J. Appl. Prob. **46,** 1209–1212 (2009)
Printed in England

AN INEQUALITY FOR VARIANCES OF THE DISCOUNTED REWARDS

EUGENE A. FEINBERG * ** AND
JUN FEI,* *** *Stony Brook University*

Abstract

We consider the following two definitions of discounting: (i) multiplicative coefficient in front of the rewards, and (ii) probability that the process has not been stopped if the stopping time has an exponential distribution independent of the process. It is well known that the expected total discounted rewards corresponding to these definitions are the same. In this note we show that, the variance of the total discounted rewards is smaller for the first definition than for the second definition.

Keywords: Total discounted reward; variance; stopping time

2000 Mathematics Subject Classification: Primary 60G40
Secondary 90C40

1. Introduction

In this note we study two definitions of discounting: (i) multiplicative coefficient in front of the rewards, and (ii) probability that the process has not been stopped if the stopping time has an exponential distribution independent of the process. It is well known that the total discounted rewards corresponding to these definitions have equal expectations. However, as we will show, the second moment and variance are smaller for the first definition than for the second definition.

Since its introduction by Markowitz in his Nobel Prize winning paper [5], variance has played an important role in stochastic optimization. In particular, there is a significant amount of literature on various optimizations of Markov decision processes (MDPs); see the pioneering work by Jaquette [4] and Sobel [7]–[9], a survey by White [11], and recent references by Van Dijk and Sladký [10] and Baykal-Gürsoy and Gürsoy [1].

Our interest in the variance of total discounted rewards is motivated by constrained optimization of continuous-time MDPs. According to [2], optimization policies can be presented in different forms. In particular, they can be presented in the forms of randomized stationary and switching stationary policies. The expected total discounted rewards are equal for the corresponding randomized stationary and switching stationary policies [2, Theorem 5.1]. However, the variances of the total discounted rewards for the policies can be different. In addition, they may depend on the definition of discounting.

Received 6 July 2009; revision received 5 October 2009.
* Postal address: Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA.
** Email address: eugene.feinberg@sunysb.edu
*** Email address: jun.fei@sunysb.edu

2. Main result

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space with a filtration $\mathcal{F}_t$, $t \in [0, \infty)$, where $\mathcal{F}_s \subseteq \mathcal{F}_t \subseteq \mathcal{F}$ for all $0 \le s < t < \infty$. Consider a nondecreasing sequence of stopping times T_n, $n = 1, 2, \ldots$. Let

$$\mathcal{F}_\infty = \bigcup_{t \in [0,\infty)} \mathcal{F}_t.$$

We consider an $\mathcal{F}_t$-adapted stochastic process r_t, $t \in [0, \infty)$, and an $\mathcal{F}_{T_n}$-adapted stochastic sequence R_n, $n = 1, 2, \ldots$. The process r_t can be interpreted as the reward rate at time t. In addition, a lump sum R_n is collected at time T_n.

There are two natural ways to define the total discounted rewards. One way is to interpret discounting as the coefficient in front of the reward rate. In this case, the total discounted rewards are defined as

$$J_1 = \int_0^\infty \mathrm{e}^{-\alpha t} r_t \, \mathrm{d}t + \sum_{n=1}^\infty \mathrm{e}^{-\alpha T_n} R_n,$$

where $\alpha > 0$ is the discount rate.

Another way is to define the total discounted rewards as the total rewards until a stopping time T that has an exponential distribution with rate α. Let T be independent of $\mathcal{F}_\infty$ and let $\mathrm{P}\{T > t\} = \mathrm{e}^{-\alpha t}$. Then the total discounted reward can be defined as

$$J_2 = \int_0^T r_t \, \mathrm{d}t + \sum_{n=1}^{N(T)} R_n,$$

where

$$N(t) = \sup\{n \colon T_n \le t\}, \qquad t \ge 0.$$

It is well known that

$$\mathrm{E}[J_1] = \mathrm{E}[J_2], \tag{2.1}$$

if at least one side of this equation is well defined (a random variable has a well-defined expectation if either the expectation of its positive part is finite or the expectation of its negative part is finite).

Indeed,

$$\begin{aligned}
\mathrm{E} \sum_{n=1}^{N(T)} R_n &= \sum_{n=1}^\infty \mathrm{E}\ R_n \mathbf{1}\{T \ge T_n\} \\
&= \sum_{n=1}^\infty \mathrm{E}\, \mathrm{E}[R_n \mathbf{1}\{T \ge T_n\} \mid \mathcal{F}_{T_n}] \\
&= \mathrm{E} \sum_{n=1}^\infty R_n \, \mathrm{E}[\mathbf{1}\{T \ge T_n\} \mid \mathcal{F}_{T_n}] \\
&= \mathrm{E} \sum_{n=1}^\infty R_n \, \mathrm{P}\{T \ge T_n \mid \mathcal{F}_{T_n}\} \\
&= \mathrm{E} \sum_{n=1}^\infty R_n \mathrm{e}^{-\alpha T_n}
\end{aligned}$$

https://doi.org/10.1239/jap/1261670699 Published online by Cambridge University Press

and

$$\begin{aligned}\mathrm{E}\int_0^T r_t\,\mathrm{d}t &= \mathrm{E}\left[\int_0^\infty r_t\,\mathbf{1}\{T \ge t\}\,\mathrm{d}t\right]\\ &= \int_0^\infty \mathrm{E}[r_t\,\mathbf{1}\{T \ge t\}]\,\mathrm{d}t\\ &= \int_0^\infty \mathrm{E}[r_t]\,\mathrm{E}[\mathbf{1}\{T \ge t\}]\,\mathrm{d}t\\ &= \int_0^\infty \mathrm{E}\ r_t\,\mathrm{P}\{T \ge t\}\,\mathrm{d}t\\ &= \mathrm{E}\int_0^\infty \mathrm{e}^{-\alpha t} r_t\,\mathrm{d}t.\end{aligned}$$

In particular, (2.1) holds for deterministic functions r and R, and, therefore,

$$\mathrm{E}[J_1 \mid \mathcal{F}_\infty] = \mathrm{E}[J_2 \mid \mathcal{F}_\infty] \quad \mathrm{P}\text{-a.s.}, \tag{2.2}$$

if either $\mathrm{E}[|J_1| \mid \mathcal{F}_\infty] < \infty$ or $\mathrm{E}[|J_2| \mid \mathcal{F}_\infty] < \infty$ P-a.s. However, the second moments can be different. Indeed, we have the following statement.

Theorem 2.1. *If either* $\mathrm{E}[|J_1| \mid \mathcal{F}_\infty] < \infty$ *or* $\mathrm{E}[|J_2| \mid \mathcal{F}_\infty] < \infty$ P*-a.s., then*

$$\mathrm{var}(J_1) \le \mathrm{var}(J_2),$$

and the equality holds if and only if $\mathrm{var}(J_2 \mid \mathcal{F}_\infty) = 0$ P*-a.s.*

Proof. By the total variance formula (see [6, p. 83] or [3, p. 454]), for $i = 1, 2$,

$$\mathrm{var}(J_i) = \mathrm{E}[\mathrm{var}(J_i \mid \mathcal{F}_\infty)] + \mathrm{var}(\mathrm{E}[J_i \mid \mathcal{F}_\infty]).$$

Therefore, because of (2.2),

$$\mathrm{var}(\mathrm{E}[J_1 \mid \mathcal{F}_\infty]) = \mathrm{var}(\mathrm{E}[J_2 \mid \mathcal{F}_\infty]).$$

In addition, $\mathrm{E}[\mathrm{var}(J_1 \mid \mathcal{F}_\infty)] = 0$ and $\mathrm{E}[\mathrm{var}(J_2 \mid \mathcal{F}_\infty)] \ge 0$. Hence, $\mathrm{var}(J_2) - \mathrm{var}(J_1) = \mathrm{E}[\mathrm{var}(J_2 \mid \mathcal{F}_\infty)] \ge 0$, i.e. $\mathrm{var}(J_1) \le \mathrm{var}(J_2)$.

Example 2.1. Consider a continuous-time Markov chain with two states: 1 and 0, where 0 is an absorbing state. Let state 1 be the initial state. The process spends an exponential time $T_1 \sim \exp(\lambda)$ at state 1 and then jumps to state 0. At state 1 the reward rate is 1 and at the jump epoch there is no lump sum reward. At state 0 the process collects no rewards. Let the discount factor be α and let $T \sim \exp(\alpha)$.

The total discounted rewards under the two definitions are

$$J_1 = \int_0^{T_1} \mathrm{e}^{-\alpha t}\,\mathrm{d}t = \frac{1}{\alpha}(1 - \mathrm{e}^{-\alpha T_1}), \qquad J_2 = \int_0^{T \wedge T_1} \mathrm{d}t = T \wedge T_1.$$

For the first definition,

$$\mathrm{var}(J_1) = \frac{1}{\alpha^2}\,\mathrm{var}(\mathrm{e}^{-\alpha T_1}) = \frac{1}{\alpha^2}(M_{T_1}(-2\alpha) - (M_{T_1}(-\alpha))^2) = \frac{\lambda}{(\lambda+\alpha)^2(\lambda+2\alpha)},$$

where $M_X(s)$ is the moment generating function of a random variable X. In particular, $M_{T_1}(s) = \lambda/(\lambda - s)$.

https://doi.org/10.1239/jap/1261670699 Published online by Cambridge University Press

Since $T \wedge T_1$ is an exponential random variable with intensity $\lambda + \alpha$,

$$\operatorname{var}(J_2) = \frac{1}{(\lambda + \alpha)^2}.$$

Thus, $\operatorname{var}(J_1) < \operatorname{var}(J_2)$.

Example 2.2. Consider a discrete-time Markov chain where at each jump the process receives a lump sum reward of 1. Let the time interval between jumps be 1 unit of time. The discount factor is α and $T \sim \exp(\alpha)$.

The total discounted rewards under the two definitions are respectively

$$J_1 = \sum_{n=1}^{\infty} \mathrm{e}^{-\alpha n} = \frac{\mathrm{e}^{-\alpha}}{1 - \mathrm{e}^{-\alpha}}, \qquad J_2 = \sum_{n=1}^{N(T)} 1 = N(T).$$

Note that J_1 is a deterministic number and J_2 is a random variable depending on T. Thus, $\operatorname{var}(J_1) = 0 < \operatorname{var}(J_2)$. In fact, direct calculation shows that

$$\operatorname{var}(J_2) = \frac{\mathrm{e}^{-\alpha}}{(1 - \mathrm{e}^{-\alpha})^2}.$$

Acknowledgement

This research was supported by NSF grants CMMI-0600538 and CMMI-0928490.

References

[1] Baykal-Gürsoy, M. and Gürsoy, K. (2007). Semi-Markov decision processes: nonstandard criteria. *Prob. Eng. Inf. Sci.* **21,** 635–657.
[2] Feinberg, E. A. (2004). Continuous time discounted jump Markov decision processes: a discrete-event approach. *Math. Operat. Res.* **29,** 492–524.
[3] Fristedt, B. and Gray, L. (1997). *A Modern Approach to Probability Theory*. Birkhäuser, Boston, MA.
[4] Jaquette, S. C. (1975). Markov decision processes with a new optimality criterion: continuous time. *Ann. Statist.* **3,** 547–553.
[5] Markowitz, H. M. (1952). Portfolio selection. *J. Finance* **7,** 77–91.
[6] Shiryaev, A. N. (1996). *Probability*, 2nd edn. Springer, New York.
[7] Sobel, M. J. (1982). The variance of discounted Markov decision processes. *J. Appl. Prob.* **19,** 794–802.
[8] Sobel, M. J. (1985). Maximal mean/standard deviation ratio in an undiscounted MDP. *Operat. Res. Lett.* **4,** 157–159.
[9] Sobel, M. J. (1994). Mean-variance tradeoffs in an undiscounted MDP. *Operat. Res.* **42,** 175–183.
[10] Van Dijk, N. M. and Sladký, K. (2006). On the total reward variance for continuous-time Markov reward chains. *J. Appl. Prob.* **43,** 1044–1052.
[11] White, D. J. (1988). Mean, variance, and probabilistic criteria in finite Markov decision processes: a review. *J. Optimization Theory Appl.* **56,** 1–29.