Markov decision processes and interval Markov chains: exploiting the connection

Mingmei Teo
Supervisors: Prof. Nigel Bean, Dr Joshua Ross

University of Adelaide

July 10, 2013

THE UNIVERSITY of ADELAIDE

Intervals and interval arithmetic

- We use the notation

$$X = [\underline{X}, \overline{X}]$$

to represent an interval

- Interval arithmetic allows us to perform arithmetic operations on intervals and can be represented as follows

$$X \odot Y = \{x \odot y : x \in X, y \in Y\}$$

where X and Y represent intervals and $\odot$ is the arithmetic operator

Intervals and interval arithmetic

Let $X = [-1, 1]$. Then we have

$$X^2 = \{x^2 : x \in [-1, 1]\} = [0, 1]$$

whilst

$$X \cdot X = \{x_1 \cdot x_2 : x_1 \in [-1, 1], x_2 \in [-1, 1]\} = [-1, 1].$$

So here, we have the idea of 'one-sample' and 're-sample'.

Computation with interval arithmetic

- Computational software, e.g. INTLAB
  - Performs arithmetic operations on interval vectors and matrices
  - Solves systems of linear equations with intervals

Why might interval arithmetic be useful?

- Point estimate of parameters with sensitivity analysis
- Can we avoid the need for sensitivity analysis?
- Is it possible to directly incorporate the uncertainty of parameter values into our model?
- Intervals can be used to bound our parameter values,

$$[x - error, x + error]$$

Markov chains + intervals = ?

- Consider a discrete time Markov chain with $n+1$ states, $\{0,\ldots,n\}$, and state 0 an absorbing state
- Interval transition probability matrix

$$\mathbb{P} = \left[\begin{array}{c|ccc} [1,1] & [0,0] & \cdots & [0,0] \\ \hline [\underline{P}_{10}, \overline{P}_{10}] & & & \\ \vdots & & \mathbb{P}_s & \\ [\underline{P}_{n0}, \overline{P}_{n0}] & & & \end{array}\right]$$

Conditions on the interval transition probability matrix

- Bounds are valid probabilities,

$$0 \leq \underline{P}_{ij} \leq \overline{P}_{ij} \leq 1$$

- Row sums must satisfy the following,

$$\sum_j \underline{P}_{ij} \leq 1 \leq \sum_j \overline{P}_{ij}$$

Time homogeneity

- Standard Markov chains:
  - One-step transition probability matrix, P, constant over time
- Interval Markov chains:
  - Time inhomogeneous interval matrix
  - Time homogeneous interval matrix
    - One-sample (Time homogeneous Markov chain)
    - Re-sample (Time inhomogeneous Markov chain)

Hitting times and mean hitting times

- N_i is the random variable describing the number of steps required to hit state 0 conditional on starting in state i
- $\nu_i = E[N_i]$ is expected number of steps needed to hit state 0 conditional on starting in state i

Hitting times problem

We want to calculate an interval hitting times vector, $[\underline{\boldsymbol{\nu}}, \overline{\boldsymbol{\nu}}]$, for our interval Markov chain. That is, we want to solve

$$[\underline{\boldsymbol{\nu}}, \overline{\boldsymbol{\nu}}] = (I - \mathbb{P}_s)^{-1}\mathbf{1}$$

where I is the identity matrix, $\mathbf{1}$ is vector of ones, $\mathbb{P}_s$ is sub-matrix of the interval matrix $\mathbb{P}$ and $\underline{\boldsymbol{\nu}}$ and $\overline{\boldsymbol{\nu}}$ represent the lower and upper bounds of the hitting times vector.

Can we solve the system of equations directly?

- Can we just use INTLAB and interval arithmetic to solve the system of equations?
- INTLAB uses an iterative method to solve the system of equations
    - Problem: ensuring the same realisation of the interval matrix is chosen at each iteration
- Problem: ensuring $\sum_j P_{ij} = 1$

Hitting times interval

We seek to calculate the interval hitting times vector of an interval Markov chain by minimising and maximising the hitting times vector,

$$\boldsymbol{\nu} = (I - P_s)^{-1} \mathbf{1},$$

where

$$P_s = \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{1n} & \cdots & P_{nn} \end{bmatrix}$$

is a realisation of the interval $\mathbb{P}_s$ matrix with the row sums condition obeyed.

Maximisation case

We wanted to solve the following maximisation problem for $k = 1, \ldots, n$.

$$\max \nu_k = \left[(I - P_s)^{-1} \mathbf{1} \right]_k$$

subject to

$$\sum_{j=0}^{n} P_{ij} = 1, \qquad \text{for } i = 1, \ldots, n,$$

$$\underline{P}_{ij} \leq P_{ij} \leq \overline{P}_{ij}, \qquad \text{for } i = 1, \ldots, n;\ j = 0, \ldots, n.$$

New formulation of the problem

$$\max \nu_k = \left[(I - P_s)^{-1} \mathbf{1} \right]_k$$

subject to

$$\sum_{j=1}^{n} P_{ij} = 1 - \underline{P}_{i0}, \qquad \text{for } i = 1, \ldots, n,$$

$$\underline{P}_{ij} \leq P_{ij} \leq \overline{P}_{ij}, \qquad \text{for } i, j = 1, \ldots, n.$$

Feasible region of maximisation problem

- Constraints are row-based
- Let F_i be the feasible region of row i, for $i = 1, \ldots, n$
- Represents the possible vectors for the i^{th} row of the P_s matrix
- F_i is defined by bounds and linear constraints which form a convex hull

What can we do with this?

- Numerical experience suggests the optimal solution occurs at a vertex of the feasible region
- Look to prove this conjecture using Markov decision processes (MDPs)
- We want to be able to represent our maximisation problem as an MDP and exploit existing MDP theory

What are Markov decision processes?

- A way to model decision making processes to optimise a pre-defined objective in a stochastic environment
- Described by decision times, states, actions, rewards and transition probabilities
- Optimised by decision rules and policies

Mapping

Lemma

Our maximisation problem is a Markov decision process restricted to only consider Markovian decision rules and stationary policies.

- Prove this by representing our maximisation problem as an MDP

Proof: states, decision times and rewards

- States
  - Both representations involve the same underlying Markov chain
- Decision times
  - Every time step of the underlying Markov chain
  - Infinite-horizon MDP as we allow the process to continue until absorption
- Reward $= 1$
  - Each step increases the time to absorption by one

Proof: actions

- Recall, F_i is the feasible region of row i
- We choose to let each vertex in F_i correspond to an action of the MDP when in state i
- To recover the full feasible region, need convex combinations of vertices $\Rightarrow$ convex combinations of actions

Proof: transition probabilities

- Let $\mathbf{P}_i^{(a)}$ be the associated probability distribution vector for an action a
- When an action a is chosen in state i, the corresponding $\mathbf{P}_i^{(a)}$ is inserted into the i^{th} row of the matrix, P_s
- Considering all states $i = 1, \ldots, n$, we get the P_s matrix

Proof: Markovian decision rules and stationary policy

- Markovian decision rules
  - Maximisation problem involves choosing the transition probabilities of a Markov chain
- Stationary policy
  - We have a time homogeneous (one-sample) interval Markov chain
  - Means optimal P_s matrix remains constant over time
  - Hence the choice of decision rule is independent of time

Optimal at vertex

Theorem

There exists an optimal solution of the maximisation problem where row i of the optimal matrix, P_s^*, represents a vertex of F_i for all $i = 1, \ldots, n$.

- Need to show there is no extra benefit from having randomised decision rules as opposed to deterministic decision rules

Why do we care about randomised and deterministic?

- Randomised decision rules ⇒ convex combination of actions ⇒ non-vertex of F_i
- Deterministic decision rules ⇒ single action ⇒ vertex of F_i
- Want deterministic decision rules!

Proof

Proposition (Proposition 6.2.1. of Puterman[1])

For all $v \in V$,

$$\sup_{d \in D^{MD}} \{r_d + P_d v\} = \sup_{d \in D^{MR}} \{r_d + P_d v\}.$$

- This proposition from Puterman[1] gives us that there is nothing to be gained from randomised decision rules
- So there exists an optimal is obtained for deterministic decision rules

[1] M.L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming

Conclusions

- Proven that an optimal solution occurs at a vertex of the feasible region
- This theorem provides us with a useful analytic property which we can exploit when obtaining the optimal solution through numerical methods

What else?

- Determine if interval analysis can be used to investigate model sensitivity
- Vary width of intervals for parameters and see effect on mean hitting times intervals

Questions

Questions?

Counter-example for an analytic solution

Consider the following interval transition probability matrix,

$$\mathbb{P} = \begin{bmatrix} [1,1] & [0,0] & [0,0] & [0,0] \\ [0.3,0.35] & [0,1] & [0,0] & [0,0.1] \\ [0.2,0.3] & [0,1] & [0,1] & [0,1] \\ [0.1,0.2] & [0,1] & [0,0.3] & [0,0] \end{bmatrix}.$$

Counter-example for an analytic solution

Our proposed analytic solution:

$$P_s = \begin{bmatrix} 0.6 & 0 & 0.1 \\ 0 & 0 & 0.8 \\ 0.6 & 0.3 & 0 \end{bmatrix}.$$

Optimal solution obtained numerically from MATLAB:

$$P_s^* = \begin{bmatrix} 0.6 & 0 & 0.1 \\ 0 & 0.8 & 0 \\ 0.6 & 0.3 & 0 \end{bmatrix}.$$