MARLIM: Multi-Agent Reinforcement Learning for Inventory Management

Rémi Leluc[1], Elie Kadoche[2], Antoine Bertoncello[2], Sébastien Gourvénec[2]




[1] LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[2] TotalEnergies OneTech, France

NeurIPS Workshop on Reinforcement Learning for Real Life (2022)

Inventory Management: Goal and Contributions

GOAL: Find the right balance between the **supply** and **demand** of products by **optimizing replenishment decisions** and **minimizing costs**.

Benefits: better inventory accuracy, insights to cost savings, avoidance of stock-outs

Main issue: Environment uncertainty , demands and lead-times are stochastic with potentially high volatility, controller may exceedingly or insufficiently order, leading to unnecessary *ordering*, *holding* and *shortage* costs.

Contributions:

(1) We develop a novel reinforcement learning framework, called **MARLIM**, to address the inventory management problem for a single-echelon multi-products supply chain on a production line with stochastic demands and lead-times.

(2) We provide the methodology to train agents in different scenarios for fixed or shared capacity constraints with specific handling of storage overflows.

(3) We perform various numerical experiments on real-world data to demonstrate the benefits of our method over classical baselines.

Inventory Management: Goal and Contributions

GOAL: Find the right balance between the **supply** and **demand** of products by **optimizing replenishment decisions** and **minimizing costs**.

Benefits: better inventory accuracy, insights to cost savings, avoidance of stock-outs

Main issue: Environment uncertainty , demands and lead-times are stochastic with potentially high volatility, controller may exceedingly or insufficiently order, leading to unnecessary *ordering*, *holding* and *shortage* costs.

<u>Contributions:</u>

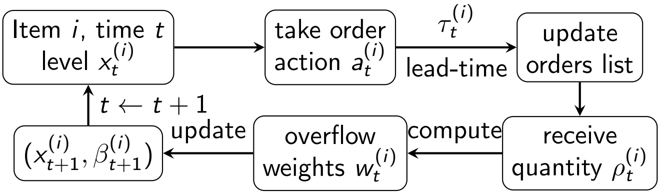
(1) We develop a novel reinforcement learning framework, called **MARLIM**, to address the inventory management problem for a single-echelon multi-products supply chain on a production line with stochastic demands and lead-times.

(2) We provide the methodology to train agents in different scenarios for fixed or shared capacity constraints with specific handling of storage overflows.

(3) We perform various numerical experiments on real-world data to demonstrate the benefits of our method over classical baselines.

Inventory Dynamics

At time t, for product $i = 1, \ldots, n$, inventory controller decides about the order $a_t^{(i)}$ to take based on the current inventory level $x_t^{(i)}$ and the stochastic demand $\delta_t^{(i)}$. This order arrives after a stochastic lead-time $\tau_t^{(i)}$.



After receiving the replenishment quantity $\rho_t^{(i)}$ the inventory levels are temporarily updated through $\lambda_t^{(i)} = x_t^{(i)} + \lfloor w_t^{(i)} \rho_t^{(i)} \rfloor$. The levels and backlogs β_t are updated to evaluate the inventory costs $C_t^{(i)}$.

$$x_{t+1}^{(i)} = \left(\lambda_t^{(i)} - \delta_t^{(i)}\right)_+ \qquad \beta_{t+1}^{(i)} = \beta_t^{(i)} + \left(\lambda_t^{(i)} - \delta_t^{(i)}\right)_-$$

$$C_t^{(i)} = \alpha_o \underbrace{a_t^{(i)} C_o^{(i)}}_{\text{ordering}} + \alpha_h \underbrace{x_t^{(i)} C_h^{(i)}}_{\text{holding}} + \alpha_s \underbrace{\beta_{t+1}^{(i)} C_s^{(i)}}_{\text{shortage}} \quad (\alpha_o + \alpha_h + \alpha_s = 1)$$

Numerical Results

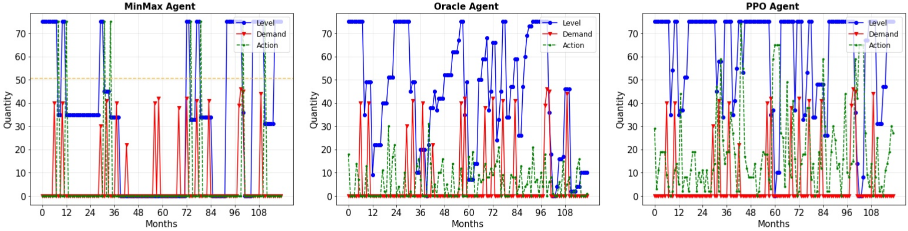


Figure: Inventory level (*blue*) of an item over $T = 120$ months with different agents: MinMax (left), Oracle (center), PPO (right). The demand is plotted in *red* and the order actions are plotted in *green*. The safety stock MinMax agent is displayed in *orange*.

- PPO Agent avoids stock-outs of items !
- Whenever demand spike or a demand plateau for a long time, there is an incentive to order.

Conclusion and take-home message

• Rigorous methodological and practical RL framework to address the inventory management problem for a single-echelon multi-products supply chain on a production line with stochastic demands and lead-times.

• We illustrated our method with extensive numerical experiments on real data for both single and multi-agent algorithms.

• Future work will focus on the statistical properties of the developed framework by further exploring links with mean-field approximation theory and the effects of exogenous variables in reinforcement learning.

Thank you and see you at the conference !