Probability in the Engineering and Informational Sciences, **28**, 2014, 335–352.
doi:10.1017/S0269964814000035

CHARACTERIZATIONS OF OPTIMAL POLICIES IN A GENERAL STOPPING PROBLEM AND STABILITY ESTIMATING

Evgueni Gordienko and Andrey Novikov
Department of Mathematics, UAM-Iztapalapa,
San Rafael Atlixco 186,
col. Vicentina, C.P. 11320, Mexico City, Mexico
E-mail: gord@xanum.uam.mx and an@xanum.uam.mx

We consider an optimal stopping problem for a general discrete-time process $X_1, X_2, \ldots, X_n, \ldots$ on a common measurable space. Stopping at time n $(n = 1, 2, \ldots)$ yields a reward $R_n(X_1, \ldots, X_n) \geq 0$, while if we do not stop, we pay $c_n(X_1, \ldots, X_n) \geq 0$ and keep observing the process. The problem is to characterize all the optimal stopping times τ, i.e., such that maximize the mean net gain:

$$E(R_\tau(X_1, \ldots, X_\tau) - \sum_{n=1}^{\tau-1} c_n(X_1, \ldots, X_n)).$$

We propose a new simple approach to stopping problems which allows to obtain not only sufficient, but also necessary conditions of optimality in some natural classes of (randomized) stopping rules.

In the particular case of Markov sequence $X_1, X_2, \ldots$ we estimate the stability of the optimal stopping problem under perturbations of transition probabilities.

1. INTRODUCTION

Starting with pioneer works on sequential analysis in statistics (such as [3,35]), the theory of optimal stopping already has a rather long history. Basic methods were summarized in the books [6] (for general discrete-time processes) and [16,32] (for Markov processes). In recent decades new important applications came into view, for example, in the risk theory and financial mathematics (see, e.g., [18,24,33]), in the change detection theory (see, e.g., [5,14]), in software testing (see, for instance, [23]), just to name some.

Since the time of the general theory foundation in the 60s—early seventies the major part of new results dealt with optimal stopping of independent random variables and Markov chains and processes. Along with the main dynamic programming method some alternative approaches were proposed, for example, the linear programming technique (see, e.g., recent works [8,17]). In the Markov case the problem of optimal stopping was embedded in a general context of controllable Markov chains (see, e.g., [16,30,31]). Concerning optimal stopping in

 0269-9648/14 $25.00

the non-Markovian case there were some new results dealing with some particular classes of processes (see, e.g., [1,2,20]).

In the present paper, we consider the following discrete-time optimal stopping problem. Let $X_1, X_2, \ldots, X_n, \ldots$ be a general *discrete-time stochastic process* with a measurable state space $(\mathfrak{X}, \mathscr{X})$. It is supposed that $X_1, X_2, \ldots$ are observed on the one-by-one basis getting successively the data $x_1, x_2, \ldots$ (values of $X_1, X_2, \ldots$). At each stage n, $n = 1, 2, \ldots$, after the data $x_1, \ldots, x_n$ have been observed, we may stop, and if we stop, we acquire a *reward* $R_n(x_1, \ldots, x_n) \geq 0$. If we do not stop, we have to pay $c_n(x_1, \ldots, x_n) \geq 0$ as a *maintenance fee* and keep observing for $X_{n+1}, X_{n+2}, \ldots$ in order to stop at a later time.

In Section 2, we introduce some natural classes $\mathscr{F}$ of almost surely finite stopping times τ (in general, randomized). The functional to be maximized is

$$G(\tau) := E[R_\tau(X_1, \ldots, X_\tau) - \sum_{n=1}^{\tau-1} c_n(X_1, \ldots, X_n)]. \tag{1}$$

A stopping time τ_* is optimal in $\mathscr{F}$ if

$$G(\tau_*) = G_* := \sup_{\tau \in \mathscr{F}} G(\tau). \tag{2}$$

One goal of the first part of the paper is the characterization of the structure of all optimal stopping times in any class $\mathscr{F}$ satisfying Assumption 1 (see Section 2). To achieve this goal we propose a new simple and direct method of "upper bounds" for stopping time optimization (in some situations equivalent to the corresponding dynamic programming technique). In comparison with the approach used in [6], Ch. 4, we do not exploit essential supremums over "future stopping", but make use of the direct approximation of the infinite horizon optimization problem by means of the corresponding problems with finite time intervals. Such method allows to prove not only sufficient but also necessary conditions of optimality in $\mathscr{F}$ and (under additional restrictions on the functions R_n, c_n) in wider classes $\mathscr{F}_0$, $\mathscr{F}_1$, where $\mathscr{F}_0$ consists of all almost surely finite stopping times. Along with "dynamic programming arguments" our optimality conditions involve certain "tail conditions" (known in the Markov case as "equalizing property"; see, e.g., [16]). To prove our results we impose certain boundedness conditions on $\{R_n, n \geq 1\}$, which are rather standard, but, for example, the assumption (14) in Section 2 is less restrictive than the related conditions in Theorems 4.4 and 4.5 in [6]. Note that sufficient conditions of optimality given in Theorems 2 and 3 of Section 2 are in some way distinct from their counterparts in [6].

A potential advantage of the method used in this paper is the fact that it makes feasible to aggregate any policies of control of $X_1, X_2, \ldots$ which operate before a stopping time τ. In the paper [25], such scheme was realized for independent $X_1, X_2, \ldots$. However, it is easy to see that the independence conditions is not essential for the main results of [25].

There is another known approach to the treatment of optimal stopping problems: enlarging the state space including in it all "histories" of the process under consideration. This allows one to reduce the problem to the Markovian case (see a discussion on this topic, for instance, in [32]). For some control problems this method was used in a number of engineering papers related to artificial intelligence (see, for instance, [4], [34]). These papers offer some heuristic procedures of optimizations without mathematical justification of their optimality. For a general random sequence $X_1, X_2, \ldots$, the above-mentioned enlarging of a state space leads to a non-homogeneous Markov chain on a complicated space. There is little hope to take advantage of the existing theory of Markov decision processes in an effective way (which in the case of infinite horizon mostly operates with homogeneous processes; see, e.g.,

[15], [11] for non-homogeneous controllable Markov sequences, where strong assumptions on "reward-cost" are made to handle infinite horizon optimization problems).

In Section 3, we deal with quantitative estimation of "stability" (continuity) of the stopping time optimization for a Markov chain $X_0, X_1, \ldots$ with a transition probability p. Such estimations are useful in many applied control problems, where either p is unknown and it is approximated by en available statistical estimate $\tilde{p}$, or when the transition probability p is replaced by some "theoretical approximation" $\tilde{p}$, to obtain a chain with a simpler structure (see, e.g., [20]). In our setting, it is supposed that the available "stopping rule" corresponding to the stopping time $\tilde{\tau}_*$ optimal for the chain with the transition probability $\tilde{p}$ is applied to the "original" chain governed by p. Under certain "drift conditions" we prove the following inequality (see (1) and (2)):

$$0 \le \Delta := G_* - G(\tilde{\tau}_*) \le K \sup_{x \in \mathfrak{X}} ||p(\cdot|x) - \tilde{p}(\cdot|x)||, \tag{3}$$

where $||\cdot||$ is the total variation norm.

To support the usage of the drift conditions given in Assumption 3 we offer a counterexample where the right-hand side of (3) tends to zero, but "the stability index" Δ on the left-hand side of (3) is infinite. We also present two examples showing how to bound $||p(\cdot|x) - \tilde{p}(\cdot|x)||$, when $\tilde{p}$ is a certain statistical estimator of p, and how inequality (3) works in one queuing model.

For geometrically ergodic Markov chains, a different bound for Δ was found in [37]. Some relevant results (without any quantitative estimation) on approximation for optimal stopping problem can be found in [20,22].

It is worth noting that the method developed in Section 2 has been applied (for particular processes) to solving several general problems in statistical sequential analysis (see [26–28]). Further possible applications of our results could appear in models of statistical sequential analysis with dependent observations, in detecting changes in non-Markovian discrete-time processes, and in some models of risk theory.

2. CHARACTERIZATIONS OF OPTIMAL STOPPING RULES

2.1. Basic Assumptions and Notation

We work in this Section with stopping *rules* rather than stopping times, which facilitates the task of characterizing the optimality we pursue.

A *(randomized) stopping rule* ψ is a family of functions $(\psi_1, \psi_2, \ldots, \psi_n, \ldots)$ with $\psi_n : \mathfrak{X}^n \mapsto [0,1]$ measurable with respect to the n-fold product $\mathscr{X}^n$ of the σ-algebra $\mathscr{X}$ by itself, $n = 1, 2, \ldots$. Any $\psi_n(x_1, \ldots, x_n)$ is interpreted as the conditional probability to stop, given that the process came to stage n, and that the data observed were $x_1, x_2, \ldots, x_n$. In a usual way, every stopping rule ψ generates a *(randomized) stopping time* τ_ψ, with respect to the sequence of σ-algebras $\sigma(X_1, \ldots, X_n)$, $n = 1, 2, \ldots$.

Let us also suppose that the functions of maintenance cost $c_n : \mathfrak{X}^n \mapsto \mathbb{R}^+$ and of the final reward $R_n : \mathfrak{X}^n \mapsto \mathbb{R}^+$ are measurable with respect to $\mathscr{X}^n$ and such that $Ec_n(X_1, \ldots, X_n) < \infty$ and $ER_n(X_1, \ldots, X_n) < \infty$, for all $n = 1, 2 \ldots$.

We denote: $C_n(x_1, \ldots, x_n) = \sum_{i=1}^{n-1} c_i(x_1, \ldots, x_i)$ ($C_1(x_1) \equiv 0$, by definition).

For any stopping rule ψ let

$$t_n^\psi = t_n^\psi(x_1, \ldots, x_n) := (1 - \psi_1(x_1)) \ldots (1 - \psi_{n-1}(x_1, \ldots, x_{n-1})),$$

and

$$s_n^\psi = s_n^\psi(x_1, \ldots, x_n) := t_n^\psi(x_1, \ldots, x_n)\psi_n(x_1, \ldots, x_n),$$

https://doi.org/10.1017/S0269964814000035 Published online by Cambridge University Press

$n = 1, 2, \dots$ (by definition, $t_1^{\psi} = 1$). Then, obviously, for all $n = 1, 2, \dots$

$$P(\tau_{\psi} = n) = Es_n^{\psi}, \quad P(\tau_{\psi} \geq n) = Et_n^{\psi}. \tag{4}$$

We interpret s_n and t_n as $s_n(X_1, \dots, X_n)$ and $t_n(X_1, \dots, X_n)$, respectively, when under the expectation (or probability) sign (as in (4)). So do we with any function of observations F_n, supposing throughout the paper that F_n stands for $F_n(X_1, \dots, X_n)$, if F_n is under the probability or the expectation sign, and for $F_n(x_1, \dots, x_n)$, otherwise.

Let $\mathscr{F}_0$ be the set of the stopping rules ψ which stop with probability one, i.e., such that

$$P(\tau_{\psi} < \infty) = \sum_{n=1}^{\infty} Es_n^{\psi} = 1.$$

For any stopping rule ψ, we define the expected reward and the expected cost $\bar{C}(\psi)$ from using ψ as

$$\bar{R}(\psi) := \sum_{n=1}^{\infty} Es_n^{\psi} R_n \quad \text{and} \quad \bar{C}(\psi) := \sum_{n=1}^{\infty} Es_n^{\psi} C_n,$$

respectively.

The conditions we impose below on stopping rules guarantee that either $\bar{R}(\psi)$ or $\bar{C}(\psi)$ is finite for all $\psi \in \mathscr{F}$. Naturally, in such a case the *mean net gain* from using ψ is defined as

$$G(\psi) = \bar{R}(\psi) - \bar{C}(\psi) = G(\psi) = \sum_{n=1}^{\infty} Es_n^{\psi}(R_n - C_n),$$

which is a generalization of (1) to the class of randomized stopping rules we consider here.

2.2. The Structure of Optimal Finite-Horizon Stopping Rules

In this Section, we briefly revisit the classical "backward induction" case of optimal stopping (see [6], for example). Largely, this is the result of Theorem 3.2 in [6], but complemented with the necessity of the structure of optimal rules. This case will serve as the basis for the general infinite-horizon problem in Section 2.3 below.

Let N be any natural number, and let $\mathscr{F}^N$ be the class of (finite-horizon) stopping rules ψ such that

$$(1 - \psi_1)(1 - \psi_2) \cdots (1 - \psi_N) \equiv 1.$$

Let

$$G_N(\psi) := G(\psi) = \sum_{n=1}^{N} Es_n^{\psi}(R_n - C_n), \quad \psi \in \mathscr{F}^N. \tag{5}$$

Let us define the family of functions $V_n^N = V_n^N(x_1, \dots, x_n)$, $(x_1, \dots, x_n) \in \mathfrak{X}^n$, $n = 1, 2, \dots, N$ in the following way ("backward induction"):

Start with

$$V_N^N := R_N(X_1, \dots, X_N). \tag{6}$$

Then, for $n = N - 1, \dots, 2, 1$, define recursively

$$V_n^N := \max\{R_n(X_1, \dots, X_n), Q_n^N - c_n(X_1, \dots, X_n)\}, \tag{7}$$

where

$$Q_n^N := E\{V_{n+1}^N | X_1, \dots, X_n\}, \tag{8}$$

for $n = N - 1, \dots, 1$, and $Q_0^N = EV_1^N(X_1)$.

The following assertion is essentially Theorem 3.2 in [6] complemented with the necessity part.

THEOREM 1: *For all* $\psi \in \mathscr{F}^N$

$$G_N(\psi) \leq Q_0^N. \tag{9}$$

If there is an equality in (9) for some $\psi \in \mathscr{F}^N$*, then*

$$I_{\{R_n > Q_n^N - c_n\}} \leq \psi_n(X_1, \dots, X_n) \leq I_{\{R_n \geq Q_n^N - c_n\}} \tag{10}$$

almost surely on

$$T_n^\psi := \{t_n^\psi(X_1, \dots, X_n) > 0\}$$

for all $n = 1, 2, \dots, N-1$.

If $\psi = (\psi_1, \dots, \psi_N)$ *is such that (10) (almost surely) holds on* T_n^ψ *for all* $n = 1, 2, \dots, N-1$*, and* $\psi_N \equiv 1$*, then* $G_N(\psi) = Q_0^N$.

The proof of Theorem 1 can be found in Section A.1 of the Appendix.

By Theorem 1,

$$Q_0^N = \sup_{\psi \in \mathscr{F}^N} G(\psi),$$

and $\psi \in \mathscr{F}^N$ is optimal if and only if it satisfies (10). The optimality of the non-randomized rule with $\psi_n(X_1, \dots, X_n) = I_{\{R_n \geq Q_n^N - c_n\}}$, $n = 1, 2, \dots, N-1$ (a particular case of (10)) also follows Theorem 3.2 [6].

2.3. General Stopping Rules

In this section, we treat the case of general stopping times $\tau \in \{1, 2, \dots\}$. Unlike the general case in [6], we prefer not to use the "dynamic programming" approach based on the essential supremums of the conditional mean gain from "acting optimally in the future" (see the definition of γ_n in (4.2′) [6]), but directly deal with the limits of V_n^N ($N \to \infty$) from the preceding section (this corresponds to γ_n' in [6], Chapter 4). This makes the problem of characterizing the structure of optimal stopping rules especially clear and easy (almost as simple as the backward induction in the preceding Section—see the proof of Theorem 2 below). Thus, we prefer to work with stopping times we call "truncatable" (this means that the mean net gain of the truncated, at time N, stopping time is close enough to that of the non-truncated one, whenever N is large). This idea has been largely exploited in some problems of statistical sequential analysis for discrete-time stochastic processes (see, e.g., [26], [27] or [28]), and corresponds to the very usual procedure of "finite-horizon" approximation in sequential analysis (see, e.g., [3,9,10], among many others). In particular, below in this section we show that, under rather general assumptions about the cost structure (suitable for statistical applications) every finite stopping time is truncatable.

Let for any stopping rule ψ

$$G_N(\psi) := \sum_{n=1}^{N-1} E s_n^\psi (R_n - C_n) + E t_n^\psi (R_N - C_N).$$

It is easy to see that $G_N(\psi)$ (see (5)) coincides with $G(\psi^N)$, where $\psi^N = (\psi_1, \dots, \psi_{N-1}, 1, \dots)$ is the rule ψ truncated at time N.

https://doi.org/10.1017/S0269964814000035 Published online by Cambridge University Press

Because ψ^N has finite horizon N, we can use the result of the preceding section; in particular, the inequality (9) is valid. The idea of what follows is to pass to the limit, as $N \to \infty$, in (9) in order to get some upper bound for $G(\psi)$.

Let us start with the behavior of V_n^N in (7), as $N \to \infty$, $n = 1, 2, \ldots$.

It is easy to see that for all $N \geq n$, $n = 1, 2, \ldots$, $V_n^N \leq V_n^{N+1}$ with probability 1 (see, e.g., the proof of Lemma 3.3 [27]). It follows from this that with probability 1 there exists $V_n = \lim_{N\to\infty} V_n^N$, $n = 1, 2, \ldots$.

By the Lebesgue monotone convergence theorem we immediately have that the right-hand side of (9) converges to $Q_0 = EV_1(X_1)$ (see the definition of Q_0^N next to (8)).

In the same way, with probability 1

$$Q_n = \lim_{N\to\infty} Q_n^N = E\{V_{n+1}|X_1, \ldots, X_n\}.$$

Let $\mathscr{F}_0$ be the class of all stopping rules ψ such that $P(\tau_\psi < \infty) = 1$. We only work with classes $\mathscr{F} \subset \mathscr{F}_0$ of stopping rules (we call truncatable) which satisfy the following assumption.

Assumption 1:

1. *For all $\psi \in \mathscr{F}$, $G(\psi)$ is well-defined (in the sense that either $\bar{R}(\psi)$ or $\bar{C}(\psi)$ is finite),*
2. *For all natural N, $\mathscr{F}^N \subset \mathscr{F}$.*
3. *For every $\psi \in \mathscr{F}$ such that $G(\psi) > -\infty$ it holds that $\lim_{N\to\infty} G_N(\psi) = G(\psi)$.*

The classes satisfying Assumption 1 do exist, for example, the class $\bigcup_{N\geq 1} \mathscr{F}^N$ of all finite-horizon rules obviously satisfies it. Below in this section, under more specific conditions, we give less trivial (and more interesting from the theoretical and practical points of view) examples of $\mathscr{F}$ satisfying Assumption 1.

Lemma 1: *For any class $\mathscr{F}$ satisfying Assumption 1*

$$\sup_{\psi\in\mathscr{F}} G(\psi) = Q_0.$$

The proof of Lemma 1 is almost identical to that of Lemma 3.5 in [27].

Our first main theorem below gives a necessary and sufficient structure of optimal stopping rules in any class $\mathscr{F}$ satisfying Assumption 1.

Theorem 2: *Suppose that $Q_0 < \infty$ and that $\mathscr{F} \subset \mathscr{F}_0$ is a class of stopping rules for which Assumption 1 is fulfilled.*

If $\psi \in \mathscr{F}$ is optimal in $\mathscr{F}$, that is, if

$$G(\psi) = \sup_{\psi'\in\mathscr{F}} G(\psi'), \tag{11}$$

then

$$I_{\{R_n > Q_n - c_n\}} \leq \psi_n \leq I_{\{R_n \geq Q_n - c_n\}} \tag{12}$$

almost surely on T_n^ψ for all n = 1,2,..., and

$$Et_n^\psi(V_n - R_n) \to 0, \quad as \quad n \to \infty. \tag{13}$$

On the other hand, if a stopping rule $\psi \in \mathscr{F}$ satisfies (12) (almost surely on T_n^ψ) for all n=1,2, ..., and satisfies (13), then it is optimal in $\mathscr{F}$.

We place the proof of Theorem 2 in Section A.2 of the Appendix.

The aim of what follows in this section is to give examples of applications of Theorem 2 to some explicitly defined classes $\mathscr{F}$.

If the family $\{R_n,\, n \geq 1\}$, is uniformly integrable, the following class $\mathscr{F}_1$ of stopping rules can be used as $\mathscr{F}$ in Theorem 2:

$$\mathscr{F}_1 := \{\psi \in \mathscr{F}_0 : \bar{R}(\psi) < \infty\}.$$

LEMMA 2: *If the sequence $\{R_n,\, n \geq 1\}$ is such that*

$$\sup_{n\geq 1} ER_n I_{\{R_n \geq k\}} \to 0, \quad as \quad k \to \infty, \tag{14}$$

then for the class $\mathscr{F}_1$ Assumption 1 is fulfilled.

(See Section A.3 of the Appendix for the proof.)

In the particular case when $E\sup_{n\geq 1} R_n < \infty$, the whole class $\mathscr{F}_0$ of all stopping rules, which terminate with probability 1 satisfies Assumption 1.

LEMMA 3: *If*

$$E\sup_{n\geq 1} R_n < \infty, \tag{15}$$

then for the class $\mathscr{F}_0 = \{\psi : P(\tau_\psi < \infty) = 1\}$ Assumption 1 is fulfilled.

PROOF: It follows from Lemma 2 that the class $\{\psi : P(\tau_\psi < \infty) = 1,\, \bar{R}(\psi) < \infty\}$ satisfies Assumption 1. But under the supposition of (15) we have: $\bar{R}(\psi) = \sum_{n=1}^{\infty} E s_n^{\psi} R_n \leq E\sup_{n\geq 1} R_n < \infty$, therefore $\mathscr{F}_1 = \mathscr{F}_0$ in this particular case. ∎

Under the condition of Lemma 3 there are some weaker sufficient conditions for the optimality in $\mathscr{F}_0$.

THEOREM 3: *Suppose that (15) is fulfilled.*

Then $\psi \in \mathscr{F}_0$ is optimal in $\mathscr{F}_0$, i.e.

$$G(\psi) = \sup_{\psi' \in \mathscr{F}_0} G(\psi'),$$

if and only of for the stopping rule ψ the inequalities (12) hold (almost surely on T_n^{ψ}) for all $n = 1, 2, \ldots$.

Suppose, additionally to (15), that $\{C_n, n \geq 1\}$, is such that for every $k > 0$

$$P(C_n(X_1, \ldots, X_n) < k) \to 0 \quad as \quad n \to \infty.$$

Then every ψ for which (12) holds (almost surely on T_n^{ψ}) for all $n = 1, 2, \ldots$ belongs to $\mathscr{F}_0$ and is optimal in $\mathscr{F}_0$.

The proof of Theorem 3 can be found in Section A.4 of the Appendix.

Remark 1: Under the conditions of Theorem 3, if a stopping rule ψ satisfying (12) generates a finite stopping time τ_ψ, then this rule is optimal. The optimality of a non-randomized τ_ψ (corresponding to $\psi_n = I_{\{R_n \geq Q_n - c_n\}}$, $n = 1, 2, \ldots$, cf. (12)) under the condition of finiteness of τ_ψ can be derived from Theorems 4.4 and 4.5 in [6] under these conditions.

Remark 2: The method we use in this paper (it may be called the "method of upper bounds", see Theorems 1 and 2 and their proofs, especially Lemma 4 in the Appendix) can be applied to much more general problems than those considered here. For example, in [25] essentially the same method was used for a sequential statistical problem with control. It is not difficult to see that it can be generalized to finding optimal stopping and control policies in problems similar to those considered here, but with the control involved.

3. MARKOV CASE: STABILITY ESTIMATING

Let now $\{X_n\} \equiv \{X_n,\ n = 0, 1, \dots\}$ and $\{\widetilde{X}_n\} \equiv \{\widetilde{X}_n,\ n = 0, 1, \dots\}$ be two homogeneous discrete-time Markov processes on a common state space $\mathfrak{X}$, with corresponding *transition probabilities* $p \equiv p(B|x)$ and $\tilde{p} \equiv \tilde{p}(B|x)$, $x \in \mathfrak{X}$, $B \in \mathscr{B}(\mathfrak{X})$. We assume that $\mathfrak{X}$ is either a Borel subset of a finite-dimensional Euclidean space or some subset of $\{0, 1, 2, \dots\}$, and that $\mathscr{B}(\mathfrak{X})$ is the Borel σ-algebra of subsets of $\mathfrak{X}$.

In our stability ("continuity") estimating setting it is supposed that one aims to figure (or to approximate) the stopping rule ψ_* optimal for the "original" process $\{X_n\}$ when its transition probability p is (at least partly) *unknown*, and it is approximated by some *available* (known) transition probability $\tilde{p}$ (for instance, obtained by means of statistical estimation). Also this includes the situations where the known p is replaced by certain $\tilde{p}$ in order to get a Markov chain with a simpler structure (making easier the stopping rule optimization).

We suppose that the stopping rule $\tilde{\psi}_*$ optimal for $\{\tilde{X}_n\}$ is applied to the process $\{X_n\}$ (in place of the inaccessible stopping rule ψ_*). Theorem 4 given below provides an upper bound for the decrease of the mean gain when replacing ψ_* by $\tilde{\psi}_*$.

In this section, we assume that the reward and cost functions $R_n(x_1, \dots, x_n) \equiv R(x_n)$, $c_n(x_1, \dots, x_n) \equiv c(x_n)$ are *bounded* and that they depend only on the last stage of the process (and do not depend on n).

For each initial state $x \in \mathfrak{X}$ of the process let as before $G(x, \psi)$ be the mean net gain obtained applying to $\{X_n\}$ a *non-randomized* stopping rule $\psi \in \mathscr{F}_0$ (with a finite stopping time). Similarly (using the same R and c), the mean net gain $\tilde{G}(x, \psi)$ is defined for the process $\{\tilde{X}_n\}$. It is well-known (see, e.g., [16,30–32]) that (particularly under the assumption given below) the corresponding *value functions*:

$$G_*(x) := \sup_{\psi \in \mathscr{F}_0} G(x, \psi),\ \tilde{G}_*(x) := \sup_{\psi \in \tilde{\mathscr{F}}_0} \tilde{G}(x, \psi), x \in \mathfrak{X}$$

($\tilde{\mathscr{F}}_0$ includes rules generating almost surely finite for $\{\tilde{X}_n\}$ stopping times) satisfy the optimality equations:

$$G_*(x) = \max\left\{R(x), \int_{\mathfrak{X}} G_*(y)p(dy|x) - c(x)\right\},\ x \in \mathfrak{X}, \tag{16}$$

(and the similar equation with $\tilde{p}$ for $\tilde{G}_*$).

If we define $S := \{x \in \mathfrak{X} : R(x) = G_*(x)\}$ and $\tilde{S} := \{x \in \mathfrak{X} : R(x) = \tilde{G}_*(x)\}$ then an optimal rule ψ_* consists in stopping on the first entrance of the process $\{X_n\}$ in S (respectively, on the first entrance of $\{\tilde{X}_n\}$ in $\tilde{S}$ for the optimal rule $\tilde{\psi}_*$). On the other hand, the application of the rule $\tilde{\psi}_*$ to the "original" process $\{X_n\}$ means stopping $\{X_n\}$ on the first entrance in $\tilde{S}$. In the last case, the mean gain $G(x, \tilde{\psi}_*)$ is obtained.

The stability index Δ is defined as follows (see, e.g., [12,13]):

$$\Delta(x) := G_*(x) - G(x, \tilde{\psi}_*) \equiv G(x, \psi_*) - G(x, \tilde{\psi}_*) \geq 0, \quad x \in \mathfrak{X}. \tag{17}$$

The index Δ measures the decrease of the average gain (in comparison with the maximal value G_*) when one applies the stopping rule $\tilde{\psi}_*$ (optimal for $\{\widetilde{X}_n\}$) to the "original" process $\{X_n\}$.

Let

$$d(p, \tilde{p}) := \sup_{x \in \mathfrak{X}} ||p(\cdot|x) - \tilde{p}(\cdot|x)||, \tag{18}$$

where $||\nu - \tilde{\nu}||$ is the *total variation norm* of the signed measure $\nu - \tilde{\nu}$.

We refer to the stopping optimization problem as *stable* if

$$\Delta(x) \to 0 \quad \text{as} \quad d(p, \tilde{p}) \to 0, \quad \text{for all} \quad x \in \mathfrak{X}.$$

First, we give an example of an unstable problem.

Example 3.1: Let $\mathfrak{X} = \{0, 1, 2, \dots\}$ and $\epsilon \in (0, 1)$ be arbitrary but fixed, and let the transition probability matrices $p = (p_{i,k}),\ \tilde{p} = (\tilde{p}_{i,k})$ (for $\{X_n\}$ and $\{\widetilde{X}_n\}$, respectively) be defined as follows:

$$p_{0,0} = 1; \qquad p_{i,i-1} = 1 \quad \text{for} \quad i \geq 2;$$
$$p_{1,0} = 1 - \epsilon; \qquad p_{1,k} = \epsilon \left(\frac{\pi^2}{6} - 1\right)^{-1} \frac{1}{k^2}, \quad k = 2, 3, \dots;$$
$$\tilde{p}_{0,0} = 1; \qquad \tilde{p}_{i,i-1} = 1 \quad \text{for} \quad i \geq 1.$$

Let the initial state $x = 1,\ c(i) = 1,\ \ i = 0, 1, 2, \dots;\ \ R(0) = 2$ and $R(i) = 0$ for $i \geq 1$.

It is evident that the optimal for $\{\widetilde{X}_n\}$ stopping rule $\tilde{\psi}_*$ is to stop entering into $\widetilde{S} = \{0\}$, and that $\widetilde{G}(\tilde{\psi}_*) = 1$. Applying this rule to $\{X_n\}$ we get: $G(\tilde{\psi}_*) = -\infty$ since $E\tau = \infty$ for τ being the time of the first entrance of $\{X_n\}$ to $\{0\}$. Thus, for *every* $\epsilon \in (0, 1)$ in (17) $\Delta := \Delta(1) = \infty$. In the same time, it is easy to check for this example (see (18)) that

$$d(p, \tilde{p}) = \frac{1}{2} \sum_{k=0}^{\infty} |p_{1,k} - \tilde{p}_{1,k}| = \epsilon \to 0 \quad \text{as} \quad \epsilon \to 0.$$

Remark 3: It is clear that the reason behind the instability in this example is that the process $\{X_n\}$ reaches the set $\widetilde{S} = \{0\}$ for an infinite, in average, time. We can easily modify this example providing entrance times to $\{0\}$ with finite mean. Indeed, we can choose $p_{1,k} = \epsilon p_{1,k}(\epsilon),\ k \geq 2$ in such a way that $3 \leq E\tilde{\tau} < \infty$. Then $\Delta = \Delta(\epsilon) \geq 1$, for all $\epsilon \in (0, 1)$. (It is even possible to make $\Delta(\epsilon) \to \infty$ as $\epsilon \to 0$ keeping $E\tau < \infty$.)

The above arguments suggest that to prove "stability" we need some upper bounds on certain power moments of entrance time into corresponding "stopping sets".

For random variables ξ and τ with values in $\{0, 1, 2, \dots\}$ we write: $\tau \prec \xi$ (τ is less than ξ in distribution) if

$$P(\tau > k) \leq P(\xi > k) \quad \text{for} \quad k = 0, 1, 2, \dots.$$

For a measurable subset $M \subset \mathfrak{X}$ we define the following entrance times:

$$\tau_Y^M := \inf\{n > 0 :\ Y_n \in M\}, \tag{19}$$

where $Y = \{Y_n\} \in \left\{\{X_n\}, \{\widetilde{X}_n\}\right\} \equiv \{X, \tilde{X}\}$.

Assumption 2: *Suppose that a pair of measurable subsets $Q, \tilde{Q} \subset \mathfrak{X}$ can be chosen in such a way that they satisfy Assumption 3 below, and that*

$$\tau_X^S \prec \tau_X^Q; \quad \tau_X^{\tilde{S}} \prec \tau_X^{\tilde{Q}}; \quad \tau_{\tilde{X}}^S \prec \tau_{\tilde{X}}^Q; \quad \tau_{\tilde{X}}^{\tilde{S}} \prec \tau_{\tilde{X}}^{\tilde{Q}}.$$

The next assumption involves so-called "drift conditions" (expressed in terms of the "stochastic Lyapunov function"). Such sort of conditions is widely used in the theory of Markov processes (see, e.g., [21]). We borrow the drift conditions that provide "accessibility" from [19], Ch. 5. The functions and constants implicated in Assumption 3 below can be explicitly calculated for many specific (applied) Markov chains (see [19,21] and Example 3.3 below).

Assumption 3: *There exists a measurable function $V : \mathfrak{X} \to \mathbb{R}^+$, numbers $s > 1$, $\varkappa > 0$, $l < \infty$ such that:*

(a) $\sup_{x\in Q} V(x) < \infty$.

(b) $\sup_{x\in\mathfrak{X}} E|V(X_1|X_0 = x) - V(x)|^s \le l$.

(c) $\sup_{x\notin Q} E[V(X_1|X_0 = x) - V(x)] \le -\varkappa$.

(d) *The conditions (a)–(c) hold for the process $\{\widetilde{X}_n\}$ (with the same s, but possibly different $\widetilde{V}, \tilde{\varkappa}, \tilde{l}$).*

(e) *The same conditions are satisfied for both processes $\{X_n\}$ and $\{\widetilde{X}_n\}$ if we replace Q by $\widetilde{Q}$ in (a)–(d) (with the same exponent $s > 1$, but possibly different other parameters).*

Suppose that for Markov chains $\{X_n\}$ and $\{\tilde{X}_n\}$ the same initial state $x = x_0 \notin Q \cup \tilde{Q}$ is fixed.

Theorem 4: *Let Assumptions 2 and 3 hold. Then*

$$\Delta(x_0) \le K(x_0)\ [d(p, \tilde{p})]^{\frac{s-1}{s+1}}, \tag{20}$$

where the distance d is defined in (18), and $K(x_0)$ is an explicitly calculated constant.

As one can see from the proof of (20), the constant $K(x_0)$ is completely determined by x_0, by $b := \max\{\sup_{x\in\mathfrak{X}} R(x), \sup_{x\in\mathfrak{X}} c(x)\}$ and by the quantities involved in Assumption 3.

Not to make formulas too cumbersome we give the expression for $K(x_0)$ only in the special case when in Assumptions 2, 3 $Q = \tilde{Q}$, and the "test function" V and constants κ, l are the same for both processes $\{X_n\}$ and $\{\tilde{X}_n\}$ (see Example 3.3 for such a case). Let

$$\alpha_s = \left[\frac{2^s}{s\kappa}\left(1 + \left(\frac{2}{\kappa}\right)^s l\right)\right]^{\frac{1}{s-1}} \text{ for } 1 < s \le 2, \tag{21}$$

$$\alpha_s = \max\left\{1, \frac{2^{s-3}(s-1)}{\kappa}\left[\left(\frac{2}{\kappa}\right)^2 l^{\frac{2}{s}} + 2^s\left(1 + \left(\frac{2}{\kappa}\right)^s l\right) - 1\right]\right\}, \text{ for } s > 2. \tag{22}$$

Then in (20)

$$K(x_0) = 4b\left[4 + \left(\alpha_s + \frac{2V(x_0)}{\kappa}\right)^s\right]. \tag{23}$$

https://doi.org/10.1017/S0269964814000035 Published online by Cambridge University Press

Remark 4: It can be shown that inequality (20) holds true (with a constant different from $K(x_0)$) if we replace the boundedness of the maintenance cost c with the condition: $c(x) \geq \alpha > 0$, $x \in \mathfrak{X}$.

Concerning possible applications of the "stability inequality" (20), a natural questions arises: "How can one know (or estimate) the distance $d(p, \tilde{p})$ in (20)? We can give a sound answer at least in two important cases. Number one. In some applied model, the transition probability p can be known, but it leads to optimality equation (16), which is beyond the hope to be solved. A possible way to get around (i.e., to approximate ψ_*) is to replace p with some transition probability $\tilde{p}$ with a simpler structure. Such approach was realized, for instance, in the paper [20].

Number two. An upper bound of $d(p, \tilde{p})$ can be frequently found when $\tilde{p}$ is certain statistical estimation of the unknown transition probability p. We offer only the simplest example to illustrate this point.

Example 3.2: Suppose that $\mathfrak{X}$ is some measurable subset of $\mathbb{R}$ and that

$$X_n = F(X_{n-1}, \xi_n),$$

$n = 1, 2, \ldots$, and $\xi_1, \xi_2, \ldots$ are i.i.d. random variables with an *unknown density* f.

Supposing that $\xi_1, \xi_2, \ldots$ are observable, let $\tilde{f}_m$ be a kernel-type statistical estimation of f based on a sample $\xi_1, \ldots, \xi_m$, $m = 1, 2, \ldots$ (see, e.g., [7]). The according "approximating process" is $\tilde{X}_n = F(\tilde{X}_{n-1}, \tilde{\xi}_n)$, $n = 1, 2, \ldots$, where $\tilde{\xi}_1, \tilde{\xi}_2, \ldots$ are i.i.d. random variables with the density $\tilde{f}_m$. Under certain restrictions (for instance, assuming that $\frac{\partial}{\partial s} F(x, s) > 0$ for all $x, s \in \mathbb{R}$) we obtain for $d(p, \tilde{p})$ in (18) the following expression:

$$d(p, \tilde{p}) = \sup_{x \in \mathbb{R}} \int_{-\infty}^{\infty} |f_{F(x,\xi)}(t) - f_{F(x,\tilde{\xi})}(t)| dt = \int_{-\infty}^{\infty} |f(t) - \tilde{f}_m(t)| dt. \qquad \textbf{(24)}$$

If it is known (or supposed) that the unknown density f belongs to a certain class (for instance, with integrable second derivatives and "fast enough" vanishing tails) then the estimators $\tilde{f}_m$ in (24) can be constructed in such a manner that (see, e.g., [7])

$$E \int_{-\infty}^{\infty} |f(t) - \tilde{f}_m(t)| dt \leq B m^{-\gamma}, \quad m = 1, 2, \ldots,$$

where the constants $B < \infty$ and $\gamma > 0$ depend only on the characteristics of the considered class of densities.

Consequently, from (20) we get:

$$E\Delta(x_0) \leq K(x_0) B^{\frac{s-1}{s+1}} m^{-\gamma \frac{s-1}{s+1}}, \quad m = 1, 2, \ldots.$$

Finally, we examine an example of stopping an applied Markov process for which the constant $K(x_0)$ in (20) can be written in terms of simple parameters of the processes.

Example 3.3: Consider the Markov chain $\{X_n\}$ on $\mathfrak{X} = \{0, 1, 2, \ldots\}$, where X_n represents the number of customers occupying the $M_\lambda |GI|\ 1\ |\infty$ queuing system just after the departure of the nth customer. In this model the input flow is a Poisson process with parameter λ, and successive service times $\eta_1, \eta_2, \ldots$ are nonnegative i.i.d. random variables with a common distribution function F. The transition probabilities p_{ik}, $i, k \in \mathfrak{X}$ are easily calculated in terms of λ and F.

Suppose that the parameter λ is known, but F is not, and the latter is approximated by an admissible distribution function $\widetilde{F}$ (for instance, obtained from statistical estimates). The i.i.d. random variables with the distribution function $\widetilde{F}$ we denote by $\tilde{\eta}_1, \tilde{\eta}_2, \ldots$. Thus the chain $\{X_n\}$ is approximated by a Markov chain $\{\widetilde{X}_n\}$ with the transition probabilities $\tilde{p}_{ik}$ calculated using λ and $\tilde{F}$.

Along with the boundedness of R and c we assume the following:

(a) the distribution functions F, $\widetilde{F}$ have densities f and $\tilde{f}$;

(b) there are $s > 1$ and constants $h, \tilde{h} < \infty$ and $\gamma, \tilde{\gamma} < 1$ such that

$$\lambda\, E\eta_1 \le \gamma, \quad \lambda\, E\tilde{\eta}_1 \le \tilde{\gamma}, \quad E\eta_1^s \le h, \quad E\tilde{\eta}_1^s \le \tilde{h}; \tag{25}$$

(c) there is an integer $L \ge 0$, such that $R(k) = 0$ for $k > L$, and $\inf_{k\ge 0} c(k) > 0$;

(d) the initial state $x_0 > L$;

(e) $\sup_\psi G(x_0, \psi) > 0, \quad \sup_\psi \widetilde{G}(x_0, \psi) > 0$.

Condition (c) implies the existence of optimal stopping rules for $\{X_n\}$ and $\{\widetilde{X}_n\}$. Moreover by (c) and (e) we can state that $S, \widetilde{S} \subset [0, L]$, where S and $\widetilde{S}$ are the corresponding "optimal stopping sets".

As it was shown in [19], Section 5.3, conditions (25) yield the fulfillment of Assumption 3 for $Q = \widetilde{Q} = \{0\}$ with $V(k) = k,\ k = 0, 1, 2, \ldots$. Assumption 2 also holds true. (Passing from x_0 to 0 the chains can not avoid any state $x_0 - 1, x_2 - 2, \ldots, 1$ and so the stopping sets $S, \widetilde{S}$.) Therefore, we can apply inequality (20), where by simple calculations we get that

$$d(p, p) = \frac{1}{2} \sup_{i\ge 0} \sum_{k=0}^{\infty} |p_{i,k} - \tilde{p}_{i,k}| \le \frac{1}{2} \int_0^\infty |f(t) - \tilde{f}(t)| dt.$$

Finally,

$$\Delta \le K(x_0)\, 2^{\frac{s+1}{s-1}} \left[\int_0^\infty |f(t) - \tilde{f}(t)| dt \right]^{\frac{s-1}{s+1}}.$$

Remark 5: Assuming additionally (for the sake of simplicity) that $\gamma = \tilde{\gamma}$, $h = \tilde{h}$, and that, for example, $s = 2$, we can show that $K(x_0) = 4b \left\{ 4 + \left[\frac{2}{1-\gamma}(1 + \frac{4}{(1-\gamma)^2} h) + \frac{2x_0}{1-\gamma} \right]^2 \right\}$

Acknowledgements

The authors thank Mexican CONACyT (National Council for Science and Technology) for a partial support of this work.

The authors are very grateful to the referees and to the Associate Editor for very helpful comments and suggestions which contributed to a considerable improvement of an early version of this paper.

References

1. Allart, P. (2004). Optimal stopping rules for correlated random walks with a discount. *Journal of Applied Probability* 41: 483–496.
2. Allart, P. & Montecino, M. (2008). Optimal buy/sell rules for correlated random walks. *Journal of Applied Probability* 45: 33–44.
3. Arrow, K.J., Blackwell, D. & Girshik, M.A. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica* 17 (3/4): 213–244.
4. Bacchus, F., Boutilier, C. & Grove, A. (1997). Structured Solution Methods for non-markovian decision processes. Proc. 14th National Conf. AI (AAAI-97), Providence, August.

5. Baird, H.S. & Mallows, C.L. (1997). The evolution of a problem. *Statistica Sinica* 7: 211–220.
6. Chow, Y.S., Robbins, H. & Siegmund, D. (1971). *Great Expectations: The Theory of Optimal Stopping.* Boston: Houghton Mifflin Company.
7. Devroye, L. (1985). *Nonparametric Density Estimation. The L_1 View*. New York: Wiley.
8. Dufour, F. & Piunovski, A.B. (2010). Multiobjective stopping problem for discrete-time Markov processes: convex analytic approach. *Journal of Applied Probability* 47: 947–966.
9. Ferguson, T.S. (1967). *Mathematical Statistics: A Decision Theoretic Approach*. New York: Academic Press.
10. Ghosh, M., Mukhopadhyay, N. & Sen, P.K. (1997). *Sequential Estimation*. New York: Wiley.
11. Gikhman, I.I. & Skorokhod, A.V. (1977). *Controllable Stochastic Processes* (in Russian). Kiev: Naukova dumka.
12. Gordienko, E.I. (1992). An estimate of the stability of optimal control of certain stochastic and deterministic systems. *Journal of Soviet Mathematics* 59: 891–899 (Translated from Russian publication of 1989).
13. Gordienko, E.I. & Yushkevich A.A. (2003). Stability estimates in the problem of average optimal switching of a Markov chain. *Mathematical Methods of Operations Research* 57: 345–365.
14. Han, D. & Tsung, F. (2009). The optimal stopping time for detecting changes in discrete Markov processes. *Sequential Analysis* 28: 115–135.
15. Hinderer, K. (1970). *Foundations of non-stationary dynamic programming with discrete time parameter*. Lecture Notes in Operations Research and Mathematical Systems, **33**, Berlin: Springer.
16. Hordijk, A. (1974). *Dynamic Programming and Markov Potential Theory*, 2nd ed., Mathematical Centre Tracts **51**, Amsterdam: Mathematisch Centrum.
17. Horiguchi, M. (2001). Markov decision processes with a stopping time constraint. *Mathematical Methods of Operations Research* 53: 279–295.
18. Jensen, V. (1997). An optimal stopping problem in risk theory. *Scand Actuarial J.* 2:149–159.
19. Kalashnikov, V.V. (1994). *Mathematical Methods in Quening Theory*. Dordrecht: Kluwer Academic.
20. Kühne, R. & Rüschendorf, L. (2004). Approximate optimal stopping of dependent sequences. *Theory of Probability and its Application*, 48: 465–480.
21. Meyn, S.P. & Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. London: Springer-Verlag.
22. Matskyavichyus, V. (1973). Passing to the limit in optimal stopping problem for Markov processes. *Translated from Litovskii Matematicheskii Sbornik*, 13(1): 115–128.
23. Morali, N. & Soyer, R. (2002). Optimal stopping in software testing. *Naval Research Logistics*, 50(1): 88–104.
24. Muciek, B.K. (2002). Optimal stopping of a risk process model with interest rates. *Journal of Applied Probability* 39: 261–270.
25. Novikov, A. (2009). Optimal sequential multiple hypothesis testing in presence of control variables. *Kybernetika* 45(3): 507–528.
26. Novikov, A. (2009). Optimal sequential multiple hypothesis tests. *Kybernetika* 45(2): 309–330.
27. Novikov, A. (2009). Optimal sequential tests for two simple hypotheses. *Sequential Analysis* 28(2): 188–217.
28. Novikov, A. (2010). Optimal sequential procedures with Bayes decision rules. *Kybernetika* 46(4): 754–770.
29. Novikov, A. & Novikov P. (2010). Locally most powerful sequential tests of a simple hypothesis vs. one-sided alternatives. *Journal of Statistical Planning and Inference* 140(3): 750–765.
30. Rieder, U. (1975). On stopped decision processes with discrete time parameter. *Stochastic Processes and their Applications* 3: 365–383.
31. Ross, Sh.M. (1970). *Applied probability models with optimization applications*, New York: Dover Publications, Inc. plans and sequential probability ratio tests. *Journal of the American Statistical Association* 65(329): 431–437.
32. Shiryaev, A.N. (1978). *Optimal Stopping Rules*. New York: Springer-Verlag.
33. Shiryaev, A.N. (1999). *Essential of Stocastic Finance: Facts, Models, Theory*. River Edge, NJ: World Scientific.
34. Thiebaux, S., Gretton, Ch., Slaney, J. & Price. D. (2006). Decision-theoretic planning with non-Markovian rewards. *Journal of Artificial Intelligence Research* 25: 17–74.
35. Wald, A. (1947). *Sequential Analysis*. New York: Wiley.
36. Wald, A. & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Annals of Mathematical Statistics* 19: 326–339.
37. Zaitseva, E. (2008). Stability estimating in optimal stopping problem. *Kybernetika* 44: 400–415.

APPENDIX A. PROOFS OF THE MAIN RESULTS OF SECTION 2

A.1 Proof of Theorem 1

The proof is based on the simple lemma below which is essentially a reformulation of Lemma A.2 in [29].

For any random variable v with $E|v| < \infty$ let us define:

$$B_k v = E\{v|X_1, \ldots, X_k\}.$$

Let also

$$D_k v = \max\{R_k, B_k v - c_k\}.$$

Lemma 4: *Let v be any random variable such that $E|v| < \infty$. Then*

$$Es_k^{\psi}(R_k - C_k) + Et_{k+1}^{\psi}(v - C_{k+1}) \leq Et_k^{\psi}(D_k v - C_k), \tag{A.1}$$

where $E|D_k v| < \infty$.

There is an equality in (A.1) if and only if

$$I_{\{R_k > B_k v - c_k\}} \leq \psi_k(X_1, \ldots, X_k) \leq I_{\{R_k \geq B_k v - c_k\}}$$

with probability one on $T_k^{\psi} = \{t_k^{\psi}(X_1, \ldots, X_k) > 0\}$.

Let now for any natural $1 \leq n \leq N$

$$\widetilde{G}_n^N(\psi) = \sum_{i=1}^{n-1} Es_i^{\psi}(R_i - C_i) + Et_n^{\psi}(V_n^N - C_n). \tag{A.2}$$

It is easy to see that $G_N(\psi) = \widetilde{G}_N^N$ (see (6)) and that $\widetilde{G}_1^N(\psi) = Q_0^N$ (see (8)). Applying Lemma 4 initially to $v \equiv V_N^N$ we have

$$G_N(\psi) \leq \sum_{n=1}^{N-2} Es_n^{\psi}(R_n - C_n) + Et_N^{\psi}(V_{N-1}^N - C_{N-1}) = \widetilde{G}_{N-1}^N, \tag{A.3}$$

where the equality is attained if and only if

$$I_{\{R_{N-1} > Q_{N-1}^N - c_{N-1}\}} \leq \psi_{N-1}(X_1, \ldots, X_{N-1}) \leq I_{\{R_{N-1} \geq Q_{N-1}^N - c_{N-1} f^{N-1}\}}$$

with probability one.

Starting from (A.3), it is easy to see by induction, using Lemma 4 again, that for all $n = N-1, \ldots, 1$

$$\widetilde{G}_{n+1}^N(\psi) \leq \widetilde{G}_n^N(\psi) \tag{A.4}$$

with an equality if and only if

$$I_{\{R_n > Q_n^N - c_n\}} \leq \psi_n(X_1, \ldots, X_n) \leq I_{\{R_n \geq Q_n^N - c_n\}} \tag{A.5}$$

μ^n-almost everywhere on T_n^{ψ}. Thus, we have, in particular,

$$G_N(\psi) = \widetilde{G}_N^N(\psi) \leq \widetilde{G}_1^N(\psi) = Q_0^N$$

with an equality if and only if (A.5) holds with probability one on T_n^{ψ} for all $n = N-1, \ldots, 1$.

A.2. Proof of Theorem 2

Let for any $\psi \in \mathscr{F}$

$$\widetilde{G}_n(\psi) = \sum_{i=1}^{n-1} Es_i^{\psi}(R_i - C_i) + Et_n^{\psi}(V_n - C_n).$$

Passing to the limit, as $N \to \infty$, on both sides of (A.4), we have

$$\widetilde{G}_{n+1}(\psi) \leq \widetilde{G}_n(\psi), \tag{A.6}$$

and passing to the limit, as $N \to \infty$, in (7), we obtain

$$V_n = \max\{R_n, Q_n - c_n\},$$

where $Q_n = \lim_{N\to\infty} Q_n^N$ satisfies the equation

$$Q_n = E\{V_{n+1}|X_1, \ldots, X_n\},$$

$n = 0, 1, 2, \ldots$. Obviously, $\widetilde{G}_1(\psi) = Q_0$, so it follows from (A.6) that for all natural n

$$G(\psi) \leq \widetilde{G}_n(\psi) \leq \widetilde{G}_{n-1}(\psi) \leq \ldots \widetilde{G}_1(\psi) = Q_0. \tag{A.7}$$

Let us suppose that there exists some stopping rule $\psi \in \mathscr{F}$ such that $G(\psi) = Q_0 < \infty$. Then there are equalities in all of the inequalities in (A.7).

Taking into account the necessary condition in Lemma 4, we see that it is only possible when

$$I_{\{R_n > Q_n - c_n\}} \leq \psi_n \leq I_{\{R_n \geq Q_n - c_n\}} \tag{A.8}$$

almost surely on T_n^{ψ}, for all natural n. In addition, we have that for all natural n

$$\widetilde{G}_n(\psi) = \sum_{i=1}^{n-1} Es_i^{\psi}(R_i - C_i) + Et_n^{\psi}(V_n - C_n) = Q_0. \tag{A.9}$$

Because, by supposition, $\psi \in \mathscr{F}$, it follows that $G_n(\psi) \to G(\psi) = Q_0$, as $n \to \infty$, thus (13) follows from (A.9).

On the other hand, suppose now that a stopping rule $\psi \in \mathscr{F}$ satisfies (A.8) almost surely on T_n^{ψ}, for all natural n, and that (13) holds.

Then, by Lemma 4, there are equalities in all of the inequalities (A.7) except, probably, for the first one. In particular, we have that (A.9) holds for all natural n. Then, it follows from (A.9) and (13) that $G_n(\psi) \to Q_0$, as $n \to \infty$. Because, by supposition, $\psi \in \mathscr{F}$, we have that $G(\psi) = \lim_{n\to\infty} G_n(\psi) = Q_0$, or

$$G(\psi) = \sup_{\psi' \in \mathscr{F}} G(\psi').$$

A.3. Proof of Lemma 2

Let us suppose that (14) is satisfied. Then for all $\psi \in \mathscr{F}_1$ $G(\psi)$ is well-defined, because, by definition, $\bar{R}(\psi) < \infty$. Thus, property 1 of Assumption 1 is satisfied. Properties 2 and 3 of Assumption 1 are also satisfied by the definition of $\mathscr{F}_1$.

Let us prove that for $\psi \in \mathscr{F}_1$ property 4 of Assumption 1 holds.

https://doi.org/10.1017/S0269964814000035 Published online by Cambridge University Press

Let $\psi \in \mathscr{F}_1$ be such that $G(\psi) > -\infty$. It follows that $\bar{C}(\psi) < \infty$, thus,

$$\sum_{n=N}^{\infty} Es_n^{\psi} C_n \to 0, \quad \text{as} \quad N \to \infty. \tag{A.10}$$

Because $C_n(x_1, \ldots, x_n) \geq C_N(x_1, \ldots, x_N)$ for all $(x_1, \ldots, x_n) \in \mathfrak{X}^n$, $(x_1, \ldots, x_N) \in \mathfrak{X}^N$ for all $n \geq N$, it follows from (A.10) that

$$\sum_{n=N}^{\infty} Es_n^{\psi} C_N = Et_N^{\psi} C_N \to 0 \quad \text{as} \quad N \to \infty. \tag{A.11}$$

On the other hand,

$$\sum_{n=N}^{\infty} Es_n^{\psi} R_n \to 0, \tag{A.12}$$

as $N \to \infty$, because $\bar{R}(\psi) = \sum_{n=1}^{\infty} Es_n^{\psi} R_n < \infty$ by the definition of $\mathscr{F}_1$.

We have now

$$G(\psi) - G_N(\psi) = \sum_{n=N}^{\infty} Es_n^{\psi} R_n - \sum_{n=N}^{\infty} Es_n^{\psi} C_n - Et_N^{\psi} R_N + Et_N^{\psi} C_N,$$

where the first two summands tend to 0 as $N \to \infty$ by virtue of (A.10) and (A.12), the last one tends to 0 by (A.11), and so does the third summand, because $t_N^{\psi} R_N$ is uniformly integrable (due to $t_N^{\psi} \leq 1$), and it tends to 0 in probability (because $Et_N^{\psi} = P(\tau_{\psi} \geq N) \to 0$, due to the fact that $\psi \in \mathscr{F}_1$), so $Et_N^{\psi} R_N \to 0$ as $N \to \infty$.

A.4. Proof of Theorem 3

Let $R_N^* = \sup_{1 \leq n \leq N} R_n(X_1, \ldots, X_n)$ and $R^* = \sup_{N \geq 1} R_N^*$.

First of all, it is easy to see, by induction, that under the conditions of the Theorem $V_n^N \leq E\{R_N^* | X_1, \ldots, X_n\}$ with probability 1, for all $n \leq N$, for all natural N (see the definitions in (6), (7) and (8)). Thus,

$$V_n \leq E\{R^* | X_1, \ldots, X_n\},$$

with probability 1 for all natural n. Because of this, for every ψ such that $P(\tau_{\psi} < \infty) = 1$

$$Et_n^{\psi} V_n \leq Et_n^{\psi} R^* \to 0 \tag{A.13}$$

as $n \to \infty$ (because $ER^* < \infty$ and $Et_n^{\psi} \to 0$ as $n \to \infty$). Analogously,

$$Et_n^{\psi} R_n \to 0 \tag{A.14}$$

as $n \to \infty$. It follows from (A.13) and (A.14) that (13) is satisfied, thus, by Theorem 2, (11) holds.

Let us suppose now, additionally, that (3) holds and let ψ satisfy (12). Again, we can use (A.9), which is valid for all natural n. Because $\sum_{i=1}^{n-1} Es_i^{\psi} R_i + Et_n^{\psi} R_n \leq ER^* = M < \infty$, and (see (A.13) and (A.14)) $Et_n^{\psi}(V_n - R_n) \to 0$ as $n \to \infty$, it follows from (A.9), in particular, that the last summand in (A.9)

$$Et_n^{\psi} C_n \leq M - Q_0 + 1$$

for all sufficiently large n.

Let us prove that $\lim_{n\to\infty} Et_n^{\psi} = 0$. Suppose the contrary, i.e., that $Et_n^{\psi} \to \epsilon > 0$ as $n \to \infty$. Let $k > 0$ be any real number. Then

$$
\begin{aligned}
M - Q_0 + 1 &\geq Et_n^{\psi} C_n = Et_n^{\psi} C_n I_{\{C_n \geq k\}} + Et_n^{\psi} C_n I_{\{C_n < k\}} \geq k Et_n^{\psi} I_{\{C_n \geq k\}} \\
&\geq k(Et_n^{\psi} - P(C_n < k)) \geq k(\epsilon - P(C_n < k)) \geq k\epsilon/2
\end{aligned}
$$

if n is sufficiently large, by (3). This implies that $k \leq 2(M - Q_0 + 1)/\epsilon$, which is a contradiction, because k is arbitrarily large.

Thus, we showed that $\lim_{n\to\infty} Et_n^{\psi} = \lim_{n\to\infty} P(\tau_{\psi} \geq n) = P(\tau_{\psi} = \infty) = 0$, that is, $P(\tau_{\psi} < \infty) = 1$. The rest follows from the already proved part of the theorem.

APPENDIX B. PROOF OF THEOREM 4

In a standard manner, we introduce auxiliary Markov decision processes $\{Z_n\} \equiv \{Z_n,\ n = 0, 1, \dots\}$ and $\{\widetilde{Z}_n\} \equiv \{\widetilde{Z}_n,\ n = 0, 1, \dots\}$, which correspond to stopping $\{X_n\}$ and $\{\widetilde{X}_n\}$. These processes are defined on the state space $\mathscr{X} = \mathfrak{X} \cup \{*\}$, where "$*$" stands for absorbing state where processes move out after being stopped. The action space is $A = \{0, 1\}$, where the action $a = 0$ results in stopping, and the action $a = 1$ means continuation of observations.

The corresponding one-step return $r(z, a)$, $z \in \mathscr{X}$, $a \in A$ is defined by the functions c and R.

The transition probability function for $\{Z_t\}$ is

$$
q(*|z, 0) = 1, \quad z \in \mathscr{X},
$$

$$
q(D|z, 1) := \begin{cases} p(D \setminus \{*\}|x), & if \quad z = x \in \mathfrak{X}, \\ \quad 1 & if \quad z = * \quad and \quad * \in D, \\ \quad 0 & if \quad z = * \quad and \quad * \notin D;\ D \in \mathscr{B}(\mathscr{X}). \end{cases} \qquad \textbf{(B.1)}
$$

Replacing in (B.1) p by $\tilde{p}$, we define the transition probability $\tilde{q}$ for the process $\{\widetilde{Z}_n\}$.

The "stopping sets" S and $\widetilde{S}$ corresponding to the optimal stopping rules ψ_* and $\tilde{\psi}_*$ determine the *stationary policies of control* $f = (f, f, \dots)$ and $\tilde{f} = (\tilde{f}, \tilde{f}, \dots)$, which prescribe to stop on first entrance in S (in $\tilde{S}$, respectively).

Under Assumption 3 from Theorem 2 in Section 5.2 [19], we get that for each processes $\{X_n\}, \{\widetilde{X}_n\}$ and for each sets $M = S$ or $M = \widetilde{S}$ the sth power moment of the corresponding entrance time defined in (19) are bounded by certain calculable constant. Thus in view of Assumption 2, we can find a constant $\mu < \infty$, such that μ is an upper bound for sth power moment of the first entrance times of $\{X_n\}$ and of $\{\tilde{X}_n\}$ in each set S and $\tilde{S}$.

Using the last fact and the boundedness of the return function $r(x, a)$, by standard arguments we can show that in (17) $G(x_0, \psi_*) = w(x_0, f)$; $G(x_0, \tilde{\psi}_*) = w(x_0, \tilde{f})$, where

$$
w(x, \varphi) := E_x^{\varphi}[\sum_{n=0}^{\infty} r(Z_n, \varphi(Z_n))] \qquad \textbf{(B.2)}
$$

(φ is a stationary policy), and that the above mentioned policies f and $\tilde{f}$ are optimal stationary policies with respect to criterion (B.2) and its counterpart $\tilde{w}(x, \varphi)$ defined by (B.2) replacing $\{Z_n\}$ by $\{\tilde{Z}_n\}$.

Consequently, the stability index in (17) can be rewritten as follows (omitting the fixed initial state x_0 in the notation): $\Delta = w(f) - w(\tilde{f})$.

Now,

$$\begin{aligned} 0 &\le \Delta = w(f) - \tilde{w}(\tilde{f}) + \tilde{w}(\tilde{f}) - w(\tilde{f}) \\ &= \max_{\varphi\in\{f,\tilde{f}\}} w(\varphi) - \max_{\varphi\in\{f,\tilde{f}\}} \tilde{w}(\varphi) + \tilde{w}(\tilde{f}) - w(\tilde{f}) \\ &\le 2\ \max_{\varphi\in\{f,\tilde{f}\}} |w(\varphi) - \tilde{w}(\varphi)|. \end{aligned} \qquad \textbf{(B.3)}$$

We will estimate the quantity $I := |w(f) - \tilde{w}(f)|$. (For $\varphi = \tilde{f}$ the same calculations are valid.) For any given $N \ge 1$ using (B.2) we write (simplifying the notation for E_x^{φ}):

$$I \le \left| E \sum_{n=0}^{N-1} [r(Z_n, f(Z_n)) - r(\widetilde{Z}_n, f(\widetilde{Z}_n))] \right| \qquad \textbf{(B.4)}$$

$$+ \left| E \sum_{n\ge N} r(Z_n, f(Z_n)) \right| + \left| E \sum_{n\ge N} r(\widetilde{Z}_n, f(\widetilde{Z}_n)) \right| =: I_N + I'_N + I''_N. \qquad \textbf{(B.4)}$$

By the definition of the total variation norm, $I_N \le 2b \sum_{n=0}^{N-1} ||\mathscr{D}(Z_n) - \mathscr{D}(\widetilde{Z}_n)||$, where $b := \max\{\sup_{x\in\mathfrak{X}} R(x),\ \sup_{x\in\mathfrak{X}} c(x)\}$, and $\mathscr{D}(Z)$ is the distribution of the random element Z. It is easy to show (see, e.g., [37]) that for $n \ge 1$

$$||\mathscr{D}(Z_n) - \mathscr{D}(\widetilde{Z}_n)|| \le 2n \sup_{x\in\mathfrak{X}} ||p(\cdot|x) - \tilde{p}(\cdot|x)||.$$

Therefore,

$$I_N \le 2\, b\, N^2\, \epsilon, \qquad \textbf{(B.5)}$$

where $\epsilon := \sup_{x\in\mathfrak{X}} ||p(\cdot|x) - \tilde{p}(\cdot|x)|| \in [0, 1]$.

Recalling that the application of the policy f means stopping in the set S, let T be the time of the first entrance of $\{X_n\}$ in S. Then, since $r\big(Z_n, f(Z_n)\big) = 0$ for $n > T$, we get for the second summand in (B.4) the following inequalities:

$$\begin{aligned} I'_N &= \left| E \sum_{n=N}^{T} r(Z_n, f(Z_n))\ I_{\{T\ge N\}} \right| \le b\ E \sum_{n=N}^{T} I_{\{T\ge N\}} \le b\ E(T-N)\ I_{\{T\ge N\}} \\ &\le b\ ET\ I_{\{T\ge N\}} \le b\ \big(ET^s\big)^{1/s}\ [P(T\ge N)]^{\frac{s-1}{s}} \le b\big(ET^s\big)^{1/s} \frac{(ET^s)^{\frac{s-1}{s}}}{N^{s-1}} \le \frac{b\mu}{N^{s-1}}, \end{aligned} \qquad \textbf{(B.6)}$$

due to the Hölder and Markov inequalities, where μ is the above mentioned upper bound of ET^s. The last summand I''_N in (B.4) is estimated similarly. Therefore, by (B.4), (B.5) and (B.6),

$$I \le b\left(2\ N^2\epsilon + \frac{2\mu}{N^{s-1}}\right). \qquad \textbf{(B.7)}$$

We choose $N = N(\epsilon) := \left[\frac{1}{\epsilon^{\alpha}}\right] + 1$, where $\alpha = \frac{1}{s+1}$ and $[\cdot]$ denotes the integer part. Then by simple calculations from (B.7) it follows that

$$I \le\ \epsilon^{\frac{s-1}{s+1}}\ 2\, b\, (4+\mu).$$

Finally, from (B.3) we obtain that

$$\Delta \le 4\, b\, (4+\mu)\, \epsilon^{\frac{s-1}{s+1}}. \qquad \textbf{(B.8)}$$

Under the “drift conditions” given in Assumption 3, the constant μ in (B.8) is bounded by the second summand in brackets in (23), with α_s defined in (21) and (22). It follows from Theorem 2 on page 116 in [19].