Statistics, Probability and Game Theory
IMS Lecture Notes – Monograph Series (1996) Volume 30

BLACKWELL OPTIMAL POLICIES IN COUNTABLE DYNAMIC PROGRAMMING WITHOUT APERIODICITY ASSUMPTIONS

ALEXANDER A. YUSHKEVICH[1]
University of North Carolina at Charlotte

Abstract

The existence of stationary Blackwell optimal policies is proved in denumerable dynamic programming models satisfying compactness-continuity conditions and a uniform Doeblin condition. The latter is given in terms compatible with periodic chains.

1. Introduction and Summary. Dekker and Hordijk (1988), (1992) studied the existence of stationary Blackwell optimal policies in denumerable Markov decision chains with compact action sets and continuous in action transition probabilities and rewards. In the first of the cited works, their most complete result is proved under a uniform geometric convergence condition which excludes periodic chains. Recently Tijms (1994) developed an elegant way to treat the periodicity in connection with the average optimality equation: namely, to substitute the given controlled chain by a perturbed chain with a same geometric sitting time at all states. In this short note we show that Tijms' (1994) idea, combined with the Dekker and Hordijk's (1988) results, can be used in connection with the Blackwell optimality too. In the second of the cited publications, Dekker and Hordijk substitute the uniform geometric convergence condition by a uniform geometric recurrence condition which does not exclude periodic chains. However, the relation between the latter condition and the Tijms' condition we use here is not clear. It is a great pleasure to contribute this paper to a volume in the honor of David Blackwell, whom the author considers as his teacher in dynamic programming.

A dynamic programming model is determined by a state space X, action sets $A(x)$, a transition function $p(x, a, B)$, and a real-valued reward function $r(x, a)$, $a \in A(x)$, $x \in X$, $B \subset X$ (we omit measurability assumptions and other formalities in this preliminary paragraph). The selection of an initial state x and a policy π defines a probability distribution $\mathbf{P}_x^\pi$ in the space of sequences $x_0a_1x_1a_2\dots$ of consecutively visited states x_t and actions

[1]Supported by NSF Grant DMS-9404177.