Solving act-state independent imprecise decision processes

Ricardo Shirota Filho, Matthias C. M. Troffaes, Nathan Huntley

University of São Paulo, Brasil; Durham Univeristy, UK

WPMSIIP - September 2009

From the University of São Paulo

Ricardo Shirota Filho

- Ph.D. student at the University of São Paulo
- Supervised by Prof. Fabio G. Cozman
- Currently visiting Durham University
  - Markov Decision Processes (sequential decision making)
  - Artificial Intelligence Planning
  - Imprecise probabilities
  - Algorithms

From Durham University

Matthias C. M. Troffaes

- Lecturer at Durham University
  - Foundations of statistics
  - Sequential decision making
  - Imprecise probabilities
  - Reliability, fault trees

Nathan Huntley

- Ph.D. student at Durham University
- Supervised by Matthias Troffaes
  - Sequential decision making
  - Imprecise probabilities

After a few discussions...

A common topic of interest

Investigate conditions for optimality in sequential decision making with imprecision under fairly general assumptions

- Ricardo: Markov Decision Processes, applications
- Matthias and Nathan: decision trees, arbitrary choice functions, gambles

The menu

In this presentation

- General description and current results
- A simple illustrative example

Later...

- Matthias: Locality property and implications for foundations
- Nathan: Implications for backward induction

The menu

In this presentation

- General description and current results
- A simple illustrative example

Later...

- Matthias: Locality property and implications for foundations
- Nathan: Implications for backward induction

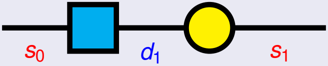
Sequential decision making

Single decision

Sequential decisions

Sequential decision making

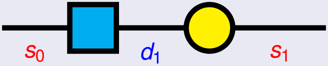
Single decision



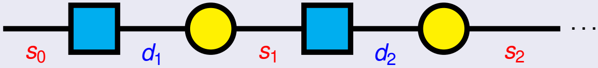
Sequential decisions

Sequential decision making

Single decision



Sequential decisions



Assumptions

For more general results

- No probabilities are assumed and rewards do not need to be expressed in terms of utility, instead we use arbitrary choice functions
- Rewards can depend on full state history
- State and action spaces can depend on the stage

However, not everything is perfect...

Limiting condition

- Act-state independence could not be avoided

Assumptions

For more general results

- No probabilities are assumed and rewards do not need to be expressed in terms of utility, instead we use arbitrary choice functions
- Rewards can depend on full state history
- State and action spaces can depend on the stage

However, not everything is perfect...

Limiting condition

- Act-state independence could not be avoided

Assumptions

For more general results

- No probabilities are assumed and rewards do not need to be expressed in terms of utility, instead we use arbitrary choice functions
- Rewards can depend on full state history
- State and action spaces can depend on the stage

However, not everything is perfect...

Limiting condition

- Act-state independence could not be avoided

Assumptions

For more general results

- No probabilities are assumed and rewards do not need to be expressed in terms of utility, instead we use arbitrary choice functions
- Rewards can depend on full state history
- State and action spaces can depend on the stage

However, not everything is perfect...

Limiting condition

- Act-state independence could not be avoided

Assumptions

For more general results

- No probabilities are assumed and rewards do not need to be expressed in terms of utility, instead we use arbitrary choice functions
- Rewards can depend on full state history
- State and action spaces can depend on the stage

However, not everything is perfect...

Limiting condition

- Act-state independence could not be avoided

Assumptions

For more general results

- No probabilities are assumed and rewards do not need to be expressed in terms of utility, instead we use arbitrary choice functions
- Rewards can depend on full state history
- State and action spaces can depend on the stage

However, not everything is perfect...

Limiting condition

- Act-state independence could not be avoided

Assumptions

For more general results

- No probabilities are assumed and rewards do not need to be expressed in terms of utility, instead we use arbitrary choice functions
- Rewards can depend on full state history
- State and action spaces can depend on the stage

However, not everything is perfect...

Limiting condition

- Act-state independence could not be avoided

Our main result

Locality property

$$X + \bigoplus_{s_k} E_{s_k} Y(s_k) \in \operatorname{opt}\left(\mathcal{X} + \bigoplus_{s_k} E_{s_k} \mathcal{Y}(s_k) \middle| h_{k-1}\right) \iff$$
$$X \in \operatorname{opt}(\mathcal{X}|h_{k-1}) \text{ and } Y(s_k) \in \operatorname{opt}(\mathcal{Y}(s_k)|h_{k-1}s_k) \text{ for all } s_k.$$

It can be shown that

The locality property is necessary and sufficient for normal form solutions to reduce to a sequence of single stage normal form solutions.

Our main result

Locality property

$$X + \bigoplus_{s_k} E_{s_k} Y(s_k) \in \text{opt}\left(\mathcal{X} + \bigoplus_{s_k} E_{s_k} \mathcal{Y}(s_k) \middle| h_{k-1}\right) \iff$$

$X \in \text{opt}(\mathcal{X}|h_{k-1})$ and $Y(s_k) \in \text{opt}(\mathcal{Y}(s_k)|h_{k-1}s_k)$ for all s_k.

It can be shown that

The locality property is necessary and sufficient for normal form solutions to reduce to a sequence of single stage normal form solutions.

Additional results

When considering:

- Lower previsions
- Rewards expressed in terms of utility

Maximality and E-admissibility

- Strictly positive lower probabilities
- Marginal extension

Additional results

When considering:

- Lower previsions
- Rewards expressed in terms of utility

Maximality and E-admissibility

- Strictly positive lower probabilities
- Marginal extension

The simple coin problem

An agent bets sequentially on a coin.

He recieves one utile on correct prediction and loses one otherwise.

The agent's objective is to perform optimally, thus maximizing the expected profit over the sequence of coin tosses.

The simple coin problem

An agent bets sequentially on a coin.

He recieves one utile on correct prediction and loses one otherwise.

The agent's objective is to perform optimally, thus maximizing the expected profit over the sequence of coin tosses.

The simple coin problem

An agent bets sequentially on a coin.

He recieves one utile on correct prediction and loses one otherwise.

The agent's objective is to perform optimally, thus maximizing the expected profit over the sequence of coin tosses.

The simple coin problem

Some extra assumptions:

- The bias of the coin is not known (Bayesian agent is no longer an obvious choice)
- The toss is not affected by the decision (act-state independence)

And to make things more interesting...

- Learning is considered (full state history is available)

The simple coin problem

Some extra assumptions:

- The bias of the coin is not known (Bayesian agent is no longer an obvious choice)
- The toss is not affected by the decision (act-state independence)

And to make things more interesting...

- Learning is considered (full state history is available)

The simple coin problem

Some extra assumptions:

- The bias of the coin is not known (Bayesian agent is no longer an obvious choice)
- The toss is not affected by the decision (act-state independence)

And to make things more interesting...

- Learning is considered (full state history is available)

The simple coin problem

Some extra assumptions:

- The bias of the coin is not known (Bayesian agent is no longer an obvious choice)
- The toss is not affected by the decision (act-state independence)

And to make things more interesting...

- Learning is considered (full state history is available)

Formulation

1. States (possible outcomes)
   - heads
   - tails
2. Decisions (possible bets)
   - heads
   - tails
3. Rewards (expressed in utility)
   - if bet $=$ outcome, receive 1
   - if bet $\neq$ outcome, pay 1
4. Transition probabilities (IDM)
   - Vacuous prior
   - Updated using observed transition

Formulation

1. States (possible outcomes)
   - heads
   - tails
2. Decisions (possible bets)
   - heads
   - tails
3. Rewards (expressed in utility)
   - if bet $=$ outcome, receive 1
   - if bet $\neq$ outcome, pay 1
4. Transition probabilities (IDM)
   - Vacuous prior
   - Updated using observed transition

Formulation

1. States (possible outcomes)
    - heads
    - tails
2. Decisions (possible bets)
    - heads
    - tails
3. Rewards (expressed in utility)
    - if bet $=$ outcome, receive 1
    - if bet $\neq$ outcome, pay 1
4. Transition probabilities (IDM)
    - Vacuous prior
    - Updated using observed transition

Formulation

1. States (possible outcomes)
   - heads
   - tails
2. Decisions (possible bets)
   - heads
   - tails
3. Rewards (expressed in utility)
   - if bet $=$ outcome, receive 1
   - if bet $\neq$ outcome, pay 1
4. Transition probabilities (IDM)
   - Vacuous prior
   - Updated using observed transition

Formulation

Notice that this example is similar to an MDPIP, because

- We assume that our uncertainty is expressed by a credal set
- Our rewards are expressed in terms of utilities

However, it is *not* an MDPIP because

- The Markov assumption does not hold (probabilities depend on the full state history)

Formulation

Notice that this example is similar to an MDPIP, because

- We assume that our uncertainty is expressed by a credal set
- Our rewards are expressed in terms of utilities

However, it is *not* an MDPIP because

- The Markov assumption does not hold (probabilities depend on the full state history)

The Imprecise Dirichlet Model

Predictive lower prevision

$$\underline{E}(X|h_k) = \sum_i \left(\frac{n_i}{N+s} X(i) \right) + \frac{s}{N+s} \inf_{t \in \Delta} \left(\sum_i t_i X(i) \right)$$

Optimality criteria:

- Γ-maximin ($\underline{E}(X) > \underline{E}(Y)$)
- Interval dominance ($\underline{E}(X) > \overline{E}(Y)$)
- Maximality ($\underline{E}(X - Y) > 0$)
- E-admissibility

The Imprecise Dirichlet Model

Predictive lower prevision

$$\underline{E}(X|h_k) = \sum_i \left(\frac{n_i}{N+s} X(i) \right) + \frac{s}{N+s} \inf_{t \in \Delta} \left(\sum_i t_i X(i) \right)$$

Optimality criteria:

- Γ-maximin ($\underline{E}(X) > \underline{E}(Y)$)
- Interval dominance ($\underline{E}(X) > \overline{E}(Y)$)
- Maximality ($\underline{E}(X - Y) > 0$)
- E-admissibility

Implementation

- Python
- Bayesian agent
  - Precise probabilities (distribution)
  - Uniform prior
- "IDM" agent
  - Vacuous prior
  - Maximality
  - E-admissibility

Implementation

- Python
- Bayesian agent
  - Precise probabilities (distribution)
  - Uniform prior
- "IDM" agent
  - Vacuous prior
  - Maximality
  - E-admissibility

Implementation

- Python
- Bayesian agent
  - Precise probabilities (distribution)
  - Uniform prior
- “IDM” agent
  - Vacuous prior
  - Maximality
  - E-admissibility

Implementation

- Python
- Bayesian agent
  - Precise probabilities (distribution)
  - Uniform prior
- “IDM” agent
  - Vacuous prior
  - Maximality
  - E-admissibility

Implementation

We considered

- 10 coin tosses
- 10,000 experiments
- Average gain

In particular

- Bayesian = Γ-maximin
- Interval dominance = maximality = E-admissibility

Implementation

We considered

- 10 coin tosses
- 10,000 experiments
- Average gain

In particular

- Bayesian = Γ-maximin
- Interval dominance = maximality = E-admissibility

Implementation

We considered

- 10 coin tosses
- 10,000 experiments
- Average gain

In particular

- Bayesian = Γ-maximin
- Interval dominance = maximality = E-admissibility

Implementation

We considered

- 10 coin tosses
- 10,000 experiments
- Average gain

In particular

- Bayesian = Γ-maximin
- Interval dominance = maximality = E-admissibility

Implementation

We considered

- 10 coin tosses
- 10,000 experiments
- Average gain

In particular

- Bayesian = Γ-maximin
- Interval dominance = maximality = E-admissibility

Implementation

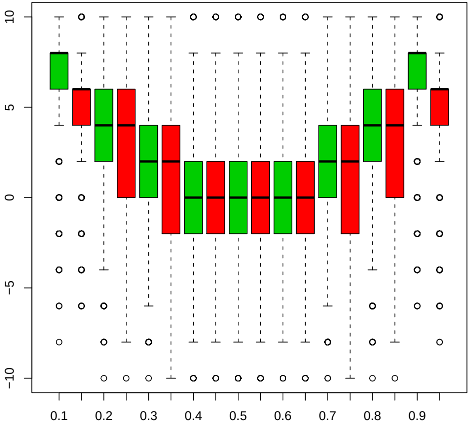
We considered

- 10 coin tosses
- 10,000 experiments
- Average gain

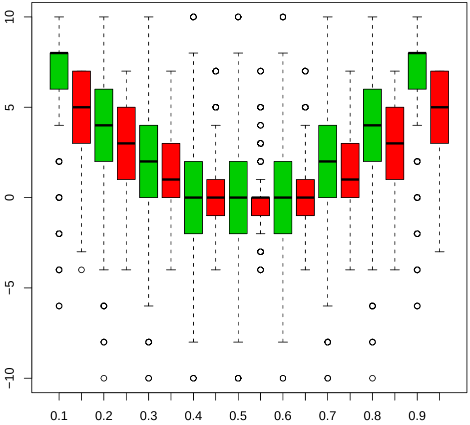
In particular

- Bayesian = Γ-maximin
- Interval dominance = maximality = E-admissibility

Bayesian vs Maximal agent



Bayesian vs Maximal agent with no-bet option



Conclusions

1. For best performance, forget about imprecise probabilities and simply be a happy Bayesian
2. For robustness, avoid Γ-maximin (or Bayesian agent) and instead adopt maximality or E-admissibility

Conclusions

1. For best performance, forget about imprecise probabilities and simply be a happy Bayesian
2. For robustness, avoid Γ-maximin (or Bayesian agent) and instead adopt maximality or E-admissibility

Future steps

In a near future:

1. Γ-maximin and interval dominance?
2. More complex problems (e.g. Peter Walley's bag of marbles)...
3. Real applications?

Ultimate goal

1. Act-state dependence

Questions? Comments?

Thank you for your attention!

Future steps

In a near future:

1. Γ-maximin and interval dominance?
2. More complex problems (e.g. Peter Walley's bag of marbles)...
3. Real applications?

Ultimate goal

1. Act-state dependence

Questions? Comments?

Thank you for your attention!

Future steps

In a near future:

1. Γ-maximin and interval dominance?
2. More complex problems (e.g. Peter Walley's bag of marbles)...
3. Real applications?

Ultimate goal

1. Act-state dependence

Questions? Comments?

Thank you for your attention!

Future steps

In a near future:

1. Γ-maximin and interval dominance?
2. More complex problems (e.g. Peter Walley's bag of marbles)...
3. Real applications?

Ultimate goal

1. Act-state dependence

Questions? Comments?

Thank you for your attention!

Future steps

In a near future:

1. Γ-maximin and interval dominance?
2. More complex problems (e.g. Peter Walley's bag of marbles)...
3. Real applications?

Ultimate goal

1. Act-state dependence

Questions? Comments?

Thank you for your attention!

Future steps

In a near future:

1. Γ-maximin and interval dominance?
2. More complex problems (e.g. Peter Walley's bag of marbles)...
3. Real applications?

Ultimate goal

1. Act-state dependence

Questions? Comments?

Thank you for your attention!

Future steps

In a near future:

1. Γ-maximin and interval dominance?
2. More complex problems (e.g. Peter Walley's bag of marbles)...
3. Real applications?

Ultimate goal

1. Act-state dependence

Questions? Comments?

Thank you for your attention!

Future steps

In a near future:

1. Γ-maximin and interval dominance?
2. More complex problems (e.g. Peter Walley's bag of marbles)...
3. Real applications?

Ultimate goal

1. Act-state dependence

Questions? Comments?

Thank you for your attention!