Cross-Entropy Method for Importance Sampling

Tom Bosc

03/02/17

Introduction

- "Almost all off-policy learning methods require importance sampling"
- How to choose behavior μ?
- Minimize the variance of the estimator directly if possible.
- Minimize the KL divergence between the ideal estimator and a parametric density. **Cross-Entropy method [1]**

Importance sampling

We want to estimate an expectation.

$$\ell = \mathbb{E}_f[H(\mathbf{X})] = \int H(\mathbf{x})\, f(\mathbf{x})\, \mathrm{d}\mathbf{x}\,, \tag{5.39}$$

Ordinary importance sampling gives us:

$$\ell = \int H(\mathbf{x})\, \frac{f(\mathbf{x})}{g(\mathbf{x})}\, g(\mathbf{x})\, \mathrm{d}\mathbf{x} = \mathbb{E}_g\left[H(\mathbf{X})\, \frac{f(\mathbf{X})}{g(\mathbf{X})}\right], \tag{5.40}$$

Variance minimization

We want to minimize the variance of the importance sampling estimator:

$$\min_g \mathrm{Var}_g \left(H(\mathbf{X}) \frac{f(\mathbf{X})}{g(\mathbf{X})} \right). \tag{5.44}$$

The solution is:

$$g^*(\mathbf{x}) = \frac{|H(\mathbf{x})| f(\mathbf{x})}{\int |H(\mathbf{x})| f(\mathbf{x}) \, \mathrm{d}\mathbf{x}} . \tag{5.45}$$

Where the denominator is I (the expectation that we want to estimate) if $H(X) \geq 0$...

Variance minimization

If we have chosen the importance sampling density in the same family of distribution as the target, we can rewrite the variance minimisation problem as:

$$\min_{\mathbf{v}\in\mathscr{V}} V(\mathbf{v}) \,, \tag{5.51}$$

where

$$V(\mathbf{v}) = \mathbb{E}_{\mathbf{v}}[H^2(\mathbf{X})\, W^2(\mathbf{X};\mathbf{u},\mathbf{v})] = \mathbb{E}_{\mathbf{u}}[H^2(\mathbf{X})\, W(\mathbf{X};\mathbf{u},\mathbf{v})] \,. \tag{5.52}$$

We can often solve analytically:

$$\mathbb{E}_{\mathbf{u}}[H^2(\mathbf{X})\, \nabla W(\mathbf{X};\mathbf{u},\mathbf{v})] = \mathbf{0} \tag{5.55}$$

Cross-Entropy method

Sometimes, it is difficult to find the min, so we can minimize the KL divergence between the min and a convenient parametric distribution.

$$\min_{\mathbf{v}} \mathcal{D}\left(g^*, f(\cdot; \mathbf{v})\right).$$

With:

$$g^*(\mathbf{x}) = \frac{|H(\mathbf{x})|\, f(\mathbf{x})}{\int |H(\mathbf{x})|\, f(\mathbf{x})\, \mathrm{d}\mathbf{x}}\,. \tag{5.45}$$

The minus entropy of the min is constant in the minimisation. Thus, the solution maximizes minus cross-entropy

$$\max_{\mathbf{v}} D(\mathbf{v}) = \max_{\mathbf{v}} \mathbb{E}_{\mathbf{u}}\left[H(\mathbf{X}) \ln f(\mathbf{X}; \mathbf{v})\right]. \tag{5.61}$$

Cross-Entropy method

Similarly to variance minimisation, we can solve analytically.

$$\mathbb{E}_{\mathbf{u}} \left[H(\mathbf{X}) \nabla \ln f(\mathbf{X}; \mathbf{v}) \right] = \mathbf{0} \,, \tag{5.62}$$

But we can also use another distribution parametrized by w and iterate:

$$\max_{\mathbf{v}} D(\mathbf{v}) = \max_{\mathbf{v}} \mathbb{E}_{\mathbf{w}} \left[H(\mathbf{X}) \, W(\mathbf{X}; \mathbf{u}, \mathbf{w}) \ln f(\mathbf{X}; \mathbf{v}) \right] , \tag{5.64}$$

In our case:

In the RL framework:

- The expectation I that we want to estimate is a value function such as $Q(s, a)$
- We have one target density μ_s for each state (symbolized by f and later parametrized by u in previous equations)
- The behavior density is re-estimated at each policy evaluation in a GPI.
- If we use the distribution parametrized by w, we can compute the cross entropy minimiser using w and use the result as the next iteration's w

Experiments: Gambler's problem: 50k iterations

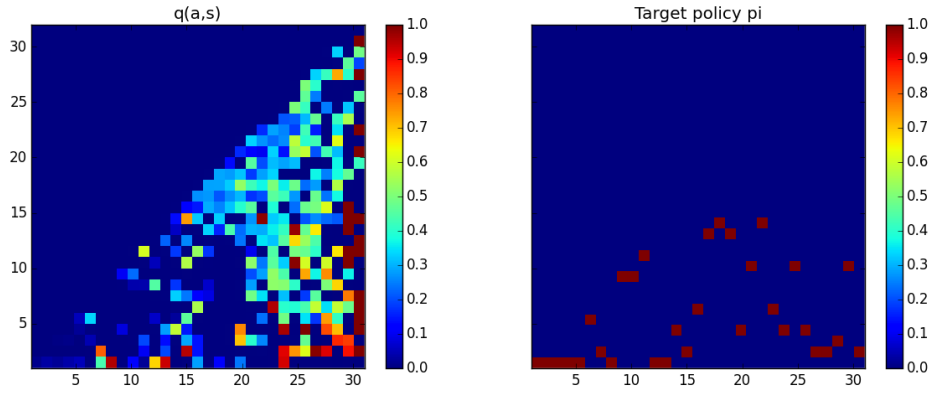

Figure 1: Cross-entropy behavior

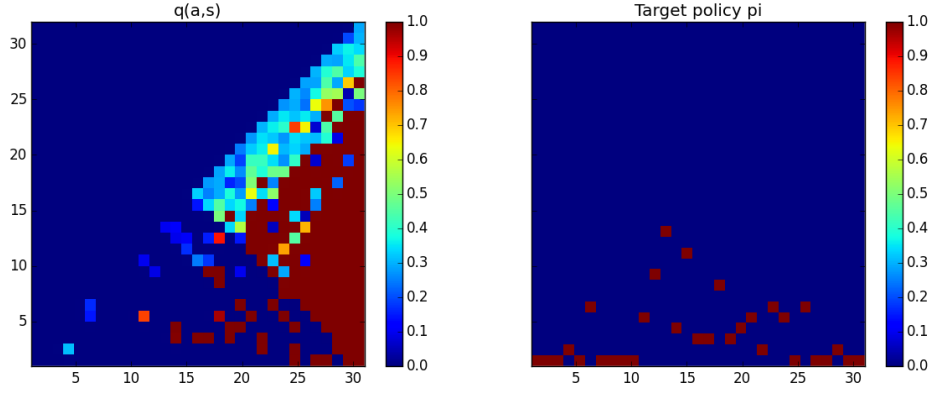

Figure 2: Uniform behavior

Experiments: Gambler's problem: 200k iterations

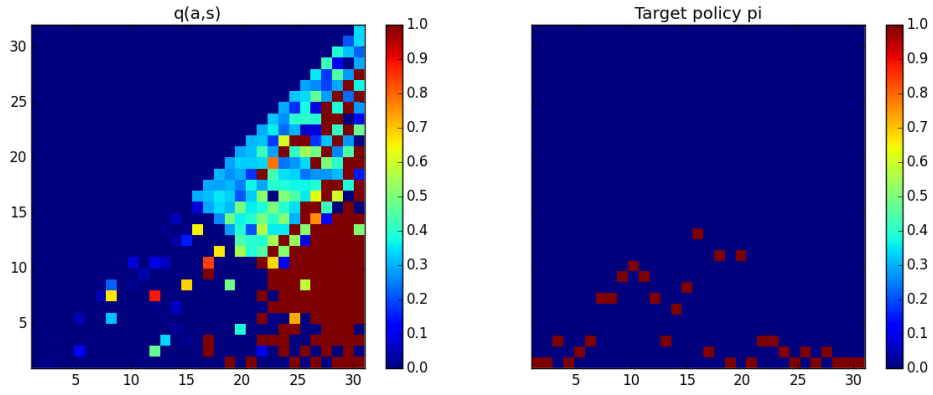

Figure 3: Cross-entropy behavior

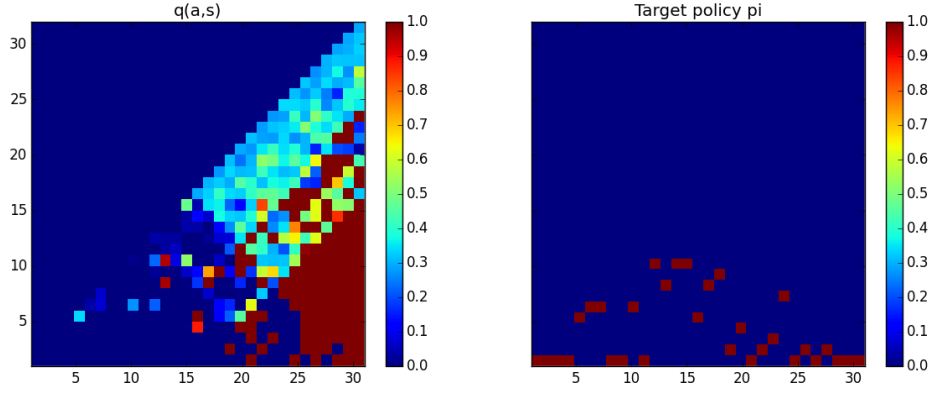

Figure 4: Uniform behavior

Bibliography

- [1] Simulation and the Monte-Carlo method, subsection 5.6, Rubinstein, Kroese, 2007