



# Discrete stop-or-go games

János Flesch<sup>1</sup> · Arkadi Predtetchinski<sup>2</sup> · William Sudderth<sup>3</sup>

Accepted: 24 January 2021 / Published online: 24 March 2021  
© The Author(s) 2021

## Abstract

Dubins and Savage (How to gamble if you must: inequalities for stochastic processes, McGraw-Hill, New York, 1965) found an optimal strategy for limsup gambling problems in which a player has at most two choices at every state  $x$  at most one of which could differ from the point mass  $\delta(x)$ . Their result is extended here to a family of two-person, zero-sum stochastic games in which each player is similarly restricted. For these games we show that player 1 always has a pure optimal stationary strategy and that player 2 has a pure  $\epsilon$ -optimal stationary strategy for every  $\epsilon > 0$ . However, player 2 has no optimal strategy in general. A generalization to  $n$ -person games is formulated and  $\epsilon$ -equilibria are constructed.

**Keywords** Stochastic game · Optimal strategy · Equilibrium · Limsup payoff · Liminf payoff

## 1 Introduction

A general theory of when to stop playing a sequence of games was developed toward the middle of the previous century. This theory of “optimal stopping” is presented in the monographs of Chow et al. (1971) and Shiryaev (1973). At roughly the same

---

The authors are grateful for the comments of the associate editor and an anonymous reviewer.

---

✉ János Flesch  
j.flesch@maastrichtuniversity.nl  
Arkadi Predtetchinski  
a.predtetchinski@maastrichtuniversity.nl  
William Sudderth  
bill@stat.umn.edu

<sup>1</sup> Department of Quantitative Economics, Maastricht University, P.O. Box 616, Maastricht 6200 MD, The Netherlands

<sup>2</sup> Department of Economics, Maastricht University, P.O. Box 616, Maastricht 6200 MD, The Netherlands

<sup>3</sup> School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA

time, Dubins and Savage (1965) were formulating their theory of gambling problems which encompassed a theory of when to stop playing. The relationship between the two theories is treated by Dubins and Sudderth (1977a).

A class of two-person, zero-sum stopping problems was defined by Dynkin (1969), who also showed his games to have a value. A number of mathematicians have extended these results in various directions such as to non-zero sum games and  $n$ -person games. (See, for example, Rosenberg et al. (2001), Shmaya et al. (2003, 2004), Mashiah-Yaakovi (2014), or Solan and Laraki (2013).)

Here we propose a theory of two-person, zero-sum games related to Dubins–Savage gambling theory. In the formulation of Dynkin, exactly one player was able to halt play at every stage. Here it is possible that one or both or neither of the players can stop at some states. The payoff for Dynkin’s game was zero if play was never stopped whereas we use the Dubins–Savage payoff corresponding to the limsup of the values of a utility function. These games are shown to have a value and good strategies are found for the players. A related class of  $n$ -person games is defined and shown to have approximate equilibria.

## 2 The two-person model

### 2.1 The game

A two-person *stop-or-go game*  $G = (S, A, B, q, u)$  is a two-person, zero-sum stochastic game such that  $S$  is a countable non-empty state space; for each  $x \in S$  there are action sets  $A(x) = \{g, s\}$  or  $\{g\}$  and  $B(x) = \{g, s\}$  or  $\{g\}$  for players 1 and 2, respectively; the law of motion  $q$  satisfies  $q(\cdot|x, g, g) = \alpha(x)$  where  $\alpha(x)$  is a countably additive probability measure defined on all subsets of  $S$  and  $q(\cdot|x, a, b) = \delta(x)$ , the point mass at  $x$ , if either  $a = s$  or  $b = s$ ; the utility function  $u : S \mapsto \mathbb{R}$  is assumed to be bounded. Note that each player has available the “go” action  $g$  at every state but may only have the “stop” action  $s$  at certain assigned states. At any given state, the action  $s$  may be available to one or both of the players or to neither of them. It is possible that  $\alpha(x) = \delta(x)$  for some states  $x$ . Also a player may “stop” temporarily by playing action  $s$  at one stage and action  $g$  at the next stage.

The game is played at stages in  $\mathbb{N} = \{0, 1, \dots\}$ . Play begins at an initial state  $x = x_0 \in S$ . At every stage  $n \in \mathbb{N}$ , the play is in a state  $x_n \in S$ . In this state, player 1 chooses an action  $a_n \in A(x_n)$  and simultaneously player 2 chooses an action  $b_n \in B(x_n)$ . The next state  $x_{n+1}$  has distribution  $q(\cdot|x_n, a_n, b_n)$ . Thus, play of the game generates a random infinite history  $h = (x_0, a_0, b_0, x_1, a_1, b_1, \dots)$ . The payoff from player 2 to player 1 is  $u^*(h) = \limsup_n u(x_n)$ . Of course, this limsup is just  $u(x_k)$  if either player chooses the stop action  $s$  from stage  $k$  onwards.

### 2.2 Strategies and expected payoffs

The set of histories ending at stage  $n$  is denoted by  $H_n$ . Let  $Z = \{(x, a, b)|x \in S, a \in A(x), b \in B(x)\}$ . Then  $H_0 = S$  and  $H_n = Z^n \times S$  for every stage  $n \geq 1$ . Let

$H = \cup_{n \in \mathbb{N}} H_n$  denote the set of all finite histories. For each history  $h \in H$ , let  $x_h$  denote the final state in  $h$ . Let  $H_\infty = Z \times Z \times \dots$  be the collection of all infinite histories  $h = (x_0, a_0, b_0, x_1, a_1, b_1, \dots)$ .

A strategy for player 1 is a map  $\pi$  that to each finite history  $h \in H$  assigns a probability distribution (that is, a mixed action) on  $A(x_h)$ . Similarly, a strategy for player 2 is a map  $\sigma$  that to each history  $h \in H$  assigns a probability distribution on  $B(x_h)$ . Let  $\Pi$  and  $\Sigma$  denote the sets of strategies for players 1 and 2 respectively. A strategy is called stationary if the assigned mixed actions only depend on the history through its final state.

Beginning at some initial state  $x = x_0$ , player 1 chooses a strategy  $\pi$  and player 2 chooses a strategy  $\sigma$ . The strategies together with the law of motion and the initial state  $x$  determine the distribution  $P_{x,\pi,\sigma}$  of the infinite history  $h \in H_\infty$ . The payoff from player 2 to player 1 is the expected value

$$u(x, \pi, \sigma) = E_{x,\pi,\sigma} u^*$$

where  $u^* : H_\infty \rightarrow \mathbb{R}$  is given by  $u^*(x_0, a_0, b_0, x_1, a_1, b_1, \dots) = \limsup_n u(x_n)$ . Player 1's objective is to maximize this expected payoff and player 2 seeks to minimize it.

### 2.3 Value and optimality

A two-person stop-or-go game is a special limsup stochastic game of the type treated in Maitra and Sudderth (1992) and it follows from their results, or from the more general result of Martin (1998), that it has a value  $V(x)$  for every initial state  $x \in S$ , i.e.

$$V(x) = \sup_{\pi \in \Pi} \inf_{\sigma \in \Sigma} u(x, \pi, \sigma) = \inf_{\sigma \in \Sigma} \sup_{\pi \in \Pi} u(x, \pi, \sigma).$$

For  $\epsilon \geq 0$ , a strategy  $\pi \in \Pi$  for player 1 is called  $\epsilon$ -optimal for initial state  $x$  if  $u(x, \pi, \sigma) \geq V(x) - \epsilon$  for every strategy  $\sigma \in \Sigma$  for player 2. Similarly, a strategy  $\sigma \in \Sigma$  for player 2 is called  $\epsilon$ -optimal for initial state  $x$  if  $u(x, \pi, \sigma) \leq V(x) + \epsilon$  for every strategy  $\pi \in \Pi$  for player 1. A strategy is called  $\epsilon$ -optimal if it is  $\epsilon$ -optimal for every initial state. A 0-optimal strategy is called optimal.

### 2.4 n-person games

The two-person model of this section will be generalized to games with an arbitrary finite number of players in Sect. 8 below. The results on two-person games in the earlier sections will be used to construct  $\epsilon$ -equilibria for  $n$ -person games.

### 3 Theorems for two-person stop-or-go games

Our first result is that player 1 always has an easily described optimal stationary strategy.

**Theorem 1** *A pure optimal stationary strategy for player 1 is, at every state  $x$ , to play action  $s$  if  $u(x) = V(x)$  and  $s \in A(x)$ , and to play action  $g$  otherwise.*

If player 2 is a dummy with only one action at every state, then this theorem specializes to give a version of the original result of Dubins and Savage (1965, page 61) for one-person problems. It was their result which led us to the study of stop-or-go games.

Player 2 need not have an optimal strategy, much less a stationary optimal strategy, as the following example, adapted from Sudderth (1983), shows.

**Example 1** Let  $S = \{1, 2, \dots\}$ ;  $u(n) = n^{-1} - 1$  for  $n$  odd,  $u(n) = 0$  for  $n$  even;  $A(n) = \{g\}$ ,  $B(n) = \{s, g\}$ ;  $q(n+1|n, g, g) = 1$  and, by definition of action  $s$ ,  $q(n|n, g, s) = 1$ . The value of the game is  $-1$  at every state because player 2 can play  $g$  a large number of times and then stop at a large odd number. However, no strategy for player 2 can achieve the value. Note that player 1 is a dummy with only one action at every state.

However, player 2 does have nearly optimal stationary strategies.

**Theorem 2** *If  $\epsilon > 0$ , then a pure  $\epsilon$ -optimal stationary strategy for player 2 is to play action  $s$  if  $u(x) \leq V(x) + \epsilon$  and  $s \in B(x)$ , and to play action  $g$  otherwise.*

If the state space is finite, then player 2 has an optimal stationary strategy.

**Corollary 1** *If  $S$  is finite, then a pure optimal stationary strategy for player 2 is to play action  $s$  if  $u(x) \leq V(x)$  and  $s \in B(x)$ , and to play action  $g$  otherwise.*

**Proof** For each positive integer  $n$ , let  $\sigma_n$  be the strategy of Theorem 2 when  $\epsilon = 1/n$ . So, by the theorem,  $\sigma_n$  is  $1/n$ -optimal for player 2. Now each  $\sigma_n$  is a pure stationary strategy and, because  $S$  is finite, there are only finitely many pure stationary strategies. So some strategy, say  $\sigma^*$ , must occur infinitely often in the sequence of the  $\sigma_n$ . It follows that  $\sigma^*$  is optimal and also that  $\sigma^*$  is the strategy described in the statement of the corollary.  $\square$

**Remark 1** Suppose that the rules of a stop-or-go game are made more restrictive in the sense that, whenever a player uses the stop action, the game ends and the payoff to player 1 from player 2 is the utility at the current state. Then no temporary stops are available and a player must at each stage either play the go action  $g$  or stop permanently. The stationary strategies of Theorems 1 and 2 are still available because any stationary strategy that plays  $s$  at a state  $x$  must continue to do so. Thus the theorems still hold and the value of the game is unchanged.

**Remark 2** A drawback of Theorems 1 and 2 is that the stationary strategies they specify depend on the value function  $V$ . However, an algorithm for calculating  $V$  is given in

Maitra and Sudderth (1992). In general, this algorithm requires iterating an operator for a number of steps up to an arbitrary countable ordinal. In the special case when the state space is finite, the algorithm terminates at the first countable ordinal (Theorem 11.13, page 201, Maitra and Sudderth 1996).

**Remark 3** The limsup payoff  $u^*$  is more general than it may first appear. Suppose the state at stage  $n$  is defined to be

$$y_n = (x_0, a_0, b_0, x_1, \dots, x_{n-1}, a_{n-1}, b_{n-1}, x_n)$$

and the utility  $u$  is taken to be a bounded function of the  $y_n$ . For example, the utility function could be of the form

$$u(y_n) = \frac{1}{n} \sum_{k=0}^{n-1} r(x_k, a_k, b_k), \quad n \geq 1$$

where  $r$  is a bounded real-valued function. With this change of variable, the notion of a stationary strategy loses interest. However, Theorems 1 and 2 also tell us that there exist pure subgame perfect strategies for player 1 and pure subgame  $\epsilon$ -perfect strategies for player 2.

The next section has some preliminary results. Section 5 is devoted to the proof of Theorem 1. Section 6 treats the special case of games with 0-1 valued utility functions. Section 7 is for the proof of Theorem 2. Section 8 is on  $n$ -person games. The final section mentions possible generalizations.

## 4 Preliminaries

### 4.1 The optimality equation

For each  $x \in S$ , let  $M(x)$  be the one-shot game with action sets  $A(x)$  for player 1 and  $B(x)$  for player 2 and with payoff for actions  $a \in A(x)$  and  $b \in B(x)$  equal to  $\sum_{y \in S} V(y)q(y|x, a, b)$ .

**Lemma 1** For each  $x \in S$ , the value of  $M(x)$  is  $V(x)$ ; that is,

$$\begin{aligned} V(x) &= \sup_{\mu} \inf_v \sum_{a,b} \sum_{y \in S} V(y)q(y|x, a, b)\mu(a)v(b) \\ &= \inf_v \sup_{\mu} \sum_{a,b} \sum_{y \in S} V(y)q(y|x, a, b)\mu(a)v(b), \end{aligned}$$

where  $\mu$  and  $v$  range over the probability measures on  $A(x)$  and  $B(x)$ , respectively.

This is a standard result for stochastic games. See, for example, Flesch et al. (2018).

### 4.2 Stopping times and stop rules

A stopping time  $\tau$  is a mapping from the space  $H_\infty$  of infinite histories into  $\{0, 1, \dots\} \cup \{\infty\}$  such that, if  $\tau(h) = n$  and  $h'$  agrees with  $h$  up to and including stage  $n$ , then  $\tau(h') = n$ . A stop rule  $t$  is a stopping time such that  $t(h) < \infty$  for all infinite histories  $h \in H_\infty$ .

If  $\tau$  is a stopping time and  $z = (x, a, b) \in Z$ , let  $\tau[z]$  be the function on  $H_\infty$  defined by  $\tau[z](h) = \tau(zh) - 1$  where  $zh$  is the concatenation of  $z$  and  $h$ . If  $\tau(zh) \geq 1$  for some  $h \in H_\infty$ , then the same inequality holds for all  $h \in H_\infty$  and  $\tau[z]$  is easily seen to be a stopping time. Clearly  $\tau[z]$  is a stop rule if  $\tau$  is.

### 4.3 Continuation strategies

Given a finite history  $h = (x_0, a_0, b_0, \dots, x_n)$  and a strategy  $\pi$  for player 1, the continuation strategy of  $\pi$  at  $h$  is the map  $\pi[h]$  that to each finite history  $h' = (x_n, a_n, b_n, \dots, x_m)$  assigns the probability distribution  $\pi(hh')$ , where  $hh' = (x_0, a_0, b_0, \dots, x_n, a_n, b_n, \dots, x_m)$ . Intuitively, in the subgame at  $h$ , this is the strategy induced by  $\pi$ . For a strategy  $\sigma$  of player 2, the continuation strategy  $\sigma[h]$  is defined analogously.

### 4.4 A useful equality

Let  $\tau$  be a stopping time. The following equality can be thought of as separating the payoff into that earned before time  $\tau$  and that after time  $\tau$ :

$$u(x, \pi, \sigma) = \limsup_t \int_{\tau \geq t} u(x_t) dP_{x,\pi,\sigma} + \int_{\tau < \infty} u(x_\tau, \pi[h_\tau], \sigma[h_\tau]) dP_{x,\pi,\sigma}. \tag{1}$$

Here  $t$  varies over the directed set of stop rules. The limsup of a real-valued function of stop rules  $r(t)$  is defined by  $\limsup_t r(t) = \inf_s \sup_{t \geq s} r(t)$  where both  $s$  and  $t$  range over the collection of stop rules. Similarly,  $\liminf_t r(t) = \sup_s \inf_{t \geq s} r(t)$ . The symbol  $h_\tau$  denotes that part of the infinite history  $h \in H_\infty$  up to time  $\tau$  and  $\pi[h_\tau]$  and  $\sigma[h_\tau]$  are the continuation strategies for  $\pi$  and  $\sigma$ . Equality (1) first appeared in Dubins and Sudderth (1977b) and is also on page 66 of Maitra and Sudderth (1996).

It will be convenient to have a slight variation of equality (1) in which  $u(x, \pi, \sigma)$  is replaced by  $u_*(x, \pi, \sigma) = E_{x,\pi,\sigma} u_*$  where  $u_* : H_\infty \rightarrow \mathbb{R}$  is given by  $u_*(x_0, a_0, b_0, x_1, a_1, b_1, \dots) = \liminf_n u(x_n)$ :

$$u_*(x, \pi, \sigma) = \liminf_t \int_{\tau \geq t} u(x_t) dP_{x,\pi,\sigma} + \int_{\tau < \infty} u_*(x_\tau, \pi[h_\tau], \sigma[h_\tau]) dP_{x,\pi,\sigma}. \tag{2}$$

This equality is easily obtained from equality (1) by replacing  $u$  with  $-u$ .

## 5 The proof of Theorem 1

For  $x \in S$ , let  $a(x) = s$  if  $u(x) = V(x)$  and  $s \in A(x)$  and let  $a(x) = g$  otherwise. Let  $\pi$  be the stationary strategy for player 1 that plays action  $a(x)$  at each state  $x$ . Thus  $\pi$  is the strategy that is asserted in Theorem 1 to be optimal for player 1. (Note that if  $u(x) \geq V(x)$  and  $s \in A(x)$ , then  $u(x) = V(x)$  because player 1 can guarantee a payoff of  $u(x)$  by playing  $s$  forever.) We will now prove that  $\pi$  is optimal.

### 5.1 The strategy $\pi$ conserves $V$

The next lemma shows that, if player 1 uses the strategy  $\pi$ , then the value function cannot decrease in expectation.

**Lemma 2** *For every initial state  $x = x_0$  and every strategy  $\sigma$  for player 2, the process  $V(x_n)$  is a submartingale under  $P_{x_0, \pi, \sigma}$ ; that is,*

$$E_{x_0, \pi, \sigma}[V(x_{n+1})|h_n] \geq V(x_n)$$

for every finite history  $h_n = (x_0, a_0, b_0, \dots, a_{n-1}, b_{n-1}, x_n)$ .

**Proof** Because  $\pi$  is the stationary strategy that plays action  $a(x)$  at each state  $x$ , it suffices to show that, for each  $x \in S$ ,  $a(x)$  is optimal in the one-shot game  $M(x)$ . Let  $x \in S$ .

**Case 1.**  $u(x) = V(x)$  and  $s \in A(x)$ .

Here  $a(x) = s$  gives payoff  $V(x)$  in the one-shot game  $M(x)$  and is clearly optimal in  $M(x)$ .

**Case 2.**  $u(x) < V(x)$  and  $s \notin A(x)$ .

In this case,  $a(x) = g$  is the unique action available to player 1 and must therefore be optimal in  $M(x)$ .

**Case 3.**  $u(x) < V(x)$  and  $s \in A(x)$ .

In this case,  $s \notin B(x)$ . (If  $s \in B(x)$ , player 2 can guarantee a payoff no larger than  $u(x)$  by playing  $s$  forever.) So  $B(x) = \{g\}$ . Also  $a(x) = g$  in this case. So we need to show that  $\sum_{y \in S} V(y) \cdot \alpha(x)(y) = \sum_{y \in S} V(y)q(y|x, g, g) \geq V(x)$ .

For an argument by contradiction, assume that  $\sum_{y \in S} V(y) \cdot \alpha(x)(y) < V(x)$ . Choose  $\epsilon > 0$  such that  $u(x) < V(x) - \epsilon$  and  $\sum_{y \in S} V(y) \cdot \alpha(x)(y) < V(x) - \epsilon$ , and choose  $\epsilon_1$  such that  $0 < \epsilon_1 < \epsilon/2$ . Assume also that  $x_0 = x$  is the initial state.

Let  $\pi_1$  be an  $\epsilon_1$ -optimal strategy for player 1. Define the stopping time  $\tau$  as the first time (if any) when  $\pi_1$  uses the action  $g$ ; that is, for each infinite history,  $h = (x_0, a_0, b_0, x_1, a_1, b_1, \dots) \in H_\infty$ ,  $\tau(h) = \inf\{n | \pi_1(h_n) = g\}$  where  $h_n = (x_0, a_0, b_0, \dots, a_{n-1}, b_{n-1}, x_n)$ . Note that prior to time  $\tau$  the strategy  $\pi_1$  is playing  $s$  and so the process of states remains at  $x_0 = x$ . In particular,  $x_\tau = x$  with probability 1 if  $\tau < \infty$ . Thus at time  $\tau$  player 1 plays  $g$  and player 2 must play  $g$ , so the conditional distribution of  $x_{\tau+1}$  is  $q(\cdot|x, g, g) = \alpha(x)$ .

Let  $\sigma$  be a strategy for player 2 such that the continuation strategy  $\sigma[h_{\tau+1}]$  is  $\epsilon_1$ -optimal at  $x_{\tau+1}$  whenever  $\tau < \infty$ .

We set  $P = P_{x,\pi_1,\sigma}$  in the calculations below. By equation (1), we have

$$\begin{aligned}
 V(x) - \epsilon_1 &\leq u(x, \pi_1, \sigma) \\
 &= \int_{\tau < \infty} u(x_\tau, \pi_1[h_\tau], \sigma[h_\tau])dP + \limsup_t \int_{\tau \geq t} u(x_t)dP.
 \end{aligned}$$

If  $\tau(h) = n$ , then, with probability one,  $h_n$  is of the form

$$h_n = (x, s, g, x, s, g, x, \dots, s, g, x)$$

and  $\pi_1(h_n) = \sigma(h_n) = g$ . Hence, the next state  $y = x_{n+1}$  has distribution  $\alpha(x)$ . So, by the choice of the strategy  $\sigma$ ,

$$\begin{aligned}
 u(x, \pi_1[h_n], \sigma[h_n]) &= \sum_y u(y, \pi_1[h_{n+1}], \sigma[h_{n+1}]) \cdot \alpha(x)(y) \\
 &\leq \sum_y V(y) \cdot \alpha(x)(y) + \epsilon_1 \\
 &\leq V(x) - \epsilon + \epsilon_1.
 \end{aligned}$$

Hence,

$$\int_{\tau < \infty} u(x_\tau, \pi_1[h_\tau], \sigma[h_\tau])dP \leq P[\tau < \infty] \cdot [V(x) - \epsilon - \epsilon_1].$$

Now  $x_t = x$  with probability one on the set  $[\tau \geq t]$ . So

$$\begin{aligned}
 \limsup_t \int_{\tau \geq t} u(x_t)dP &= (\limsup_t P[\tau \geq t]) \cdot u(x) \\
 &= P[\tau = \infty] \cdot u(x) \\
 &\leq P[\tau = \infty] \cdot [V(x) - \epsilon].
 \end{aligned}$$

Combining these inequalities, we have

$$\begin{aligned}
 V(x) - \epsilon_1 &\leq u(x, \pi_1, \sigma) \\
 &\leq P[\tau < \infty] \cdot [V(x) - \epsilon + \epsilon_1] + P[\tau = \infty] \cdot [V(x) - \epsilon] \\
 &= V(x) - \epsilon + \epsilon_1 P[\tau < \infty] \leq V(x) - \epsilon + \epsilon_1,
 \end{aligned}$$

which contradicts our choice of  $\epsilon_1 < \epsilon/2$ . This completes the proof. □

**Lemma 3** For every initial state  $x$ , every strategy  $\sigma$  for player 2, and every stop rule  $t$ ,  $E_{x,\pi,\sigma} V(x_t) \geq V(x)$ .

**Proof** This follows from Lemma 2 and the optional sampling theorem. □

### 5.2 Player 1 reaches good states by using $\pi$

The objective in this section is to show that, if player 1 plays the strategy  $\pi$  then, for every strategy of player 2, it is almost certain that states  $x$  will be reached where the utility  $u(x)$  is almost as large as the value  $V(x)$ . To be precise, for  $\epsilon > 0$ , define  $\tau_\epsilon(h) = \inf\{n \mid u(x_n) \geq V(x_n) - \epsilon\}$  for  $h = (x_0, a_0, b_0, x_1, a_1, b_1, \dots) \in H_\infty$ .

**Lemma 4** *For every initial state  $x$ , every strategy  $\sigma$  for player 2 and all  $\epsilon > 0$ ,  $P_{x,\pi,\sigma}[\tau_\epsilon < \infty] = 1$ .*

**Proof** Fix  $\epsilon > 0$ . Suppose first that player 2 plays the stationary strategy  $\sigma_1$  which always plays the action  $g$ . In this special case, player 1 faces a one-person problem that is equivalent to a stop-or-go gambling problem as defined in Section 5.4 of Maitra and Sudderth (1996). Let  $W$  be the value function for this one-person problem. Then it follows from Corollary 4.7, page 99 in Maitra and Sudderth (1996) that an optimal strategy for player 1 versus  $\sigma_1$  is the strategy  $\pi_1$  that plays action  $s$  at state  $x$  if  $u(x) \geq W(x)$  and  $s \in A(x)$ , and plays action  $g$  otherwise. Now  $W \geq V$  because player 2 has been restricted to play  $\sigma_1$  in the one-person problem. So  $\pi_1$  certainly plays  $g$  at state  $x$  if  $u(x) < V(x)$  and thereby agrees with  $\pi$  on this set. By Theorems 7.2 and 7.7, pages 76-78 in Maitra and Sudderth (1996), the probability is one under  $\pi_1$  of reaching the set  $[u \geq W - \epsilon] \subseteq [u \geq V - \epsilon]$ . Since  $\pi_1$  agrees with  $\pi$  on the set  $[u < V]$ , the  $P_{x,\pi,\sigma_1}$ -probability of reaching the set  $[u \geq V - \epsilon]$  is also one. That is, the conclusion of the lemma holds for the special case when  $\sigma = \sigma_1$ .

Now let  $\sigma$  be an arbitrary strategy for player 2 and consider a state  $x$  such that  $u(x) \leq V(x) - \epsilon < V(x)$ . Then  $B(x)$  must be the singleton  $\{g\}$ . (If  $s \in B(x)$ , then player 2 can guarantee a payoff no larger than  $u(x)$  by playing  $s$  repeatedly.) Thus the strategy  $\sigma$  must agree with  $\sigma_1$  on the set  $[u \leq V - \epsilon]$  and therefore  $P_{x,\pi,\sigma}[\tau_\epsilon < \infty] = P_{x,\pi,\sigma_1}[\tau_\epsilon < \infty] = 1$ . □

**Lemma 5** *For all strategies  $\sigma$  for player 2, all  $\epsilon > 0$ , and every stop rule  $r$ , there is a stop rule  $t \geq r$ , such that  $P_{x,\pi,\sigma}[u(x_t) \geq V(x_t) - \epsilon] \geq 1 - \epsilon$ .*

**Proof** Fix  $\sigma$ ,  $\epsilon$ , and assume first that  $r$  is the identically 0 stop rule. By countable additivity,  $P_{x,\pi,\sigma}[\tau_\epsilon < \infty] = \sup_{n \in \mathbb{N}} P_{x,\pi,\sigma}[\tau_\epsilon \leq n]$ . So, by Lemma 4, there is a constant stop rule  $t_0$  such that  $P_{x,\pi,\sigma}[\tau_\epsilon \leq t_0] \geq 1 - \epsilon$ . Let  $t = \min\{t_0, \tau_\epsilon\}$ . Then

$$P_{x,\pi,\sigma}[u(x_t) \geq V(x_t) - \epsilon] \geq P_{x,\pi,\sigma}[\tau_\epsilon = t] = P_{x,\pi,\sigma}[\tau_\epsilon \leq t_0] \geq 1 - \epsilon.$$

So the lemma is proved for the special case when  $r = 0$ .

Now let  $r$  be an arbitrary stop rule. By the previous case, there is, for each infinite history  $h \in H_\infty$ , a stop rule  $t_h$  depending on the finite history  $h_r(h)$  such that

$$P_{x_r(h),\pi,\sigma[h_r(h)]}[u(x_{t_h}) \geq V(x_{t_h}) - \epsilon] \geq 1 - \epsilon.$$

Define

$$t(h) = r(h) + t_h(x_r(h), a_r(h), b_r(h), x_{r(h)+1}, a_{r(h)+1}, b_{r(h)+1}, \dots).$$

For the conditional probability given  $h_r(h)$ , we have the inequality

$$P_{x,\pi,\sigma}[u(x_t) \geq V(x_t) - \epsilon | h_r(h)] = P_{x_r(h),\pi,\sigma[h_r(h)]}[u(x_{t_h}) \geq V(x_{t_h}) - \epsilon] \geq 1 - \epsilon$$

for every value of  $h_r(h)$ . Hence, the inequality also holds unconditionally. □

### 5.3 Completion of the proof that $\pi$ is optimal

Let  $x \in S$  and let  $\sigma$  be a strategy for player 2. We need to prove that  $u(x, \pi, \sigma) \geq V(x)$ . By a ‘‘Fatou equation’’ (Maitra and Sudderth (1996), Theorem 2.2, page 60)

$$u(x, \pi, \sigma) = E_{x,\pi,\sigma} u^* = \limsup_t E_{x,\pi,\sigma} u(x_t) = \inf_r \sup_{t \geq r} E_{x,\pi,\sigma} u(x_t).$$

The final equality is just the definition of the limsup over the directed set of stop rules.

So it suffices to show that, for every stop rule  $r$ ,  $\sup_{t \geq r} E_{x,\pi,\sigma} u(x_t) \geq V(x)$ . To that end, fix  $r$  and  $\epsilon > 0$ . By Lemma 5, there exists a stop rule  $t \geq r$  such that  $P_{x,\pi,\sigma}[u(x_t) \geq V(x_t) - \epsilon] \geq 1 - \epsilon$ . Let  $k$  be an upper bound on the absolute value of  $u$  and therefore also an upper bound on the absolute value of  $V$ . An elementary calculation then shows that

$$E_{x,\pi,\sigma} u(x_t) \geq E_{x,\pi,\sigma} V(x_t) - \epsilon(1 + 2k),$$

and, hence by Lemma 3,

$$E_{x,\pi,\sigma} u(x_t) \geq V(x) - \epsilon(1 + 2k).$$

Since  $\epsilon$  is an arbitrary positive number, the proof of Theorem 1 is complete.

**Remark 4** In the language of Exercise 18.13, page 224 in Maitra and Sudderth (1996), Lemma 3 shows that  $\pi$  is uniformly thrifty and Lemma 5 shows that  $\pi$  is uniformly equalizing. The exercise is to show, as is done above, that these two conditions imply that  $\pi$  is optimal. These notions have their origin in the Dubins–Savage theory of thrifty and equalizing strategies for gambling problems (Dubins and Savage (1965), pages 46–54).

## 6 Games with 0–1 utility functions

Before proceeding to the proof of Theorem 2, we study a stop-or-go game  $G_K$  in which the utility function  $u$  is the indicator function of a subset  $K$  of  $S$ . The results obtained for  $G_K$  will be used to show that optimal strategies allow to reach and stay in good states (see Sect. 7.2).

In the game  $G_K$ , the limsup payoff  $u^*$  from player 2 to player 1 is the indicator of the set  $L$  of those  $h = (x_0, a_0, b_0, x_1, a_1, b_1, \dots) \in H_\infty$  such that  $x_n \in K$  for infinitely many  $n$ . Thus  $L = \bigcap_n \bigcup_{m \geq n} [x_m \in K]$ .

In this special case, there are optimal strategies of a very simple form. Indeed, let  $\pi_0$  be the pure stationary strategy for player 1 that plays action  $s$  at state  $x$  if  $x \in K$  and  $s \in A(x)$ , and plays action  $g$  otherwise; and let  $\sigma_0$  be the pure stationary strategy for player 2 that plays  $s$  at  $x$  if  $x \notin K$  and  $s \in B(x)$ , and plays  $g$  otherwise.

Unlike the strategies of Theorems 1 and 2, the strategies  $\pi_0$  and  $\sigma_0$  can be played without knowledge of the value function  $V$ . Note that  $V(x) = 1$  if  $x \in K$  and  $s \in A(x)$ , because player 1 can guarantee payoff 1 by always playing action  $s$ , and  $V(x) = 0$  if  $x \notin K$  and  $s \in B(x)$ , because player 2 can guarantee payoff 0 by always playing action  $s$ , but in all other cases we only know that  $V(x) \in [0, 1]$  and determining  $V(x)$  is not immediate. Indeed, the simplicity of the utility function  $u$  does not seem to lead to essential simplifications in the transfinite induction provided in Maitra and Sudderth (1992) to determine the value.

**Theorem 3** *In the game  $G_K$ , the pure stationary strategy  $\pi_0$  is optimal for player 1, and the pure stationary strategy  $\sigma_0$  is optimal for player 2.*

The proof will use three lemmas. In these lemmas we prove even more: the strategies in Theorem 3 are optimal responses to all stationary strategies of the opponent.

**Lemma 6** *In the game  $G_K$ , the strategy  $\pi_0$  is an optimal response to every stationary strategy for player 2.*

**Proof** Fix a stationary strategy  $\sigma$  for player 2. Then player 1 faces a one-person problem of the type treated in Section 5.4 of Maitra and Sudderth (1996). Let  $Q$  be the value function for this one-person problem; that is,  $Q(x) = \sup_{\pi} u(x, \pi, \sigma)$  for all  $x$ . By Corollary 4.7, page 99 in Maitra and Sudderth (1996), an optimal strategy  $\pi'$  in this problem is to play action  $s$  if  $x \in K$  or  $Q(x) = 0$ , and  $s \in A(x)$ , and to play action  $g$  otherwise. (The optimality of  $\pi'$  also follows from Theorem 1 specialized to the one-person case.) Notice that  $\pi_0$  and  $\pi'$  differ only at states  $x$  such that  $x \notin K$ ,  $Q(x) = 0$ , and  $s \in A(x)$ . Let  $D$  be the collection of such states. Since the value function  $Q$  equals 0 on the set  $D$ , the strategy  $\pi_0$  is obviously optimal at every  $x \in D$ . If  $x \notin D$ , then  $\pi_0$  and  $\pi'$  agree up to time  $\tau(h) = \inf\{n \mid x_n \in D\}$  for  $h = (x_0, a_0, b_0, x_1, a_1, b_1, \dots) \in H_{\infty}$ . If  $\tau(h) < \infty$  then  $x_{\tau(h)} \in D$  and the continuation of strategy  $\pi_0$  is optimal at  $x_{\tau(h)}$ . By Lemma 2.3, page 92 in Maitra and Sudderth (1996),  $\pi_0$  is optimal.  $\square$

**Lemma 7** *In the game  $G_K$ , the strategy  $\sigma_0$  is an optimal response to every stationary strategy for player 1.*

Theorem 3 will follow from Lemmas 6 and 7. Indeed, these lemmas imply that  $(\pi_0, \sigma_0)$  is a Nash equilibrium in the game  $G_K$ . Since  $G_K$  is a zero-sum game, it means that  $\pi_0$  and  $\sigma_0$  are optimal strategies.

For the application of this result in Sect. 7 below, it will be convenient to restate and prove Lemma 7 for the equivalent liminf stop-or-go problem  $G'_C$ , where  $C$  is the complement of the set  $K$ , the players are reversed and the payoff  $u_*$  from player 2 to player 1 is the indicator of the set  $E$  of those infinite histories  $h$  such that  $x_n \in C$  for all but finitely many  $n$ . Thus  $E = \cup_n \cap_{m \geq n} [x_m \in C]$ . Now let  $\pi_1$  be the stationary strategy for player 1 in the game  $G'_C$  that plays action  $s$  at state  $x$  if  $x \in C$  and  $s \in A(x)$ , and plays action  $g$  otherwise. The next lemma is equivalent to Lemma 7.

**Lemma 8** *In the game  $G'_C$ , the strategy  $\pi_1$  is an optimal response to every stationary strategy for player 2.*

**Proof** Fix a stationary strategy  $\sigma$  for player 2. Since  $E_{x,\pi_1,\sigma} u_* = P_{x,\pi_1,\sigma}(E)$ , we need to show that

$$P_{x,\pi_1,\sigma}(E) \geq P_{x,\pi,\sigma}(E) \quad \text{for all } x, \pi. \tag{3}$$

For each nonnegative integer  $n$ , let  $C_n = \bigcap_{m \geq n} [x_m \in C]$ . Then the set  $E$  is the increasing union of the  $C_n$  and, by countable additivity,  $P_{x,\pi_1,\sigma}(E) = \sup_n P_{x,\pi_1,\sigma}(C_n)$ . Here is an intermediate step toward proving (3).

**Step 1**  $P_{x,\pi_1,\sigma}(C_0) \geq P_{x,\pi,\sigma}(C_0)$  for all  $x, \pi$ .

Observe that the set  $C_0$  is the decreasing intersection of the sets  $F_n$ , where  $F_n = \bigcap_{0 \leq m \leq n} [x_m \in C]$  for each nonnegative integer  $n$ . Thus  $P_{x,\pi,\sigma}(C_0) = \inf_n P_{x,\pi,\sigma}(F_n)$  for all  $x, \pi, \sigma$ . So, to verify Step 1, it suffices to show

$$P_{x,\pi_1,\sigma}(F_n) \geq P_{x,\pi,\sigma}(F_n) \quad \text{for all } x, \pi, n. \tag{4}$$

The proof of (4) is by induction on  $n$ . The case  $n = 0$  follows from the fact that the quantities  $P_{x,\pi,\sigma}(F_0)$  are all equal to 1 if  $x \in C$  and are all equal to 0 if not.

So assume that  $n > 0$  and that the desired inequalities hold for  $n - 1$ . If the initial state  $x \notin C$ , then  $P_{x,\pi,\sigma}(F_n) = 0$  for all  $\pi, \sigma$ . So assume that  $x \in C$ . If  $s \in A(x)$ , then  $\pi_1$  plays  $s$  forever and  $P_{x,\pi_1,\sigma}(F_n) = 1 \geq P_{x,\pi,\sigma}(F_n)$ . So assume that  $s \notin A(x)$ . Then every strategy  $\pi$  for player 1 must play  $g$  at  $x$ . Condition on  $h_1 = (x, g, b, x_1)$  and use the inductive hypothesis to calculate as follows:

$$\begin{aligned} P_{x,\pi,\sigma}(F_n) &= \sum_{b,x_1} P_{x,\pi,\sigma}(F_n|h_1) \cdot q(x_1|x, g, b) \cdot \sigma(x)(b) \\ &= \sum_{b,x_1} P_{x_1,\pi,\sigma}(F_{n-1}) \cdot q(x_1|x, g, b) \cdot \sigma(x)(b) \\ &\leq \sum_{b,x_1} P_{x_1,\pi_1,\sigma}(F_{n-1}) \cdot q(x_1|x, g, b) \cdot \sigma(x)(b) \\ &= P_{x,\pi_1,\sigma}(F_n). \end{aligned}$$

This completes the proof of Step 1.

**Step 2**  $P_{x,\pi_1,\sigma}(E) \geq P_{x,\pi,\sigma}(C_n)$  for all  $x, \pi, n$ .

As noted above,  $P_{x,\pi,\sigma}(E) = \sup_n P_{x,\pi,\sigma}(C_n)$ . So this step will complete the proof of the inequalities in (3). The proof of Step 2 is again by induction on  $n$ . The case  $n = 0$  follows from Step 1 because  $C_0 \subseteq E$  and therefore  $P_{x,\pi_1,\sigma}(E) \geq P_{x,\pi_1,\sigma}(C_0) \geq P_{x,\pi,\sigma}(C_0)$  for all  $x, \pi$ .

So assume that  $n > 0$  and the desired inequalities hold for  $n - 1$ . Condition on  $h_1 = (x, a_0, b_0, x_1)$  and use the inductive hypothesis to get

$$\begin{aligned} P_{x,\pi,\sigma}(C_n) &= \sum_{a_0,b_0,x_1} P_{x,\pi,\sigma}(C_n|h_1) \cdot q(x_1|x, a_0, b_0) \cdot \pi(x)(a_0) \cdot \sigma(x)(b_0) \\ &= \sum_{a_0,b_0,x_1} P_{x_1,\pi[h_1],\sigma[h_1]}(C_{n-1}) \cdot q(x_1|x, a_0, b_0) \cdot \pi(x)(a_0) \cdot \sigma(x)(b_0) \\ &\leq \sum_{a_0,b_0,x_1} P_{x_1,\pi_1,\sigma}(E) \cdot q(x_1|x, a_0, b_0) \cdot \pi(x)(a_0) \cdot \sigma(x)(b_0). \end{aligned}$$

The fact that  $\sigma[h_1] = \sigma$ , which was used in the line above, holds because  $\sigma$  is stationary. Now from the shift invariance of the set  $E$ , it follows that

$$\begin{aligned} P_{x,\pi_1,\sigma}(E) &= \sum_{a_0,b_0,x_1} P_{x_1,\pi_1,\sigma}(E|h_1) \cdot q(x_1|x, a_0, b_0) \cdot \pi_1(x)(a_0) \cdot \sigma(x)(b_0) \\ &= \sum_{a_0,b_0,x_1} P_{x_1,\pi_1,\sigma}(E) \cdot q(x_1|x, a_0, b_0) \cdot \pi_1(x)(a_0) \cdot \sigma(x)(b_0). \end{aligned}$$

So it suffices to show that

$$\begin{aligned} &\sum_{a_0,b_0,x_1} P_{x_1,\pi_1,\sigma}(E) \cdot q(x_1|x, a_0, b_0) \cdot \pi(x)(a_0) \cdot \sigma(x)(b_0) \\ &\leq \sum_{a_0,b_0,x_1} P_{x_1,\pi_1,\sigma}(E) \cdot q(x_1|x, a_0, b_0) \cdot \pi_1(x)(a_0) \cdot \sigma(x)(b_0). \end{aligned}$$

If  $A(x) = \{g\}$  is a singleton, then both  $\pi$  and  $\pi_1$  must play  $g$  at  $x$  so that the two quantities above are the same. So assume that  $A(x) = \{s, g\}$ . If  $x \in C$ , then  $\pi_1$  will play  $s$  forever and thus  $P_{x,\pi_1,\sigma}(E) = 1$ . So assume that  $x \notin C$  in which case  $\pi_1$  plays  $g$  at  $x$ . If  $\pi$  also plays  $g$ , then the two quantities above are equal. If  $\pi$  plays  $s$ , then by the inductive hypothesis with  $h_1 = (x, g, s, x)$ ,

$$P_{x,\pi,\sigma}(C_n) = P_{x,\pi[h_1],\sigma}(C_{n-1}) \leq P_{x,\pi_1,\sigma}(E).$$

Whether  $\pi$  plays  $g$  or  $s$ , the desired inequality holds. Consequently, it also holds if  $\pi$  plays a mixture of the two. The proofs of Step 2 and the lemma are now complete.  $\square$

Let  $\sigma_1$  be the stationary strategy for player 2 in the game  $G'_C$  that plays action  $s$  at state  $x$  if  $x \notin C$  and  $s \in B(x)$ , and plays  $g$  otherwise. The following lemma is equivalent to Lemma 6, which we have already proved.

**Lemma 9** *In the game  $G'_C$ , the strategy  $\sigma_1$  is an optimal response to every stationary strategy for player 1.*

Lemmas 8 and 9 imply the next theorem, which amounts to a restatement of Theorem 3.

**Theorem 4** *In the game  $G'_C$ , the pure stationary strategy  $\pi_1$  is optimal for player 1, and the pure stationary strategy  $\sigma_1$  is optimal for player 2.*

## 7 Stop-or-go games with liminf payoff

Let  $G' = (S, A, B, q, u)$  be a stop-or-go game as defined in Sect. 2 except that the payoff is now taken to be  $u_*$  where

$$u_*(h) = u_*(x_0, a_0, b_0, x_1, a_1, b_1, x_2, \dots) = \liminf_n u(x_n).$$

Because  $-\liminf_n u(x_n) = \limsup_n (-u(x_n))$ , this liminf stop-or-go game is equivalent to a limsup stop-or-go game as in Sect. 2 with the players reversed and  $u$  replaced by  $-u$ . Let  $u_*(x, \pi, \sigma) = E_{x, \pi, \sigma} u_*$  denote the expected payoff at state  $x$  when the players choose the strategies  $\pi$  and  $\sigma$ , and let  $W$  be the value function for the liminf stop-or-go game  $G'$ . We will now prove the following theorem, which is equivalent to Theorem 2.

**Theorem 5** *If  $\epsilon > 0$ , then a pure  $\epsilon$ -optimal stationary strategy for player 1 in the game  $G'$  is to play action  $s$  if  $u(x) \geq W(x) - \epsilon$  and  $s \in A(s)$ , and to play action  $g$  otherwise.*

Fix  $\epsilon > 0$ . Let  $a'(x) = s$  if  $u(x) \geq W(x) - \epsilon$  and  $s \in A(x)$  and let  $a'(x) = g$  otherwise. Let  $\pi'$  be the stationary strategy that plays action  $a'(x)$  at each state  $x$ . To prove Theorem 5, and thereby Theorem 2, we need to show that  $\pi'$  is  $\epsilon$ -optimal for player 1 in the game  $G'$ .

### 7.1 The strategy $\pi'$ conserves $W$

Here is the analogue to Lemma 2 for the strategy  $\pi'$ .

**Lemma 10** *For every initial state  $x = x_0$  and every strategy  $\sigma$  for player 2, the process  $W(x_n)$  is a submartingale under  $P_{x_0, \pi', \sigma}$ ; that is,*

$$E_{x_0, \pi', \sigma} [W(x_{n+1}) | h_n] \geq W(x_n)$$

for every finite history  $h_n = (x_0, a_0, b_0, x_1, \dots, a_{n-1}, b_{n-1}, x_n)$ .

**Proof** The proof is similar to the proof of Lemma 2.

Let  $M'(x)$  be the one-shot game with action sets  $A(x)$ ,  $B(x)$  and payoff for actions  $a \in A(x)$ ,  $b \in B(x)$  equal to  $\sum_{y \in S} W(y) q(y|x, a, b)$ . It suffices to show that, for each state  $x$ ,  $a'(x)$  is optimal for player 1 in  $M'(x)$ . Note that the value of  $M'(x)$  is  $W(x)$  by the optimality equation for the game  $G'$ .

**Case 1:**  $u(x) \geq W(x) - \epsilon$  and  $s \in A(x)$ .

In this case,  $a'(x) = s$  and is clearly optimal in  $M'(x)$ .

**Case 2:**  $u(x) < W(x) - \epsilon$  and  $s \notin A(x)$ .

In this case,  $a'(x) = g$  is the unique action available to player 1 and is therefore optimal.

**Case 3:**  $u(x) < W(x) - \epsilon$  and  $s \in A(s)$ .

In this case,  $s \notin B(x)$ . (If  $s \in B(s)$ , player 2 can guarantee a payoff no larger than  $u(x)$  by playing  $s$  forever.) So  $B(x) = \{g\}$  and player 2 must play  $g$ . Also  $a'(x) = g$  in this case. So we need to show that  $\sum_y W(y)\alpha(x)(y) = \sum_y W(y)q(y|x, g, g) \geq W(x)$ .

For an argument by contradiction, assume that  $\sum_y W(y) \cdot \alpha(x)(y) < W(x)$ . Choose  $\epsilon_0$  such that  $0 < \epsilon_0 < \epsilon$ ,  $u(x) < W(x) - \epsilon_0$ , and  $\sum_y W(y) \cdot \alpha(x)(y) < W(x) - \epsilon_0$ . Next choose  $\epsilon_1$  such that  $0 < \epsilon_1 < \epsilon_0/2$  and a strategy  $\pi_1$  for player 1 that is  $\epsilon_1$ -optimal at  $x$ . Let  $\tau$  be the first time (if any) when  $\pi_1$  plays action  $g$ . Note that prior to time  $\tau$  the strategy  $\pi_1$  is playing  $s$  and so  $x_n = x$  for  $0 \leq n \leq \tau$ . At time  $\tau$  both players play  $g$  and the distribution of  $x_{\tau+1}$  is therefore  $\alpha(x) = q(\cdot|x, g, g)$ . Let  $\sigma$  be a strategy for player 2 such that  $\sigma[h_{\tau+1}]$  is  $\epsilon_1$ -optimal at  $x_{\tau+1}$  when  $\tau < \infty$ .

Repeat the calculation in case 3 of the proof of Lemma 2 using equality (2), rather than (1), with  $W$  in place of  $V$ , and  $\epsilon_0$  in place of  $\epsilon$  to find that  $W(x) - \epsilon_1 \leq W(x) - \epsilon_0 + \epsilon_1$  contradicting our choice of  $\epsilon_1 < \epsilon_0/2$ . □

The next lemma follows from Lemma 10 as Lemma 3 did from Lemma 2.

**Lemma 11** *For every initial state  $x$ , every strategy  $\sigma$  for player 2, and every stop rule  $t$ ,  $E_{x,\pi',\sigma} W(x_t) \geq W(x)$ .*

### 7.2 Reaching and staying in good states

When the payoff is the limsup, a good strategy must, with high probability, reach states with utility close to the value infinitely often. However, a good strategy for the liminf payoff must, with high probability, reach such states and eventually stay in the collection of them.

Let  $\epsilon > 0$ . Define  $C = \{x \in S \mid u(x) \geq W(x) - \epsilon\}$  and, as in Sect. 6, let  $E$  be the set of infinite histories  $h = (x_0, a_0, b_0, x_1, a_1, b_1, \dots)$  such that  $x_n \in C$  for all but finitely many  $n$ . Here is the main result of this section.

**Theorem 6** *For all  $x \in S$  and strategies  $\sigma$  for player 2,  $P_{x,\pi',\sigma}(E) = 1$ .*

The first step in the proof of Theorem 6 is to show that, for every  $x$  and every stationary strategy  $\sigma$ , there exist strategies  $\pi$  such that  $P_{x,\pi,\sigma}(E)$  is close to one.

**Lemma 12** *For all  $x \in S$  and stationary strategies  $\sigma$  for player 2,  $\sup_{\pi} P_{x,\pi,\sigma}(E) = 1$ .*

**Proof** With  $\sigma$  fixed, player 1 faces a liminf one-person stop-or-go game equivalent to a liminf gambling problem as treated in Sudderth (1983). From position  $h_n = (x_0, a_0, b_0, x_1, \dots, x_n)$  in the one-person problem the distribution of  $a_n$  is selected by player 1 as a distribution on  $A(x_n)$  and  $b_n$  has distribution  $\sigma(x_n)$ . Then  $x_{n+1}$  has conditional distribution  $q(\cdot|x_n, a_n, b_n)$ . The payoff to player 1 from a strategy  $\pi$  at

state  $x$  is then  $u_*(x, \pi, \sigma)$ . Let  $Q$  be the value function for this one-person problem. Then, for each  $x \in S$ ,

$$Q(x) = \sup_{\pi} u_*(x, \pi, \sigma) \geq \inf_{\sigma'} \sup_{\pi} u_*(x, \pi, \sigma') = W(x),$$

where  $W$  is the value function for the two-person game.

Choose  $\epsilon_1$  such that  $0 < \epsilon_1 < \epsilon$  and choose  $\pi_1$  such that

$$E_{x, \pi_1, \sigma} u_* = u_*(x, \pi_1, \sigma) \geq \sup_{\pi} u_*(x, \pi, \sigma) - \epsilon_1^2 = Q(x) - \epsilon_1^2.$$

By Lemma 1 in Sudderth (1983),

$$Q(x) \geq E_{x, \pi_1, \sigma} Q^* \quad \text{and} \quad P_{x, \pi_1, \sigma}[u_* \leq Q^*] = 1,$$

where  $Q^*(h) = \limsup_n Q(x_n) = \liminf_n Q(x_n)$  with probability one. Hence,

$$\epsilon_1^2 \geq E_{x, \pi_1, \sigma}[Q^* - u_*] \geq \int_{Q^* - u_* \leq \epsilon_1} \epsilon_1 \, d P_{x, \pi_1, \sigma} \geq \epsilon_1 \cdot P_{x, \pi_1, \sigma}[Q^* - u_* \leq \epsilon_1].$$

So

$$P_{x, \pi_1, \sigma}[Q^* - u_* < \epsilon_1] \geq 1 - \epsilon_1.$$

Thus

$$\begin{aligned} P_{x, \pi_1, \sigma}(E) &\geq P_{x, \pi_1, \sigma}[W^* - u_* < \epsilon] \\ &\geq P_{x, \pi_1, \sigma}[Q^* - u_* < \epsilon] \\ &\geq P_{x, \pi_1, \sigma}[Q^* - u_* < \epsilon_1] \geq 1 - \epsilon_1. \end{aligned}$$

The result follows because  $\epsilon_1$  is arbitrarily small. □

Consider now the liminf stop-or-go game  $G'_C$  of Sect. 6 in which the utility function is the indicator of the set  $C$ . Let  $\pi_1$  and  $\sigma_1$  be the strategies as in Sect. 6. The strategy  $\pi'$  is the same as the strategy  $\pi_1$ . By Lemma 8,  $\pi'$  is an optimal response to the stationary strategy  $\sigma_1$ . So, by Lemma 12,

$$P_{x, \pi', \sigma_1}(E) = \sup_{\pi} P_{x, \pi, \sigma_1}(E) = 1, \quad x \in S.$$

Also, by Theorem 4, the game  $G'_C$  has the value function  $V'_C(x) = P_{x, \pi', \sigma_1}(E) = 1$  and  $\pi'$  is optimal for player 1 in the game. Hence, for all  $x \in S$  and all strategies  $\sigma$  for player 2,  $P_{x, \pi', \sigma}(E) \geq V'_C(x) = 1$ . This completes the proof of Theorem 6.

### 7.3 Completion of the proof that $\pi'$ is optimal

Fix  $x \in S$  and a strategy  $\sigma$  for player 2. By Lemma 11 and the Fatou equation (Maitra and Sudderth (1996), Theorem 2.2, page 60),

$$E_{x,\pi',\sigma}(W_*) = E_{x,\pi',\sigma}(\liminf_n W(x_n)) = \liminf_t E_{x,\pi',\sigma}(W(x_t)) \geq W(x), \quad (5)$$

where  $W_*(h) = \liminf_n W(x_n)$  for  $h = (x_0, a_0, b_0, x_1, a_1, b_1, \dots) \in H_\infty$ . By Lemma 10, the process  $\{W(x_n)\}$  is a bounded submartingale under  $P_{x,\pi',\sigma}$  and therefore converges  $P_{x,\pi',\sigma}$ -almost surely to its liminf  $W_*$ . Also, by Theorem 6,  $P_{x,\pi',\sigma}(E) = 1$ , which implies that  $P_{x,\pi',\sigma}[u_* \geq W_* - \epsilon] = 1$ . Therefore,

$$u_*(x, \pi', \sigma) = E_{x,\pi',\sigma}u_* \geq E_{x,\pi',\sigma}(W_*) - \epsilon. \quad (6)$$

Theorem 5, and consequently Theorem 2 now follow from (5) and (6).

## 8 n-person stop-or-go games

Let  $n \geq 2$  be a positive integer. An  $n$ -person stop-or-go game  $G = (S, A, q, u)$  is a stochastic game with players  $I = \{1, \dots, n\}$ , a countable non-empty state space  $S$ ; for each  $x \in S$ ,  $A(x) = A_1(x) \times \dots \times A_n(x)$  is the product of the action sets  $A_1(x), \dots, A_n(x)$  for the players, where for each player  $i \in I$ ,  $A_i(x) = \{g, s\}$  or  $\{g\}$ ; the law of motion  $q$  satisfies  $q(\cdot|x, a) = q(\cdot|x, a_1, \dots, a_n) = \alpha(x)$  if  $a_1 = \dots = a_n = g$  where  $\alpha(x)$  is a countably additive probability measure defined on all subsets of  $S$  and  $q(\cdot|x, a) = \delta(x)$  if at least one of the actions  $a_i$  equals  $s$ ; the function  $u = (u_1, \dots, u_n)$  is the vector of the bounded real-valued utility functions  $u_i, i \in I$ , for the players. Note that if any player plays the stop action  $s$ , then the state remains the same.

Play is similar to that in two-player games. At each stage  $k$ , the play is at some state  $x_k$ , the players simultaneously choose actions  $a_k = (a_{k1}, \dots, a_{kn}) \in A(x_k)$ , and the next state has distribution  $q(\cdot|x_k, a_k)$ . A play from an initial state  $x_0$  generates an infinite history  $h = (x_0, a_0, x_1, a_1, \dots)$ . Each player  $i$  is either a limsup player and has payoff  $\limsup_k u_i(x_k)$  or a liminf player with payoff  $\liminf_k u_i(x_k)$ .

For each finite history  $h = (x_0, a_0, \dots, x_{k-1}, a_{k-1}, x_k)$ , let  $x_h = x_k$  denote the final state in  $h$ . A strategy  $\sigma_i$  for player  $i \in I$  is a mapping that assigns to each finite history  $h$  a mixed action  $\sigma_i(h)$  on  $A_i(x_h)$ . A strategy profile  $\sigma = (\sigma_1, \dots, \sigma_n)$  consists of a strategy  $\sigma_i$  for each player  $i \in I$ . An initial state  $x$  together with a profile  $\sigma$  and the law of motion  $q$  determine a distribution  $P_{x,\sigma}$  for the infinite history  $h = (x_0, a_0, x_1, a_1, \dots)$  and the expected payoff  $u_i(x, \sigma) = E_{x,\sigma}[\limsup_k u_i(x_k)]$  or  $E_{x,\sigma}[\liminf_k u_i(x_k)]$  according to whether player  $i$  is a limsup or a liminf player.

For  $\epsilon \geq 0$ , an  $\epsilon$ -equilibrium is a strategy profile  $\sigma$  such that, for every  $i$ ,  $\sigma_i$  is an  $\epsilon$ -optimal strategy for player  $i$  versus the remaining strategies denoted by  $\sigma_{-i}$ . A 0-equilibrium is called simply an equilibrium.

For the purpose of constructing an  $\epsilon$ -equilibrium, the game  $G$  can be viewed as a game of perfect information; that is, a game in which at most one player has a choice

of actions at each stage. The reason is that, if two or more players have the action  $s$  available at some state, then an equilibrium is attained trivially when these players play action  $s$ . Such states can be viewed as absorbing. At all other states, at most one player can have the action  $s$  available and such a player we call the active player at the state.

**Theorem 7** *If  $G$  is an  $n$ -person stop-or-go game, then  $G$  has an  $\epsilon$ -equilibrium for every  $\epsilon > 0$ . If the state space  $S$  is finite or if every player has a limsup payoff function, then  $G$  has an equilibrium.*

**Proof** The proof is based on an idea of Mertens and Neyman (cf. Mertens 1987, which also appears in Thuijsman and Rahgavan (1997),<sup>1</sup> and uses an auxiliary game for each player  $i \in I$ .

**Step 1: The auxiliary game  $G_i$**

Consider the auxiliary zero-sum game  $G_i$  in which player  $i$  maximizes his own payoff and the other players minimize player  $i$ 's payoff. The game  $G_i$  can be viewed as being a two-person stop-or-go game. This is because the players  $-i$  can be viewed as a single player.

For each finite history  $h$  the continuation game has a value  $v_i(h)$  and player  $i$  has a pure stationary strategy  $\sigma_i$  that is  $\epsilon/2$ -optimal in every subgame. This follows from Theorem 1 if player  $i$  is a limsup maximizer and from Theorem 5 if  $i$  is a liminf maximizer. Similarly, the players  $-i$  have a pure stationary profile  $\sigma_{-i}^i$  that is  $\epsilon/2$ -optimal in every subgame. Let  $u_i(h, \sigma_i, \sigma_{-i})$  be the expected payoff to player  $i$  from playing  $\sigma_i$  versus  $\sigma_{-i}$ .

Then

$$u_i(h, \sigma_i, \sigma_{-i}) \geq v_i(h) - \frac{\epsilon}{2} \quad \forall \sigma_{-i}, \forall h. \tag{7}$$

Note that if player  $i$  is the active player, then none of his actions can increase his value in expectation. Indeed, suppose that  $h$  is the history, ending in state  $x$ . Then:

$$v_i(h) = \max_{a \in A(x)} \sum_{x' \in S} q(x'|x, a) \cdot v_i(h, a, x'), \tag{8}$$

where  $(h, a, x')$  is the history when after  $h$  action  $a$  is played and the new state is  $x'$ .

**Step 2: In the original game**

Let  $\sigma$  be the pure strategy profile  $\sigma = (\sigma_i)_{i \in I}$ . Consider the strategy profile  $\sigma^*$  such that:

- The players follow  $\sigma$  as long as there is no deviation from  $\sigma$ . Note that because  $\sigma$  consists of pure strategies, a deviation is immediately noticed.
- If a player  $i$  deviates from  $\sigma_i$ , then all his opponents “punish” player  $i$  in the remaining game, by switching to the strategy profile  $\sigma_{-i}^i$  from the next stage.

<sup>1</sup> Both papers consider perfect information games, but their models differ somewhat from ours. Mertens and Neyman consider deterministic transitions, which makes it possible to discretize the payoffs. Thuijsman and Rahgavan consider the average payoff on finitely many states.

We argue that  $\sigma^*$  is an  $\epsilon$ -equilibrium. Let  $H'$  denote the set of histories in which no deviation from  $\sigma$  has taken place. (So, according to  $\sigma^*$ , the players should still follow  $\sigma$ .) Let  $h \in H'$ , ending in some state  $x$ . Assume that player  $i$  is active in state  $x$ .

If player  $i$  does not deviate, and follows  $\sigma_i$  in the remaining game, then by (7), his payoff in the subgame at  $h$  is at least

$$v_i(h) - \frac{\epsilon}{2}.$$

On the other hand, if player  $i$  deviates at  $h$  from  $\sigma_i(h)$  to any action  $a \neq \sigma_i(h)$ , then player  $i$ 's payoff in the subgame at  $h$  is at most

$$\sum_{x' \in S} q(x'|x, a) \cdot \left( v_i(h, a, x') + \frac{\epsilon}{2} \right) \leq v_i(h) + \frac{\epsilon}{2},$$

because players  $-i$  will “punish” player  $i$  from the next stage and because of (8).

Thus, no player can profitably deviate, up to  $\epsilon$ , at any history in  $H'$ . This means that  $\sigma^*$  is an  $\epsilon$ -equilibrium indeed. The proof of the first assertion is complete.

To verify the second assertion, note that, by Theorem 1 and Corollary 1, players have optimal pure stationary strategies in zero-sum stop-or-go games when  $S$  is finite or if the players are limsup maximizers. Thus the argument above can be repeated with  $\epsilon$  taken to be zero. □

The strategies of the  $\epsilon$ -equilibria constructed in the proof of Theorem 7 are stationary except when punishments occur. So one might suspect that there exist stationary  $\epsilon$ -equilibria. Here is a two-person example where stationary  $\epsilon$ -equilibria do not exist for small  $\epsilon$ .

**Example 2** Let  $S = \{1, 2, t_1, t_2\}$ . States  $t_1$  and  $t_2$  are absorbing while players 1 and 2 are active in states 1 and 2, respectively. So  $A_1(1) = \{s, g\}$ ,  $A_2(1) = \{g\}$ ;  $A_1(2) = \{g\}$ ,  $A_2(2) = \{s, g\}$ . The motion at states 1 and 2 is given by

$$q(\cdot|1, g, g) = \frac{1}{2}\delta(2) + \frac{1}{2}\delta(t_1), \quad q(\cdot|2, g, g) = \frac{1}{2}\delta(1) + \frac{1}{2}\delta(t_2).$$

The values of the utility function  $u$  are  $(1, \epsilon)$ ,  $(3, 1)$ ,  $(0, 0)$ ,  $(0, 2 - 2\epsilon)$  at states 1, 2,  $t_1$ ,  $t_2$  respectively, where  $\epsilon \in (0, \frac{1}{4})$ . Both players seek to maximize the limsup of their utilities.

There are four pure stationary profiles:  $(s, s)$  where both active players play  $s$ ,  $(g, g)$  where both play  $g$ ,  $(s, g)$  where player 1 plays  $s$  and 2 plays  $g$ ,  $(g, s)$  where 1 plays  $g$  and 2 plays  $s$ . We need not consider mixed strategies because repeated play of a mixed action with positive mass on  $g$  is equivalent to playing  $g$ .

The profile  $(s, s)$  is not an equilibrium because it gives player 1 payoff 1 at state 1 while player 1 would get payoff  $\frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 3 > 1$  by playing  $g$ .

The profile  $(g, g)$  is not an equilibrium because it gives player 1 payoff 0 at state 1 while player 1 would get payoff 1 by playing  $s$ .

The profile  $(s, g)$  is not an equilibrium because it gives player 2 payoff  $\frac{1}{2} \cdot (2 - 2\epsilon) + \frac{1}{2} \cdot \epsilon < 1$  at state 2 while player 2 would get payoff 1 by playing  $s$ .

The profile  $(g, s)$  is not an equilibrium because it gives player 2 payoff 1 at state 2 while, by playing  $g$ , player 2 would get payoff

$$\left[ \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \dots \right] \cdot (2 - 2\epsilon) = \frac{2}{3} \cdot (2 - 2\epsilon) > 1.$$

In the special case when every player has a 0-1 valued utility function, there is a very simple stationary equilibrium.

**Theorem 8** *Suppose that, for every  $i \in I$ , the utility function  $u_i$  is the indicator of a subset  $K_i$  of  $S$ . Then an equilibrium is the profile  $\sigma = (\sigma_1, \dots, \sigma_n)$  where each strategy  $\sigma_i$  plays  $s$  at each state  $x$  if  $x \in K_i$  and  $s \in A_i(x)$ , and plays  $g$  otherwise.*

**Proof** The strategies  $\sigma_{-i}$  can be viewed as a single stationary strategy in a two-person game versus player  $i$ . By Lemma 6,  $\sigma_i$  is optimal versus  $\sigma_{-i}$  if player  $i$  is a limsup maximizer. The same holds by Lemma 8 if player  $i$  is a liminf maximizer.  $\square$

## 9 Extensions

It seems plausible that the results proved here for stop-or-go games with countable state space and a countably additive law of motion can be generalized to finitely additive games and also to Borel measurable games as in Maitra and Sudderth (1993a, b). The existence of the value follows in both cases from general theorems. Generalizations of some results of probability theory would be needed in the finitely additive case, and there are likely to be measurability obstacles in the Borel setting.

In Sect. 8 we could reduce  $n$ -player stop-or-go games to perfect information games. Such a reduction would also be possible in the zero-sum case for the proofs of Theorems 1 and 2, but would not lead to essential simplifications.

Theorem 7 was stated for games with finitely many players, but it also holds for a set of players of arbitrary cardinality, with a very similar proof.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Chow YS, Robbins H, Siegmund D (1971) Great expectations: the theory of optimal stopping. Houghton Mifflin, Boston
- Dubins L, Savage L (1965) How to gamble if you must: inequalities for stochastic processes. McGraw-Hill, New York (Dover editions in 1976 and 2014)
- Dubins L, Sudderth W (1977a) On countably additive gambling and optimal stopping theory. *Zf Wahrscheinlichkeitstheorie* 41:5972

- Dubins L, Sudderth W (1977b) Persistently  $\epsilon$ -optimal strategies. *Math Oper Res* 2:125–134
- Dynkin E (1969) Game variant of a problem of optimal stopping. *Dokl Akad Nauk SSSR* 10:270–274
- Flesch J, Predtetchinski A, Sudderth W (2018) Simplifying optimal strategies in positive and negative stochastic games. *Discrete Appl Math* 251:40–56
- Maitra A, Sudderth W (1992) An operator solution of stochastic games. *Isr J Math* 78:33–49
- Maitra A, Sudderth W (1993a) Finitely additive and measurable stochastic games. *Int J Game Theory* 22:201–223
- Maitra A, Sudderth W (1993b) Borel stochastic games with limsup payoff. *Ann Probab* 21:861–885
- Maitra A, Sudderth W (1996) *Discrete gambling and stochastic games*. Springer, New York
- Martin D (1998) The determinacy of Blackwell games. *J Symb Logic* 63:1565–1581
- Mashiah-Yaakovi A (2014) Subgame perfect equilibria in stopping games. *Int J Game Theory* 43:89–135
- Mertens J-F (1987) Repeated games. In: *Proceedings of the International Congress of Mathematicians*, Berkeley
- Rosenberg D, Solan E, Vieille N (2001) Stopping games with randomized strategies. *Probab Theory Relat Fields* 119:433–451
- Shiryayev AN (1973) *Statistical sequential analysis: optimal stopping rules*. American Mathematical Society (translated from Russian)
- Shmaya E, Solan E, Vieille N (2003) An application of Ramsey theorem to stopping games. *Games Econ Behav* 42:300–306
- Shmaya E, Solan E, Vieille N (2004) Two-player non zerosum stopping games in discrete time. *Ann Probab* 38:2733–2764
- Solan E, Laraki R (2013) Equilibrium in two-player non-zero-sum Dynkin games in continuous time. *Stochastics* 85:997–1014
- Sudderth W (1983) Gambling problems with a limit inferior payoff. *Math Oper Res* 8:287–297
- Thuijsman F, Rahgavan TES (1997) Perfect information stochastic games and related classes. *Int J Game Theory* 26:403–408

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.