Note on the Equivalence of Risk-Sensitive Average Criteria

Magaly Arisbe Aguilera-González[1*], Rolando Cavazos-Cadena[2], Mario Cantú-Sifuentes[2]

Programa de Maestría en Estadística Aplicada[1*]. Universidad Autónoma Agraria Antonio Narro, Calzada Antonio Narro 1923. Buenavista, Saltillo, Coah., México. CP 25315, magalyarisbe@hotmail.com. Departamento de Estadística y Cálculo[2], Universidad Autónoma Agraria Antonio Narro. rcavazos@uaaan.mx, mcansif@gmail.com

ABSTRACT

This work is concerned with finite-state Markov decision chains. It is supposed that the system is driven by a decision-maker assessing random cost via a utility function U. The main objective is to provide explicit examples of utility functions such that, in spite of representing different risk perceptions, render the same optimal average index and share the same average optimal stationary policies.

Key Words: Risk aversion, Risk attraction, Superior and inferior average cost criteria, Optimal stationary policy, Risk-neutral optimality equation.
Running Head: Equivalence of Risk-sensitive average criteria.
AMS Subject Classifications: 90C40, 91B06

INTRODUCTION

This work deals with discrete-time finite-state Markov decision processes (MDPs) endowed with a risk-sensitive average cost criterion. Besides mild continuity-compactness assumptions, the class of models analyzed below is characterized by two main features concerning

(i) the dynamics of the system, and (ii) the way in which the controller measures the performance of a control policy:

(i) It is assumed that if the system is driven by a stationary policy, then the sate space is *irreducible*. This means that, regardless of the initial sate x, every state y is visited by the system with positive probability.

(ii) The controller assesses a random cost via a utility function U, which is used to measure the performance of a control policy by the long-run U-average cost criterion, an idea that will be formally introduced in Section 2.

In this framework, *the main objective* of this note is *to highlight an interesting phenomenon*, namely:

• Controllers with *different* attitudes before a random cost, may end-up with *the same* optimal average index.

This fact is illustrated for specific utility functions by performing directly the necessary computations, which involve the particular properties of the functions. However, when the functions are 'combined' to obtain a new utility, such properties are not necessarily inherited by the new mapping. For instance, consider two controllers with utility functions U_i, $i = 1,2$, given by

$$U_1(x)=x^2,\ U_2(x)=\sqrt{x},\ x\geq 0.$$

These functions are homogeneous of degrees 2 and 1/2, respectively and, as it will be shown below,

they render the same average criterion. However,

$$U = aU_1 + (1-a)U_2 \quad a \in (0,1), \qquad (1.1)$$

is not homogeneous of any degree, and it is interesting to see whether or not the average index associated to U coincides with the common average criterion induced by U_1 and U_2. In this direction it will be shown that

• The class U of utility function whose average criteria coincide with a given index J is a cone, that is, $U_1, U_2 \in \mathrm{U} \Rightarrow U_1 + U_2 \in \mathrm{U}$ and $cU_1 \in \mathrm{U}$ for every $c > 0$.

In particular, this conclusion that if U_1 and U_2 render the same average index J, then the average criterion associated to the function U in (1.1) also coincides with J.

The theory and applications of MDPs have been extensively studied; see, for instance, Hernández Lerma (1988), Puterman (1994), Arapostathis *et al.* (1993), Sennott (1998), Bäuerle and Rieder (2011). Concerning the idea of risk-sensitive-average optimality, it was initiated in Howard and Matheson (1972) for exponential utilities, and the interest in other type of utilities was recently sparkled in Bäuerle and Rieder (2013).

The organization of the subsequent material is as follows: In Section 2 the decision model is introduced and the idea of certainty equivalent of a random cost with respect to a general utility U is briefly discussed. Next, in Section 3 the notions of risk-aversion and risk-attraction are introduced and these concepts are illustrated using utility functions frequently used in economics. The exposition continues in Section 4 where the risk-sensitive average criteria are formulated, and then in Section 5 it is shown that the three different utilities considered in the paper render the same optimal average cost and share the same stationary policies. The exposition concludes in Section 6 showing the class of utilities that determine the same optimal average criteria have the cone property.

2. Decision Model and Utility Functions

Let $\mathrm{M} = (S, A, \{A(\mathrm{x})\}_{\mathrm{x}\in S}, C, P)$ be an MDP, where the state space S is a finite set endowed with the discrete topology, the action set A is a metric space and, for each $x \in S$, $A(x) \subset A$ is the nonempty subset of admissible actions at x, whereas

$$C{:}\mathrm{IK} \to (0,\infty)$$

is the *positive* cost function, where $\mathrm{IK} := \{(x,a) | x \in S, a \in A(x)\}$ is the space of admissible pairs. On the other hand, $P = [p_{xy}(\cdot)]$ is the controlled transition law on S given IK, that is, for all $(x,a) \in \mathrm{IK}$ and $y \in S$, the relations $p_{xy}(a) \geq 0$ and ${}^{\mathrm{P}}{}_{y\in S} p_{xy}(a) = 1$ are satisfied. This model model M represents a dynamical system driven by a decision maker (controller) applying actions A_t as follows: At each time $t \in \mathrm{IN} := \{0,1,2,3,\ldots\}$ the controller observes the current state, say $X_t = x \in S$, and knows the previous states and actions. Using that information, the decision maker chooses the action (control) $A_t = a \in A(x)$ to be applied, and such an intervention has two consequences: a cost $C(x,a)$ is incurred, and the evolution of the system is influenced in such a way that the new state at $t + 1$ will be $X_{t+1} = y \in S$ with probability $p_{xy}(a)$.

Assumption 2.1. (*i*) *For each $x \in S$, $A(x)$ is a compact subset of A.*
(*ii*) *For every $x,y \in S$, the mappings $a \mapsto C(x,a)$ and $a \mapsto p_{xy}(a)$ are continuous in $a \in A(x)$.*

Observe that, since C is positive, this assumption yields that

$$0 < \min_{k\in \mathrm{K}} C(k) \leq \max_{k\in \mathrm{K}} C(k) \equiv \|C\| < \infty. \qquad (2.1)$$

Policies. A policy π is a rule for choosing actions which, at each time $t \in \mathrm{IN}$, may depend on the current state as well as on the record of previous states and actions; see, for instance, Puterman (1994), or Bäuerle and Reider (2011) for details. The class of all policies is denoted by P and, given the initial state $x \in S$ and the policy π being used for choosing actions, the distribution of the state-action process $\{(X_t, A_t)\}$ is uniquely determined; such a distribution and the corresponding expectation operator are denoted by P_x^π and E_x^π, respectively. Next, define $\mathrm{IF} := {}^{\mathrm{Q}}{}_{x\in S} A(x)$ and notice that IF is a compact metric space, which consists of all functions $f{:}S \to A$ such that $f(x) \in A(x)$ for each $x \in S$. A policy π is *stationary* if there exists $f \in \mathrm{IF}$ such that the equality $A_t = f(X_t)$ is always valid under π; the class of stationary policies is naturally identified with IF. Observe that, when the system is driven by $f \in \mathrm{IF}$, the state process $\{X_t\}$ is a Markov chain with time-invariant transition matrix $[p_{xy}(f(x))]_{x,y\in S}$.

Utility Functions. A basic assumption in this work is that the attitude of the decision-maker before a random cost Y is determined by a utility function U. This means that Y is assessed via $E[U(Y)]$, where it is supposed that the expectation is well-defined. Thus, given two random costs Y and Y_1, the decision maker will prefer to pay Y_1 when $E[U(Y_1)] < E[U(Y)]$, and will be indifferent between both costs if $E[U(Y)] = E[U(Y_1)]$; observe that the preferen-

ces of the controller do not change when an affine transformation with positive slope is applied to U. In the sequel, all the utilities in the discussion are supposed to be continuous and strictly increasing functions on $[0,\infty)$.

Now, suppose that a decision maker with utility function U receives the offer to avoid the random cost Y by paying a fixed amount c. In this case, the offer will be definitively accepted if $U(c) < E[U(Y)]$ and will be refused when $U(c) > E[U(Y)]$. The threshold value c^* satisfying $U(c^*) = E[U(Y)]$—so that the decision maker is indifferent between paying the fixed amount c^*or the random cost Y—is the *certainty equivalent of Y with respect to U*.

Definition 2.1. [Certainty Equivalent.] *Let U be a utility function defined on* $[0,\infty)$ *If Y is a random cost, the certainty equivalent of Y is the number* $\mathrm{E}_U(Y)$ *such that*

$$U(\mathrm{E}_U(Y)) = E[U(Y)].$$

According to this definition, when the controller faces a random cost Y, he/she will (gladly) pay $\mathrm{E}_U(Y)$ in order to avoid the random cost Y ; note that the certainty equivalent is explicitly given by

$$\mathrm{E}_U(Y) = U^{-1}[E[U(Y)]]. \qquad (2.2)$$

Since U is strictly increasing, the inverse function U^{-1} exists and then $\mathrm{E}_U(Y)$ is well-defined if $U(Y)$ has finite expectation, as it is the case when Y takes values on a compact interval contained in $[0,\infty)$, a condition that is supposed to hold for all of the random costs Y under consideration. The certainty equivalent $\mathrm{E}_U(Y)$, represents the controller's assessment of Y in terms of a single number and may be thought of as *a kind of average* of Y in terms of the preferences of the decision maker.

Example 2.1. Let Y be a random variable taking values in a compact interval contained in $[0,\infty]$.

(i) For each $x \geq 0$, let the power utility U_γ be given by

$$U_\gamma(x) = x^\gamma,$$

where $\gamma > 0$. In this case $U_\gamma^{-1}(y) = x^{1/\gamma}$ and then

$$\varepsilon_{U_\gamma}(Y) = U_\gamma^{-1}\left(E\left[U_\gamma(Y)\right]\right) = \left(E\left[Y^\gamma\right]\right)^{1/\gamma} = \|Y\|_\gamma$$

so that $\mathrm{E}_{U_\gamma}(Y)$ is the usual γ-mean of Y . Note that

$U_1(x) = x$ is the idenity function and $\mathrm{E}_{U1}(Y) = E[Y]$ is the usual expectation of Y ;

$U_2(x) = x^2$, and $\mathrm{E}_{U2}(Y) = E[Y^2]^{1/2}$ is the quadratic mean of Y .

(ii) The logarithmic utility is given by

$$U_L(x) = \log(x),\ x \geq 0.$$

In this case, $U_L^{-1}(y) = e^y$ and

$$\mathrm{E}UL(Y) = UL{-}1\ (E[UL(Y)]) = eE[\log(Y)]$$

is the logarithmic mean of Y.

(iii) Consider now the utility U given by

$$U(x) = (x - a)^3,$$

where a is a positive number. In this case $U^{-1}(y) = a + y^{1/3}$ and then the corresponding certainty equivalent is given by.

$$\varepsilon_u(Y) = U^{-1}\left(E\left[U(Y)\right]\right) = a + \left(E\left[(Y-a)^3\right]\right)^{1/3}$$

The above utilities above are widely used in economics (Stokey and Lucas, 1989).

3. Risk-Aversion and Attraction

The attitude of a controller before a random cost Y is determined by its certainty equivalent, which as already mentioned, is a kind of average. A decision-maker with utility function U is *risk-neutral* if

$$\mathrm{E}_U(Y) = E[Y]$$

for every random cost Y. In this case, the certainty equivalent has a physical interpretation which does not depend on the observer, namely, $E[Y]$ is the average of the observed values of Y in a long-series of identical random experiments generating the random cost. By comparing the certain equivalents of a controller with $E[Y]$, a classification of the attitude before a random cost is obtained.

Definition 3.1. [Risk Aversion and Attraction.] *Consider a decision-maker with utility function* $U{:}[0,\infty) \to \mathrm{IR}$ *and let* $I \subset [0,\infty)$.

(i) The controller is risk-averse on I if

$$\mathrm{E}_U(Y) \geq E[Y]$$

for every random variable Y taking values on I with probability 1.

(ii) The controller is risk-seeking I if

$$\mathrm{E}_U(Y) \leq E[Y] \qquad \textit{when } P[Y \in I] = 1.$$

(iii) The risk-premium associated to Y is given by

$$\Delta_U(Y) = \mathrm{E}_U(Y) - E[Y].$$

In words, the controller is risk-averse (resp. risk-seeking) on an interval I if his/her assessment of a random cost Y taking values on I is higher (resp. lower) than $E[Y]$. Of course, the controller knows that $E[Y]$ is the average of Y in a long series of identical trials, but also realizes that in a specific instance the value attained by Y does not generally coincide with $E[Y]$. When a controller is risk-averse, he/she is 'afraid' of the occurrence of costs exceeding the expected value, whereas if the controller is risk-

seeking then the possible occurrence of a value less that $E[Y]$ is more relevant for his/her perceptions. For instance, consider the owner of an expensive brand new car paying \$500 for an insurance policy guaranteeing that, in case of a crash in the next year, he/she will receive an identical vehicle. The cost of the car is \$300,000 and the owner feels that there is a small probability equal to 0.001 of participating in a crash. What the owner foresees for the next year, is a random cost Y that can take the values \$0 and \$300,000 with probabilities 0 and 0.001, respectively, so that $E[Y] = \$300$; however, \$500 were gladly paid to avoid facing the random cost Y, so that Y is assessed higher than its expectation $E[Y]$, that is, $E_U(Y) \geq \$500 > E[Y]$, indicating that the owner is risk-averse. On the other hand, if a \$200 insurance policy is rejected by the owner of the car then $E_U(Y) < \$200 < E[Y]$, indicating that the owner is risk-seeking.

Recalling that all of the utility functions considered in the paper are increasing,Definitions 2.1 and 3.1 together yield that a controller with utility function U is risk-averse on I if

$$E[U(Y)] \geq U(E[Y]) \text{ when } P[Y \in I] = 1,$$

a requirement that, by Jensen's inequality, is equivalent to the *convexity* of U on the interval I. Similarly, the controller is risk-seeking on I whenever

$$E[U(Y)] \leq U(E[Y]) \text{ if } P[Y \in I] = 1,$$

a relation that is valid exactly when U is *concave* on I. On the other hand, the equality $E_U(Y) = E[Y]$ holds for every random cost taking values in I if, and only if, the controller is both riskaverse and risk-seeking, that is, when U is a linear function, an, without loss of generality, in this case it can be assumed that U is the identity function.

Example 3.1. *The risk-aversion and attraction will be analyzed for each one of the utilities in Example 2.1.*

(i) For each $\gamma > 0$, the power utility $U_\gamma(x) = x^\gamma$ satisfies that

$$U'_\gamma(x) = \gamma x^{\gamma-1}, \text{ and } U''_\gamma(x) = \gamma(\gamma-1)x^{\gamma-2}, x > 0.$$

Therefore,

- If $\gamma < 1$, then $U''_\gamma(x)$ is always negative, and U_γ is concave on $[0,\infty)$, so that a controller with this utility function is risk-seeking;
- If $\gamma > 1$, then $U''_\gamma(x) > 0$ for every $x > 0$. It follows that U_γ is convex on $[0,\infty)$, indicating that U_γ pertains to a risk-averse controller.

Of course, when $\gamma = 1$, so that $U_\gamma(x) = x$, the utility function is both convex and concave, and the controller is risk-neutral.

(ii) The logarithmic utility $U_L(x) = \log(x)$ satisfies that

$$U'_L(x) = 1/x, \text{ and } U''_L(x) = -1/x^2, x > 0.$$

Thus, U_L is concave on $[0,\infty)$ and pertains to a risk-seeking controller.

(iii) For a positive number a, the utility $U(x) = (x-a)^3$ satisfies

$$U^0(x) = 3(x-a)^2, \text{ and } U^{00}(x) = 6(x-a), x \geq 0,$$

and then U is concave on $[0,a]$ and convex on $[a,\infty)$, Thus, a controller with utility function U is risk-seeking in the interval $[0,a]$ and risk-averse on $[a,\infty)$.

4. Average Criteria

In this section, the (long-run) average cost criterion associated to a given utility is introduced, and a characterization of the risk-neutral average index is presented in terms of the optimality equation. Let M be the MDP introduced in Section 2, and suppose that a controller with utility function $U \in \mathrm{U}$ drives the system using a policy $\pi \in \mathrm{P}$ starting at $X_0 = x$. In this context $J_{U,n}(\pi,x)$ stands for the certainty equivalent of the total cost $\sum_{t=0}^{n-1} C(X_t, A_t)$ incurred before time n > 0, that is,!#!

$$J_{U,n}(\pi,x) = U^{-1}\left(E_x^\pi\left[U\left(\sum_{t=0}^{n-1} C(X_t,A_t)\right)\right]\right); \quad (4.1)$$

see (2.2). With this notation, the (long-run superior limit) U-average cost at state x under policy π is given by

$$J_U(\pi,x) := \limsup_{n\to\infty} \frac{1}{n} J_{U,n}(\pi,x), \quad (4.2)$$

and the corresponding optimal value function is specified as

$$J_{U^*}(x) := \inf_{\pi \in \mathrm{P}} J_U(\pi,x), x \in S, \quad (4.3)$$

whereas a policy $\pi^* \in \mathrm{P}$ is U-average optimal if $J_U(\pi^*,x) = J_{U^*}(x)$ for each $x \in \mathrm{S}$. In certain sense, the criterion (4.2) represents a pessimistic point of view, since it measures the performance of the policy π in terms of the largest (worst) limit point of the averages $J_{U,n}(\pi,x)/n$. The optimistic perspective to the average index is given by the inferior limit U-average criterion specified by

$$J_U-(\pi,x) := \liminf_{n\to\infty} \frac{1}{n} J_{U,n}(\pi,x), \quad (4.4)$$

with corresponding optimal value function

$$J_{U^*}-(x) := \inf_{\pi \in \mathrm{P}} J_U-(\pi,x), x \in S \quad (4.5)$$

note that (4.2)—(4.5) immediately yield that

$$J_{U^*-}(\cdot) \le J_{U^*}(\cdot). \tag{4.6}$$

As it will be shown below, for the utilities analyzed in Examples 2.1 and 3.1, the equality holds in the above display under the following communication condition.

Assumption 4.1. *For each stationary policy f the state space is communicating, that is, given x,y ∈ S, there exists a positive integer n ≡ n(x,y) such that* $P_x^f[X_n = y] > 0$.

Under this condition the Markov chain induced by a stationary policy f has an invariant distribution ρ_f:S → (0,1], that is, $^{P}_{x\in S}\rho_f(x) = 1$ and $^{P}_{x\in S}\rho_f(x)p_{xy}(f(x)) = \rho_f(y)$ for each y ∈ S. In this case, the classical ergodic theorem yields that for every initial state $X_0 = x$,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n-1}C\left(X_t,A_t\right)=\sum_{y\in S}Pf(y)C\left(y,f(y)\right)=:\alpha_f,P_x^f$$

-a.s. (4.7)

The average criteria in (4.2) and (4.4) have been widely studied in two cases: When the utility function is exponential, that is, $U(x) = e^{\lambda x}$ for some λ 6= 0, or when $U(x) = x$, which corresponds to a risk-neutral controller. In this latter case the subindex U will not be explicitly indicated in (4.1)–(4.6), and the analysis of the corresponding risk-neutral average criteria is based on the following result (Pueterman, 1994).

Theorem 4.1. *Under Assumptions 2.1 and 4.1 the following assertions (i)-(iv) are valid:*
(*i*) There exist $g \in$ IR as well as a function h:S → IR such that the following (risk-neutral average cost) optimality equation holds:

$$g+h(x)=\min_{a\in A(x)}\left[C(x,a)+\sum_{y\in S}P_{xy}(a)h(y)\right],\quad x\in S, \tag{4.8}$$

(*ii*) The risk-neutral superior and inferior average criteria render the same optimal value function, and the optimal average cost is equal to g:

$$J_{*-}(x)=J^*(x)=g,\quad x\in S,$$

where $J_{*-}(x)$ and $J^*(x)$ are given in (4.3) and (4.5) with the identity function instead of U.

(*iii*) There exists a stationary policy f ∈ IF satisfying

$$g + h(x) = C(x,f(x)) + {}^{X}_{y\in S}p_{xy}(f(x))h(y),\ x \in S, \tag{4.9}$$

and such a policy is optimal with respect to the superior and inferior average cost criteria, that is,

$$J^*(x)=J(x;f)=g=J_-(x;f)=J^*_-(x),\quad x\in S. \tag{4.10}$$

Note that (4.10) is equivalent to the following relations:

$$g\le\liminf_{n\to\infty}\frac{1}{n}E_x^{\pi}\left[\sum_{k=0}^{n-1}C\left(X_t,A_t\right)\right],\ \ x\in S\ \ \pi\in P$$

$$g=\lim_{n\to\infty}\frac{1}{n}E_x^{f}\left[\sum_{k=0}^{n-1}C\left(X_t,A_t\right)\right],\ \ x\in S. \tag{4.11}$$

According to Theorem 4.1, a risk-neutral average optimal policy can be always found in the class IF of stationary policies. For some models, for instance in inventory theory (Bertsekas, 2004), it can be frequently determined *a priori* that the optimal stationary policy has a special structure, and the quest of an optimal policy can be restricted to a 'small' subset IF^0 of IF. In this case, instead of finding a solution (g,h(·)) of the *nonlinear* optimality equation (4.8) to determine the optimal policy $f \in$ IF in (4.10), it may be interesting (and more efficient) to compute the risk-neutral average cost associated to each policy φ ∈ IF^0, and then pick the one with smallest average cost. Note that for φ ∈ IF, (4.7) and the bounded convergence theorem together imply that

$$J^*(x,\phi)=\lim_{n\to\infty}\frac{1}{n}E_x^{\phi}\left[\sum_{k=0}^{n-1}C\left(X_t,A_t\right)\right]=\sum_{x\in S}\rho_\phi(x)\,C\left(x,\phi(x)\right),$$

so that $J(\cdot,\varphi)$ is determined in terms of ρ_φ, the invariant distribution of the Markov chain associated to φ. The following result shows that ρ_φ can be found using any computing program that solves a linear system of equations with *nonsingular coefficient matrix.*

Theorem 4.2. *Given* φ ∈ IF, *let P be the transition matrix of the Markov chain associated to* φ, *that is, P* = $[p_{xy}(\varphi(x)]_{x,y\in S}$, *and let the row vector* ρ = (ρ_x, x ∈ S) *be the unique invariant distribution of the matrix P, so that* ρP = ρ.
(i) The matrix I–P +J is invertible, where J is the square matrix on S with all of its components are equal to 1.
(ii) The invariant distribution ρ *is the unique solution to the linear system*

$$x[I - P + J] = 1l$$

where 1l *is the row vector with all of its components equal to 1.*

Proof. By Assumption 4.1 the matrix P is communicating, so that the Perron-Frobenious theorem ensures that the (left) Kernel of the $I - P$ is generated by ρ, that is,

$$\mathrm{x}[I - P] = 0 \text{ if and only if } \mathrm{x} = t\rho \text{ for some } t \in \mathrm{IR}. \qquad (4.12)$$

Also, note that the definitions of J and 1l yield that $\mathrm{xJ} = s(\mathrm{x})1\mathrm{l}$, where $s(\mathrm{x}) = {}^{\mathrm{X}}\mathrm{x}_{\mathrm{i}}$.

i

(i) Observe that

$$\mathrm{x}[\mathrm{I} - \mathrm{P} + J] = \mathrm{x}[\mathrm{I} - \mathrm{P}] + \mathrm{xJ} = \mathrm{x}[\mathrm{I} - \mathrm{P}] + s(\mathrm{x})1\mathrm{l}. \qquad (4.13)$$

Now, combine this relation with the equality $[\mathrm{I}-\mathrm{P}]1\mathrm{l}^0 = 0$ (which occurs because P is a stochastic matrix), to obtain

$$\mathrm{x}[I - P + J]1\mathrm{l}^0 = s(\mathrm{x})1\mathrm{l}1\mathrm{l}^0.$$

Next, suppose that $\mathrm{x}[\mathrm{I} - \mathrm{P} + \mathrm{J}] = 0$. In this case the above display yields that $\mathrm{s}(\mathrm{x}) = 0$, so that(4.13) implies that $\mathrm{x}[\mathrm{I} - \mathrm{P}] = 0$, a relation that via (4.12) leads to $\mathrm{x} = \mathrm{t}\rho$ for some $\mathrm{t} \in \mathrm{IR}$; since $0 = \mathrm{s}(\mathrm{x}) = \mathrm{s}(\mathrm{t}\rho) = \mathrm{t}$, it follows that $\mathrm{x} = \mathrm{t}\rho = 0$. In short,

$$\mathrm{x}[I - P + J]1\mathrm{l}^0 = 0 \Rightarrow \mathrm{x} = 0,$$

and then the matrix $[I - P + J]$ is invertible.

(*ii*) Since $\rho = \rho P$ and $s(\rho) = 1$, (4.13) shows that $\rho[I - P + J] = 1\mathrm{l}$; since $\mathrm{I} - \mathrm{P} + \mathrm{J}$ is invertible, by part (i), it follows that ρ is the unique solution of the equation $x[I - P + J] = 1\mathrm{l}$.

In the following section, the average cost criteria corresponding to the utilities in Example 2.1 and 3.1 will be studied, and the analysis will be based on the following result which, together with (4.7), shows that the relations (4.11) remain valid if the expected averages are replaced by observed averages along the sample trajectories of the state-action process.

Theorem 4.3. Under Assumptions 2.1 and 4.1 the following assertions (i) and (ii) holds:

(i) For each $\mathrm{x} \in \mathrm{S}$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} C(X_t, A_t) = g, \quad P_x^f\text{-a.s.}$$

where f is the stationary policy in (4.9).

(ii) For every $\pi \in \mathrm{P}$ and $\mathrm{x} \in \mathrm{S}$,

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} C(X_t, A_t) \geq g, \quad P_x^\pi\text{-a.s.}$$

Via the bounded convergence theorem, the first part follows combining the ergodic property (4.7) with the risk-neutral average optimality of the policy f. As for the second assertion, a proof can be found in Araphostatis *et al.* (1996).

5. Equality of Optimal Average Cost Functions

In this section the optimal average cost functions J_{U^*} and $J_{U^*_-}$ will be determined for each one of the utilities in Example 2.1. As already noted, such utilities represent different assessments of a random costs. However, the rather surprising conclusion stated below establishes that both the superior and inferior average optimal value functions J_{U^*} and $J_{U^*_-}$ coincide with the optimal risk-neutral average cost. This conclusion is stated in the corollary at the end of the section, and relies on the following result.

Theorem 5.1. Let U be any one of the utilities in Example 2.1, and let $(g,h(\cdot))$ be a solution of the risk-neutral average cost optimality equation (4.8).

Let $\mathrm{x} \in \mathrm{S}$ be arbitrary. Under Assumptions 2.1 and 4.1,

$$J_{U_-}(\pi,x) \geq g, \quad \pi \in \mathrm{P}, \qquad (5.1)$$

and

$$J_U(f,x) = g, \qquad (5.2)$$

where f is the stationary policy in (4.9).

Proof. Keeping in mind that every utility U in Example 2.1 is continuous and strictly increasing on the nonnegative ray, it follows that for every bounded sequence $(a_n) \subset [0,\infty)$

$$\liminf_{n\to\infty} U(a_n) = U\left(\liminf_{n\to\infty} a_n\right), \qquad (5.3)$$

whereas if (a_n) is convergent, then

$$\lim_{n\to\infty} U(a_n) = U\left(\lim_{n\to\infty} a_n\right), \qquad (5.4)$$

Recall now that, when the system is driven by the policy π and $X_0 = x$ is the initial state, $J_{U,n}(\pi,x)$ is the certainty equivalent of $\sum_{k=0}^{n-1} C(X_t, A_t)$ with respect to the utility U, so that

$$U\left(J_{u,n}(\pi,x)\right) = E_x^\pi\left[U\left(\sum_{t=0}^{n-1} C(X_t, A_t)\right)\right]; \qquad (5.5)$$

see (4.1). Now, to establish the desired conclusions, a separate argument for each one of the utilities in Example 2.1 will be presented.

(a) Let $U = U_\gamma$, the power utility with parameter γ. In this case (5.5) is explicitly given by

$$\left(J_{U,n}(\pi,x)\right)^\gamma = E_x^\pi\left[\left(\sum_{t=0}^{n-1} C\left(X_t, A_t\right)\right)^\gamma\right].$$

Dividing both sides of this equality by n^γ it follows that

$$\left(\frac{J_{U,n}(\pi,x)}{n}\right)^\gamma = E_x^\pi\left[\left(\frac{\sum_{t=0}^{n-1} C\left(X_t, A_t\right)}{n}\right)^\gamma\right] \qquad (5.6)$$

an equality that, after taking the inferior limit as n goes to ∞. leads to

$$\begin{aligned}\liminf_{n\to\infty}\left(\frac{J_{U,n}(\pi,x)}{n}\right)^\gamma &= \liminf_{n\to\infty} E_x^\pi\left[\left(\frac{\sum_{t=0}^{n-1} C\left(X_t, A_t\right)}{\mathrm{n}}\right)^\gamma\right] \\ &\geq E_x^\pi\left[\liminf_{n\to\infty}\left(\frac{\sum_{t=0}^{n-1} C\left(X_t, A_t\right)}{n}\right)^\gamma\right]\end{aligned}$$

where the second inequality is due to Fatou's lemma. From this point, (4.13) and Theorem 4.3(ii) together yield that

$$\left(\liminf_{n\to\infty}\frac{J_{U,n}(\pi,x)}{n}\right)^\gamma \geq E_x^\pi\left[\left(\liminf_{n\to\infty}\frac{\sum_{t=0}^{n-1} C\left(X_t, A_t\right)}{n}\right)^\gamma\right] \geq E_x^\pi\left[(g)^\gamma\right] = g^\gamma$$
$$n\to\infty$$

,

and then

$$J_{U}-(\pi,x) = \liminf_{n\to\infty}\frac{J_{U,n(\pi,x)}}{n} \geq g,$$

establishing (5.1). Now, set $\pi = f$ in (5.6) to obtain

$$\left(\frac{J_{U,n}(f,x)}{n}\right)^\gamma = E_x^f\left[\left(\frac{\sum_{t=0}^{n-1} C\left(X_t, A_t\right)}{n}\right)^\gamma\right];$$

taking the limit as $n \to \infty$ in both sides of this equality, Theorem 4.3(i) and the bounded convergence theorem together imply that $\lim_{n\to\infty}(J_{U,n}(f,x)/n)^\gamma = g^\gamma$. It follows that $(J_{U,n}(f,x)/n)$ is a convergent sequence, and that $g = \lim_{n\to\infty} J_{U,n}(f,x) = J(f,x)$; see (4.2). This completes the proof for a power utility.

(b) Let $U = \log(x)$, the logarithmic utility. In this case (5.5) becomes

$$\log\left(J_{U,n}(\pi,x)\right) = E_x^\pi\left[\log\left(\sum_{t=0}^{n-1} C\left(X_t, A_t\right)\right)\right],$$

and adding $-\log(n)$ to both sides of this equality it follows that

$$\log\left(J_{U,n}(\pi,x)/n\right) = E_x^\pi\left[\log\left(\sum_{t=0}^{n-1} C\left(X_t, A_t\right)/n\right)\right]; \qquad (5.7)$$

from this point, taking the inferior limit as n goes to ∞, Fatou's lemma yields to

$$\begin{aligned}\liminf_{n\to\infty}\log\left(J_{U,n}(\pi,x)/n\right) &= \liminf_{n\to\infty} E_x^\pi\left[\log\left(\sum_{t=0}^{n-1} C\left(X_t, A_t\right)/n\right)\right] \\ &\geq E_x^\pi\left[\liminf_{n\to\infty}\log\left(\frac{\sum_{t=0}^{n-1} C\left(X_t, A_t\right)}{n}\right)\right].\end{aligned}$$

Combining (4.13) with Theorem 4.3(ii) it follows that

$$\log\left(\liminf_{n\to\infty}\frac{J_{U,n}(\pi,x)}{n}\right) \geq E_x^\pi\left[\log\left(\liminf_{n\to\infty}\frac{\sum_{t=0}^{n-1} C\left(X_t, A_t\right)}{n}\right)\right] \geq E_x^\pi\left[\log(g)\right] = \log(g)$$

,

and then

$$J_{U-}(\pi,x) = \liminf_{n\to\infty}\frac{J_{U,n}(\pi,x)}{n} \geq g,$$

completing the proof of (5.1). Next, take $\pi = f$ in (5.7) to obtain

$$\log\left(\frac{J_{U,n}(f,x)}{n}\right) = E_x^f\left[\log\left(\frac{\sum_{t=0}^{n-1} C\left(X_t, A_t\right)}{n}\right)\right];$$

after taking the limit as $n \to \infty$ in both sides of this equality, via Theorem 4.3(i) and the bounded convergence theorem it follows that $\lim_{n\to\infty}\log(J_{U,n}(f,x)/n) = \log(g)$, which is equivalent to $g = \lim_{n\to\infty} J_{U,n}(f,x) = J(f,x)$, concluding the argument for the logarithmic utility.

(c) Let $U = (x - a)^3$. In this framework, the equality (5.5) establishes that

$$\left(J_{U,n}(\pi,x)-a\right)^3 = E_x^\pi\left[\left(\sum_{t=0}^{n-1} C\left(X_t, A_t\right)-a\right)^3\right];$$

dividing by n^3 in both sides of this relation it follows that

$$\left(\frac{J_{U,n}(\pi,x)}{n}-\frac{a}{n}\right)^3 = E_x^\pi\left[\left(\frac{1}{n}\sum_{t=0}^{n-1} C\left(X_t, A_t\right)-\frac{a}{n}\right)^3\right], \qquad (5.8)$$

a relation that, via Fatou's lemma, leads to

$$\liminf_{n\to\infty}\left(\frac{J_{U,n}(\pi,x)}{n}-\frac{a}{n}\right)^3=\liminf_{n\to\infty}E_x^{\pi}\left[\left(\frac{1}{n}\sum_{t=0}^{n-1}C\left(X_t,A_t\right)-\frac{a}{n}\right)^3\right]$$
$$\geq E_x^{\pi}\left[\liminf_{n\to\infty}\left(\frac{1}{n}\sum_{t=0}^{n-1}C\left(X_t,A_t\right)-\frac{a}{n}\right)^3\right].$$

Using that $x \mapsto (x-a)^3$ is increasing, Theorem 4.3(ii) implies that

$$\left(\liminf_{n\to\infty}\frac{J_{U,n}(\pi,x)}{n}\right)^3\geq E_x^{\pi}\left[\left(\liminf_{n\to\infty}\frac{\sum_{t=0}^{n-1}C\left(X_t,A_t\right)}{n}\right)^3\right]\geq E_x^{\pi}\left[\left(g\right)^3\right]=g^3,$$

so that

$$J_{U-}(\pi,x)=\liminf_{n\to\infty}\frac{J_{U,n}(\pi,x)}{n}\geq g.$$

To conclude, select $\pi = f$ in (5.8) to obtain

$$\left(\frac{J_{U,n}(f,x)}{n}-\frac{a}{n}\right)^3=E_x^{f}\left[\left(\frac{\sum_{t=0}^{n-1}C\left(X_t,A_t\right)}{n}-\frac{a}{n}\right)^3\right];$$

letting n increase to ∞, Theorem 4.3(i) and the bounded convergence theorem together imply that $\lim_{n\to\infty}(J_{U,n}(f,x)/n)^3 = g^3$, which is equivalent to $g = \lim_{n\to\infty} J_{U,n}(f,x)/n = J(f,x)$.

Corollary 5.1. *For each one of the utility functions U in Example 2.1, the following assertions (i) and (ii) hold:*
(i) For every $x \in S$, $J_{U^*}(x) = J_{U^*-}(x) = g$, where g is the optimal risk-neutral average cost.
(iii) A stationary policy f is U-average optimal if, and only if, f is risk-neutral average optimal.

Proof. (i) Combining (4.5) and (5.1), it follows that

$$J_{U^*-}(x)=\inf_{\pi\in P}J_{U-}(\pi,x)\geq g, x\in S$$

On the other hand, if f is as in (4.9), the relations (4.3) and (5.2) together yield that

$$J_{U^*}(x)\leq J_U(f,x)=g, x\in S.$$

These two last displays yield that $J_{U^*-}(\cdot)\leq J_U(\cdot)=g$, and the first assertion follows via (4.6).
Next, part (ii) follows from part (i).

6. The Cone Property

The result presented in this section can be briefly described as follows: If different utilities render the same average optimal value function, say $J(\cdot)$, and share an optimal stationary policy, then the optimal average index of any combination of those utilities also coincides with J. To state this result in a precise manner, let U_0 be a fixed (continuous and strictly increasing) utility function defined on $[0,\infty)$ and assume that the following properties hold for U_0:

$$J_{U^*0-}(x)=J_{U^*0}(x)\equiv J(x), x\in S, \tag{6.1}$$

and, for some policy $f \in$ IF,

$$J_{U_0}(f,x)=\lim_{n\to\infty}\frac{1}{n}J_{U_{0,n}}(f,x)=J(x),\quad x\in S \tag{6.2}$$

Note that, by Theorem 4.1, under Assumptions 2.1 and 4.1 these conditions are satisfied if U_0 is the identity function.

Definition 6.1. *The family* U *consists of all continuous and strictly increasing utility functions U on* $[0,\infty)$ *satisfying the following requirement:*

$$J_{U^*-}(x)=J_{U^*}(x)=J(x),\quad x\in S$$

and

$$J_U(f,x)=\lim_{n\to\infty}\frac{1}{n}J_{U,n}(f,x)=J(x),\quad x\in S, \tag{6.3}$$

f is as in (6.2).

where

THEOREM 6.1. *The family U is a cone, that is,*

$$U\in\mathrm{U}\Rightarrow cU\in\mathrm{U}\ \textit{for every}\ c>0, \tag{6.4}$$

and

$$U_1U_2\in\mathrm{U}\Rightarrow U_1+U_2\in\mathrm{U}. \tag{6.5}$$

Proof. As already noted in Section 3, the certainty equivalent of a random cost is not altered if the underlying utility function is multiplied by a positive constant. Therefore, form (4.1) it follows that if $c > 0$ then $J_{cU,n}(x) = J_{U,n}(x)$ for every state x. Hence, (4.2)–(4.5) yield that $J_{cU^*-}(\cdot) = J_{U^*-}(\cdot)$ and $J_{cU^*}(\cdot) = J_{U^*}(\cdot)$, and (6.4) follows from Definition 6.1. To establish (6.5), let $U_1, U_2 \in$ U be arbitrary, and note the following facts (a)–(c):
(a) Let $\pi \in$ P and $x \in S$ be arbitrary. With respect to $U_1 + U_2$ the inferior average cost under policy π at state x satisfies

$$J_{(U1+U2)-}(\pi,x) \geq J(x). \tag{6.6}$$

To establish this assertion, note that Definition 6.1 and the inclusions $U_1, U_2 \in \mathrm{U}$ yield that, for $i = 1,2$,

$$\liminf_{n\to\infty} \frac{1}{n} J_{U_i,n}(\pi,x) \geq J^*_{U_i-}(x) = J(x).$$

Therefore, given $\varepsilon \in (0,\mathrm{kCk})$, there exists a positive integer N such that

$$\frac{1}{n} J_{U_i,n}(\pi,x) > J(x) - \varepsilon, \quad n \geq N, \quad i = 1,2. \tag{6.7}$$

Now, consider the certainty equivalent $J_{U1+U2,n}(\pi,x)$, which satisfies

$$\begin{aligned}[U_i + U_2]\left(J_{U_1+U_2,n}(\pi,x)\right) &= E_x^{\pi}\left[[U_1+U_2]\left(\sum_{k=0}^{n-1} C(X_t,A_t)\right)\right] \\ &= E_x^{\pi}\left[U_1\left(\sum_{k=0}^{n-1} C(X_t,A_t)\right)\right] + E_x^{\pi}\left[U_1\left(\sum_{k=0}^{n-1} C(X_t,A_t)\right)\right] \\ &= U_1\left(J_{U_1,n}(\pi,x)\right) + U_1\left(J_{U_2,n}(\pi,x)\right).\end{aligned} \tag{6.8}$$

Next, observe that (6.7) yields that

$$J_{U1},n(x) \geq n[J(x) - \varepsilon] \text{ and } J_{U2},_{\mathrm{n}}(x) \geq n[J(x) - \varepsilon], \quad n \geq N.$$

These two last displays yield that, for $n \geq N$,

$$\begin{aligned}[U_i + U2](J_{U1}+_{U2,n}(\pi,x)) &\geq U_1\,(n[J(x) - \varepsilon]) + U_1\,(n[J(x) - \varepsilon]) \\ &= [U_1 + U_2](n[J(x) - \varepsilon]),\end{aligned}$$

that is,

$$J_{\mathrm{U1}}+_{\mathrm{U2,n}}(\pi,x) \geq n[J(x) - \varepsilon],$$

and then

$$k{=}0 \qquad \frac{1}{\mathrm{n}} J_{U_1+U_2,n}(\pi,x) \geq J(x) - \varepsilon, \quad n \geq N,$$

so that

$$J_{(U_1+U_2)-}(\pi,x) = \liminf_{n\to\infty} \frac{1}{n} J_{U_1+U_2,\mathrm{n}}(\pi,\mathrm{x}) \geq J(x) - \varepsilon,$$

and (6.6) follows, since ε is an arbitrary number in $(0,\mathrm{kCk})$.

(b) It will be shown that

$$\lim_{n\to\infty} \frac{1}{n} J_{(U_1+U_2),n}(f,x) = J(x). \tag{6.9}$$

To establish this assertion, set $\pi = f$ in (6.8) to obtain

$$[Ui + U2](JU1+U2,n(f,x)) = U1(JU1,n(f,x)) + U2(JU2,n(f,x)).$$

On the other hand, since $U_1, U_2 \in \mathrm{U}$ the requirement (6.3) yields that, for each $\varepsilon \in (0,\mathrm{kCk})$, there exists a positive integer N such that

$$J(x) - \varepsilon \leq \frac{1}{n} J_{Ui,n}(f,x) \leq J(x) + \varepsilon, \quad i = 1,2, \quad n \geq N.$$

Recalling the functions U_1 and U_2 are increasing, this last property and the previous display together yield that, for $n \geq N$

$$\begin{aligned}&[U_i + U_2](J_{U1+U2,n}(f,x)) \geq U_1(n[J(x) - \varepsilon])) + U_2(n[J(x) - \varepsilon])) \\ &= [U_1 + U_2](n[J(x) - \varepsilon])\end{aligned}$$

as well as

$$\begin{aligned}&[U_i + U_2](J_{U1+U2,n}(f,x)) \leq U_1(n[J(x) + \varepsilon])) + U_2(n[J(x) + \varepsilon])) \\ &= [U_1 + U_2](n[J(x) + \varepsilon]).\end{aligned}$$

Therefore, since $U_1 + U_2$ is strictly increasing, for each $n \geq N$,

$$n[J(x) - \varepsilon] \leq J_{U1+U2,n}(f,x) \leq n[J(x) + \varepsilon],$$

and then

$$J(x) - \varepsilon \leq \frac{1}{n} J_{U_1+U_2,n}(f,x) \leq J(x) + \varepsilon, \quad n \geq N.$$

Since $\varepsilon \in (0,\mathrm{kCk})$ is arbitrary, this relation yields that (6.9) holds. To conclude, observe that (6.6) implies that, for every $x \in S$,

$$J^*_{(U_1+U_2)-}(x) = \inf_{\pi\in P} J_{(U_1+U_2)-}(\pi,x) \geq J(x),$$

whereas (6.9) yields that

$$J(x) = JU1+U2(f,x) \geq JU^*1+U2(x);$$

see (4.2) and (4.3). These two last display together imply that

$$J^*_{(U_1+U_2)-}(x) \geq J(x) \geq J^*_{U_1+U_2}(x);$$

since, $J^*_{(U_1+U_2)-}(x) \leq J^*_{U_1+U_2}(x)$ it follows that

$$J^*_{(U_1+U_2)-}(x) = J^*_{U_1+U_2}(x) = J(x), \quad x \in S.$$

By Definition 6.1, this relation and (6.9) together imply that $U_1 + U_2 \in \mathrm{U}$.

REFERENCES

Arapostathis, A.; V. K. Borkar, E. Fernández-Gaucherand, M. K. Ghosh, and S. I. Marcus (1993). Discrete-time controlled Markov processes with average cost

criterion: A survey, *SIAM Journal on Control and Optimization*, **31**, 282–344.

Bäuerle, N. and Reider, U. (2011), *Markov Decision Processes with Applications to Finance*, Springer, New York.

Bäuerle, N. and Reider, U. (2013), More Risk-Sensitive Markov Decision Processes, *Mathematics of Operations Research*, To appear (Available on line from INFORMS web page).

Bertesekas, D. P. (2004). Dynamic Programming and Optimal Control, Athena Scientific, Boston.

Hernández-Lerma, O. (1989), Adaptive Markov Control Processes, Springer, New York.

Howard, A. R. and Matheson, J. E. (1972), Risk-sensitive Markov decision processes, *Management Sciences*, 18, 356-369.

Sennott, L. I. (1998), Stochastic Dynamic Programming and the Control of Queueing Systems, Willey, New York.