Journal of the Korean
OR/MS Society
Vol. 13 No. 2 December 1988

— Research Paper —

On Exponential Utility Maximization

Kun-Jen Chung*

Abstract

Let B be the present value of some sequence. This paper concerns the maximization of the expected utility of the present value B when the utility function is exponential.

Key words

Markov decision process, exponential utility, optimality equation, distribution of present value.

1. Background

Let $X_1, X_2, \ldots$ be the sequence of single period rewards of a Markov decision process (hereafter called MDP). The present value of this sequence is

$$(1)\quad B=\sum_{n=1}^{\infty} \beta^{n-1} X_n$$

where $0<\beta<1$ is discount factor. We emphasize that B is a random variable.

This paper concerns maximization of $E[u_\lambda(B)]$ for the utility function $u_\lambda(x)=-e^{-\lambda x}$. Jaquette [10], [11] studies the same problem as ours, namely maximization of $E[u_\lambda(B)]$. The analysis in [10] exploits the fact that $E[u_\lambda(B)]$ is the negative of Laplace transform of B. As a result, there is a $\lambda_0>0$ and a stationary policy which is optimal for all $0<\lambda<\lambda_0$.

Let $B_n=\sum_{i=1}^{n} X_i$. Howard and Matheson [9] studied the maximization of $E[u_\lambda(B_n)]$ both for fixed n and as $n\to\infty$. Let $B_\infty = \lim_{n\to\infty} B_n$ if the limit exists. Denardo and Rothblum [4] studied the maximization of $E[u_\lambda(B_\infty)]$ in a stopping problem.

That is, the model in [4] is an MDP in which each set A_s includes an action which "stops" the decision process. The models in [4] and [9] exhibit risk-sensitivity but lack time-preference ; the absence of discounting is the principal difference between their models and ours.

2. Notation.

Consider a Markov decision process with discount factor β $(0<\beta<1)$. Let S be the state space. Let A_s be

*National Taiwan Institute of Technology

the set of actions available in state s and $C=\{(s,a) : a\in A_s, s\in S\}$. Let s_n, a_n, and X_n indicate the state, action and reward in the nth period. We assume that s_{n+1} and X_n be random variables which depend only on s_n and a_n. Suppose that there is a countable sample space K such that $P(X_n\in K)=1$ for all n. Let

$$P^a_{sjk}=P\{s_{n+1}=j,\ X_n=k \mid s_n=s,\ a_n=a\}.$$

We assume that K lies in a compact set ; this corresponds to the assumption that there is $b<\infty$ such that

$$P\{0\leq X_1\leq b \mid s_1=s,\ a_1=a\}=1 \text{ for all } (s,\ a)\in C.$$

We indicate that the present value of the single rewards is $B=\sum_{n=1}^{\infty}\beta^{n-1}X_n$ which takes values only in $[0, u]$ where $u=b/(1-\beta)$.

3. Optimality equations.

Let $B(n)=\sum_{i=1}^{n}\beta^{i-1}X_i$ with B continuing to denote $B(\infty)$. For each $s\in S$ and $\lambda\geq 0$, let $f_0(s,\lambda)=-1$,

$$(2)\qquad \begin{aligned} f_n(s,\ \lambda)&=sup\{E(-e^{-\lambda B(n)} \mid s_1=s)\}\ (n=1,\ 2,\ \ldots),\\ f(s,\ \lambda)&=sup\{E(-e^{-\lambda B} \mid s_1=s)\}\end{aligned}$$

where the suprema are over all policies.

It can be shown that

(3) $f_n(s,\lambda)=\max\{E[e^{-\lambda X_1}f_{n-1}(s_2,\beta\lambda) \mid s_1=s,\ a_1=a]\ ;\ a\in A_s\}=\max\{\sum_{j\in S} g^a_{sj}(\lambda)\ f_{n-1}(j,\ \beta\lambda) : a\in A_s\}$

where $q^a_{sj}(\lambda)=\sum_{k\in K} p^a_{sjk}\ e^{-\lambda k}$. Let π be a policy and $v(\pi,\ s,\ \lambda)=E_\pi[-e^{-\lambda B} \mid s_1=s]$

where E_π denotes the expectation with respect to the probability distribution of B induced by π. The $\pi*$ is said to be λ-optimal if $v(\pi*,\ s,\ \lambda)\geq v(\pi,\ s,\ \lambda)$ for all $s\in S$ and π.

The main result of this paper is the following.

Theorem 1 : Suppose :

(d) For each $s\in S$, $A_s\subset R$ (the set of real numbers) and A_s is a compact set.

(e) For each $s\in S$ and $n\in I_+=\{1,\ 2,\ \ldots\}$, $J_n(s,\ a)$ is lower semi-continuous on A_s, where $J_n(s,\ a)=\sum_{j\in S} q^a_{sj}(\lambda)\ f_{n-1}(j,\ \beta\lambda)$, $(s,\ a)\in C$.

Then the following statements hold.

(a) For each $s\in S$ and $\theta>0$,

$$\lim_{n\to\infty} f_n(s,\ \theta)=f(s,\ \theta)$$

with $f_n(s, \theta) \leqq f_{n+1}(s, \theta)$ *for all n.*

(b) For each $s \in S$ and $\theta > 0$,

$$(4)\ f(s, \theta) = max\{\sum_{j \in S} q^a_{sj}(\theta)\ f(j, \beta\theta) : a \in A_s\}.$$

(c) Let $a_n = \delta_{n\lambda}(s) \in A_s$ attain the maximum in (4) when $\theta = \beta^{n-1}\lambda$ and let $\pi(\lambda) = (\delta_{1\lambda}, \delta_{2\lambda}, \cdots)$ be the policy which uses the single period rule $\delta_{n\lambda}$ in period n. Then $\pi(\lambda)$ is λ-optimal.

Before we show that Theorem 1, we need Dini's theorem.

Theorem 2 (Dini) [16] : Let $\{g_n\}$ be a sequence of upper semicontinuous real-valued functions on a countably compact space X, and suppose that for each $x \in X$, the sequence $\{g_n(x)\}$ decreases monotonically to zero. The $\{g_n\}$ converges to zero uniformly.

We specially indicate that if a real-valued function h is lower semi-continuous, then -h is upper semi-continuous.

The proof of Theorem 1 : Fix $\theta > 0$. Since $B(n) \geqq \theta$ for all n, $-1 \leq -\exp[-\theta B(n)] \leq o$, so $-1 \leqq f_n(s, \theta) \leqq 0$ for all *n*, s and θ. Therefore $fo(s, \theta) \leqq f_1(s, \theta)$. Induction leads to $f_n(s, \theta) \leqq f_{n+1}(s, \theta)$ for all $\theta \geqq 0$. It follows from

$$(5)\ p\{0 \leq X_i \leq u \text{ for all } i\} = 1$$

that $P\{0 \leq B - B(n) \leq \beta^n u/(1-\beta)\} = 1$. Therefore

$f_n(s, \theta) \leq f(s, \theta) \leq f_n(s, \theta)\ \exp[-\theta\beta^n u/(1-\beta)]$

$\leq f(s, \theta)\ exp\ [-\theta\beta^n u/(1-\beta)]$;

$$f(s, \theta) = \lim_{n \to \infty} f_n(s, \theta).$$

Now we shall show that f satisfies (4). Using $f_n(s, \theta) \leq f(s, \theta)$ for all n, s, and θ,

$$\sum_{j \in S} q^a_{sj}(\theta)\ f_n(j, \beta\theta) \leq \sum_{j \in S} q^a_{sj}(\theta)\ f_n(j, \beta\theta).$$

Fix s and θ. We get

$$f_n(s, \theta) \leq \sup\{J(s, a) : a \in A_s\}$$

were $J(s, a) = \sum_{j \in S} q^a_{sj}(\theta) f\ (j, \beta\theta)$.

In order to derive the opposite inequality, we start with $f(s, \theta) \geqq f_n(s, \theta)$ to obtain

$$f(s, \theta) \geqq f_n(s, \theta) = sup\{J_n(s, a) : a \in A_s\}.$$

By assumption, the supremum is a bounded monotone sequence (as $n \to \infty$); so it has a limit. Therefore

$$(6)\ f(s, \theta) \geqq \lim_{n \to \infty}\{J_n(s, a) : a \in A_s\}],\ s \in S.$$

Dini's theorem, assumptions (d) and (e) imply for each $s \in S$ that $J_n(s, a)$ converges uniformly to $J(s, a)$ on A_s. Therefore for each $\varepsilon > o$, there exists an integer m if $n \geq m$ so that

$$-\varepsilon \leq J_n(s, a) - J(s, a) \leqq 0 \text{ for all } a \in A_s$$

So $sup\{J(s, a) : a \in A_s\} \leq \varepsilon + sup\{J_n(s, a) : a \in A_s\}$

and $sup\{J(s, a) : a \in A_s\} \leq \lim_{n \to \infty} sup\{J_n(s, a) : a \in A_s\} + \varepsilon$

Since ε is an arbitrary positive number, $sup\{J(s, a) : a \in A_s\} \leq \lim_{n \to \infty} sup\{J_n(s, a) : a \in A_s\}$. By (6), we get

$$f(s, \lambda) \geq sup\{J(s, a) : a \in A_s\}.$$

So (4) holds.

In order to establish (c) in Theorem 1, define $\pi(\lambda)$ as in the statement of (c) in Theorem 1. An induction which employs (4) and starts at n=1 establishes

$$f(s, \lambda) = E_{\pi(\lambda)}[-e^{-\lambda B(n)} f(s_{n+1}, \beta^n) \mid s_1 = s]$$

for all n=1, 2, ... However, $f(s_{n+1}, \beta^n\lambda) \to 1$
as $n \to \infty$ because (5) implies

$$exp[-\beta^n \lambda u/(1-\beta)] \leq f(s_{n+1}, \beta^n\lambda) \leq 1 \text{ for all } n$$

(all with probability one). Therefore

$$f(s, \lambda) = E_{\pi(\lambda)}(-e^{-\lambda B}/s_1 = s).$$

This completes the proof of Theorem 1.

Comments. 1°. A result analogous to Theorem 2 is valid for minimization problems. That is, if "inf" replaces "sup" in (2), then (3) and (4) are valid with "min" replacing "max" and parts (a), (b) and (c) remain true when A_s is compact and $J_n(s, a)$ is lower semi-continuous on A_s.

2°. For the related results about our optimality equations, see [5], [7], [13], [17], and [19].

4. A Pointless Procedure

Let F be the set of all bounded real-valued functions on S. Fix $\lambda(\lambda>0)$. For each $u \in F$ and $v \in F$, let $d(u, v) = sup\{ \mid u(s, \lambda) - v(s, \lambda) \mid : s \in S\}$ Then $(F, d(\cdot, \cdot))$ is a complete metric space.

Without loss of generality, we assume that

$P\{X_1 > 1 \mid s_1 = s, s_2 = j, a_1 = a\} = 1$ so $\sum_{j \in S} q^a_{sj}(\theta) < 1$, $(s, a) \in C$, $\theta > 0$. Define a mapping $\Gamma : F \to F$ where

$$\Gamma u(s) = \max\{ \sum_{j \in S} q^a_{sj}(\theta) u(j, \theta) : a \in A_s\} \text{ for } s \in S.$$

Then $d(\Gamma u, \Gamma v) < e^{-\theta} d(u, v)$ for all $u, v \in F$. Hence, Γ is a constraction mapping, and the fixed-point theorem for contraction mappings guarantees that Γ has a unique fixed point. Since $\Gamma 0 = 0$ it follows that 0 is the fixed point of Γ. Therefore, the equation

$$g(s, \theta) = \max\{ \sum_{j \in S} q^a_{sj}(\theta) g(j, \theta) : a \in A_s\}$$

has the unique solution $g(s,\theta)=0$ for all $s \in S$ and $\theta>0$. Therefore, iterating Γ from any initial function does not necessarily yield improving approximations of f in (4) but iterating A in [2] from any initial function does yield improving approximations of the optimal solution of (3) in [2].

REFERENCES

[1] D. P. BERTSEKAS, *Dynamic Programming an Stochastic Control,* Academic Press, New York, 1976.

[2] E. V. DENARDO, Contraction mappings in the theory underlying dynamic programming, SIAM Review, 9(1967), 165-177.

[3] E. V. DENARDO, *Dynamic Programming,* Prentice-Hall, Englewood Cliffs, NJ, 1982.

[4] E. V. DENARDO AND U. G. ROTHBLUM, *Optimal stopping exponential utility, and linear programming,* Math. Programming, 16(1979), pp. 228-244.

[5] N. EAGLE II, *A utility criterion for the Markov decision process,* Ph. D. thesis, Stanford Univ., Stanford, CA, 1975.

[6] P. C. FISHBURN, *Utility Theory for Decision Making,* John Wiley, New York, 1970.

[7] N. FURUKAWA AND S. IWAMOTO, *Markovian decision processes with recursive reward functions,* Bull. Math. Statist., 15(1972), pp. 79-91.

[8] D. P. HEYMAN AND M. J. SOBEL, *Stochastic Models in Operation Research,* Volume II, McGraw-Hill, New York, 1984.

[9] R. S. HOWARD AND J. E. MATHESON, *Risk-sensitive Markov decision processes,* Management Sci., 8(1972), pp. 356-369.

[10] S. C. JAQUETTE, *Markov decision processes with a new optimality criterion : Discrete time, Ann. Stat., 1(1973), pp. 496-505.*

[11] ______, *A utility criterion for Markov decision processes,* Management Sci., 23(1976) pp. 43-49.

[12] D. M. KREPS, *Decision problems with expected utility criteria,* I : *upper and lower convergent utility,* Math. Oper. Res., 2(1977) pp. 45-53.

[13] E. L. PORTEUS, *On the optimality of structured policies in countable stage decision processes,* Management Sci., 22(1975) pp. 148-157.

[14] ______, *On the optimality of structured policies in countable decision processes,* Research Paper No. 141 Rev., Graduate School of Business, Stanford Univ., Stanford, CA, 1975.

[15] M. L. PUTERMAN AND S. L. BRUMELLE, *Policy iteration in stationary dynamic programming,* Math. Oper. Res., 4(1979) pp. 60-69.

[16] H. L. ROYDEN, *Real Analysis,* Macmillan, New York, 1963.

[17] M. SCHAL, *Utility functions and optimal policies in sequential decision problems,* in Game Theory and Mathematical Economics, O. Moeschlin and D. Pallaschke, eds., North-Holland, Amsterdam, 1981, pp. 357-365.

[18] M. J. SOBEL, *Ordinal dynamic programming,* Management Sci., 21(1975) pp. 967-975.

[19] ____, *The variance of discounted MDP's,* J. Appl. Probab., 19(1982), pp. 794-802.