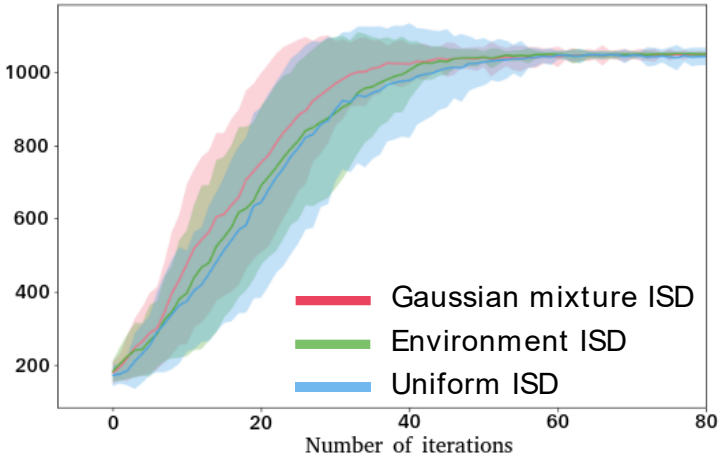
Empirical Analysis of Policy Gradient Algorithms where Starting States are Sampled accordingly to Most Frequently Visited States

Samy Aittahar[1], Raphaël Fonteneau[1], Damien Ernst[1]
[1] Department of Electrical Engineering and Computer Science, University of Liège

- Vanilla policy gradient algorithms update their policies using gradient estimates from sequences of transitions (episodes).
- Subject to high variance, especially when the number of episodes is low. Hence our solution:
- Replace the initial state distribution by a Gaussian mixture that approximates policy state visitation frequency.
- Results show that our approach works better than using an uniform or the environment ISD.



Evolution of the mean cumulative reward over iterations.