Kybernetika

Raúl Montes-de-Oca; Elena Zaitseva
About stability of risk-seeking optimal stopping

Kybernetika, Vol. 50 (2014), No. 3, 378–392

Persistent URL: http://dml.cz/dmlcz/143881

KYBERNETIKA — VOLUME *50* (2014), NUMBER 3, PAGES 378–392

ABOUT STABILITY OF RISK-SEEKING OPTIMAL STOPPING

Raúl Montes-de-Oca and Elena Zaitseva

We offer the quantitative estimation of stability of risk-sensitive cost optimization in the problem of optimal stopping of Markov chain on a Borel space X. It is supposed that the transition probability $p(\cdot|x)$, $x \in X$ is approximated by the transition probability $\widetilde{p}(\cdot|x)$, $x \in X$, and that the stopping rule $\widetilde{f}_*$, which is optimal for the process with the transition probability $\widetilde{p}$ is applied to the process with the transition probability p. We give an upper bound (expressed in term of the total variation distance: $\sup_{x\in X} \|p(\cdot|x) - \widetilde{p}(\cdot|x)\|$) for an additional cost paid for using the rule $\widetilde{f}_*$ instead of the (unknown) stopping rule f_* optimal for p.

Keywords: discrete-time Markov process, risk-seeking expected total cost, optimal stopping rule, stability index, total variation metric

Classification: 60G40, 62L15

1. MOTIVATION

In this paper we continue the study initiated in [24] of *stability* (or "robustness") of the optimal stopping problem. The "quantitative estimation of stability of optimal control" is understood in the same sense as, for example, in the works [10, 11, 12, 19], where the stability of some classes of Markov decision processes (MDP's) was investigated, and "stability inequalities" were obtained for discounted and average criteria on an infinite time interval.

Here we consider the problem of minimization of the *risk-sensitive expected total cost*:

$$W(\tau_f) := \frac{1}{\mu} \log \left[E \left(\exp \mu \left(\sum_{t=0}^{\tau_{f-1}} c_0(x_t) - r(x_{\tau_f}) \right) \right) \right], \tag{1.1}$$

where τ_f *is a stopping time corresponding to a stopping rule* f, c_0 and r are a current cost and a terminal reward, respectively, and μ *is a negative sensitivity parameter.* The last means that minimizing the functional $W(\tau_f)$ over f, we deal with a *risk-seeking situation.* In spite of the fact that the technique of this work is resembling that employed in [24], the use of (1.1) instead of the standard in optimal stopping theory criterion:

$$E \left(\sum_{t=0}^{\tau_{f-1}} c_0(x_t) - r\big(x_{\tau_f}\big) \right), \tag{1.2}$$

DOI: 10.14736/kyb-2014-3-0378

brings new aspects to stability estimation of optimal stopping rules (due to multiplicative character of the cost functional).

As far as we know, there are a few papers on the existence of stationary optimal policies for MDP's with a risk-sensitive expected total cost, and none of them covers the kind of MDP's considered in this paper (see e.g. [1, 6, 23] for finite MDP's).

On the other hand, during the last 15 years, a significant number of works appears where risk-sensitive average (per unit of time) control optimization is studied (mostly with a countable phase space, see, e.g. [3, 4, 5, 14, 16, 17]). In a part of these papers the emphasis on the risk-averse case is made . Probably this case is the most important in financial applications, (see e.g. [2, 20, 22]). Nevertheless some people in some situations prefer "risky behaviour".(This can be observed, for instance, in the "quiz show stopping problem", see [2, Ch. 10].)

The "stability estimation problem" considered in this paper is interpreted as follows (compare it with [7, 10, 11, 12, 19, 24]). Let the transition probability p of the Markov process $\{x_t\}$ for which one optimizes a stopping rules:

- either be unknown, and approximated by some statistical estimation $\widetilde{p}$;
- or be known, but approximated by a certain more simple transition kernel $\widetilde{p}$.

The transition probability $\widetilde{p}$ generates a Markov process $\{\widetilde{x}_t\}$, which can be considered as an approximation to $\{x_t\}$.

In Section 2 we give some preliminaries on the risk-seeking optimal stopping. In Section 3 we introduce assumptions under which there exist optimal stopping rules f_* and $\widetilde{f}_*$ for $\{x_t\}$ and $\{\widetilde{x}_t\}$, respectively (which minimize functionals as in (1.1)). It is supposed that $\widetilde{f}_*$ is used as an available approximation to f_*, that is, the stopping rule $\widetilde{f}_*$ is applied to the process $\{x_t\}$. (Instead of an unavailable optimal rule f_*.)

Let $\|\cdot\|$ be the *the total variation norm* of a signed bounded measure defined on the state space taken into account. Our aim is to bound (and to offer conditions to be able to do this) in terms of

$$d(p,\widetilde{p}) := \sup_{x\in X} \|p(\cdot|x) - \widetilde{p}(\cdot|x)\|, \tag{1.3}$$

the following *stability index*

$$\Delta := W(\tau_{\widetilde{f}_*}) - W(\tau_{f_*}) \equiv \ W(\tau_{\widetilde{f}_*}) - \inf_f W(\tau_f) \geq 0, \tag{1.4}$$

where W is calculated by (1.1) for the "original process" $\{x_t\}$. In Section 4 we will give simple examples of *unstable optimization problems*, when $d(p,\widetilde{p}) \to 0$, while in (1.4) $\Delta \geq M > 0$. (Here M can be arbitrarily large, but fixed).

Finally, Section 5 provides the proofs of the main results in the article.

2. RISK-SEEKING OPTIMAL STOPPING AND ITS APPROXIMATION

Let on a Borel space $(X, \mathcal{B}_X)$ be defined two discrete-time Markov processes $\{x_t\} \equiv \{x_t, t = 0, 1, \dots\}$ and $\{\widetilde{x}_t\} \equiv \{\widetilde{x}_t, t = 0, 1, \dots\}$ with the corresponding transition probabilities:

$$p = p(B|x); \quad \widetilde{p} = \widetilde{p}(B|x), \quad x \in X, \quad B \in \mathcal{B}_X.$$

For any fixed *initial state* $x \in X$, let P_x and $\widetilde{P}_x$ denote the distributions of $\{x_t\}$ and $\{\widetilde{x}_t\}$ on $(X^\infty, \mathcal{B}_X^\infty)$, respectively, and let E_x and $\widetilde{E}_x$ denote the corresponding expectations.

A stopping rule f is a sequence $f = \{f_0, f_1, \dots\}$ of measurable functions

$$f_n : (X^n, \mathcal{B}_X^n) \to A := \{0,1\}, (X^0 := \{x\}),$$

where *the action set* A consists of two actions:

- $a = 0$ means to continue observations of the process;
- $a = 1$ prescribes to stop the process.

Each stopping rule f, when applied to $\{x_t\}$ or to $\{\widetilde{x}_t\}$, generates the corresponding *stopping times*: τ_f and $\widetilde{\tau}_f$.

The current cost and the terminal reward are measurable nonnegative functions:

$$c_0 : X \to [0, \infty); \quad r : X \to [0, \infty),$$

where r is supposed to *be bounded*:

$$\bar{r} := \sup_{x \in X} r(x) < \infty. \tag{2.1}$$

We will see that under assumptions admitted in Section 3 we can restrict ourselves with the class $\Psi = \{f\}$ of stopping rules f such that

$$P_x(\tau_f < \infty) = P_x(\widetilde{\tau}_f < \infty) = 1, \quad \text{for all} \quad x \in X.$$

For every $f \in \bar{\Psi}$, the *risk-seeking costs* are defined as follows:

$$W(x,f) := -\frac{1}{\lambda} \log \left[E_x \exp\left\{ -\lambda \left[\sum_{t=0}^{\tau_f - 1} c_0(x_t) - r(x_{\tau_f}) \right] \right\} \right], \tag{2.2}$$

(here and throughout the article,

$$\sum_{t=0}^{-1} := 0,$$

by convention);

$$\widetilde{W}(x,f) := -\frac{1}{\lambda} \log \left[\widetilde{E}_x \exp\left\{ -\lambda \left[\sum_{t=0}^{\tau_f - 1} c_0(\widetilde{x}_t) - r\left(\widetilde{x}_{\tau_f}\right) \right] \right\} \right]. \tag{2.3}$$

Here $\lambda > 0$ is a given *sensitive parameter* (which is considered to be fixed in the rest of the article).

Optimal stopping rules f_* *and* $\widetilde{f}_*$ for the processes $\{x_t\}$ and $\{\tilde{x}_t\}$, respectively, are such that

$$W(x, f_*) = \inf_{f \in \Psi} W(x,f), \quad x \in X; \tag{2.4}$$

$$\widetilde{W}(x,\widetilde{f}_*) = \inf_{f\in\Psi} \widetilde{W}(x,f), \quad x \in X. \tag{2.5}$$

Now observe that, provided that f_* and $\widetilde{f}_*$ exist, they satisfy the following relations:

$$U(x,f_*) = \sup_{f\in\Psi} U(x,f) =: U_*(x), \quad x \in X, \tag{2.6}$$

$$\widetilde{U}(x,f_*) = \sup_{f\in\Psi} \widetilde{U}(x,f) =: \widetilde{U}_*(x), \quad x \in X, \tag{2.7}$$

where:

$$U(x,f) := E_x \exp\left\{ -\lambda \left[\sum_{t=0}^{\tau_{f-1}} c_0(x_t) - r(x_{\tau_f}) \right] \right\}, \tag{2.8}$$

and $\widetilde{U}$, defined as in (2.8), replacing $\{x_t\}$ by $\{\widetilde{x}_t\}$, and E_x by $\widetilde{E}_x$.

The right-hand sides of (2.6) and (2.7) are the corresponding *value functions* (which are finite because of boundeness of r).

Under assumptions of Section 3, the following affirmation will be proved.

Proposition 2.1. Let $f_* = \{f_*, f_*, \dots\}$, $\widetilde{f}_* = \{\widetilde{f}_*, \widetilde{f}_*, \dots\}$ be the stationary stopping rules such that for $x \in X$:

$$f_*(x) = \begin{cases} 0, & x \in X \backslash S, \\ 1, & x \in S; \end{cases} \qquad \widetilde{f}_*(x) = \begin{cases} 0, & x \in X \backslash \widetilde{S}, \\ 1, & x \in \widetilde{S}; \end{cases} \tag{2.9}$$

where

$$\begin{aligned} S &:= \left\{ x \in X : U_*(x) = \mathbf{e}^{\lambda r(x)} \right\}, \\ \widetilde{S} &:= \left\{ x \in X : \widetilde{U}_*(x) = \mathbf{e}^{\lambda r(x)} \right\}. \end{aligned} \tag{2.10}$$

Then the stopping rules f_* and $\widetilde{f}_*$ are *optimal* for $\{x_t\}$ and $\{\widetilde{x}_t\}$, respectively, (i. e. (2.4) and (2.5) or (2.6) and (2.7) hold).

Remark 2.1. The rules above dictate to stop the processes on the first entrance into the "stopping set" S (or $\widetilde{S}$, respectively). Analogous result is well-known for stopping problem with the cost functional (1.2) (see, e. g. [21]). Probably in the risk-sensitive case Proposition 2.1 can be shown directly. Anyway, we get its proof as a by-product of our assumptions and calculations related to the estimation of stability indices.

According to (2.4), (2.5) and (2.6) – (2.8) we define two such indices:

$$\Delta(x) := W(x,\widetilde{f}_*) - W(x,f_*) \geq 0, \tag{2.11}$$

$$\delta(x) := U(x,f_*) - U(x,\widetilde{f}_*) \geq 0. \tag{2.12}$$

3. ASSUMPTIONS AND RESULTS

Our assumptions are similar to those used in [24]. They demand "uniform geometric ergodicity" of processes under consideration.

Assumption 1.

(a) The processes $\{x_t\}$ and $\{\widetilde{x}_t\}$ have stationary probabilities π and $\widetilde{\pi}$.

(b) There exist constants η, $0 \leq \eta < 1$, and $M < \infty$, such that,

$$\sup_{x \in X} \|p^t(\cdot|x) - \pi(\cdot)\| \leq M\eta^t, t = 1, 2, \ldots, \text{and} \tag{3.1}$$

$$\sup_{x \in X} \|\widetilde{p}^t(\cdot|x) - \widetilde{\pi}(\cdot)\| \leq M\eta^t, t = 1, 2, \ldots, \tag{3.2}$$

where p^t and $\widetilde{p}^t$ are the t-step transition probabilities for the processes $\{x_t\}$ and $\{\widetilde{x}_t\}$, respectively, and $\|\cdot\|$ is the total variation norm of a signed bounded measure.

Let S and $\widetilde{S}$ be the sets defined in (2.10).

Assumption 2. There exist $\alpha > 0$ such that

$$\pi(S) \geq \alpha, \ \pi(\widetilde{S}) \geq \alpha, \ \widetilde{\pi}(S) \geq \alpha, \ \widetilde{\pi}(\widetilde{S}) \geq \alpha. \tag{3.3}$$

Remark 3.1.

(a) Conditions (3.1) and (3.2) can be ensured by the well-known Lyapunov type conditions (see e. g. [15, 18]).

(b) Assumption 2 is somewhat delicate by two reasons. Firstly, in general, the set S in (2.10) is supposed to be unknown (since p could be unknown). Secondly, the constants in the "stability inequalities" (3.4) and (3.7) below tend to infinity as in (3.3) $\alpha \downarrow 0$. On the other hand, the invariant probabilities π and $\tilde{\pi}$ in (3.3) depend (generally, in some untractable way) on the transition probabilities p and $\tilde{p}$, and therefore, on "the nearness parameter" $\|p(\cdot|x) - \widetilde{p}(\cdot|x)\|$ in (3.4) and (3.7).

In some particular cases (for instance as in Example 4.4 of Section 4, or for random walks on a compact group), it is possible, using the structure of the functions c_0 and r, to give some a priory bounds of α in (3.3) (which are valued for certain classes of the transition probabilities p and $\widetilde{p}$).

On the other hand, if one can establish the two last inequalities in (3.3) (for the "known" transition probability $\widetilde{p}$), then under Assumption 1, if $d(p, \widetilde{p})$ in (1.3) is small enough then, $\|\pi - \widetilde{\pi}\|$ can be bounded in terms of $d(p, \widetilde{p})$ (see Theorem 3.5 in [15]). Thus two first inequalities in (3.3) can be ensured.

In any case, the counterexamples given in the next section show that both Assumptions 1 and 2 are essential for the validity of the stability inequalities (3.4) and (3.7) below.

Let $d(p, \widetilde{p})$ be the distance defined in (1.3). Recall that stability indices $\Delta(x)$ and $\delta(x)$ were defined in (2.11) and (2.12).

Theorem 3.1. Let Assumptions 1 and 2 hold. Then for any $\lambda > 0$,

$$\sup_{x \in X} \delta(x) \leq K \mathbf{e}^{\lambda \bar{r}} d(p, \widetilde{p}), \tag{3.4}$$

where

$$K = \frac{2}{3} \alpha^{-1} N(N+1), \text{ and } N = \left[\frac{\log\left(\frac{\alpha}{2M}\right)}{\log(\eta)} \right] + 2. \tag{3.5}$$

Here, $[z]$ denotes the integer part of $z \in \mathbb{R}$.

Theorem 3.2. Under Assumptions 1 and 2, if

$$d(p, \widetilde{p}) \leq \frac{1}{2} \mathbf{e}^{\lambda(r(x) - \bar{r})}, \tag{3.6}$$

then

$$\Delta(x) \leq 2K\lambda^{-1} \mathbf{e}^{\lambda(\bar{r} - r(x))} d(p, \widetilde{p}). \qquad (3.7).$$

Remark 3.2. Example 4.3 below shows that the presence of the term growing exponentially with $\lambda \to \infty$ on the right-hand side of inequality (3.4) is justifiable. On the other hand, the appearance of a similar term on the right hand side of (3.7) is due to lack of sharpness in the estimation of logarithms in (2.2). Thus inequality (3.7) could be useful for small $\lambda > 0$, but it is poor for large λ.

Remark 3.3. The distance $d(p, \widetilde{p}) = \sup_{x \in X} \|p(\cdot|x) - \widetilde{p}(\cdot|x)\|$ ("discrepancy measure") on the right-hand side of (3.4) and (3.7) could be (at least theoretically) evaluated when $\widetilde{p}$ is "a simplifying approximation" to p. If p is considered to be unknown, and $\widetilde{p}$ is some statistical estimation of p, then again, sometimes the value of $d(p, \widetilde{p})$ can be bounded. For example, if $x_t = F(x_{t-1}, a_t, \xi_t)$, $t \geq 1$ with $\xi_1, \xi_2, \ldots,$ being i.i.d. random vectors in $\mathbb{R}^m$, and if an unknown density f_{ξ_1} is estimated by "kernel-type" empirical densities $\widehat{f}_n = \widehat{f}_n(\xi_1, \ldots, \xi_n)$, then in certain classes of densities f_{ξ_1} and functions F one can obtain that $Ed(p, \widetilde{p}) \equiv Ed(p, \widetilde{p}_n) \leq Ln^{-\gamma}$ with specific constants L and $\gamma > 0$ (see, e. g., [8]).

4. EXAMPLES AND COUNTEREXAMPLES

First, we consider two counterexamples which show that Assumptions 1 and 2 are essential in Theorems 3.1 and 3.2 in Section 3.

Example 4.1. (Similar to Example 3 in [24]) Let $X = \{0, 1, 2, 3\}$, $\epsilon \in (0, 1)$;

$$p(\{1\}|0) = p(\{2\}|1) = p(\{3\}|3) = 1, \quad p(\{3\}|2) = \epsilon, \quad p(\{2\}|2) = 1 - \epsilon;$$
$$\widetilde{p}(\{1\}|0) = \widetilde{p}(\{2\}|1) = \widetilde{p}(\{2\}|2) = \widetilde{p}(\{3\}|3) = 1,$$

and also, $c_0(0) = c_0(2) = c_0(3) = 0$; $c_0(1) = 1$, $r(0) = r(1) = r(2) = 0$; $r(3) = M + 1, M > 1$, where M is any positive number given.

Choose $x = 0$ as the initial state for $\{x_t\}$ and $\{\tilde{x}_t\}$. It is clear that the optimal stopping rule for $\{x_t\}$ is $f_*(j) = 0$, $j = 0, 1, 2$; $f_*(3) = 1$ (to stop on the first entrance in $j = 3$), and in (2.6), (2.8) $U(0, f_*) = \mathbf{e}^{\lambda M}$, and so in (2.2), (2.4), $W(0, f_*) = -M$ for every $\lambda > 0$.

On the other hand, the optimal rule for $\{\widetilde{x}_t\}$ is $\widetilde{f}_*(0) = 1$; $\widetilde{f}_*(1) = \widetilde{f}_*(2) = \widetilde{f}_*(3) = 0$ (to stop immediately, since "the rewarding state" $j = 3$ is not attainable). Therefore, for any $\lambda > 0$, $U(0, \widetilde{f}_*) = 1$, $W(0, \widetilde{f}_*) = 0$, and in view of (2.11), (2.12) for every $\epsilon \in (0, 1)$,

$$\Delta(0) = M; \qquad \delta(0) = \mathbf{e}^{\lambda M} - 1.$$

Meanwhile, denoting $p_\epsilon := p$, we get that $d(p_\epsilon, \widetilde{p}) \to 0$, as $\epsilon \to 0$. It is evident that the processes in this example do not satisfy Assumption 1.

Example 4.2. Let $X = [0, 1]$, $c_0 \equiv 0$; $r(x) = \begin{cases} x, \ x \in [0, 1), \\ 2, \ x = 1. \end{cases}$ For a given $\epsilon \in (0, 1)$, let $\{\xi_t\}$ be i.i.d $\sim U[0, 1]$, $\{\widetilde{\xi}_t\}$ be i.i.d $\sim U[0, 1 - \epsilon]$. Let us define: $x_0 = \widetilde{x}_0 = x \in X$; and for $t \geq 1$,

$$x_t = \begin{cases} \xi_t & \text{with probability} \quad (1 - \epsilon), \\ 1 & \text{with probability} \quad \epsilon/2, \\ 1 - \epsilon & \text{with probability} \quad \epsilon/2; \end{cases} \qquad \widetilde{x}_t = \begin{cases} \widetilde{\xi}_t & \text{with probability} \quad (1 - \epsilon), \\ 1 - \epsilon & \text{with probability} \quad \epsilon; \end{cases}$$

and suppose that $x_1, x_2, \ldots$ are independent, and also $\widetilde{x}_1, \widetilde{x}_2, \ldots$ are independent.

It is clear that Assumption 1 of Section 3 *is satisfied*. Then optimal for $\{\widetilde{x}_t\}$ stopping rule $\widetilde{f}_*$ is to stop on the first entrance in $\widetilde{S} = \{1 - \epsilon\}$, and f_*, optimal for $\{x_t\}$, is to stop on the first entrance into $S = \{1\}$. Taking into account (2.2)–(2.12) we get that (for any $x \in X, \lambda > 0, \epsilon \in (0, 1)$):

$$\delta(x) = \mathbf{e}^{2\lambda} - \mathbf{e}^{(1-\epsilon)\lambda}; \ \ \Delta(x) = 1 + \epsilon.$$

It is easy to verify that for these processes $d(p, \tilde{p}) \to 0$ as $\epsilon \to 0$. In this example, Assumption 2 does not hold since $\tilde{\pi}(S) = 0$.

Example 4.3. This example shows that under the conditions of Theorem 3.1 the left-hand side of inequality (3.4) could be of order $\ell \mathbf{e}^{\beta\lambda}$, as $\lambda \to \infty$ (with some $\ell, \beta > 0$).

Let $X = \{0, 1, 2, 3\}; c_0(0) = 10; c_0(1) = 10; c_0(2) = c_0(3) = 2000$; $r(0) = r(1) = r(3) = 0; r(2) = 1000$.

The process $\{x_t\}$ has the following transition probability matrix ($\epsilon \in (0, 1)$) :

$$\begin{array}{l} p(\{1\}|0) = 1; p(\{2\}|1) = \epsilon, p(\{3\}|1) = 1 - \epsilon; \ p(\{3\}|2) = 1; \\ p(\{0\}|3) = p(\{1\}|3) = p(\{2\}|3) = 1/3. \end{array} \tag{4.1}$$

The probability matrix $\widetilde{p}$ of the process $\{\tilde{x}_t\}$ is the same as in (4.1), but with $\epsilon = 0$.

Let $x = 0$ be the initial state (common for $\{x_t\}$ and $\{\tilde{x}_t\}$). Both Markov chains $\{x_t\}$ and $\{\tilde{x}_t\}$ are irreducible and nonperiodic. So, Assumption 1 holds (as well as Assumption 2).

It is easy to see that the stopping rule (optimal for $\{\tilde{x}_t\}$) $\widetilde{f}_*$ is to stop at $\widetilde{S} = \{0\}$ (immediately), while for every $\epsilon > 0$ fixed and all $\lambda > 0$ large enough, the rule f_* (optimal for $\{x_t\}$) is to stop on the first entrance into $S = \{2, 3\}$. By simple calculations we obtain in (2.12) that,

$$\delta(0) = \epsilon \mathbf{e}^{980\lambda} + (1-\epsilon)\mathbf{e}^{-20\lambda} - 1 \sim \epsilon \mathbf{e}^{980\lambda} \text{ as } \lambda \to \infty. \tag{4.2}$$

Particularly, this means that for large $\lambda > 0$ the rule $\widetilde{f}_*$ provides a poor approximation to the "really risky policy" f_* to stop at S (winning a lot with a small probability $\epsilon > 0$).

In spite of this circumstance, the considered example of stopping optimization is *stable*. Indeed, as $\epsilon \to 0$, for every *fixed* $\lambda > 0$ in (2.8),

$$U(0, f_*) = \epsilon \mathbf{e}^{980\lambda} + (1-\epsilon)\mathbf{e}^{-20\lambda} \to \mathbf{e}^{-20\lambda} < 1 = U(0, \widetilde{f}_*).$$

So, for all $\epsilon > 0$ small enough, the policy $\widetilde{f}_*$ is optimal for $\{x_t\}$ (rather than f_*), and the left-hand sides of inequalities (3.4) and (3.7) become zero.

Remark that in case (4.2) in inequality (3.7) $\Delta(0) \sim 980$. Therefore inequality (3.7) for large λ has the unreasonably large right-hand side.

Example 4.4. (Stability in an asset selling problem) The version of this example with habitual criterion (1.2) was considered in [24], where, in particular, Assumptions 1 and 2 were verified. We recall that in this example a sequence of successive offers $\{x_t\}$ forms geometrically ergodic process on the state space $X = [0, L]$ with a transition probability $p(\cdot|x)$, $x \in X$; c_0 is a positive constant, and $r(x) = x$, $x \in X$. The process $\{x_t\}$ is approximated by a sequence of i.i.d. random variables with a common distribution $\widetilde{F}$. Since Assumptions 1 and 2 hold true one can apply for this example inequalities (3.4) and (3.7) with

$$d(p, \widetilde{p}) = \sup_{x \in [0,L]} \sup_{B \in \mathcal{B}_{[0,L]}} |p(B|x) - \widetilde{F}(B)|$$

on their right-hand sides.

In one particular case considered in [24], when $x_t = \epsilon x_{t-1} + (1-\epsilon)\xi_t, t = 1, 2, \ldots$ with i.i.d. random variables $\{\xi_t\}$, the constant α in (3.3) of Assumption 2 can be effectively bounded from below (with a bound independent on ϵ).

5. THE PROOFS

5.1. Reduction to Markov decision processes.

Using a standard approach (see, e.g. [21]), let us define two MDP's $\{z_t\}$ and $\{\widetilde{z}_t\}$ on the space $(\mathcal{X}, \mathcal{B}_{\mathcal{X}})$ where $\mathcal{X} = X \cup \{\Theta\}$, and Θ is an absorbing state, where z_t or $\tilde{z}_t$ move at the moment of stopping of $\{x_t\}$ or $\{\tilde{x}_t\}$. Control sets are $A(x) \equiv A = \{0, 1\}$. The *one step cost function* is

$$c(x, a) := \begin{cases} c_0(x), & x \in X, a = 0, \\ -r(r) + \bar{r}, & x \in X, a = 1, \\ 0, & x = \Theta, a \in A. \end{cases} \tag{5.1}$$

Finally, the *transition probability* γ for $\{z_t\}$ is defined as follows ($x \in \mathcal{X}, B \in \mathcal{B}_{\mathcal{X}}$):

$$\gamma(B|x,0) := \begin{cases} p(B\backslash\{\Theta\}|x), & x \in X, \\ \delta_*(B), & x = \Theta; \ \gamma(B|x,1) := \delta_*(B); \end{cases} \tag{5.2}$$

where δ_* is the Dirac measure.

The transition probability $\widetilde{\gamma}$ for $\{\tilde{z}_t\}$ is defined by (5.2) replacing p by $\widetilde{p}$.

For any *control policy* π (see definitions, e. g. in [9]) and any initial state $x \in \mathcal{X}$, let P_x^π, $\widetilde{P}_x^\pi$, E_x^π, $\widetilde{E}_x^\pi$ denote probabilities (on the space of trajectories) and expectations corresponding to the kernels γ and $\widetilde{\gamma}$.

Also for every $x \in \mathcal{X}$, and *any nonrandomized policy* of control π applied to $\{z_t\}$, let τ_π denote a moment of first application of the action $a = 1$ (to stop). Similarly, $\widetilde{\tau}_\pi$ is defined for the process $\{\tilde{z}_t\}$.

Let Π denote the set of all nonrandomized control policies π such that

$$P_x(\tau_\pi < \infty) = \widetilde{P}_x(\widetilde{\tau}_\pi < \infty) = 1, \ \text{ for all } \ x \in X.$$

Now we introduce new versions of U, $\widetilde{U}$ in (2.8). For $x \in \mathcal{X}$ and $\pi = (a_0, a_1, \dots) \in \Pi$, let

$$V(x,\pi) := E_x^\pi \mathbf{e}^{-\lambda \sum_{t=0}^{\infty} c(x_t,a_t)}, \ \widetilde{V}(x,\pi) := \widetilde{E}_x^\pi \mathbf{e}^{-\lambda \sum_{t=0}^{\infty} c(\tilde{x}_t,a_t)} \tag{5.3}$$

($\mathbf{e}^{-\infty} := 0$). In view of (5.1), the function c is nonnegative. Therefore V and $\tilde{V} \in [0,1]$, also $V(\Theta,\pi) = \widetilde{V}(\Theta,\pi) = 1$ (see (5.1) and (5.2)).

The *value functions* (corresponding to functionals (5.3)) are defined as follows:

$$\begin{aligned} V_*(x) &:= \sup\nolimits_{\pi\in\Pi} V(x,\pi), x \in \mathcal{X}; \\ \widetilde{V}_*(x) &:= \sup\nolimits_{\pi\in\Pi} \widetilde{V}(x,\pi), x \in \mathcal{X}. \end{aligned} \tag{5.4}$$

Lemma 5.1. (*Optimality Equation*)

For any $x \in \mathcal{X}$,

$$V_*(x) = \max_{a\in A} \mathbf{e}^{-\lambda c(x,a)} \int_{\mathcal{X}} V_*(y)\gamma(\mathrm{d}y|x,a), \tag{5.5}$$

and also for every $x \in X$,

$$V_*(x) = \max\left\{\mathbf{e}^{\lambda r(x)-\lambda\bar{r}}, \mathbf{e}^{-\lambda c_0(x)} \int_X V_*(y)p(\mathrm{d}y|x)\right\}. \tag{5.6}$$

Similar equations (with $\widetilde{\gamma}$ and $\widetilde{p}$) hold for $\widetilde{V}_*$.

The proof of (5.5) is accomplished as the proof of Lemma 3 in [6].

Equation (5.6) follows from (5.1), (5.2), (5.5) and the fact that $V_*(\Theta) = 1$.

5.2. New kernels and corresponding operators

Define on $(\mathcal{X} \times A, \mathcal{B}_{\mathcal{X}})$ the following kernels:

$$q(B|x,a) := \mathbf{e}^{-\lambda c(x,a)}\gamma(B|x,a); \ \tilde{q}(B|x,a) := \mathbf{e}^{-\lambda c(x,a)}\widetilde{\gamma}(B|x,a). \tag{5.7}$$

For every *stationary policy* $f \in \Pi$, $f \equiv (f, f, \dots)$, $(f : X \to A)$ we write: $q(\cdot|x,f) \equiv q(\cdot|x,f(x))$, $\widetilde{q}(\cdot|x,f) \equiv q(\cdot|x,f(x)), x \in \mathcal{X}$, and define in a standard manner q^n $(n = 1,2,\dots)$:

$$q^n(B|x,f);= \int_{\mathcal{X}} q^{n-1}(B|y,f(y))q(\mathrm{d}y|x,f(x)). \tag{5.8}$$

The kernels $\tilde{q}^n$ are defined similarly.

Let D denote the space of all measurable functions $\varphi : \mathcal{X} \to [0,1]$, such that $\varphi(\Theta) = 1$. We equip D with the uniform metric:

$$\rho(\varphi,\varphi') := \sup_{x\in\mathcal{X}} |\varphi(x) - \varphi'(x)| \equiv \sup_{x\in X} |\varphi(x) - \varphi'(x)|. \tag{5.9}$$

For each $f \in \mathbb{F}$ (where $\mathbb{F}$ is a subset of stationary policies from Π), we define two operators: $T_f : D \to D$; $\widetilde{T}_f : D \to D$ as follows:

$$T_f\varphi(x) := \int_{\mathcal{X}} \varphi(y)q(\mathrm{d}y|x,f(x)), \ x \in \mathcal{X}. \tag{5.10}$$

Replacing q by $\tilde{q}$, the operator $\widetilde{T}_f$ is defined similarly. From (5.2), it follows that T_f, $\widetilde{T}_f$ indeed map D into D.

Lemma 5.2. For each $f \in \mathbb{F}$,

$$V_f = T_f V_f \quad \text{and} \quad \widetilde{V}_f = \widetilde{T}_f \widetilde{V}_f, \tag{5.11}$$

where $V_f(x) := V(x,f(x))$, $\widetilde{V}_f(x) := \widetilde{V}(x,f(x)), x \in \mathcal{X}$ and $V, \widetilde{V}$ were defined in (5.3).

The proof of Lemma 5.2 is a straightforward usage of Markov property.

5.3. Relations between control policies and stopping rules

From the definitions of the control processes $\{z_t\}$ and $\{\tilde{z}_t\}$ above it follows that each control policy $\pi \in \Pi$ generates the stopping rule $f_\pi \in \Psi$ (where Ψ is the class of all stopping rules with a.s. finite stopping times, see Section 2). Moreover, comparing (2.8) and (5.3), since $\tau_\pi, \widetilde{\tau}_\pi$ are almost surely finite, and at instance τ_π, $c(x_{\tau_\pi}) = -r(x_{\tau_\pi}) + \bar{r}$, we find that for every $x \in X$ and $\pi \in \Pi$,

$$V(x,\pi) = U(x,f_\pi)\mathbf{e}^{-\lambda\bar{r}}; \ \widetilde{V}(x,\pi) = \widetilde{U}(x,f_\pi)\mathbf{e}^{-\lambda\bar{r}}. \tag{5.12}$$

Consequently, in (2.6), (2.7) and (5.4), for each $x \in X$.

$$V_*(x) = U_*(x)\mathbf{e}^{-\lambda\bar{r}}; V_*(x) = U_*(x)\mathbf{e}^{-\lambda\bar{r}}, \tag{5.13}$$

(because *each* a.s. finite stopping rule is generated by some policy from Π). For this reason we can rewrite "the stopping sets" in (2.10) in the following equivalent form:

$$S := \left\{x \in X : V_*(x) = \mathbf{e}^{\lambda r(x) - \lambda \bar{r}}\right\}, \ \widetilde{S} := \left\{x \in X : \widetilde{V}_*(x) = \mathbf{e}^{\lambda r(x) - \lambda \bar{r}}\right\}. \tag{5.14}$$

Now we define two key stationary control policies (compare them with (2.9) in Proposition 2.1):

$$f_*(x) := \begin{cases} 0, & x \in X \backslash S, \\ 1, & x \in S, \\ 1, & x = \Theta, \end{cases} \quad \text{and} \tag{5.15}$$

$$\widetilde{f}_*(x) := \begin{cases} 0, & x \in X \backslash \widetilde{S}, \\ 1, & x \in \widetilde{S}, \\ 1, & x = \Theta. \end{cases} \tag{5.16}$$

In view of Assumptions 1 and 2 and the well-known properties of Markov processes (see e. g. [18]) we get that f_* and $\widetilde{f}_* \in \Pi$, and therefore they generate stopping rules with a.s. finite stopping times τ_{f_*} and $\widetilde{\tau}_{f_*}$.

5.4. Contracting properties of operators T_{f_*} and $T_{\widetilde{f}_*}$

For any $f \in \mathcal{F} := \{f_*, \widetilde{f}_*\}$ (see (5.15) and (5.16), let T_f and $\widetilde{T}_f$ be operators defined in (5.10), and

$$N := \left[\frac{\log\left(\frac{\alpha}{2M}\right)}{\log(\eta)}\right] + 2,$$

where M, η and α are constants involved in Assumptions 1 and 2. Let also ρ be the uniform metric in (5.9).

Lemma 5.3. For every $f \in \mathcal{F}$ and $u, v \in D$,

$$\rho\left(T_f^N u, T_f^N v\right) \le \left(1 - \frac{3}{4}\alpha\right)\rho(u,v), \quad \rho\left(\widetilde{T}_f^N u, \widetilde{T}_f^N v\right) \le \left(1 - \frac{3}{4}\alpha\right)\rho(u,v). \tag{5.17}$$

Proof. Let, for example, $f = f_*$. It is easy to see that

$$T_f^n u(x) = \int_{\mathcal{X}} u(y) q^n(\mathrm{d}y|x, f(x)) \ (x \in \mathcal{X}), u \in D, n \ge 1,$$

and q^n was defined in (5.8). Thus, by (5.9), for any $u, v \in D$,

$$\begin{aligned} \rho(T_f^n u, T_f^n v) &= \sup_{x \in \mathcal{X}} \left| \int_{\mathcal{X}} u(y) q^n(\mathrm{d}y|x, f(x)) - \int_{\mathcal{X}} v(y) q^n(\mathrm{d}y|x, f(x)) \right| \\ &\le \rho(u,v) \sup_{x \in \mathcal{X}} q^n(X|x, f(x)). \end{aligned} \tag{5.18}$$

Taking (5.7) into account, we obtain than for every $x \in X$ and $B \in \mathcal{B}_{\mathcal{X}}$,

$$q^n(B|x, f(x)) \le \gamma^n(B|x, f(x)), \tag{5.19}$$

where γ^n is the nth power of the kernel γ in (5.2). Using (5.18), (5.19), and the result proved in Lemma 4.1 in [24]:

$$q^N(X|x, f(x)) \leq \left(1 - \frac{3}{4}\alpha\right), \quad x \in X,$$

we obtain the desired inequalities (5.17). □

5.5. Optimality of stationary policies f_* and $\widetilde{f}_*$

Lemma 5.4. The stationary policies of control f_* and $\widetilde{f}_*$ are optimal correspondingly for the processes $\{z_t\}$ and $\{\widetilde{z}_t\}$, respectively, (with respect to criterion (5.3)).

Proof. Let us prove the optimality of f_*. By (5.14) and (5.15) the policy f_* maximizes the right-hand sides of the optimality equations (5.5) and (5.6). By definitions (5.14) and (5.15), for every $x \in \mathcal{X}$,

$$T_{f_*} V_*(x) = \int_{\mathcal{X}} V_*(y) q(\mathrm{d}y|x, f_*(x)) = \mathbf{e}^{-\lambda c(x, f_*(x))} \int_{\mathcal{X}} V_*(y)\gamma(\mathrm{d}y|x, f_*(x))$$

$$= \max_{a \in A} \mathbf{e}^{-\lambda c(x,a)} \int_{\mathcal{X}} V_*(y)\gamma(\mathrm{d}y|x, a) = V_*(x)$$

by virtue of (5.5). Thus $T_{f_*} V_* = V_*$. On the other hand, by Lemma 5.2

$$V(x, f_*(x)) \equiv V_{f_*}(x) = T_{f_*} V_{f_*}(x).$$

Therefore, $V(x, f_*(x)) = V_*(x)$ since, by Lemma 5.3, the operator T_{f_*} has a unique fixed point in D. □

5.6. Estimation of closeness of the operators T_f and $\widetilde{T}_{f_*}$

Lemma 5.5. For each $f \in \mathcal{F}$ and every $n \geq 1$, $\varphi \in \mathcal{D}$,

$$\rho(T_f^n \varphi, \widetilde{T}_f^n \varphi) \leq \frac{n}{2} d(p, \widetilde{p}), \tag{5.20}$$

where $d(p, \widetilde{p})$ was defined in (1.3).

Proof. For every $x \in \mathcal{X}$ fixed we have, (for example, for $f = f_*$):

$$\begin{aligned} I_n := \quad & \left| T_f^n \varphi(x) - \widetilde{T}_f^n \varphi(x) \right| \leq \left| \int_{\mathcal{X}} \varphi(y) \int_{\mathcal{X}} q(\mathrm{d}z|x, f) q^{n-1}(\mathrm{d}y|z, f) \right. \\ & \left. - \int_{\mathcal{X}} \varphi(y) \int_{\mathcal{X}} \widetilde{q}(\mathrm{d}z|x, f) q^{n-1}(\mathrm{d}y|z, f) \right| + \left| \int_{\mathcal{X}} \varphi(y) \int_{\mathcal{X}} \widetilde{q}(\mathrm{d}z|x, f) q^{n-1}(\mathrm{d}y|z, f) \right. \\ & \left. - \int_{\mathcal{X}} \varphi(y) \int_{\mathcal{X}} \widetilde{q}(\mathrm{d}z|x, f) \widetilde{q}^{n-1}(\mathrm{d}y|z, f) \right| =: I_{1,n} + I_{2,n}. \end{aligned} \tag{5.21}$$

Then by Fubini's theorem,

$$I_{1,n} = \left| \int_{\mathcal{X}} \left[q(\mathrm{d}z|x,f) - \widetilde{q}(\mathrm{d}y|z,f) \right] \int_{\mathcal{X}} \varphi(y) q^{n-1}(\mathrm{d}y|z,f) \right|,$$

where the last integral represents a function from D with values in $[0,1]$ and with value 1 at the point $x = \Theta$. Thus by (5.7) and (5.2),

$$I_{1,n} = \leq \frac{1}{2} \| p(\cdot|x) - \widetilde{p}(\cdot|x) \|. \tag{5.22}$$

Now applying the similar arguments to $I_{2,n}$ in (5.21) we obtain that

$$\begin{aligned} I_{2,n} &\leq \int_{\mathcal{X}} \widetilde{q}(\mathrm{d}z|x,f) \left| \int_{\mathcal{X}} \varphi(y) \left[q^{n-1}(\mathrm{d}y|z,f) - \widetilde{q}^{n-1}(\mathrm{d}y|z,f) \right] \right| \\ &\leq \tfrac{1}{2} \sup_{x \in X} \| p^{n-1}(\cdot|x) - \widetilde{p}^{n-1}(\cdot|x) \| \\ &\leq \tfrac{n-1}{2} \sup_{x \in X} \| p^{n-1}(\cdot|x) - \widetilde{p}(\cdot|x) \|, \end{aligned} \tag{5.23}$$

where the last inequality was shown, for example, in [24]. Combining (5.21) – (5.23), we get (5.20). □

5.7. The rest of the proof of Theorems 3.1 and 3.2

From (5.12) and Lemma 5.4 it follows that the stability index $\delta(x)$ in (2.12) is

$$\delta(x) = \mathbf{e}^{\lambda \bar{r}} \left[V(x, f_*) - V(x, \widetilde{f}_*) \right], \tag{5.24}$$

where V is defined in (5.3). Using Lemmas 5.2, 5.3 and 5.4, an upper bound for $V(x,f_*) - V(x,\widetilde{f}_*)$ is found in a similar way as in the proof of Theorem 2.1 in [24]. This results in the following inequality (valued for all $x \in X$)

$$0 \leq V(x, f_*) - V(x, \widetilde{f}_*) \leq \frac{2}{3\alpha} N(N+1) d(p, \widetilde{p}), \tag{5.25}$$

where d is from (1.3), α is from Assumption 2, and, as in Lemma 5.3, N is an integer from (3.5).

From (5.24) and (5.25) inequality (3.4) of Theorem 3.1 follows.

Now, from (2.2), (2.8) and (5.12) we find that the stability index $\Delta(x)$ in (2.11) is expressed as follows:

$$\begin{aligned} \Delta(x) &= \frac{1}{\lambda} \log \left(\mathbf{e}^{\lambda \bar{r}} V(x, f_*) \right) - \frac{1}{\lambda} \log \left(\mathbf{e}^{\lambda \bar{r}} V(x, \widetilde{f}_*) \right) \\ &= \frac{1}{\lambda} \log \frac{V(x, f_*)}{V(x, \widetilde{f}_*)} \leq \frac{1}{\lambda} \frac{V(x, f_*) - V(x, \widetilde{f}_*)}{V(x, \widetilde{f}_*)}. \end{aligned} \tag{5.26}$$

Denoting in (5.25) $K := \frac{2}{3\alpha}N(N+1)$, we obtain from the last inequality that

$$V(x,\widetilde{f}_*) = V(x,f_*) - Kd(p,\widetilde{p}) \geq \mathbf{e}^{\lambda(r(x)-\bar{r})} - Kd(p,\widetilde{p}) \geq \frac{1}{2}\mathbf{e}^{\lambda(r(x)-\bar{r})}, \tag{5.27}$$

provided that condition (3.6) holds. The first inequality in (5.27) is due to the fact that f_* gives a maximum to $V(x,\pi)$, so $V(x,f_*) \geq V\ (x,\ h)$, where h is the policy: "to stop at $t=0$". Combining (5.25) – (5.27) we prove Theorem 3.2.

ACKNOWLEDGEMENT

The authors thank all the referees for their helpful comments and suggestions, which were used to aid the improvement of this paper.

This work was supported in part by CONACYT (México) and AS CR (Czech Republic) under Grant No. 171396.

(Received May 9, 2013)

REFERENCES

[1] G. Avila-Godoy and E. Fernández-Gaucherand: Controlled Markov chains with exponential risk-sensitive criteria: modularity, structured policies and applications. In: Decision and Control 1998. Proc. 37th IEEE Conference. Vol. 1, IEEE, pp. 778–783.

[2] N. Bäuerle and U. Rieder: Markov Decision Processes with Applications to Finance. Springer-Verlag, Berlin 2011.

[3] V. S. Borkar and S. P. Meyn: Risk-sensitive optimal control for Markov decision processes with monotone cost. Math. Oper. Res. *27* (2002), 192–209.

[4] R. Cavazos-Cadena: Optimality equations and inequalities in a class of risk-sensitive average cost Markov decision chains. Math. Methods Oper. Res. *71* (2010), 47–84.

[5] R. Cavazos-Cadena and E. Fernández-Gaucherand: Controlled Markov chains with risk-sensitive criteria: Average costs, optimality equations, and optimal solutions. Math. Methods Oper. Res. *49* (1999), 299–324.

[6] R. Cavazos-Cadena and R. Montes-de-Oca: Optimal stationary policies in risk-sensitive dynamic programs with finite state space and nonegative rewards. Appl. Math. *27* (2000), 167–185.

[7] N. M. Van Dijk and K. Sladký: Error bounds for nonnegative dynamic models. J. Optim. Theory Appl. *101* (1999), 449–474.

[8] L. Devroye and L. Györfy: Nonparametric Density Estimation: The L_1 View. John Wiley, New York 1986.

[9] E. B. Dynkin and A. A. Yushkevich: Controlled Markov Processes. Springer Verlag, New York 1979.

[10] E. I. Gordienko and A. A. Yushkevich: Stability estimates in the problem of average optimal switching of a Markov chain. Math. Methods Oper. Res. *57* (2003), 345–365.

[11] E. I. Gordienko, E. Lemus-Rodríguez, and R. Montes-de-Oca: Average cost Markov control processes: stability with respect to the Kantorovich metric. Math. Methods Oper. Res. *70* (2009), 13–33.

[12] E. I. Gordienko and F. Salem: Robustness inequalities for Markov control processes with unbounded costs. Syst. Control Lett. *33* (1998), 125–130.

[13] O. Hernández-Lerma and J. B. Lasserre: Further Topics on Discrete-time Markov Control Processes. Springer-Verlag, New York 1999.

[14] A. Jaśkiewicz: Average optimality for risk-sensitive control with general state space. Ann. Appl. Probab. *17* (2007), 654–675.

[15] N. V. Kartashov: Strong Stable Markov Chains. VSP, Utrecht 1996.

[16] S. I. Marcus, E. Fernández-Gaucherand, D. E. Hernández-Hernández, S. Coraluppi, and P. Fard: Risk sensitive Markov decision processes. Progress in System and Control Theory *22* (1997), 263–280.

[17] G. B. Di Masi and L. Stettner: Infinite horizon risk sensitive control of discrete time Markov processes with small risk. Systems Control Lett. *40* (2000), 15–20.

[18] S. P. Meyn and R. L. Tweedie: Markov Chains and Stochastic Stability. Springer-Verlag, London 1993.

[19] R. Montes-de-Oca and F. Salem-Silva: Estimates for perturbations of an average Markov decision processes with a minimal state and upper bounded stochastically ordered Markov chains. Kybernetika *41* (2005), 757–772.

[20] B. K. Muciek: Optimal stopping of risk processes: model with interest rates. J. Appl. Probab. *39* (2002), 261–270.

[21] A. N. Shiryaev: Optimal Stopping Rules. Springer-Verlag, New York 1978.

[22] A. N. Shiryaev: Essential of Stochastic Finance. Facts, Models, Theory. World Scientific Publishing Co., Inc., River Edge, N. J. 1999.

[23] K. Sladký: Bounds on discrete dynamic programming recursions I. Kybernetika *16* (1980), 526–547.

[24] E. Zaitseva: Stability estimating in optimal stopping problem. Kybernetika *44* (2008), 400–415.

Raúl Montes-de-Oca, Departamento de Matemáticas, Universidad Autónoma Metropolitana-Iztapalapa, Av. San Rafael Atlixco 186, Col. Vicentina, 09340 México, D.F. México.
e-mail: momr@xanum.uam.mx

Elena Zaitseva, Instituto Tecnológico Autónomo de México, Río Hondo 1, Col. Progreso Tizapán, Del. Álvaro Obregón 01080 México, D.F. México.
e-mail: lenagordi@hotmail.com