13 INVARIANT GAMBLING PROBLEMS AND MARKOV DECISION PROCESSES

Lester E. Dubins
Ashok P. Maitra
William D. Sudderth

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.

13.1 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. From a different point of view Schäl (1989) has also studied the reward structures of gambling and Markov decision theory. 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.

13.2 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 $\sigma_0, \sigma_1, \dots$ such that $\sigma_0 \in \mathbb{P}(F)$, and , for each $n \geq 1$, σ_n is a universally measurable function from F^n into $\mathbb{P}(F)$. A strategy is *available in* Γ *at* x if $\sigma_0 \in \Gamma(x)$ and $\sigma_n(x_1, x_2, \dots, x_n) \in \Gamma(x_n)$ for every $n \geq 1$ and $x_1, x_2, \dots, x_n \in F$.

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 $X_1, X_2, \dots$ 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 $u : F \mapsto \mathbb{R}$, 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 $t(h') = k$. (In particular, if $t(h) = 0$ for some h, then t is identically 0.) 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

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