, we can conclude that for $k = 0, 1, 2, \ldots$, $n = 0, 1, 2, \ldots$

$$M_i^{(k,\pi)}(t) := \mathsf{E}_i^{\pi}(\exp(\xi(t)^k) = \frac{\mathrm{d}^k}{\mathrm{d}\gamma^k}\mathsf{E}_i^{\pi}[\exp(\gamma\xi(t))]|_{\gamma=0} \quad \text{is the } k\text{th moment of } \xi(t) \tag{5}$$

and the Taylor expansion around $\gamma = 0$ reads

$$U_i^{(\gamma)}(t, f) = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} M_i^{(k,\pi)}(t). \tag{6}$$

Similarly on introducing the moment generating function for the central moments of $\xi(t)$ by

$$\widetilde{U}_i^{(\gamma)}(t, f) := \mathsf{E}_i^{\pi}[\exp(\gamma(\xi(t) - \mathsf{E}_i^{\pi}\xi(t))]^k \quad \text{for all} \quad i \in \mathcal{I} \tag{7}$$

for the kth central moments of $\xi(t)$ we have

$$\widetilde{M}_i^{(k,\pi)}(t) := \mathsf{E}_i^{\pi}[\xi(t) - \mathsf{E}_i^{\pi}\xi(t)]^k = \frac{\mathrm{d}^k}{\mathrm{d}\gamma^k}\mathsf{E}_i^{\pi}[\exp(\gamma(\xi(t) - \mathsf{E}_i^{\pi}\xi(t))]_{\gamma=0} \tag{8}$$

and the Taylor expansion around $\gamma = 0$ for sufficiently small γ reads

$$\widetilde{U}_i^{(\gamma)}(t, f) = 1 + \sum_{k=1}^{\infty} \frac{\gamma^k}{k!} \widetilde{M}_i^{(k,\pi)}(t) \tag{9}$$

Let $M^{(k,\pi)}(t)$, $\widetilde{M}^{(k,\pi)}(t)$ be the (column) vector of the k moments, central k moments respectively, with elements $M_i^{(k,\pi)}(t)$, $\widetilde{M}_i^{(k,\pi)}(t)$ respectively.

In particular, if the system starts in state i, the expected total reward earned in state i up to the first exit of state i within time interval $[t, t+\delta)$ is equal to

$$M_i^{(1,\pi)}(t+\delta) = M_i^{(1,\pi)}(t) \cdot [1 - q_i(f_i) \cdot \delta] + \delta \cdot r(i) \cdot [1 - q_i(f_i) \cdot \delta] + o(\delta^2) \tag{10}$$

and for the s-th power of this reward it holds

$$\begin{aligned} M_i^{(s,\pi)}(t+\delta) &= \{M_i^{(1,\pi)}(t)[1 - q_i(\cdot) \cdot \delta] + [r(i) \cdot \delta] \cdot [1 - q_i(f_i) \cdot \delta]\}^s \\ &= M_i^{(s,\pi)}(t) + s \cdot M_i^{(s-1,\pi)}(t) \cdot \delta + o(\delta^2) \end{aligned} \tag{11}$$

(Observe that $M_i^{(s,\pi)}(t) = [M_i^{(1,\pi)}(t)]^s$, $[M_i^{(1,\pi)}(t) + r(i) \cdot \delta]^s = M_i^{(s,\pi)}(t) + s \cdot r(i) \cdot M_i^{(s-1,\pi)}(t) + o(\delta^2)$.)

4 Higher Moments in Continuous-time Models

On inserting from (5),(7),(10),(11) we can conclude that

$$\text{For } k=0: \; M_i^{(0,\pi)}(t+\delta) = [1 - q_i(f_i) \cdot \delta] + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \cdot \delta \cdot M_j^{(0,\pi)}(t) \quad \Rightarrow M_j^{(0,\pi)}(t) = 1 \;\; \forall j. \tag{12}$$

$$\text{For } k=1: \; M_i^{(1,\pi)}(t+\delta) = [1 - q_i(f_i) \cdot \delta] \cdot [M_i^{(1,\pi)}(t) + r(i) \cdot \delta] + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \cdot \delta \cdot [r_{ij} + M_j^{(1,\pi)}(t)] \tag{13}$$

$$\begin{aligned} \text{For } k=2: \; & M_i^{(2,\pi)}(t+\delta) = [1 - q_i(f_i) \cdot \delta] \cdot [M_i^{(2,\pi)}(t) + 2 \cdot M_i^{(1,\pi)}(t) \cdot r(i) \cdot \delta] \\ & + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \cdot \delta \cdot \left\{ [r_{ij}]^2 + 2 \cdot r_{ij} \cdot M_j^{(1,\pi)}(t) + M_j^{(2,\pi)}(t) \right\} + o(\delta^2). \end{aligned} \tag{14}$$

$$\begin{aligned} \text{For } k=3: \; & M_i^{(3,\pi)}(t+\delta) = [1 - q_i(f_i) \cdot \delta] \cdot [M_i^{(3,\pi)}(t) + 3 \cdot M_i^{(2,\pi)}(t) \cdot r(i) \cdot \delta] \\ & + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \cdot \delta \left\{ [r_{ij}]^3 + 3 \cdot [r_{ij}]^2 \cdot M_j^{(1,\pi)}(t) + 3 \cdot [r_{ij}] \cdot M_j^{(2,\pi)}(t) + M_j^{(3,\pi)}(t) \right\} + o(\delta^2). \end{aligned} \tag{15}$$

$$\begin{aligned} \text{For } k=4: \; & M_i^{(4,\pi)}(t+\delta) = [1 - q_i(f_i) \cdot \delta] \cdot [M_i^{(4,\pi)}(t) + 4 \cdot M_i^{(4,\pi)}(t)] + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \cdot \delta \times \\ & \times \left\{ [r_{ij}]^4 + 4 \cdot [r_{ij}]^3 \cdot M_j^{(1,\pi)}(t) + 6 \cdot [r_{ij}]^2 \cdot M_j^{(2,\pi)}(t) + 4 \cdot [r_{ij}] \cdot M_j^{(3,\pi)}(t) + M_j^{(4,\pi)}(t) \right\} + o(\delta^2). \end{aligned} \tag{16}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

In general:

$$\begin{aligned} M_i^{(s,\pi)}(t+\delta) &= [1 - q_i(f_i) \cdot \delta] \cdot [M_i^{(s,\pi)}(t) + s \cdot r(i) \cdot \delta] \\ &\quad + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \cdot \delta \cdot \left\{ \sum_{k=0}^{s} \binom{s}{k} \cdot [r_{ij}]^k \cdot M_j^{(s-k,\pi)}(t) \right\} + o(\delta^2) \end{aligned} \tag{17}$$

hence

$$\begin{aligned} M_i^{(s,\pi)}(t+\delta) - M_i^{(s,\pi)}(t) &= s \cdot M_i^{(s-1,\pi)}(t) \cdot r(i) \\ &\quad + \sum_{j \in \mathcal{I}, j \neq i} q_{ij}(f_i) \cdot \delta \cdot \left\{ \sum_{k=1}^{s} \binom{s}{k} [r_{ij}]^k M_j^{(s-k,\pi)}(t) \right\} + \sum_{j \in \mathcal{I}} q_{ij}(f_i) \cdot \delta \cdot M_i^{(s,\pi)}(t) + o(\delta^2) \end{aligned} \tag{18}$$

or equivalently

$$\frac{1}{\delta}\{M_i^{(s,\pi)}(t+\delta) - M_i^{(s,\pi)}(t)\} = s \cdot M_i^{(s-1,\pi)}(t) \cdot r(i) \cdot \delta + \sum_{j\in\mathcal{I}, j\neq i} q_{ij}(f_i) \times \left\{ \sum_{k=1}^{s} \left(\begin{array}{c} s \\ k \end{array} \right) \cdot [r_{ij}]^k \cdot M_j^{(s-k,\pi)}(t) \right\} + \sum_{j\in\mathcal{I}} q_{ij}(f_i) \cdot M_i^{(s,\pi)}(t) + o(\delta) \quad (19)$$

For δ tending to null from the above material we can arrive at the following systems of differential equations.

$$\text{For } k=0: \ \frac{\mathrm{d}}{\mathrm{d}t} M_i^{(0,\pi)}(t) = \sum_{j\in\mathcal{I}} q_{ij}(f_i)\, M_j^{(0,\pi)}(t) \qquad \Rightarrow M_j^{(0,\pi)}(t) = 1 \quad \forall j. \quad (20)$$

$$\text{For } k=1: \ \frac{\mathrm{d}}{\mathrm{d}t} M_i^{(1,\pi)}(t) = r(i) + \sum_{j\in\mathcal{I}, j\neq i} q_{ij}(f_i)\, r_{ij} + \sum_{j\in\mathcal{I}} q_{ij}(f_i)\, M_j^{(1,\pi)}(t). \quad (21)$$

$$\text{For } k=2: \ \frac{\mathrm{d}}{\mathrm{d}t} M_i^{(2,\pi)}(t) = 2 \cdot M_i^{(1,\pi)}(t) \cdot r(i) + \sum_{j\in\mathcal{I}, j\neq i} q_{ij}(f_i) \left\{ [r_{ij}]^2 + 2 \cdot r_{ij} \cdot M_j^{(1,\pi)}(t) \right\} + \sum_{j\in\mathcal{I}} q_{ij}(f_i) \cdot M_j^{(2,\pi)}(t). \quad (22)$$

$$\text{For } k=3: \ \frac{\mathrm{d}}{\mathrm{d}t} M_i^{(3,\pi)}(t) = 3 \cdot M_i^{(2,\pi)}(t) \cdot r(i) + \sum_{j\in\mathcal{I}, j\neq i} q_{ij}(f_i) \times \times \left\{ [r_{ij}]^3 + 3 \cdot [r_{ij}]^2 \cdot M_j^{(1,\pi)}(t) + 3 \cdot [r_{ij}] \cdot M_j^{(2,\pi)}(t) \right\} + \sum_{j\in\mathcal{I}} q_{ij}(f_i)\, M_j^{(3,\pi)}(t). \quad (23)$$

$$\text{For } k=4: \ \frac{\mathrm{d}}{\mathrm{d}t} M_i^{(4,\pi)}(t) = 4 \cdot M_i^{(3,\pi)}(t) \cdot r(i) + \sum_{j\in\mathcal{I}, j\neq i} q_{ij}(f_i) \times \times \left\{ [r_{ij}]^4 + 4 \cdot [r_{ij}]^3\, M_j^{(1,\pi)}(t) \ + \ 6 \cdot [r_{ij}]^2 \cdot M_j^{(2,\pi)}(t) + \ 4 \cdot [r_{ij}] \cdot M_j^{(3,\pi)}(t) \right\} + \sum_{j\in\mathcal{I}} q_{ij}(f_i) \cdot M_j^{(4,\pi)}(t). \quad (24)$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

In general:

$$\frac{\mathrm{d}}{\mathrm{d}t} M_i^{(s,\pi)}(t) = s \cdot M_i^{(s-1,\pi)}(t) \cdot r(i) + \sum_{j\in\mathcal{I}, j\neq i} q_{ij}(f_i) \left\{ \sum_{k=1}^{s} \left(\begin{array}{c} s \\ k \end{array} \right) \cdot M_i^{(s,\pi)}(t) \cdot \ [r_{ij}]^k\, M_j^{(s-k,\pi)}(t) \right\} + \sum_{j\in\mathcal{I}} q_{ij}(f_i) \cdot \ M_j^{(s,\pi)}(t) \quad (25)$$

5 Higher Central Moments in Continuous-time Models

Supposing that higher moments are known, the corresponding central moments can be easily computed. To this end, on recalling definition central moments, we can easily conclude that the if the system starts in state i and policy π is followed then the nth central moment at time t

$$\widetilde{M}_i^{(n,\pi)}(t) := \mathrm{E}_i^{\pi}[\xi(t) - M_i^{(1,\pi)}(t)]^n \quad (26)$$

Since $M_i^{(j,\pi)}(t) := \mathrm{E}_i^\pi[\xi(t)]^j$, after little algebra we arrive at

$$\widetilde{M}_i^{(n,\pi)}(t) := \sum_{j=0}^{n} \binom{n}{j} \cdot (-1)^{n-j} \cdot M_i^{(j,\pi)}(t) \cdot [M_i^{(1,\pi)}(t)]^{n-j} \tag{27}$$

$$= \sum_{j=0}^{n-2} \binom{n}{j} \cdot (-1)^{n-j} \cdot M_i^{(j,\pi)}(t) \cdot [M_i^{(1,\pi)}(t)]^{n-j} + (-1)^{n-1} \cdot (n-1) \cdot [M_i^{(1,\pi)}(t)]^n \tag{28}$$

In particular,

$$\widetilde{M}_i^{(1,\pi)}(t) = M_i^{(1,\pi)}(t) - M_i^{(1,\pi)}(t) = 0 \tag{29}$$

$$\widetilde{M}_i^{(2,\pi)}(t) = M_i^{(2,\pi)}(t) - [M_i^{(1,\pi)}(t)]^2 \tag{30}$$

$$\begin{aligned}\widetilde{M}_i^{(3,\pi)}(t) &= M_i^{(3,\pi)}(t) - 3 \cdot M_i^{(2,\pi)}(t) \cdot M_i^{(1,\pi)}(t) + 3 \cdot M_i^{(1,\pi)}(t) \cdot [M_i^{(1,\pi)}(t)]^2 - [M_i^{(1,\pi)}(t)]^3 \\ &= M_i^{(3,\pi)}(t) - 3 \cdot M_i^{(2,\pi)}(t) \cdot M_i^{(1,\pi)}(t) + 2 \cdot [M_i^{(1,\pi)}(t)]^3 \end{aligned} \tag{31}$$

$$\begin{aligned}\widetilde{M}_i^{(4,\pi)}(t) &= M_i^{(4,\pi)}(t) - 4 \cdot M_i^{(3,\pi)}(t) \cdot [M_i^{(1,\pi)}(t)] + 6 \cdot M_i^{(2,\pi)}(t) \cdot [M_i^{(1,\pi)}(t)]^2 \\ &\quad -4 \cdot M_i^{(1,\pi)}(t) \cdot [M_i^{(1,\pi)}(t)]^3 + [M_i^{(1,\pi)}(t)]^4 \\ &= M_i^{(4,\pi)}(t) - 4 \cdot M_i^{(1,\pi)}(t) \cdot M_i^{(3,\pi)}(t) + 6 \cdot [M_i^{(1,\pi)}(t)]^2 \cdot M_i^{(2,\pi)}(t) - 3 \cdot [M_i^{(1,\pi)}(t)]^3 \end{aligned} \tag{32}$$

Since $\widetilde{M}_i^{(1,\pi)}(t) = 0$ we shall consider $\widetilde{M}_i^{(s,\pi)}(t)$ only for $s = 2, 3, \ldots$. From (30)–(32) we immediately obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{M}_i^{(2,\pi)}(t) = \frac{\mathrm{d}}{\mathrm{d}t}M_i^{(2,\pi)}(t) - 2 \cdot [M_i^{(1,\pi)}(t)] \cdot \frac{\mathrm{d}}{\mathrm{d}t}[M_i^{(1,\pi)}(t)] \tag{33}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{M}_i^{(3,\pi)}(t) = \frac{\mathrm{d}}{\mathrm{d}t}M_i^{(3,\pi)}(t) - 3 \cdot \frac{\mathrm{d}}{\mathrm{d}t}\left\{M_i^{(2,\pi)}(t) \cdot \widetilde{M}_i^{(2,\pi)}(t)\right\} + 6 \cdot [M_i^{(1,\pi)}(t)]^2 \cdot \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{M}_i^{(1,\pi)}(t) \tag{34}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\widetilde{M}_i^{(4,\pi)}(t) &= \frac{\mathrm{d}}{\mathrm{d}t}M_i^{(4,\pi)}(t) - 4 \cdot \frac{\mathrm{d}}{\mathrm{d}t}\left\{M_i^{(2,\pi)}(t) \cdot \widetilde{M}_i^{(3,\pi)}(t)\right\} + 6 \cdot \frac{\mathrm{d}}{\mathrm{d}t}\left\{[M_i^{(2,\pi)}(t)]^2 \cdot \widetilde{M}_i^{(2,\pi)}(t)\right\} - \\ &\quad -9 \cdot [M_i^{(2,\pi)}(t)]^2 \cdot \frac{\mathrm{d}}{\mathrm{d}t}[M_i^{(1,\pi)}(t)] \end{aligned} \tag{35}$$

Acknowledgements: This work was supported by the Czech Science Foundation under Grant 18-02739S.

References

[1] Benito, F. (1982.) Calculating the variance in Markov processes with random reward. *Trabajos de Estadistica y de Investigacion Operativa* 33, pp.73–85.

[2] Cavazos-Cadena, R. and Fernandez-Gaucherand, F. (1999). *Controlled Markov chains with risk-sensitive criteria: average cost, optimality equations and optimal solutions.* Math. Methods Oper. Res. 43, pp. 121–139.

[3] Cavazos-Cadena, R. and Hernández-Hernández, D. (2005). *A characterization of the optimal risk-sensitive average cost infinite controlled Markov chains.* Ann. Appl. Probab. 15, pp. 175–212.

[4] Gantmakher, F. R. (1959). *The theory of matrices.* Chelsea, London.

[5] Guo, X. and Hernández-Lerma, O. (2009). *Continuous-Time Markov Decision Processes: Theory and Applications.* Berlin: Springer.

[6] Guo, X. and Song, X. (2009). Mean-variance criteria for finite continuous-time Markov decision processes. *IEEE Trans. Automat. Control* 54, 2151–2157.

[7] Guo, X., Ye, L. and Yin, G.A. (2012). Mean-variance optimization problem for discounted Markov decision processes. *European J. Oper. Res.* 220, 423–429.

[8] Howard, R. A. and Matheson, J. (1972). *Risk-sensitive Markov decision processes.* Manag. Sci. 23, pp. 356–369.

[9] Jaquette, S. C. (1976). *A utility criterion for Markov decision processes.* Manag. Sci. 23, pp. 43–49.

[10] Mandl, P. (1971). *On the variance in controlled Markov chains.* Kybernetika 7, pp. 1–12.

[11] Markowitz, H. (1952). Portfolio Selection. *Journal of Finance* 7, 77–92.

[12] Puterman, M. L. (1994). *Markov decision processes – discrete stochastic dynamic programming.* Wiley, New York.

[13] Ross, S. M. (1983). *Introduction to stochastic dynamic programming.* Academic Press, New York.

[14] Sladký, K. (2005). *On mean reward variance in semi-Markov processes.* Math. Methods Oper. Res. 62, pp. 387–397.

[15] Sladký, K. (2008). *Growth rates and average optimality in risk-sensitive Markov decision chains.* Kybernetika 44, pp. 206–217.

[16] Sladký, K.(2013). *Risk-sensitive and mean variance optimality in Markov decision processes.* Acta Oeconomica Pragensia 7, pp. 146–161.

[17] Sladký, K.(2018). *Central moments and risk-sensitive optimalilty in Markov reward chains.* In: Quantitave Methods in Economics – Multiple Criteria Decision Making XIX (M. Reiff, P. Gežík, Eds). University od Economics, Bratislava 2018, pp. 325–331.

[18] Sobel, M. (1982). *The variance of discounted Markov decision processes.* J. Appl. Probab. 19, pp. 794–802.

[19] van Dijk, N.M. and Sladký, K. (2006). *On the total reward variance for continuous-time Markov reward chains.* J. Appl. Probab. 43, pp. 1044–1052.

[20] Wei, Q. and Chen, X. (2016). *Continuous-time Markov decision processes under risk-sensitive average cost criterion.* Oper. Res. Lett. 44, pp. 457–462.

Author's address
Ing. Karel Sladký, CSc.
Institute of Information Theory and Automation
of the Czech Academy of Sciences
Department of Econometrics
Pod Vodárenskou věží 4, 182 08 Praha 8
Czech Republic
e-mail: sladky@utia.cas.cz