Results

- We have implemented several algorithms in this class. There are two variants for expected value, and four variants based on risk measures.

- Risk sensitive measures:
  - Instead of expected performance on the distribution of tasks, we can use percentile-based risk measures (VaR and CVaR) to specify performance on Markov decision processes (MDPs) drawn from this distribution. Intuition: safety constraints on the "worst case" MDPs in the distribution.
  - Another class of interest is episodic bounds: instead of guaranteeing average performance, we can apply percentile-based risk measures (VaR and CVaR) to the distribution of episodes. Intuition: safety constraints on the "worst case" episodes.

Approach

- The user 1) defines a minimum safe measure of performance (in terms of expected return) for the RL task and 2) specifies the probability with which the algorithm must achieve this measure of performance on Markov decision processes (MDPs) drawn from this distribution. The algorithm trains on MDPs sampled from a distribution of tasks, and returns either: 1) a solution that is guaranteed to satisfy the safety constraint with the specified probability, or 2) if a safe solution cannot be found with the specified probability, the algorithm returns "No Solution Found". By modifying the types of bounds we use, we can guarantee performance even when the target task does not come from the training distribution (assuming other assumptions are met, and that sufficient data exists).

Objective or Return

- Instead of expected performance on the distribution of tasks, we can use percentile-based risk measures (VaR and CVaR) to specify performance on MDPs drawn from the distribution. Intuition: safety constraints on the "worst case" MDPs in the distribution.
  \[
  \Pr(\text{VaR}_\alpha(J_{M_1}(a(M_{acc})) | M_1 \sim \mu, M_{acc} \sim \mu) \geq j) \geq 1 - \delta, \\
  \Pr(\text{CVaR}_\alpha(J_{M_1}(a(M_{acc})) | M_1 \sim \mu, M_{acc} \sim \mu) \geq j) \geq 1 - \delta,
  \]
  - Another class of interest is episodic bounds: instead of guaranteeing average performance, we can apply percentile-based risk measures (VaR and CVaR) to the distribution of episodes. Intuition: safety constraints on the "worst case" episodes.
  \[
  \Pr(\text{VaR}_\alpha(G_{M_1}(a(M_{acc})) | M_1 \sim \mu, M_{acc} \sim \mu) \geq j) \geq 1 - \delta, \\
  \Pr(\text{CVaR}_\alpha(G_{M_1}(a(M_{acc})) | M_1 \sim \mu, M_{acc} \sim \mu) \geq j) \geq 1 - \delta,
  \]

Conclusions

- These algorithms are adaptive, safe, and they work: we’ve laid the theoretical foundations, and initial experiments show that the safety guarantees hold in practice.

Path Forward

- Continue to develop related classes of safety guarantees.
- Gather more empirical results on different kinds of tasks.
- Work on the optimal stopping problem represented by how to split the training and safety data sets.
- Extend the algorithms, in theory and in practice, to function in the extrapolation setting.