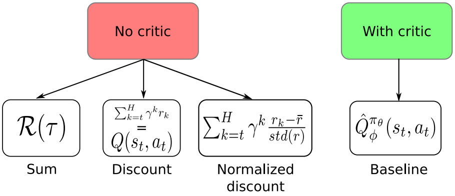
Policy Gradient in practice

Don't become an alchemist :)

Olivier Sigaud

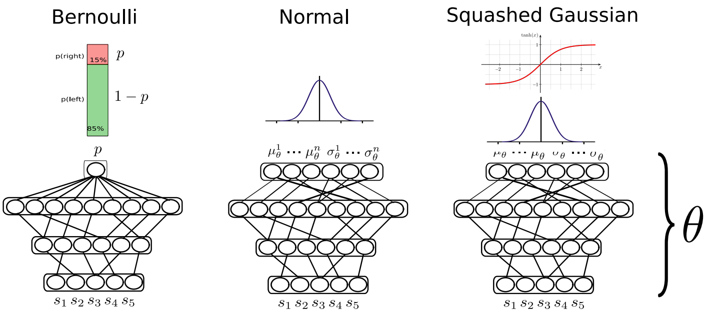
Sorbonne Université
http://people.isir.upmc.fr/sigaud

Outline



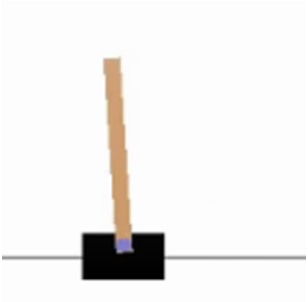
- Investigation of basic REINFORCE phenomena and issues
- Using:
  - gym "classic control": CartPole, Continuous MountainCar, Pendulum
  - Bernoulli, Normal and squashed Gaussian policies
- Visualization of policies, critics, learning curves
- A prerequisite before going to SOTA deep RL algorithms and harder benchmarks
- Understanding phenomena is better than using black-box algorithms
- Github repo: `https://github.com/osigaud/Basic-Policy-Gradient-Labs`

Stochastic policies



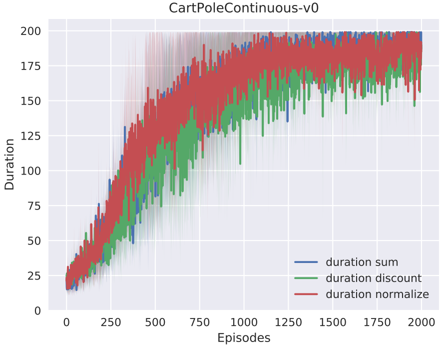
- Bernoulli: binary choice between two actions
- Normal: continuous actions, Gaussian, no bounds
- Squashed Gaussian: Normal with bounds

The CartPole-v0 environment



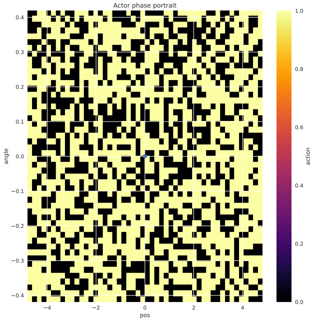
- The easiest gym classic control environment
- 4 state dimensions: $x, \dot{x}, \theta, \dot{\theta}$
- Binary action: push left or right. Use discrete or Bernoulli policy
- Custom continuous CartPole to study Gaussian policies (action in $[-1, 1]$)
- 200 steps, $+1$ at each step $\rightarrow$ utility in $[0, 200]$

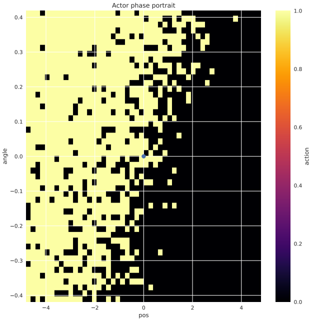
Results: Policy Gradient with Bernoulli policy and no baseline



- Variance over 10 runs
- Sum, discounted sum and normalized advantage work well
- No need for additional exploration
- Stochasticity of the binary policy is enough

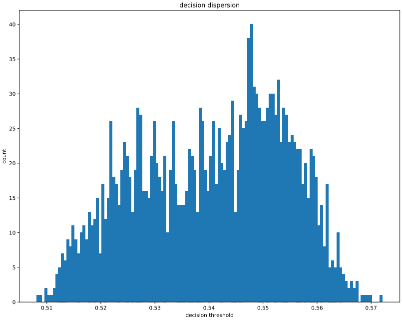
Initial/Final policy

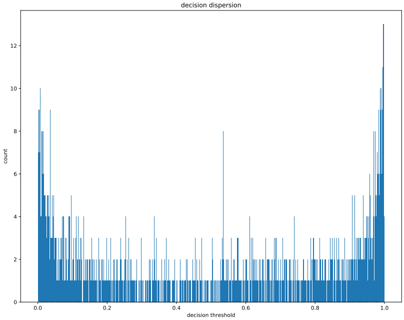




- 4 dimensions: $x, \dot{x}, \theta, \dot{\theta}$
- `FeatureInverter` wrapper to show x and θ (see video about coding)
- black = push left, yellow = push right
- General idea: push left when right, right when left, then manage pole

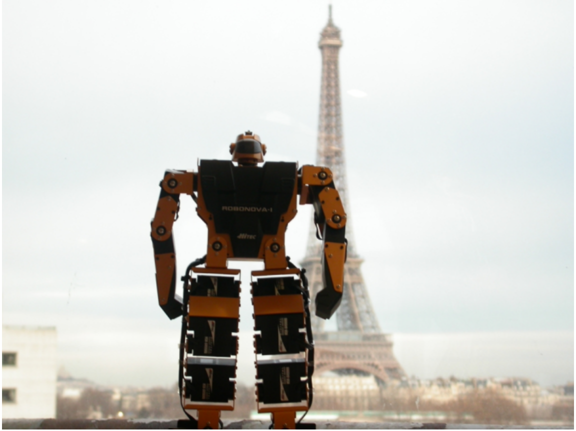
Initial/Final randomness





- Mind the scope on x-axis: initially very small (0.5 $\rightarrow$ 0.58, not centered)
- At the end of training, the policy is much less stochastic (more 0 and 1)
- Looking for optimality pushes towards less exploration

Any question?



Send mail to: Olivier.Sigaud@upmc.fr