Plan for Today

Nonparametric RL

- "Nonparametric" function approximation
- Strong guarantees across:
  Sample complexity, *space complexity*, *storage complexity*

Hindsight Learning

- Exogenous MDPs as model for OR problems
- Use of *Hindsight Planning* oracle for algorithm design
- Empirical results in VM allocation with Microsoft Azure

Tree-Partitions

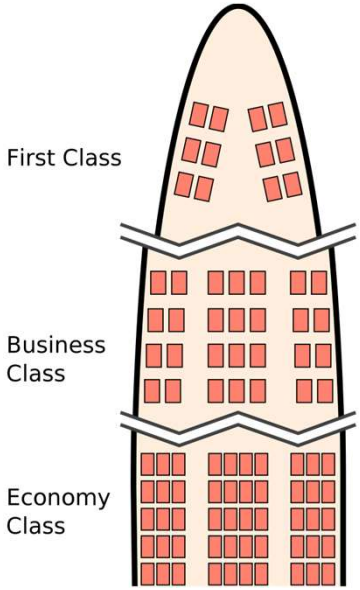
- Implement tree-based adaptive discretization from nonparametric RL algorithms
- Use ORSuite to test on "continuous Ambulance routing"

Hindsight Planning for Exo-MDPs

- Use ORSuite model for revenue management and pricing (an example of an Exo-MDP)
- Implement Bayes Selector
- Use ORSuite to run simulations to compare performance against tabular algorithms

Revenue Management Problem

Start off with fixed capacity of different item types



10 first class seats

20 business class seats

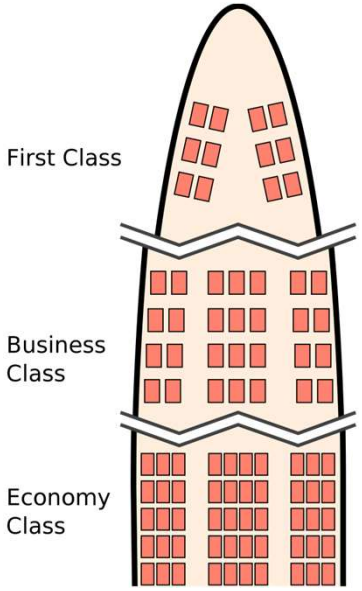
100 economy class seats

$$B = (10, 20, 100)$$

[Vera2018,2019]

Revenue Management Problem

Start off with fixed capacity of different item types



$$B = (10, 20, 100)$$

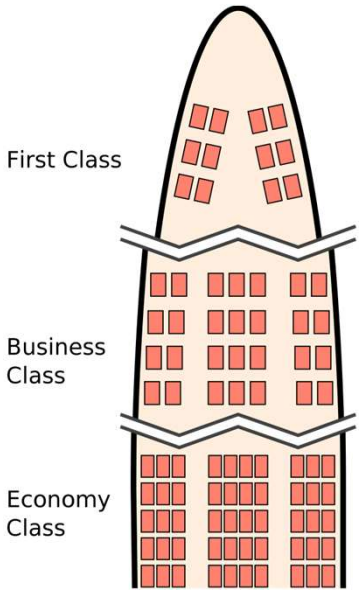
Finite set of customer types requesting part of resources with certain income:

- 2 business class seats, cost = $1000
- 1 business class seat, cost = $450
- 3 economy seats, cost = $120
- …

[Vera2018,2019]

Revenue Management Problem

Start off with fixed capacity of different item types



$$B = (10, 20, 100)$$

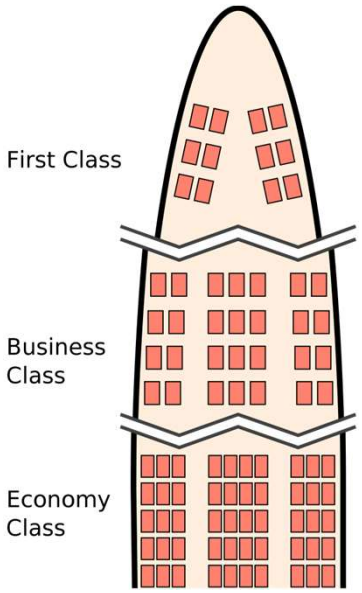
Action: Decide accept / reject for each customer type

$$\mathcal{A} = \{0, 1\}^n$$

[Vera2018,2019]

Revenue Management Problem

Start off with fixed capacity of different item types



$$B = (10, 20, 100)$$

Action: Decide accept / reject for each customer type

$$\mathcal{A} = \{0, 1\}^n$$

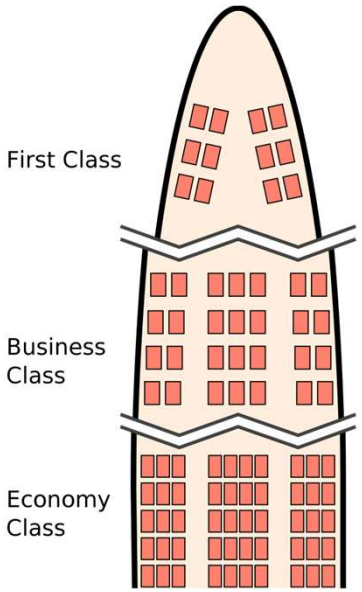
Customer arrives of specific type

2 business class seats, cost = $1000

[Vera2018,2019]

Revenue Management Problem

Start off with fixed capacity of different item types



$$B = (10, 20, 100)$$

Action: Decide accept / reject for each customer type

$$\mathcal{A} = \{0, 1\}^n$$

Customer arrives of specific type

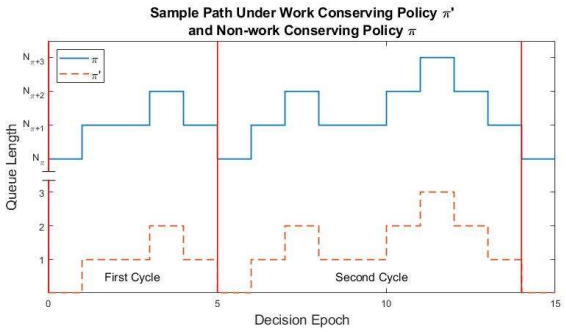
2 business class seats, cost = $1000

Request accepted if dictated by action

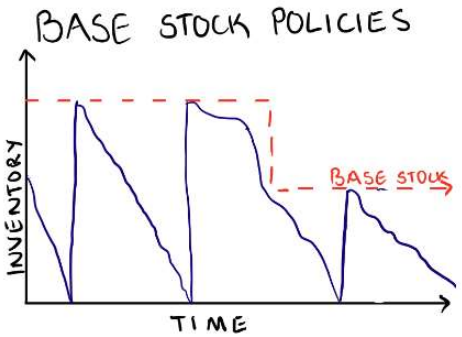
[Vera2018,2019]

Exogeneity

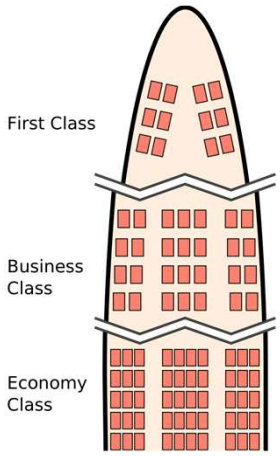
Exogenous demand governs state transition and rewards



Stochastic Networks
(patient arrivals)



Inventory Control
(demand)



Revenue Management
(fare class)

Bayes Selector

Given a fixed exogenous trace, the optimal policy can be solved via a (combinatorial) optimization problem

$$\textsc{Hindsight}(t, \boldsymbol{\xi}_{\geq t}, s) = \max_{a_t,\ldots,a_T} \sum_{\tau=t}^{T} r(s_\tau, a_\tau, \xi_\tau)$$
$$\text{s.t. } x_{\tau+1} = f(s_\tau, a_\tau, \xi_\tau), \text{ for } \tau = t, \ldots, T$$
$$s_\tau = (x_\tau, \boldsymbol{\xi}_{<\tau}), \text{ for } \tau = t, \ldots, T.$$

Requires frequent ***online*** resolves of an IP

Can develop an online policy:

- In current state, solve optimization problem replacing unknown trace with historical traces
- Execute policy by averaging over decisions aggregated for current exogenous state

[Vera2018,2019]

Bayes Selector

Solving for optimal non-anticipatory policy is hard, focus on a surrogate

$$\pi_t^\dagger(s) = \operatorname*{argmax}_{a \in \mathcal{A}} Q_t^\dagger(s, a)$$
$$Q_t^\dagger(s, a) = \mathbb{E}_{\boldsymbol{\xi}_{\geq t}}[r(s, a, \xi_t) + \textsc{Hindsight}(t+1, \boldsymbol{\xi}_{>t}, f(s, a, \xi_t))]$$
$$V_t^\dagger(s) = \mathbb{E}_{\boldsymbol{\xi}_{\geq t}}[\textsc{Hindsight}(t, \boldsymbol{\xi}_{\geq t}, s)]$$

Pick actions "optimal on average" over exogenous traces

Plan for Today

Revenue Management

- Understand Revenue Management problem as part of the ORSuite package
- Run preliminary experiments against PPO + Tabular algorithms (noticing issue of scale)

Bayes Selector

- Formulate hindsight planner using PULP – an optimization package in python
- Use hindsight planner to implement Bayes Selector algorithm

References

https://github.com/seanrsinclair/RLinOperations