MARKOV DECISION PROCESSES, DYNAMIC PROGRAMMING, AND REINFORCEMENT LEARNING IN R

JEFFREY TODD LINS
THOMAS JAKOBSEN

SAXO BANK A/S

Markov decision processes (MDP), also known as discrete-time stochastic control processes, are a cornerstone in the study of sequential optimization problems that arise in a wide range of fields, from engineering to robotics to finance, where the results of actions taken under planning may be uncertain.

An MDP is characterized by mappings for a set of states, actions, Markovian transition probabilities, and real-valued rewards within the process. An optimal planning solution seeks to maximize the sum of rewards over states under some decision policy for state-action pairs given updated transition probabilities.

The concept of dynamic programming was introduced by Bellman and is a classical solution method for approaching MDPs, however, in practice, the applicability of dynamic programming may be prohibited by the sheer size of underlying state spaces for real world problems – Bellman's so-called "curse of dimensionality" – for whereas a linear program representing an MDP can be solved in polynomial time, the degree of the polynomial may be be large enough to render theoretical algorithms inefficient in practice.

In addition, many problems do not allow for direct observations of the state space or reward functions, but rather only of some noisy information about the current state. These so-called *partially observable* MDPs constitute a class for which exact solutions may only be found efficiently for the smallest of state spaces.

Reinforcement learning extends Bellman's equations and other approaches to methods which employ robust function approximations, in order to make solutions for MDPs and POMDPs computationally tractable, and many of the wide variety of these approaches leverage statistical methods, including least squares regression, Monte Carlo methods, simulated annealing, and Markov chain methods, available in many R packages.

We demonstrate dynamic programming algorithms and reinforcement learning employing function approximations which should become available in a forthcoming R package. We highlight particularly the use of statistical methods from standard functions and contributed packages available in R, and some applications of reinforcement learning to sequential stochastic processes.

References

[Bellman, 1961] Bellman, R. (1961). *Adaptive Control Processes*. Princeton University Press.

Date: February 27, 2006.

[Littman et al., 1995] Littman, M., Cassandra, A., and Kaelbling, L. (1995). Learning policies for partially observable environments: Scaling up. In Prieditis, A. and Russell, S., editors, *Machine Learning: Proceedings of the Twelfth International Conference*, pages 362–370. Morgan Kaufmann Publishers, San Francisco, CA.

[Sutton and Barto, 1998] Sutton, R. and Barto, A. (1998). *Reinforcement learning: An introduction*. MIT Press, Cambridge, MA.

Quantitative Analysis, Saxo Bank A/S, Smakkedalen 2, 2820 Gentofte, Denmark

E-mail address: jtl@saxobank.com, tja@saxobank.com