Scenario-driven RM using simulation-based reinforcement learning

AGIFORS Revenue Management SG Meeting
04-06 May 2021

Jonas Rauch
Lead Scientist

Copyright © 2021 by PROS, Inc. Published by AGIFORS with permission.

Bio: Jonas Rauch

- Lead Scientist @PROS based in Mannheim, Germany, working on pricing and RM methods for travel and other industries
- Previously 11 years @Lufthansa working on RM methods, strategy, business processes for LH Group
- PhD in Mathematics from University of Magdeburg, Germany

Motivation

Classic RM cannot deal with epistemic uncertainty

Aleatoric uncertainty: *"[…] unknowns that differ each time we run the same experiment."*

– Wikipedia

Example:

- Mean demand $\mu = 100$
- Demand realization $X \sim \mathrm{Pois}(\mu)$
- **Dynamic program takes this into account**

Epistemic uncertainty: *"[…] things we could in principle know but do not in practice."*

– Wikipedia

Example:

- Mean demand $\mu \sim \mathcal{N}(100, 20^2)$
- Demand realization $X \sim \mathrm{Pois}(\mu)$
- **Standard RM methods cannot deal with this**

©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Example scenario

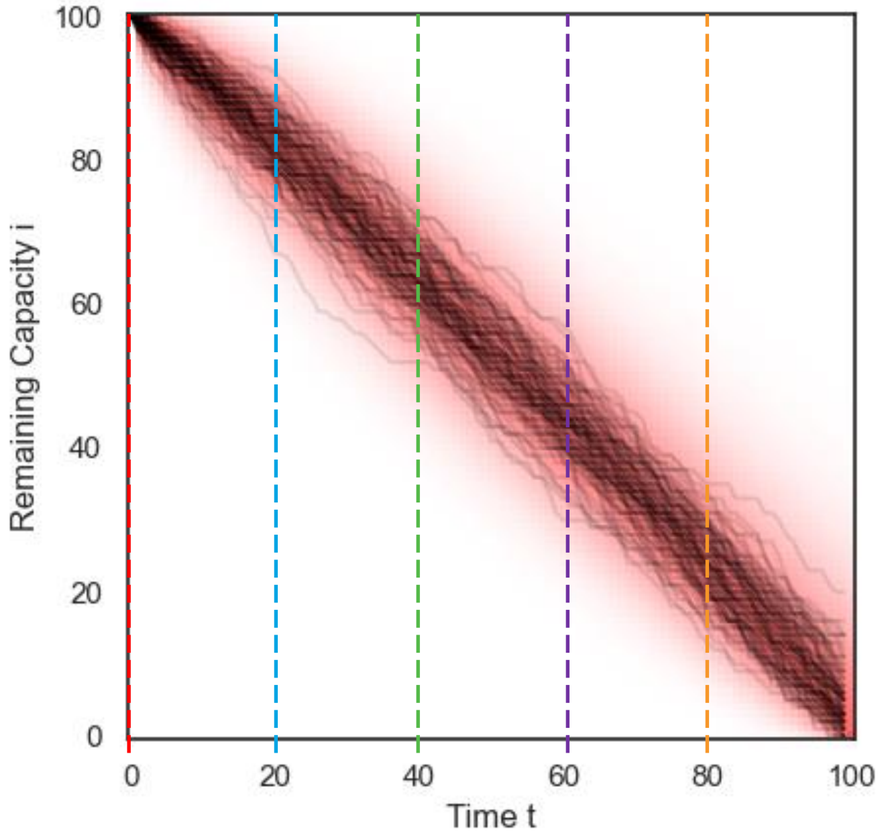
Example scenario

- Time horizon: $t \in [0, 100]$
- Capacity: 100
- Expected demand (Poisson): 120 (at revenue-optimal prices)
- Exponential WTP with two customer segments
- Dynamic pricing

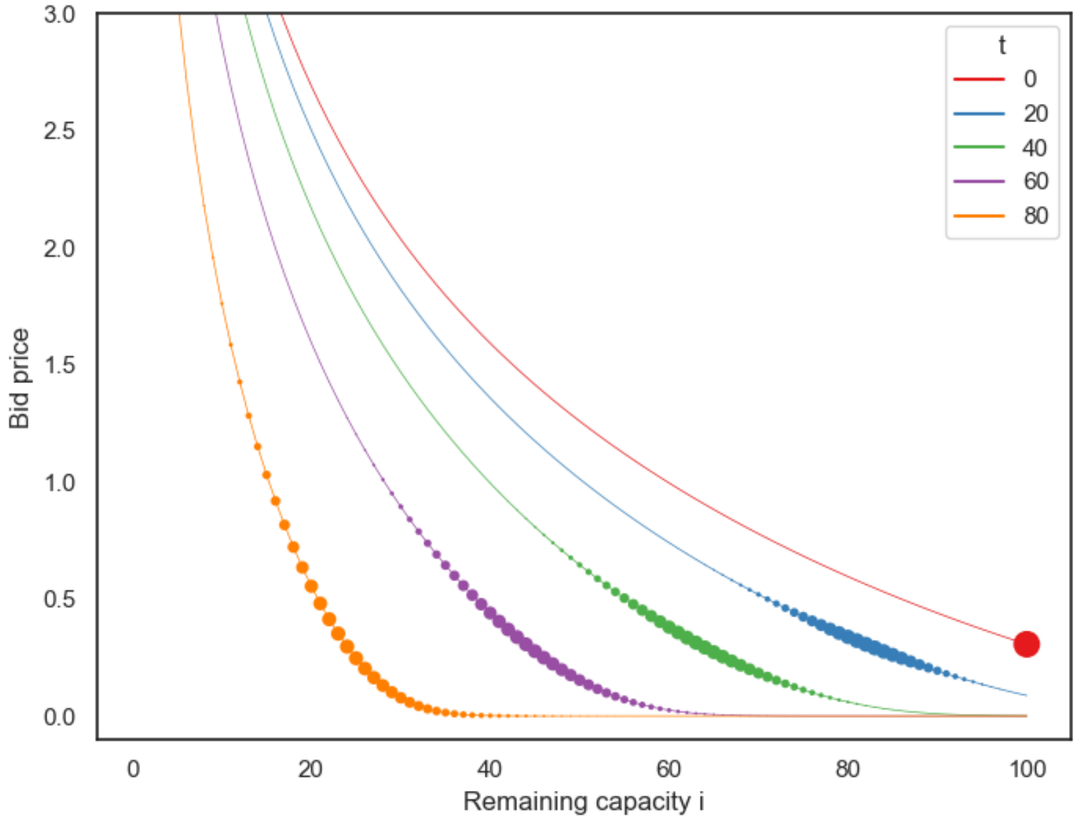
©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Sample trajectories and bid price distribution

Sample paths



Bid price vectors at different times

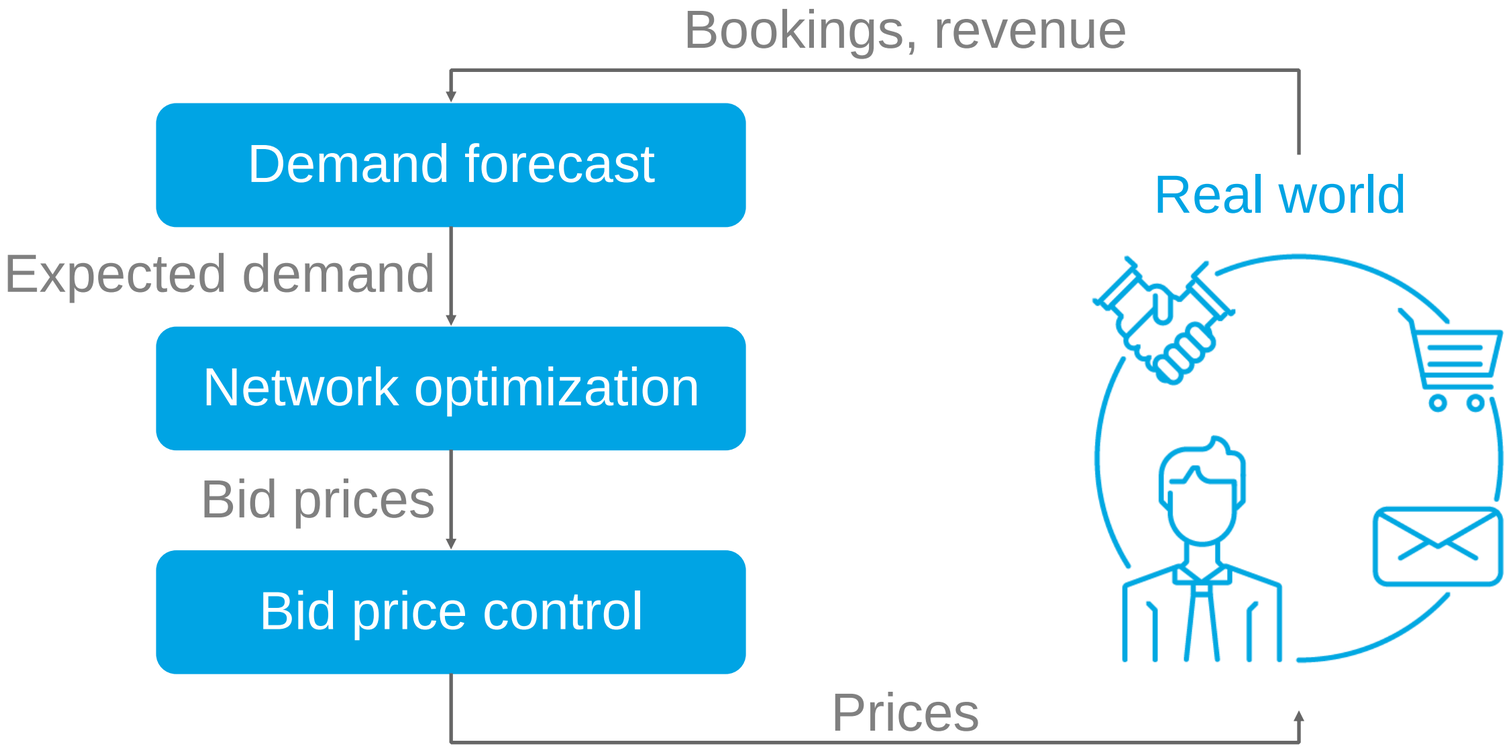


©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Simulation-based reinforcement learning

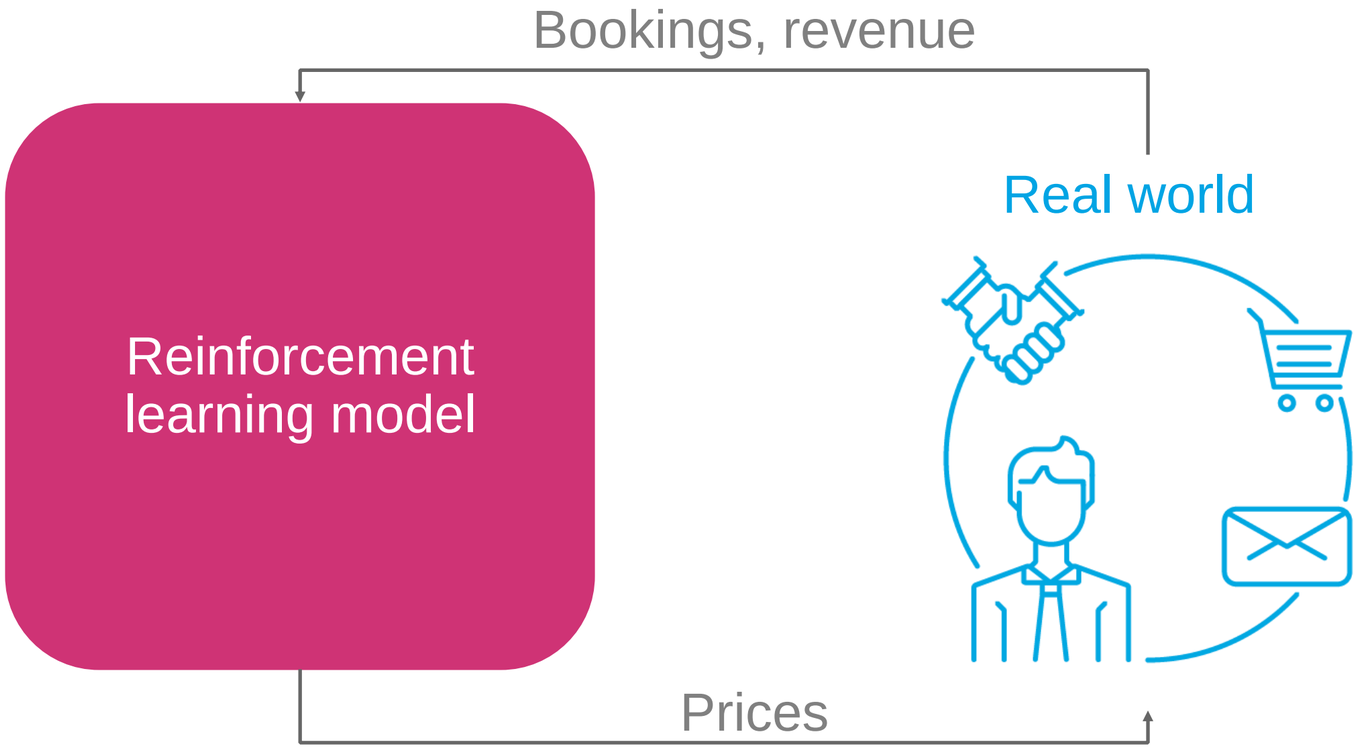
PROS

Standard RM process



©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

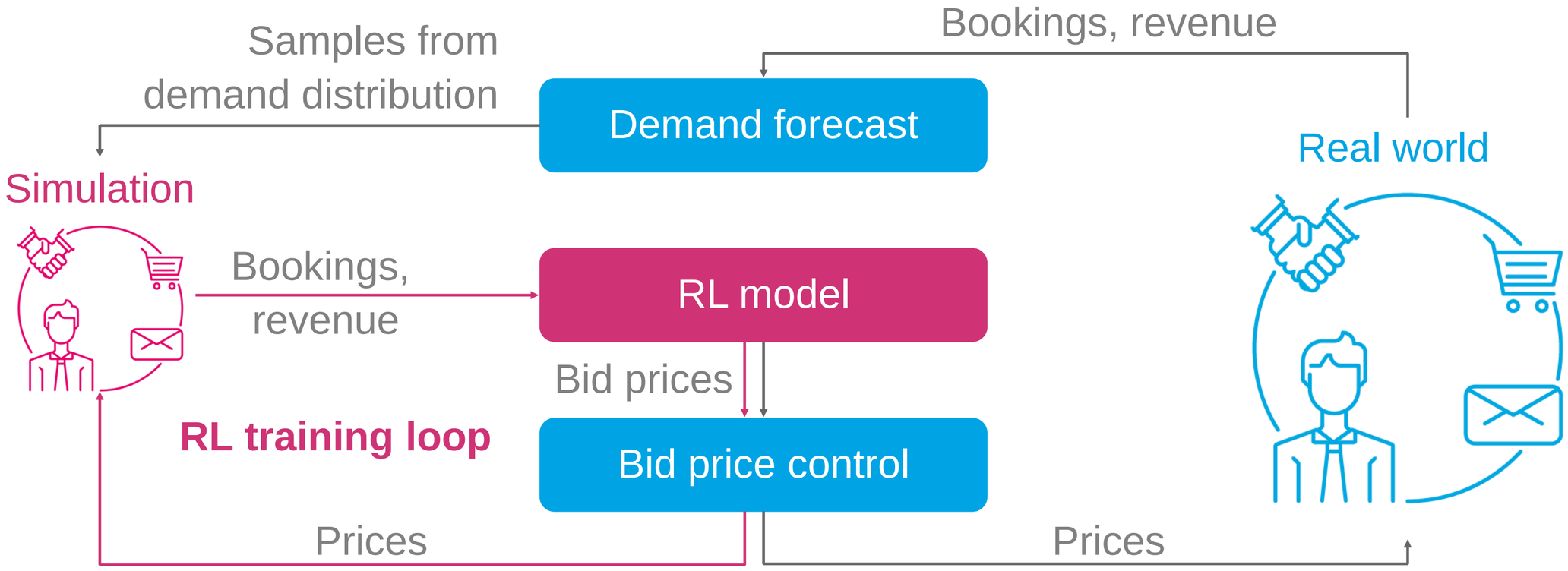
Full Reinforcement Learning



See for example: Acuna-Agost & Fiig, AGIFORS RM 2017

©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

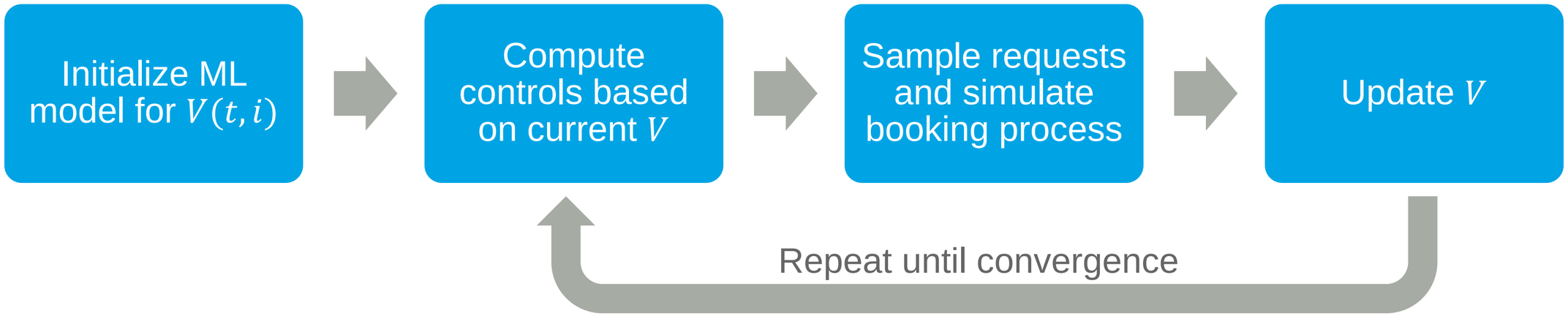
Simulation-based RL (this work)



©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Value function approximation with simulation-based RL

RL training loop



Options for updating V:

a) Monte-Carlo (MC) learning

$V(t,i) \leftarrow$ observed revenue-to-come from state (t,i) until end of run

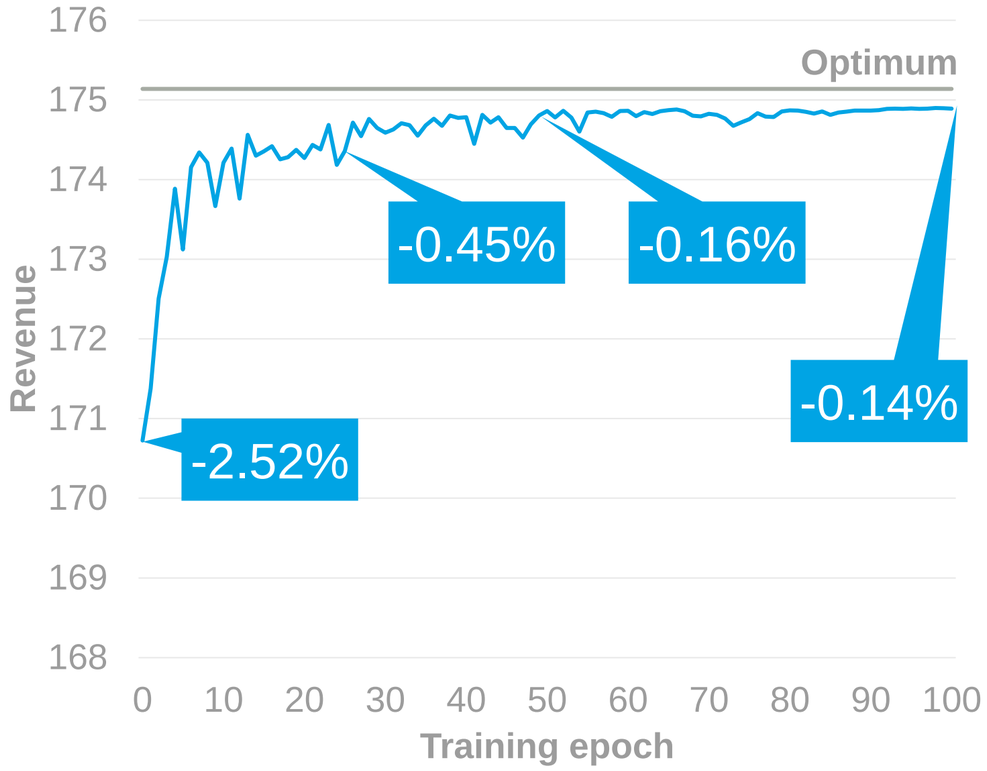
b) Temporal difference (TD) learning

$V(t,i) \leftarrow V(t+1,i') +$ observed revenue during state transition (t,i) to $(t+1,i')$

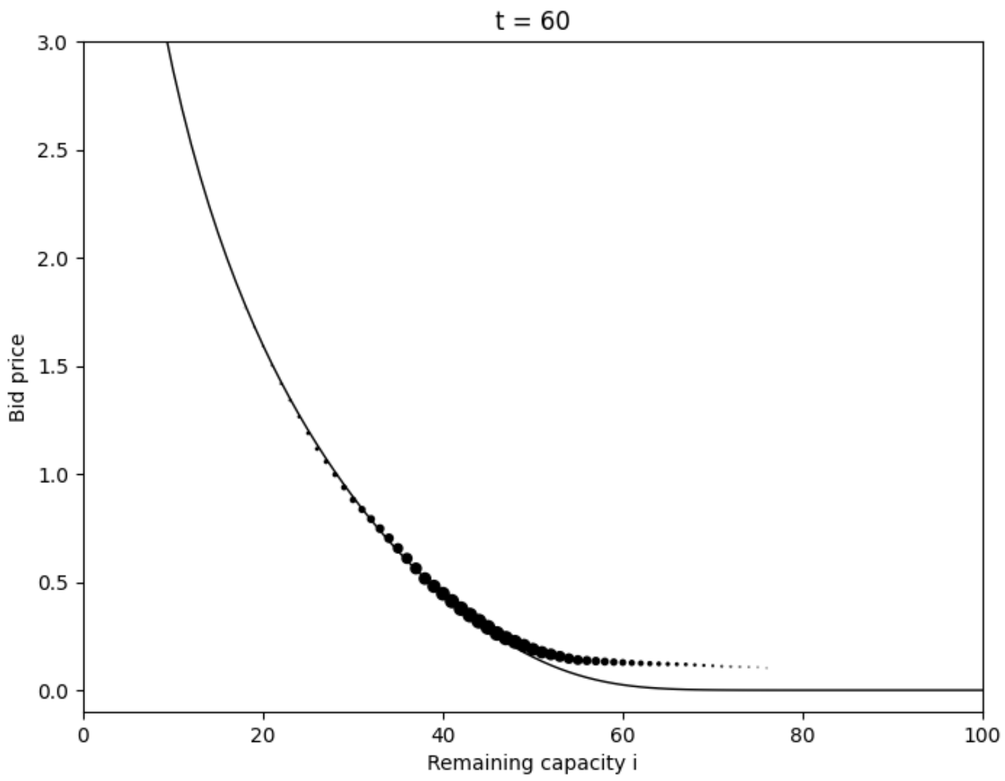
©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Monte-Carlo learning converges to a near-optimum

Revenue performance



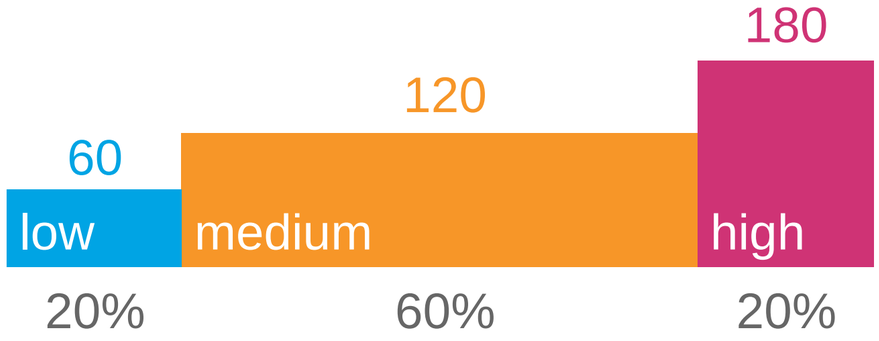
Bid price vectors ($t = 60$)

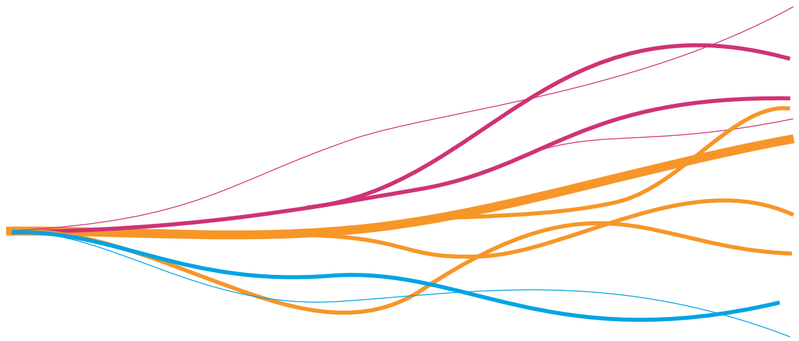


©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Scenario-based RM

Scenario-based RM (discrete case)





- Three scenarios with different demand levels
- Each occurs with given probabilities p_s
- We don't know which scenario we are in
- Price can only depend on state (t, i)
- Problem not solvable via DP
- Simple heuristics:
  - BP from medium demand scenario
  - Weighted average of BP with constant weights

©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Remaining capacity contains information about demand scenario

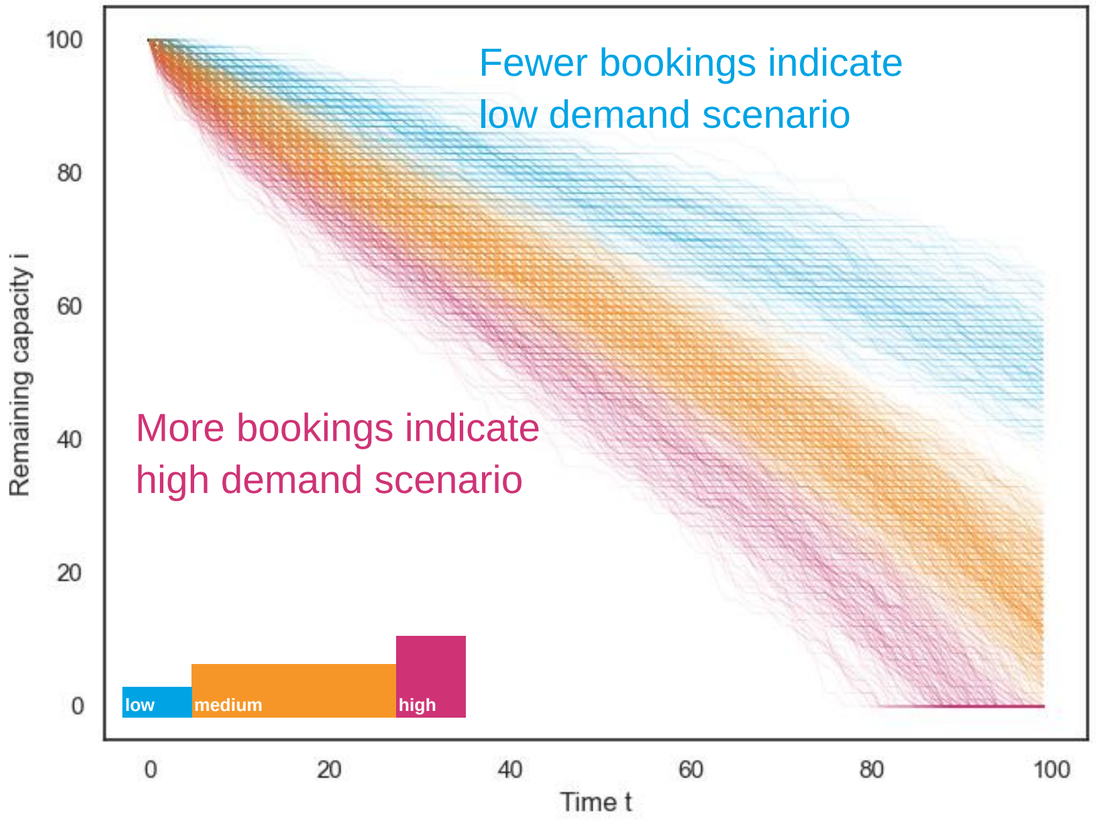
- Different state distribution between scenarios
- Prior probability p_s for scenario s
- Bayes' rule for posterior:

$$p(s \mid t, i) = \frac{p(\, i \mid t, s) \cdot p_s}{p(i \mid t\,)}$$

- Idea: Weighted average of BP

$$\pi(t, i) = \sum_s p(s \mid t, i) \cdot \pi_s(t, i)$$

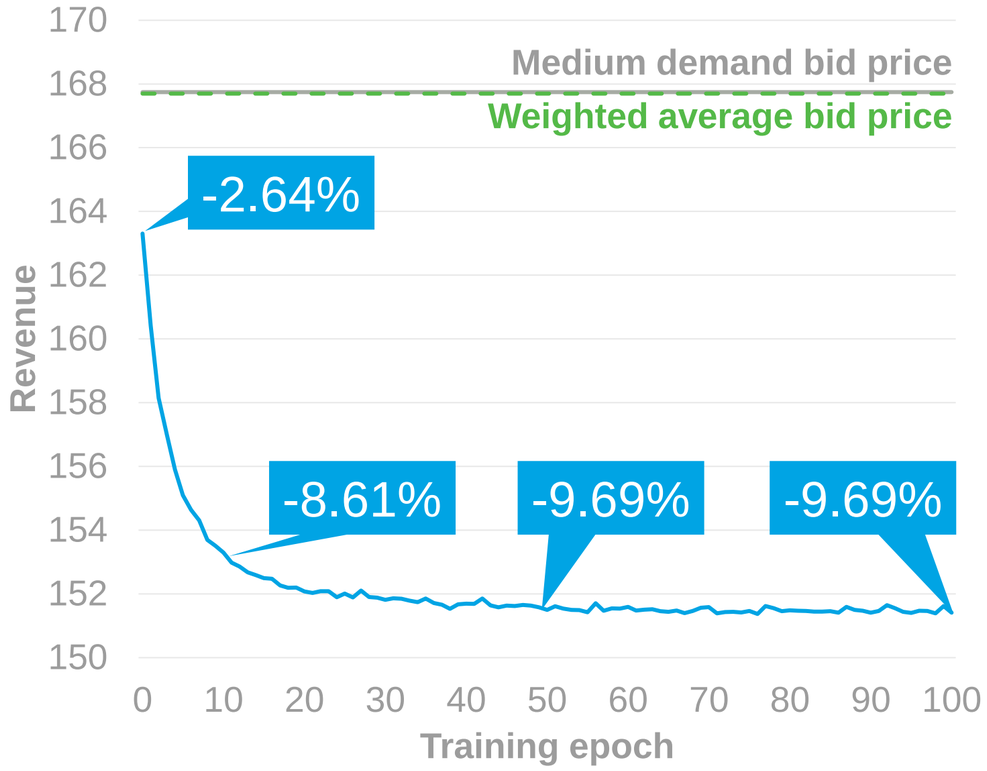
- **RL model trained on sample paths is implicitly weighted correctly!**



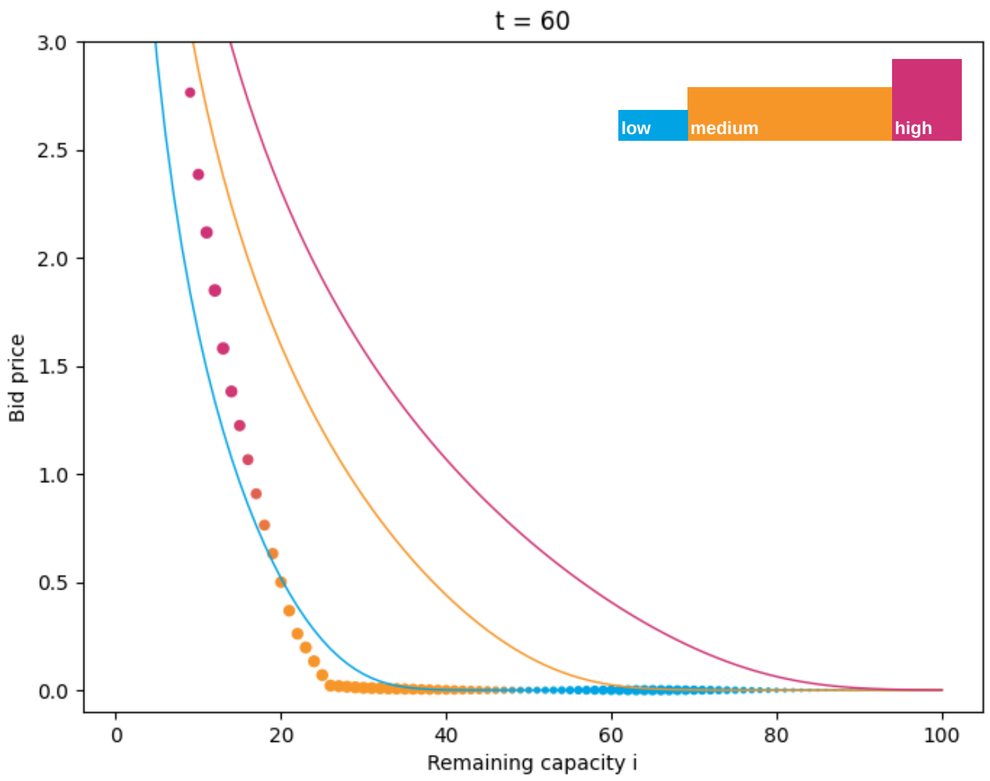
©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Monte-Carlo learning converges far from optimum

Revenue performance



Bid price vectors ($t = 60$)



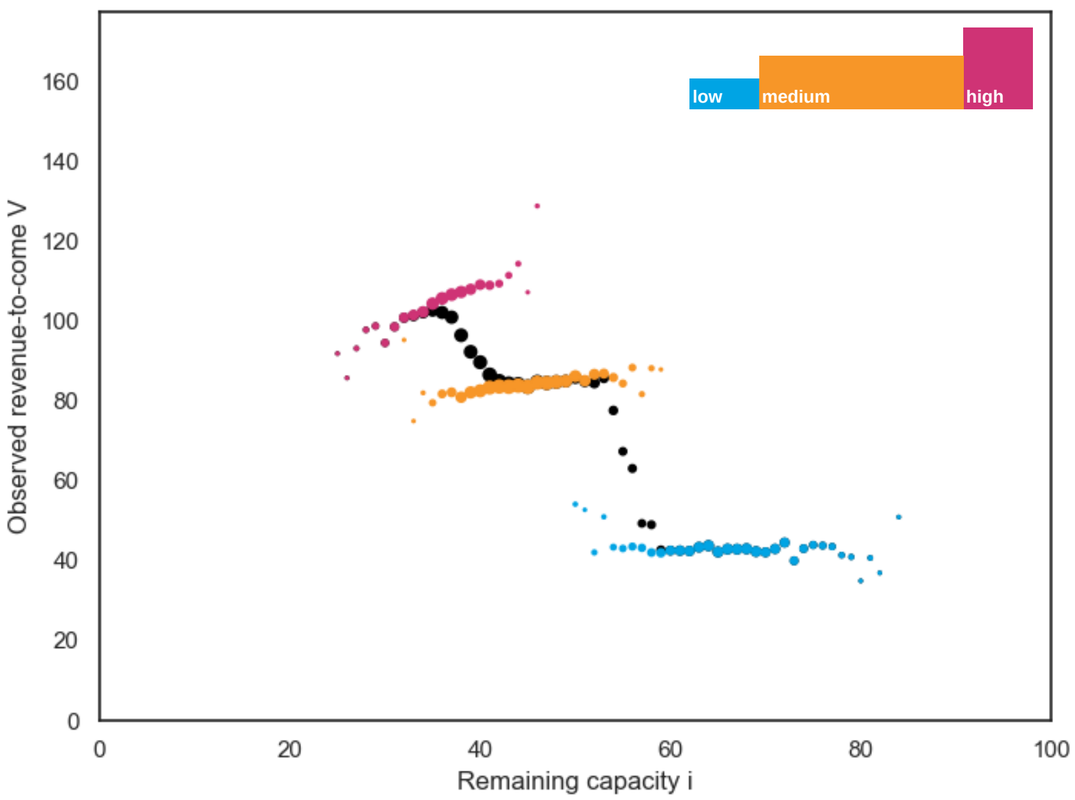
©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Simple value function approximation doesn't work

- Revenue-to-come decreases with higher remaining capacity
  - Naïve bid price is negative
  - Model restrictions push bp to zero

- Per scenario behavior is as expected
  - ➢ Compute bid price first, then average
  - ➢ Can we update the bid price instead of the value function?

Observed revenue-to-come $V\ (t = 60)$



©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

TD learning update of bid price

- Bellman equation for value function

$$\dot{V}(t,i) = -\max_f d(t,f)\big(f - \pi(t,i)\big) := -h(t,\pi(t,i))$$

- Differential equation for bid price and linearize

$$\boxed{\dot{\pi}(t,i)} = h\big(t,\pi(t,i-1)\big) - h\big(t,\pi(t,i)\big)$$
$$\approx \frac{\partial h}{\partial \pi}\big(t,\pi(t,i)\big)[\pi(t,i-1) - \pi(t,i)]$$
$$= -\boxed{d(t,f^*(t,i))} \cdot \boxed{[\pi(t,i-1) - \pi(t,i)]}$$

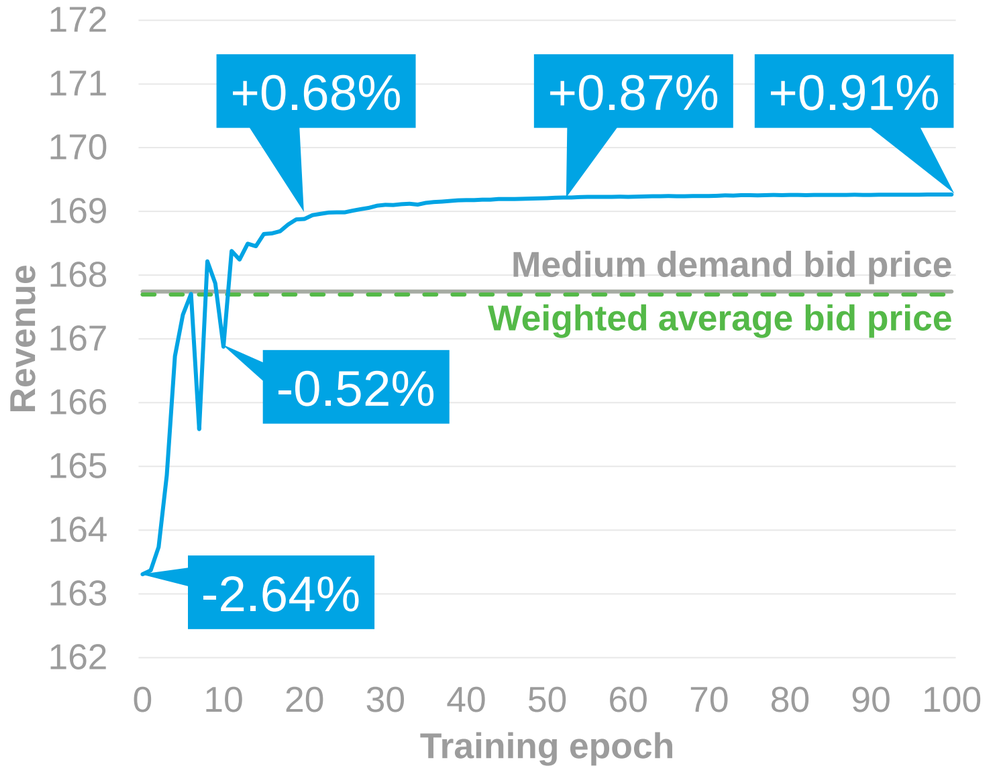
- TD learning update for bid prices:

$$\boxed{\pi(t,i) \leftarrow \pi(t+1,i)} + \boxed{[\text{bookings in state } (t,i)]} \cdot \boxed{[\pi(t,i-1) - \pi(t,i)]}$$

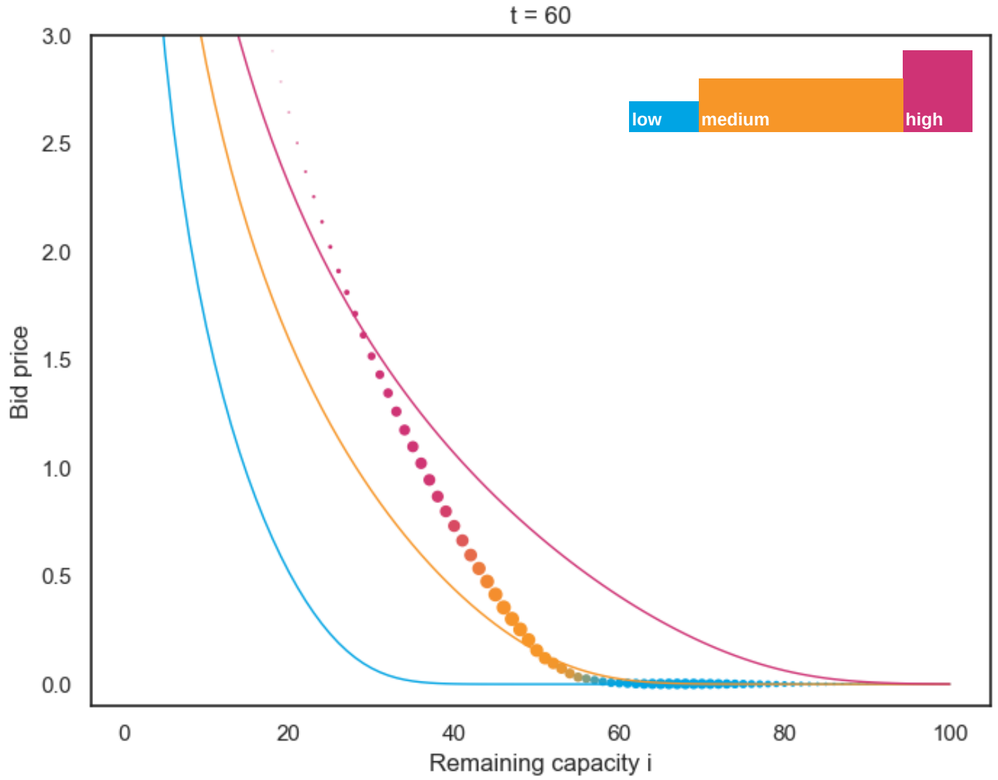
©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Bid price TD learning achieves ~1% revenue gain

Revenue performance



Bid price vectors ($t = 60$)



©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Summary

Summary

- Scenario-based RM is one way to address forecast uncertainty
- The corresponding control problem is not solvable via a DP
- Simulation-based RL works without Poisson assumption
  - Naïve value learning does not work
  - TD learning update for the bid price leads to promising first results
- Potential next step: Network case

©2021 PROS, Inc. All rights reserved. Confidential and Proprietary

Thank You

PROS Headquarters | Houston, Texas | 3100 Main Street, Suite 900 | Houston, TX 77002, USA | +1-713-335-5105 +1-800-555-3548