MARKOV DECISION PROCESSES

Lester E. Dubins
University of California
Berkeley, CA 94720

Ashok P. Maitra
University of Minnesota
Minneapolis, MN 55455

William D. Sudderth
University of Minnesota
Minneapolis, MN 55455

May 28, 2000

Chapter 1

Invariant Gambling Problems and Markov Decision Processes

Abstract

Markov decision problems can be viewed as gambling problems that are invariant under the action of a group or semi-group. It is shown that invariant stationary plans are almost surely adequate for a leavable, measurable, invariant gambling problem with a nonnegative utility function and a finite optimal reward function. This generalizes results about stationary plans for positive Markov decision models as well as measurable gambling problems.

1.1 Introduction

This paper introduces the notion of a gambling problem that is invariant, in a sense to be specified below, under the action of a group or a semigroup of transformations.

Our primary stimulus has been to understand more fully the relationship of Markov decision processes to gambling theory. It has long been known that these two theories are closely related (cf. Chapter 12 of Dubins and Savage (1965)) and perhaps each contains the other. However, it has not been possible to translate theorems about stationary plans, for example, directly from one theory to the other. It will be explained below how Markov decision processes may be viewed as invariant gambling problems. Subsequent sections will show how stationarity results in gambling theory (Dubins and Sudderth (1979)) are extended to invariant gambling problems and , in particular, to Markov decision problems. Our main interest here is in the reward structures. A comparative study of the measurable structures of the two theories was made by Blackwell (1976).

A secondary stimulus to us is the fact that there are a number of gambling problems which possess a natural group theoretic structure. Group invariance techniques have found many applications in statistical decision theory (cf. Eaton (1989) and the references therein) and could prove useful in Markov decision theory as well.

We will begin with a review of measurable gambling theory, and then introduce invariance.

1.2 Measurable Gambling Problems

Let F be a *Borel set*, that is, a Borel subset of a complete separable metric space. Let $\mathbb{P}(F)$ be the set of probability measures on the Borel sigma-field of subsets of F. Then $\mathbb{P}(F)$, equipped with its customary weak topology, is again a Borel set. A *gambling house* on F is a subset Γ of $F \times \mathbb{P}(F)$ such that each section $\Gamma(x)$ of Γ at $x \in F$ is nonempty. A *strategy* σ is a sequence $\sigma_0, \sigma_1, \ldots$ such that $\sigma_0 \in \mathbb{P}(F)$, and , for each $n \geq 1$, σ_n is a universally measurable function from X^n into $\mathbb{P}(F)$. A strategy is *available in* Γ *at* x if $\sigma_0 \in \Gamma(x)$ and $\sigma_n(x_1, x_2, \ldots, x_n) \in \Gamma(x_n)$ for every $n \geq 1$ and $x_1, x_2, \ldots, x_n \in X$.

Each strategy σ determines a unique probability measure, also denoted by σ, on the Borel subsets of the *history space* $H = F^N$, where N is the set of positive integers and H is given the product topology. Let $X_1, X_2, \ldots$

be the coordinate process on H; then, under the probability measure σ, X_1 has distribution σ_0 and , for $n \geq 1$, X_{n+1} has conditional distribution $\sigma_n(x_1, x_2, \dots, x_n)$ given $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$.

A *measurable gambling problem* is a triple (F, Γ, u), where F, the *fortune space*, is a nonempty Borel set, Γ is a gambling house on F which is an analytic subset of $F \times \mathbb{P}(F)$, and u, the *utility function*, is upper analytic which means that $\{x : u(x) > a\}$ is an analytic subset of F for every real a. Such structures with Γ and u both Borel were introduced by Strauch (1967); the extension to analytic Γ and upper analytic u is due to Meyer and Traki (1973).

In the theory of gambling (Dubins and Savage (1965)) there are two natural approaches to a gambling problem. In *leavable* problems, the gambler is allowed to stop playing at any time, whereas in *nonleavable* problems, the gambler is compelled to continue playing forever.

Consider first a leavable problem and define a *stop rule* t to be a universally measurable function from H into $\{0, 1, \dots\}$ such that whenever $t(h) = k$ and h and h' agree in their first k coordinates, then $t(h') = k$. A gambler with initial fortune x selects a strategy σ available at x and a stop rule t. The pair $\pi = (\sigma, t)$ is a *policy* available at x. The expected reward to a gambler who selects the policy π is

$$u(\pi) = \int u(X_t)\, d\sigma,$$

where $X_0 = x$. We will usually assume in this paper that u is nonnegative and we will always assume $u(\pi)$ is well-defined and finite for all available π. The *optimal reward function* for this leavable problem is

$$U(x) = \sup u(\pi),$$

where the supremum is taken over all policies π available at x.

It is also of interest to consider how well the gambler can do when playing time is restricted to at most n days, and we define $U_n(x) = \sup u(\pi)$ over all policies $\pi = (\sigma, t)$ available at x such that $t \leq n$ for positive integers n and set $U_0 = u$. The functions U_n can be calculated by backward induction as we will now explain. First define the operator S by

$$(S\phi)(x) = \sup\{\int \phi\, d\gamma : \gamma \in \Gamma(x)\} \tag{1.1}$$

for universally measurable functions ϕ on X such that $\int \phi\, d\gamma$ is well-defined for all available γ. Then

$$U_{n+1} = (SU_n) \vee u\,, \qquad n = 0, 1, \dots\,, \tag{1.2}$$

where $a \vee b$ is the maximum of a and b. For ϕ upper analytic, $S\phi$ is also (Meyer and Traki (1973); see also Dubins and Sudderth (1979)). Thus every U_n is upper analytic. Furthermore,

$$U = \lim U_n. \tag{1.3}$$

(See Section 2.15 of Dubins and Savage (1965) and also Maitra, Purves, and Sudderth (1990).) Hence, U is upper analytic as well.

If the gambling problem is nonleavable, then a gambler with initial fortune x selects a strategy σ available at x and receives the reward

$$u(\sigma) = \limsup_t \int u(X_t)\, d\sigma \ ,$$

where the lim sup is over the directed set of stop rules t. In many cases, for example when u is bounded or when the sequence $u(X_n)$ is monotone increasing or decreasing,

$$u(\sigma) = \int \{\limsup_n u(X_n)\}\, d\sigma. \tag{1.4}$$

(cf. Theorems 4.2.2 and 4.2.7 in Maitra and Sudderth (1996).) The optimal reward function for the nonleavable problem is

$$V(x) = \sup\{u(\sigma)\ :\ \sigma \,\text{available at}\, x\}.$$

The function V can be calculated by a transfinite recursive scheme based on the operator T defined for lower bounded upper analytic functions u by

$$(Tu)(x) = \sup\{\int u(X_t)\, d\sigma\ :\ \sigma \,\text{available at}\, x, t \geq 1\}.$$

It turns out that $Tu = SU$, where U is the optimal reward function for the leavable problem defined above (Maitra and Sudderth(1996, Theorem 4.8.12). Now let $V_0 = Tu$ and , for each countable ordinal $\alpha > 0$, let

$$V_\alpha = \begin{cases} T(u \wedge V_{\alpha-1}), & \text{if}\, \alpha \,\text{is a successor,} \\ \inf_{\beta<\alpha} V_\beta, & \text{if}\, \alpha \,\text{is a limit ordinal.} \end{cases}$$

(Here $a \wedge b$ is the minimum of a and b.) Then

$$V = \inf_\alpha V_\alpha \ ,$$

where the infimum is over all countable ordinals α, and V is also upper analytic (Dubins, Maitra, Purves,and Sudderth (1989), see also Maitra and Sudderth (1996)).

A gambling house Γ is said to be *leavable* if the point-mass $\delta(x) \in \Gamma(x)$ for every $x \in X$. If Γ is leavable, then U and V are the same (Corollary 3.2.2, Dubins and Savage (1965)).

1.3 Invariant Gambling Problems

Let (F, Γ, u) be a measurable gambling problem and suppose that G is a topological group or semigroup with identity that acts on the fortune space F. This means that there is a Borel mapping $a : G \times F \to F$, which we write $a(g, x) = gx$, such that $ex = x$ for e the identity element of G and $g_1(g_2 x) = (g_1 g_2)x$. For a group the mapping $x \to gx$ is necessarily one-to-one for every $g \in G$. We assume that these mappings are one-to-one in the case when G is only a semi-group. We also assume that G is a Polish space; that is, the topology on G is second countable and completely metrizable. The gambling problem will be called *invariant* (under G) if Conditions 1 and 2 below are satisfied.

To formulate the first condition, we associate to each probability measure $\gamma \in \mathbb{P}(F)$ and each $g \in G$ a measure $g\gamma \in \mathbb{P}(F)$ defined by

$$\int \phi(x)\,(g\gamma)(dx) = \int \phi(gx)\,\gamma(dx)$$

for bounded measurable functions ϕ on F. For a set $\Delta \subseteq \mathbb{P}(F)$ and $g \in G$, let

$$g\Delta = \{g\gamma : \gamma \in \Delta\}.$$

Condition 1 *For all $x \in F$ and $g \in G$, $g\Gamma(x) = \Gamma(gx)$.*

To state the second condition, let A be the group of positive, affine transformations of the real line. That is, A consists of all mappings $r \to ar + b$ for some $a > 0$ and $b \in \mathbb{R}$. Give A the topology it inherits when considered as a subset of the plane.

Condition 2 *There is a Borel measurable mapping $w : G \to A$ such that $u(gx) = w(g)(u(x))$ for every $x \in F$ and $g \in G$.*

According to Savage's theory of utility (see Theorems 5.3.2 and 5.3.3 in Savage (1954)), if u is a utility function, then $\tilde{u}$ is also if and only if $\tilde{u}$ is a positive affine transformation of u. Thus Condition 2 says that for every $g \in G$, the function $u_g(x) = u(gx)$ is a utility function.

Lemma 1 *If the function u is not constant, then w is a homomorphism from G into A.*

Proof. For $g_1, g_2 \in G$ and $x \in F$, $w(g_1g_2)(u(x)) = u(g_1g_2x) = w(g_1)(w(g_2)(u(x)))$. Now use the fact that if two affine transformations agree on two points, then they agree everywhere. ∎

As the case of a constant utility function u is uninteresting, we will assume from now on that w is a homomorphism.

Two notions of an invariant function will be used below. Let ψ be a function with domain F and suppose that G acts on the range of ψ. Then ψ will be called simply *invariant* if $\psi(gx) = g\psi(x)$ for all $g \in G, x \in F$. Thus Condition 1 says that the house Γ is invariant. To define the other notion, let w be the mapping given by Condition 2 and let ϕ be a mapping from F to the real numbers. Call ϕ *w-invariant* (under G) if $\phi(gx) = w(g)(\phi(x))$ for all $x \in F$ and $g \in G$. So Condition 2 says that the utility function u is w-invariant.

Lemma 2 *If ϕ is w-invariant and $S\phi$ is well-defined, then $S\phi$ is w-invariant also.*

Proof. For $x \in F$ and $g \in G$,

$$\begin{aligned}
(S\phi)(gx) &= \sup\{\int \phi(x)\,\gamma(dx) \;:\; \gamma \in \Gamma(gx)\} \\
&= \sup\{\int \phi(x)\,(g\gamma)(dx) \;:\; \gamma \in \Gamma(x)\} \\
&= \sup\{\int \phi(gx)\,\gamma(dx) \;:\; \gamma \in \Gamma(x)\} \\
&= \sup\{\int w(g)(\phi(x))\,\gamma(dx) \;:\; \gamma \in \Gamma(x)\} \\
&= w(g)(\sup\{\int \phi(x)\,\gamma(dx) \;:\; \gamma \in \Gamma(x)\}) \\
&= w(g)(S\phi(x)).
\end{aligned}$$

∎

Corollary 3 *The functions $U_n, n \geq 1$, and U are w-invariant, as is the function V when u is bounded below.*

Proof. It is easy to check that that suprema, infima, and pointwise limits of w-invariant functions are again w-invariant. Hence the corollary is a consequence of the lemma and the recursive schemes for calculating U_n, U, and V sketched in the previous section. ∎

Example 1 *Positive Dynamic Programming. Suppose that S is a nonempty Borel set, $\mathbb{R}$ is the set of real numbers, and that $F = S \times \mathbb{R}$. Thus a fortune x is a pair (s, c) whose first coordinate s is regarded as the* state *and whose second coordinate c as the* cash *accumulated up to the present time. The additive group G_0 of real numbers acts on F thus*

$$g(s, c) = (s, c + g)\,, \qquad g \in G_0.$$

Assume that Condition 1 holds. Then, in particular,

$$\Gamma(s, c) = c\Gamma(s, 0)\,, \qquad c \in \mathbb{R}.$$

Thus, if $\gamma \in \Gamma(s, 0)$ and the random pair (s_1, r_1) has distribution γ, then $c\gamma \in \Gamma(s, c)$ and $(s_1, r_1 + c)$ has distribution $c\gamma$. Let the utility function be $u(s, c) = c$. Condition 1 is immediate and, for each g, the affine transformation $w(g)$ is translation by g.

Suppose that a gambler has initial fortune $x = (s, 0)$. The gambler's successive fortunes can be written as $x_1 = (s_1, r_1), x_2 = (s_2, r_1 + r_2), \ldots, x_n = (s_n, r_1 + r_2 + \cdots + r_n), \ldots,$ and $u(x_n) = r_1 + r_2 + \cdots + r_n$. The utility of a strategy σ is

$$u(\sigma) = \limsup_t \int (\sum_{i=1}^{t} r_i)\, d\sigma.$$

Positive dynamic programming corresponds to the special case where, for every $s \in S$ and $\gamma \in \Gamma(s, 0)$, $\gamma\{(s_1, r_1) : r_1 \geq 0\} = 1$. In this case, the gambler's fortune is almost sure to increase in utility at every stage and the utility of σ becomes

$$u(\sigma) = \int (\sum_{i=1}^{\infty} r_i)\, d\sigma.$$

It is natural in the case of positive dynamic programming to restrict the fortune space to be $S \times [0, \infty)$ and to take G to be the semigroup of nonnegative real numbers under addition.

Negative dynamic programming can be formulated by analogy with Example 1. However, the situation is more complicated in the case of discounted dynamic programming and we are indebted to David Heath for the crucial idea in the formulation below.

Example 2 *Discounted Dynamic Programming. Let $0 < \beta < 1$ be the discount factor and let $F = S \times \mathbb{R} \times \mathbb{N}$, where S is a nonempty Borel set, $\mathbb{R}$ the* *set of real numbers, and* $\mathbb{N}$ *the set of nonnegative integers. A fortune* x *is now a triple* (s, c, n)*, where the new coordinate* n *represents the number of days of play. A gamble* $\gamma \in \Gamma(x)$ *determines the distribution of the next fortune* $x_1 = (s_1, \beta^{-1}c + r, n+1)$ *where* s_1 *is the new state and* r *is the reward earned at the current stage of play. The utility function is* $u(x) = u(s, c, n) = \beta^n c$*.*

To get the idea, consider an initial fortune $x = (s, c, 0)$*. The successive fortunes of a gambler can be written as* $x_1 = (s_1, \beta^{-1}c + r, 1), \ldots, x_n = (s_n, \beta^{-n}c + \beta^{-(n-1)}r_1 + \cdots + r_n, n), \ldots$ *. The utility of* x_n *is* $u(x_n) = c + \beta r_1 + \cdots + \beta^n r_n$*.*

Suppose that the daily reward is bounded by a constant b *in the sense that, for every* $s \in S$ *and* $\gamma \in \Gamma(s, 0, 0)$*,* $\gamma\{(s_1, r_1, 1) : |r_1| \leq b\} = 1$*. Then the utility of a strategy* σ *available at* $(s, 0, 0)$ *can be written*

$$u(\sigma) = \int (\sum_{i=1}^{\infty} \beta^i r_i) d\sigma.$$

The additive group of reals $\mathbb{R}$ *and the additive semigroup* $\mathbb{N}$ *both act on* F *as follows:*

$$g(s, c, n) = (s, c + \beta^{-n}g, n)\,, \qquad g \in \mathbb{R}$$

$$m(s, c, n) = (s, c, n + m)\,, \qquad m \in \mathbb{N}.$$

Condition 2 holds with $w(g)(y) = y + g$ *and* $w(m)(y) = \beta^m y$*. Condition 1 is assumed to hold for every* g *and* m*, or, equivalently, for the semigroup* G *of transformations on* F *that they generate.*

Here is a simple class of examples on the real line.

Example 3 *Proportional Houses. Let* $F = [0, \infty)$ *and let* G *be the group of strictly positive real numbers under multiplication. The action of* G *on* F *is just the usual multiplication. Condition 1 now says that a gambler's opportunities are proportional to the size of his fortune. In particular,* $\Gamma(x) = x\Gamma(1)$ *for* $x \geq 0$*. (A famous special case is the red-and-black casino where* $\Gamma(x) = \{w\delta(x + s) + (1 - w)\delta(x - s) : 0 \leq s \leq x\}$*.) A utility function that satisfies Condition 2 is* $u(x) = \log x$*. Now* $u(gx) = \log g + \log x$ *so that the affine function* $w(g)$ *is translation by* $\log g$*. By Corollary 2, the optimal* n*-day return is* w*-invariant and, hence,* $U_n(x) = U_n(x \cdot 1) = w(x)(U_n(1)) = U_n(1) + \log x$*. In fact, it is easy to show, by backward induction, that* $U_n(x) = nU_1(1) + \log x$*. It therefore follows from (1.3) that either* $U(x) = \log x$ *or* $U(x) = \infty$ *according to whether* $U_1(1) = 0$ *or* $U_1(1) > 0$*.*

1.4 Invariant Selectors

A Γ-*selector* is a function $\gamma : F \to \mathbb{P}(F)$ such that $\gamma(x) \in \Gamma(x)$ for every $x \in F$. The von Neumann selection theorem guarantees the existence of a universally measurable Γ-selector γ. Such a selector determines a *stationary* family of strategies γ^∞ which uses the gamble $\gamma(x)$ whenever the current fortune is x. A selector γ is *invariant* if $\gamma(gx) = g\gamma(x)$ for all $x \in F$, $g \in G$ and, in this case, the corresponding stationary family is also called *invariant*. In the next section, we address the question of when invariant stationary families are adequate. This section is devoted to the preliminary problem of the construction of measurable, invariant selectors. For brevity, we often write "u.m." for "universally measurable" below.

Assume first that G is a group. (We will consider the case of a semigroup at the end of the section.) Define R to be the equivalence relation

$$R = \{(x, y) \in F \times F : (\exists g \in G)(gx = y)\}.$$

The equivalence class of an $x \in F$ is its *orbit* $[x] = Gx$. An *orbit selector* is a function $\phi : F \to F$ such that (i) $\phi(x) \in [x]$ for every x, and (ii) $\phi(x) = \phi(y)$ whenever xRy.

For the construction of a measurable, invariant Γ-selector, we will need a measurable orbit selector. The following condition guarantees its existence.

Condition 3 *There is a sequence $\{E_n\}$ of Borel subsets of F such that $xRy \leftrightarrow (\forall n)(x \in E_n \leftrightarrow y \in E_n)$.*

The equivalence relation R is said to be *smooth* when Condition 3 holds.

Lemma 4 *If R is smooth, then R admits a Borel measurable orbit selector.*

Proof. By Theorem 5.2.1 in Becker and Kechris (1996), there is a Polish topology on F with the same Borel sets as the original topology and such that the action $(g, x) \to gx$ is continuous. So we can assume without loss of generality that the action is continuous. The existence of a Borel orbit selector then follows from results of Burgess (1979) or Srivastava (1998). ■

It can be shown that if R is Borel and admits even a universally measurable orbit selector, then R is smooth (Harrington et al, 1990). Thus Condition 3 is necessary as well as sufficient for the existence of a Borel orbit selector.

One additional condition is needed. For each $x \in F$, the *stabilizer subgroup* of x is defined to be $G_x = \{g \in G : gx = x\}$.

Condition 4 *For every $x \in F$, the stabilizer subgroup G_x is either the singleton $\{e\}$ or the whole group G.*

Notice that this condition is satisfied by the proportional houses of Example 3. Indeed, for these houses, $G_0 = G$ and $G_x = \{1\}$, for $x \neq 0$.

Lemma 5 *Under Condition 4, there is a unique Borel function $f : R \to G$ such that, for all $x, y \in F$, (i)$f(x,x) = e$ and (ii)$f(x,y)x = y$ whenever $y \in [x]$.*

Proof. If $G_x = \{e\}$ and $y \in [x]$, then there is a unique $g \in G$ such that $gx = y$ and we take $f(x,y) = g$. If $G_x = G$, then $[x]$ is a singleton and we take $f(x,x) = e$. Then f is Borel because its graph is

$$\{(x,y,g) \in R \times G \,:\, gx = y, x \neq y\} \cup \{(x,x,e) \in R \times G \,:\, x \in F\},$$

a Borel subset of $R \times G$ ∎

Here is the basic lemma on the construction of invariant Γ-selectors.

Lemma 6 *Suppose that G is a group and that Conditions 3 and 4 hold. Let η be a u.m. Γ-selector, ϕ be a Borel measurable orbit selector, and let f be the function of Lemma 4. Then the mapping $\gamma : F \to \mathbb{P}(F)$ defined by*

$$\gamma(x) = f(\phi(x), x)\eta(\phi(x)), \qquad x \in F$$

is a u.m. invariant Γ-selector. Conversely, every u.m. invariant Γ-selector satisfies this equality for some u.m. Γ-selector η.

Proof. For $x \in F$, and $g \in G$,

$$\begin{aligned}
\gamma(gx) &= f(\phi(gx), gx)\eta(\phi(gx)) \\
&= f(\phi(x), gx)\eta(\phi(x)) \\
&= gf(\phi(x), x)\eta(\phi(x)) \\
&= g\gamma(x).
\end{aligned}$$

This proves the first assertion.

Now suppose that γ is a u.m. invariant Γ-selector. Let ϕ be any Borel orbit selector and let $\eta = \gamma$. Then

$$\gamma(x) = \gamma(f(\phi(x), x)\phi(x)) = f(\phi(x), x)\gamma(\phi(x)) = f(\phi(x), x)\eta(\phi(x)).$$

■

Suppose now that G is a semigroup with identity, but not necessarily a group. Call a subset C of F *closed under* G if (i) $x \in C$ and $g \in G$ imply that $gx \in C$ and (ii) $g \in G$ and $gx \in C$ imply that $x \in C$. The *orbit* $[x]$ of $x \in F$ is defined to be the smallest set closed under G and containing x. It is easy to check that

$$R = \{(x, y) \in F \times F \,:\, y \in [x]\}$$

is an equivalence relation on F which is the same as the one previously defined when G is a group. We will not attempt here to develop a general theory for semigroups, but will instead limit ourselves to a special class of invariant problems which includes dynamic programming models and for which invariant selectors are readily available.

The semigroup G is called *special* if (i) there exists a Borel measurable orbit selector ϕ, (ii) for every $x \in F$ and $y \in [x]$, there is a $g = f(x, y)$ such that $f(x, y)\phi(x) = y$, and (iii) the function $f : R \to G$ is Borel measurable. Under Conditions 3 and 4, a group G is a special semigroup.

Lemma 7 *If G is special and η is a u.m. Γ-selector, then*

$$\gamma(x) = f(\phi(x), x)\eta(\phi(x)), \qquad x \in F$$

is a u.m. invariant Γ-selector.

Proof. The proof is the same as for the first half of Lemma 5. ■

For dynamic programming models as in Example 1, the orbit of any fortune $x = (s, c)$ is the set of all fortunes with the same first coordinate and we can take $\phi(x) = (s, 0)$. Then $f(\phi(x), x) = c$ and $\gamma(x) = c\eta(\phi(x))$.

1.5 Invariant Stationary Families of Strategies

Each u.m. Γ-selector γ determines a *stationary family of strategies* $\bar{\sigma}(x) = \gamma^\infty(x),\ x \in F$, where for each x, $\bar{\sigma}(x)$ is the strategy given by

$$\bar{\sigma}(x)_0 = \gamma(x)\,, \qquad \bar{\sigma}(x)_n(x_1, \ldots, x_n) = \gamma(x_n)$$

for all $n \geq 1$ and all $x, x_1, \ldots, x_n \in F$. The stationary family γ^∞ is called *invariant* if the selector γ is invariant.

Dubins and Savage (1965) first posed the problem as to the existence of good stationary families. They obtained a positive result for F finite and Γ leavable which was extended to the countable case by Ornstein (1969). A number of authors (cf. Barbosa-Dantas (1966), Dellacherie and Meyer (1983), Dubins and Sudderth (1979), Frid (1976), Schäl and Sudderth (1987), and Sudderth (1969)) used the techniques of Ornstein to get results about good stationary families in Borel measurable settings. Perhaps all or most of these results can be extended to give results about good invariant stationary families for invariant problems. We present two such extensions in this section and mention some questions that remain open.

Assume that the gambling problem (F, Γ, u) is invariant under a special semi-group G. Assume also that Γ is leavable so that, in particular, the two optimal reward functions U and V are the same.

The assumption that Γ is leavable is necessary for the existence of good stationary families except in the case when F is finite and every set of gambles $\Gamma(x)$, $x \in F$, is finite (Dubins and Savage (1965), see also Maitra and Sudderth (1996)). This assumption can be made without loss of generality for positive dynamic programming problems because, with positive daily rewards, a player gains nothing by terminating the game. This is not the case for negative dynamic programming problems where, as is well-known, good stationary plans need not exist even when the state space is finite (Strauch (1966)).

Recall that to each $g \in G$ is associated an affine mapping $w(g)$ which we can write as

$$w(g)(r) = a_g r + b_g, \qquad r \in \mathbb{R},$$

for some $a_g > 0, b_g \in \mathbb{R}$. We say that the action of G is *additive* if $a_g = 1$ for every $g \in G$ and we call the action *multiplicative* if $b_g = 0$ for every $g \in G$.

Here is an extension of Theorem 2.3 of Sudderth (1969) to the invariant case.

Theorem 8 *Suppose that the function $U - u$ is bounded and that the action of G is additive. Then, for each $\epsilon > 0$ and probability measure $\alpha \in \mathbb{P}(F)$, there is a u.m. invariant Γ-selector γ such that*

$$\alpha\{x \in F : u(\gamma^\infty(x)) \geq U(x) - \epsilon\} = 1.$$

Here is a similar extension of Proposition 7.1 of Dubins and Sudderth (1979).

Theorem 9 *Assume $u \geq 0$ and that the action of G is multiplicative. Then, given $0 < \epsilon < 1$, there is a u.m. invariant Γ-selector γ such that*

$$\alpha\{x \in F \,:\, u(\gamma^{\infty}(x)) \geq (1-\epsilon)U(x)\} = 1.$$

We will explain in the next section how the proof of Dubins and Sudderth (1979) can be modified for Theorem 9. A similar modification of the proof in Sudderth (1969) works for Theorem 8.

Theorem 8 implies the corresponding result of Barbosa-Dantas (1966) for positive dynamic programming problems with bounded optimal reward function. However, neither Theorem 8 nor Theorem 9 includes the result of Frid (1976) for positive dynamic programming problems with unbounded optimal reward function. (Theorem 9 does not apply because the action of the semi-group G in Example 1 is additive rather than multiplicative.) It would be interesting to have an extension of Frid's result to a more general invariant setting.

Dubins and Savage (1965) showed that an optimal stationary family of strategies exists for any gambling problem with F finite and $\Gamma(x)$ finite for all x. The analogous result holds for dynamic programming problems with finite state space and finite action sets. A common generalization would be that there is an optimal invariant stationary family for invariant gambling problems with a finite set of orbits and finite $\Gamma(x)\,, x \in F$. We do not know whether this is true.

1.6 The Proof of Theorem 9

For a proof of Theorem 9, we will show how to make the necessary changes in the arguments given for Proposition 7.1 in Dubins and Sudderth (1979). We will make reference to results in that paper by adding an asterisk. For example, Proposition 7.1* will denote Proposition 7.1 of Dubins and Sudderth (1979).

Throughout this section we fix a Borel measurable orbit selector ϕ as constructed in Section 4. Also, for $\gamma \in \mathbb{P}(F)$, and ψ a γ-integrable function, we will often write $\gamma\psi$ for the integral $\int \psi \, d\gamma$.

Lemma 10 *(cf. Lemma 6.4*) For each $\epsilon > 0$, there is a u.m. invariant Γ-selector γ such that*

$$\gamma(x)u \geq (1-\epsilon)U_1(x) \text{ for all } x \tag{1.5}$$

and

$$\gamma(x) = \delta(x) \Rightarrow u(x) = U_1(x). \tag{1.6}$$

Proof. By Lemma 6.4*, there is a u.m. Γ-selector η that satisfies (1.5) and (1.6), but η need not be invariant. So define

$$\gamma(x) = f(\phi(x), x)\eta(\phi(x))$$

as in Lemmas 6 and 7. Then γ is u.m. and invariant by those lemmas.

To verify (1.5), let $y = \phi(x)$ and $g = f(y, x)$ so that $x = gy$. Then

$$\gamma(x)u = \gamma(gy)u = (g\gamma(y))u = \int u(gz)\,\gamma(y)(dz) = w(g)(\int u(z)\,\gamma(y)(dz))$$

$$\geq w(g)((1-\epsilon)U_1(y)) = (1-\epsilon)w(g)(U_1(y)) = (1-\epsilon)U_1(gy) = (1-\epsilon)U_1(x).$$

The second equality is by the invariance of γ and the next to last is by the w-invariance of U_1 as in Corollary 2.

For (1.6), suppose $\gamma(x) = \delta(x)$, where $x = gy$ as above. Then

$$g\delta(y) = \delta(gy) = \gamma(gy) = g\gamma(y).$$

Hence, $\delta(y) = \gamma(y)$ because the mapping $x \to gx$ is one-to-one by assumption. But $y = \phi(x)$ and $\gamma(y) = \eta(y)$. Now η satisfies condition (1.6). So we have $u(y) = U_1(y)$ and

$$u(x) = w(g)(u(y)) = w(g)(U_1(y)) = U_1(x).$$

■

Define the set $Y = \phi(X)$. Then Y is Borel because $Y = \{x \in F : \phi(x) = x\}$. Also Y intersects each orbit $[x]$ in the singleton $\{\phi(x)\}$.

Corollary 11 *(cf. Corollary 6.1*) For each $\epsilon > 0$ and $\alpha \in \mathbb{P}(F)$, there is an invariant Borel Γ-selector γ such that*

$$\gamma(x)u \geq (1-\epsilon)U_1(x) \tag{1.7}$$

for α-almost every x.

Proof. Use Lemma 10 to get an invariant u.m. Γ-selector $\bar{\gamma}$ such that (1.5) holds when γ is replaced by $\bar{\gamma}$. Define α' to be the probability measure $\alpha\phi^{-1}$ so that, in particular, $\alpha'(Y) = \alpha(\phi^{-1}(Y)) = 1$. Because $\bar{\gamma}$ is u.m. and Γ is leavable, there is a Borel Γ-selector γ' such that $\alpha'\{y \in Y : \gamma'(y) = \bar{\gamma}(y)\} = 1$. Define

$$\gamma(x) = f(\phi(x), x)\gamma'(\phi(x)), \qquad x \in F.$$

Then γ is Borel measurable because f, ϕ, and γ' are. Also γ is an invariant Γ-selector by Lemma 6 or Lemma 7. Note that $\gamma'(\phi(x)) = \bar{\gamma}(\phi(x)) \Rightarrow \gamma(x) = \bar{\gamma}(x)$ as $\bar{\gamma}$ is invariant. Hence,

$$\begin{aligned}\alpha(\{x \in F : \gamma(x) = \bar{\gamma}(x)\}) &\geq \alpha(\{x \in F : \gamma'(\phi(x)) = \bar{\gamma}(\phi(x))\} \\ &= \alpha'(\{y \in Y : \gamma'(y) = \bar{\gamma}(y)\} \\ &= 1.\end{aligned}$$

Thus (1.7) holds for γ α-almost surely since it holds for $\bar{\gamma}$.

∎

The next result is that a gambler can get uniformly close (in the multiplicative sense) to the optimal n-day return U_n using an invariant stationary family of strategies. The same is not true for the optimal reward function U even if the gambler uses a stationary family that is not invariant (Blackwell and Ramakrishnan, 1988).

Lemma 12 *(cf. Proposition* 6.1**) For* $n \geq 1$ *and* $\epsilon > 0$*, there is a u.m. invariant* Γ*-selector such that*

$$u(\gamma^\infty(x)) \geq (1-\epsilon)U_n(x)$$

for all $x \in F$*.*

Proof. Choose β so that $0 < \beta < 1$ and $\beta^n > 1 - \epsilon$. By Lemma 10, we can find u.m. invariant Γ-selectors $\gamma_k, k = 1, \ldots, n$ such that

$$\gamma_k(x)U_{k-1} \geq \beta^{1/2}U_k(x), \qquad x \in F.$$

Let $k(x)$ be the least natural number $k \leq n$ such that

$$\beta^k U_k(x) = \max_{0 \leq j \leq n} \beta^j U_j(x).$$

Since the U_k's are w-invariant and u.m., it follows that $k(x)$ is u.m. and invariant. Consequently, γ defined by

$$\gamma(x) = \begin{cases} \gamma_{k(x)}(x), & k(x) = 1, \ldots, n \\ \delta(x), & k(x) = 0 \end{cases}$$

is a u.m. invariant Γ-selector. Proposition 5.1* now applies to yield the desired result. ∎

To overcome certain measurability difficulties, Dubins and Sudderth found it useful to replace the original gambling problem in their paper by a simpler Borel problem. The next two lemmas enable us to do the same.

Lemma 13 *Given $u \geq 0$, u.m., and w-invariant, and $\alpha \in \mathbb{P}(F)$, there is a Borel w-invariant function v on F such that $0 \leq v \leq u$ and $v = u$ α-almost surely.*

Proof. As in the proof of Corollary 11, let $\alpha' = \alpha\phi^{-1}$ so that $\alpha'(Y) = 1$. Choose a Borel function $v' : Y \to [0, \infty)$ such that $0 \leq v' \leq u$ on Y and $\alpha'(\{y \in Y : v'(y) = u(y)\}) = 1$. Recall from Section 4 that each $x \in F$ can be written as $x = f(\phi(x), x)\phi(x)$ for a unique group element $f(\phi(x), x)$ and set

$$v(x) = w(f(\phi(x), x))(v'(\phi(x)), \qquad x \in F.$$

Then v is Borel measurable because all of the functions appearing in its definition are Borel. To check that v is w-invariant, let $x \in F, g \in G$ and calculate:

$$\begin{aligned} v(gx) &= w(f(\phi(gx), gx))(v'(\phi(gx))) \\ &= w(f(\phi(x), gx))(v'(\phi(x))) \\ &= w(gf(\phi(x), x))(v'(\phi(x))) \\ &= w(g)(w(f(\phi(x), x))(v'(\phi(x)))) \\ &= w(g)(v(x)). \end{aligned}$$

Here the invariance of ϕ is used for the second equality and Lemma 1 for the next to last equality.

Next note that $v'(\phi(x)) = u(\phi(x)) \Rightarrow v(x) = u(x)$, since u is w-invariant. Now $\alpha'(\{x \in F : v'(\phi(x)) = u(\phi(x))\}) = 1$. So it follows that $v = u$ α-almost surely. Finally, $v(x) = w(f(\phi(x), x))(v'(\phi(x)) \leq w(f(\phi(x), x))(u(\phi(x))) = u(x)$. ■

A leavable house Γ' is *(Borel) countably parametrized* if there exist Borel Markov kernels $\gamma_1, \gamma_2, \ldots$ such that $\Gamma'(x) = \{\gamma_1(x), \gamma_2(x), \ldots\}$ for every $x \in F$. The house Γ' is clearly invariant if each of the functions γ_k is invariant.

Lemma 14 *(cf. Lemma 7.1*) For each $\alpha \in \mathbb{P}(F)$, there is an invariant, countably parametrized subhouse Γ' of Γ and a nonnegative Borel function u' on F such that*

(i) $u' \leq u$ on F, and

(ii) $U' \geq U$ α-almost surely,

where U' is the optimal reward function for (F, Γ', u').

Proof. Repeat the proof of Lemma 7.1* using Corollary 11 instead of Corollary 6.1* so that the Γ-selectors γ_k turn out to be both invariant and Borel measurable. To get the function u', use Lemma 13. ∎

In view of Lemma 14 it now suffices to prove Theorem 9 under the additional assumptions that Γ is countably parametrized and u is Borel. These assumptions will be in force for the remainder of the proof.

Lemma 15 *(cf. Lemma 7.2*) The functions $U_1, U_2, \dots, U$ are Borel measurable. For each $\epsilon > 0$ and $n \geq 1$, there is an invariant, Borel Γ-selector γ such that $u(\gamma^\infty(\cdot))$ is w-invariant, Borel measurable, and*

$$u(\gamma^\infty(x)) \geq (1 - \epsilon/2)U_n(x) \text{ for all } x. \tag{1.8}$$

Hence, for each $\alpha \in \mathbb{P}(F)$, there exists such a Borel Γ-selector γ for which

$$u(\gamma^\infty(x)) \geq (1 - \epsilon)U(x) \tag{1.9}$$

with α-probability at least $1 - \epsilon$.

Proof. The first assertion is easy to prove because Γ is countably parametrized (see Theorem 4.1 of Sudderth (1969)). Next choose $\beta \in (0, 1)$ such that $\beta^n > 1 - \epsilon/2$. Then the Γ-selector γ given by Lemma 12 is invariant and satisfies (1.8). Also, because the functions $U_1, U_2, \dots$ are Borel, it is clear from the construction in the proof of Lemma 12 that γ is Borel. The function $u(\gamma^\infty(\cdot))$ is the V for the invariant house Γ' where $\Gamma'(x) = \{\gamma(x)\}, x \in F$. Thus $u(\gamma^\infty(\cdot))$ is w-invariant by Corollary 3. It is not clear that the function $u(\gamma^\infty(\cdot))$ is Borel, but the argument for Lemma 7.2* shows that γ can be chosen so that this is true. Since $U_n \uparrow U$, the final assertion follows from (1.8). ∎

Consider now a leavable, Borel, *stop-or-go house* Σ, which means that, for some Borel Markov kernel η and all x, $\Sigma(x) = \{\eta(x), \delta(x)\}$. Assume also that η is invariant and let W be the optimal reward function for the invariant problem (F, Σ, u).

Lemma 16 *The stationary family γ^∞, where*

$$\gamma(x) = \begin{cases} \eta(x), & u(x) < W(x) \\ \delta(x), & u(x) = W(x), \end{cases}$$

is Borel measurable, invariant, and optimal for (F, Σ, u).

Proof. The house Σ is Borel and countably parametrized. Hence, W is Borel by Lemma 15 and w-invariant by Corollary 3. It follows that γ is Borel and invariant. The optimality of γ^∞ is a consequence of Proposition 4.1*. ∎

The idea of the remainder of the proof, based on ideas of Ornstein (1969), is to find a Borel invariant stop-or-go subhouse of the original Γ whose optimal reward function is almost surely as large as $(1-\epsilon)U$. Theorem 9 will then follow from Lemma 16.

The construction of the stop-or-go house is in a countable number of steps and each step will use the following operation.

To each stop-or-go house Σ, associate the house $\Gamma \cdot \Sigma$ defined by

$$(\Gamma \cdot \Sigma)(x) = \begin{cases} \Sigma(x), & \text{if } \Sigma(x) \text{ contains two elements} \\ \Gamma(x), & \text{otherwise.} \end{cases}$$

Plainly, $\Sigma \subseteq \Gamma \Rightarrow \Sigma \subseteq \Gamma \cdot \Sigma \subseteq \Gamma$ and $\Sigma \subseteq \Sigma' \Rightarrow \Gamma \cdot \Sigma' \subseteq \Gamma \cdot \Sigma$.

If η is Borel measurable and $\Sigma(x) = \{\eta(x), \delta(x)\}$ for all x, then Σ is Borel and countably parametrized, as is $\Gamma \cdot \Sigma$. So, by Lemma 15, the optimal return function W for Σ is Borel measurable, as is the optimal return function R for $\Gamma \cdot \Sigma$.

Lemma 17 *(cf. Lemma 7.5*) Suppose Σ is a leavable, invariant, Borel, stop-or-go subhouse of Γ, $\alpha \in \mathbb{P}(F)$, and $\epsilon > 0$. Then there is a leavable, invariant, Borel, stop-or-go house Σ' such that*

(i) $\Sigma \subseteq \Sigma' \subseteq \Gamma$,

(ii) $\alpha[W' \geq (1-\epsilon)R] \geq 1-\epsilon$, *and*

(iii) $R' \geq (1-\epsilon)R$,

where W' and R' are the optimal return functions for Σ' and $\Gamma \cdot \Sigma'$, respectively.

Proof. The house $\Gamma \cdot \Sigma$ is invariant, and hence, by Lemma 15, there is an invariant, Borel $\Gamma \cdot \Sigma$-selector γ such that $u(\gamma^\infty(\cdot))$ is Borel, w-invariant, and $\alpha(S) > 1-\epsilon$, where $S = \{x : u(\gamma^\infty(x)) \geq (1-\epsilon^2/2)R(x)\}$. Set $T = \{x : u(\gamma^\infty(x)) \geq (1-\epsilon)R(x)\}$. Then the set T is Borel and, because the functions u and R are w-invariant and the action of G is multiplicative,

T also has the property that $x \in T \Leftrightarrow gx \in T$ for all $x \in F, g \in G$. Now define the house Σ' by

$$\Sigma'(x) = \begin{cases} \{\eta(x), \delta(x)\}, \text{ if } x \in T \text{ and } \Sigma(x) = \{\delta(x)\}, \\ \Sigma(x), \text{ if } \Sigma(x) \text{ contains two elements,} \\ \{\delta(x)\}, \text{ otherwise.} \end{cases}$$

The proof that Σ' has the desired properties, with the exception of invariance, is exactly the same as in the proof of Lemma 7.5*. The proof that Σ' is invariant is straightforward if one uses the property of T mentionned above and our standing assumption that the mappings $x \to gx$ are one-to-one. ∎

The assumption that G acts multiplicatively was used in the proof above for Lemma 17, but is used nowhere else in the proof of Theorem 9.

Lemma 18 *(cf. Lemma 7.6*) Let $\alpha \in \mathbb{P}(F)$ and $\epsilon > 0$. There is a sequence $\Sigma_0 \subseteq \Sigma_1 \subseteq \cdots$ of leavable, invariant, Borel stop-or-go subhouses of Γ whose optimal return functions $W_0, W_1, \ldots$ satisfy*

$$\alpha[W_n \geq (1-\epsilon)U] \uparrow 1.$$

Proof. The proof is the same as that of Lemma 7.6* except that Lemma 17 is used instead of Lemma 7.5*. ∎

Define the house Σ by setting

$$\Sigma(x) = \cup_n \Sigma_n(x), \qquad x \in F,$$

where the Σ_n are from Lemma 18. Then Σ is a leavable, invariant, stop-or-go subhouse of Γ whose optimal return function W satisfies $\alpha[W \geq (1-\epsilon)U] = 1$. Theorem 9 now follows from Lemma 16.

1.7 Acknowledgement

The research of Sudderth was partially supported by National Science Foundation Grant DMS 97-03285.

Bibliography

Barbosa-Dantas [1] C.A. Barbosa-Dantas, *The Existence of Stationary Optimal Plans*, Ph.D. Dissertation, University of California, Berkeley, 1966.

Becker [2] H. Becker and A.S. Kechris, *The Descriptive Set Theory of Polish Group Actions*, LMS Lecture Notes, Cambridge University Press, 1996.

Blackwell [3] D. Blackwell, " The stochastic processes of Borel gambling and dynamic programming," *Annals of Statistics* **4**, 370-374, 1976.

B-R [4] D. Blackwell and S. Ramakrishnan, "Stationary plans need not be uniformly optimal for leavable, Borel gambling problems," *Proceedings of the American Mathematical Society* **102**, 1024-1027, 1988.

Burgess [5] J.P. Burgess, "A selection theorem for group actions," *Pacific Journal of Mathematics* **80**, 333-336, 1979.

Dellacherie [6] C. Dellacherie and P.A. Meyer, *Probabilités et Potentiel*, Chapitres IX à XI, Hermann, Paris, 1983.

DMPS [7] L. Dubins, A. Maitra, R. Purves, and W. Sudderth, "Measurable, nonleavable gambling problems," *Israel Journal of Mathematics* **67**, 257-271, 1989.

DubinsSavage [8] L.E. Dubins and L.J. Savage, *How to Gamble If You Must: Inequalities for Stochastic Processes*, McGraw, New York, 1965.

DubinsSud [9] L.E. Dubins and W.D. Sudderth, "On stationary strategies for absolutely continuous houses," *Annals of Probability* **7**, 461-476, 1979.

Eaton [10] M. Eaton, *Invariance Applications in Statistics*, Regional Conference Series in Probability and Statistics I, Institute of Mathematical Statistics, Hayward, California, 1989.

Frid [11] E.B. Frid, “On a problem of D. Blackwell from the theory of dynamic programming,” *Theory of Probability and Applications* **15**, 719-722, 1976.

Harrington [12] L. Harrington, A.S. Kechris, and A. Louveau, “A Glimm-Effros dichotomy for Borel equivalence relations,” *Journal of the American Mathematical Society* **3** 903-928, 1990.

MPS [13] A. Maitra, R. Purves, and W. Sudderth, “Leavable gambling problems with unbounded utilities,” *Transactions of the American Mathematical Society* **333**, 543-567, 1990.

MS [14] A. Maitra and W. Sudderth, *Discrete Gambling and Stochastic Games*, Springer-Verlag, New York, 1996.

Meyer [15] P.A. Meyer and M. Traki, “Réduites et jeux de hasard,” *Séminaire de Probabilité XII*, Lecture Notes in Mathematics 321, Springer-Verlag, Berlin, 155-171, 1973.

Ornstein [16] D. Ornstein, “On the existence of stationary optimal strategies,” *Proceedings of the American Mathematical Society* **20**, 563-569, 1969.

Savage [17] L.J. Savage, *The Foundations of Statistics*, Wiley, New York, 1954.

SS [18] M. Schäl and W. Sudderth, “Stationary policies and Markov policies in Borel dynamic programming,” *Probability Theory and Related Fields* **74**, 91-111, 1987.

Srivastava [19] S.M. Srivastava, *A Course on Borel Sets*, Springer-Verlag, Berlin, 1998.

Strauch [20] R.E. Strauch, “Negative dynamic programming,” *Annals of Mathematical Statistics* **37**, 871-890, 1966.

Strauch2 [21] R.E. Strauch, “Measurable gambling houses,” *Transactions of the American Mathematical Society* **126**, 64-72, 1967.

Sud [22] W.D. Sudderth, “On the existence of good stationary strategies,” *Transactions of the American Mathematical Society* **135**, 399-414, 1969.