Max Integral Operator: A Probabilistic Numeric Approach

Nishad Gothoskar

Learning and Intelligent Systems - MIT

August 25 2018

Motivation

Bellman Update in Continuous State-Action Spaces

$$V^{\pi}(x) = \max_{u \in A} \int_{x' \in S} p(x'|u)\big(R(x'|x,u) + \gamma^{\Delta t} V^{\pi}(x')\big)\ \mathrm{d}x'$$

Use a GP to model the Value Function

Max Integral Operator

In this work, we will explore techniques to evaluate expressions of the following form:

$$\max_a \int f(s', a) p(s', a) \, \mathrm{d}s'$$

We find the Max Integral expression in a variety of computational problems.
Intuitively, maximizing over some parameters while integrating out others is a useful operation.

Problem Formulation

- We consider functions f for which the following integral converges for all a.

$$\int_{-\infty}^{\infty} f(a, s) p(s) \, \mathrm{d}s$$

- s is a continuous parameter
- a may be either continuous or discrete

Preliminaries

- Bayesian Optimization
- Bayesian Quadrature

Bayesian Quadrature

Most work in BQ has focused on integrals of the form:

$$\int f(x)p(x)\,\mathrm{d}x$$

First presented as Bayes-Hermite Quadrature [1], BQ uses active sampling to estimate an integral's value. This is done by using a Gaussian Process to model the integrand and then integrating the GP.

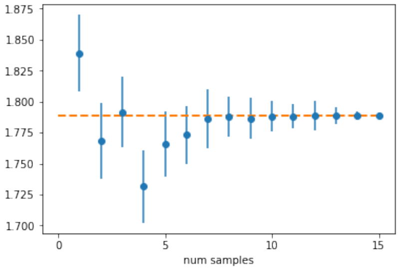
This gives us an estimated mean and variance for the integral's value.

Bayesian Quadrature - Acquisition Function

To sample function evaluations we optimize:

$$\bar{f} = \int f(x)p(x)dx$$

$$x^* = \underset{x}{\operatorname{argmin}} \mathbb{V}(\bar{f}|\mathcal{D}, x)$$



Bayesian Quadrature - Closed Form

$$\int (k(s,x)K^{-1}\boldsymbol{f})\ p(s)ds = z^T K^{-1}\boldsymbol{f}$$

$$z_i = \int exp(-0.5\sum_{d=1}^{D}\frac{(s_d - x_d^{(i)})^2}{w_d^2})\ exp(-0.5\sum_{d=1}^{D}\frac{s_d^2}{\sigma_d^2})\prod_{d=1}^{D}2\pi\sigma_d ds$$

$$z_i = \int exp(-0.5\sum_{d=1}^{D}\frac{\sigma_d^2(s_d - x_d^{(i)})^2 + w_d^2 s_i^2}{w_d^2\sigma_d^2})\prod_{d=1}^{D}2\pi\sigma_d ds$$

$$z_i = \int exp(-0.5\frac{(x_d^{(i)})^2}{w_d^2 + \sigma_d^2})\prod_{d=1}^{D}(2\pi\frac{w_d^2 + \sigma_d^2}{w_d^2\sigma_d^2})^{-1/2}\prod_{d=1}^{D}2\pi\sigma_d ds$$

$$z_i = \int exp(-0.5\sum_{d=1}^{D}\frac{(x_d^{(i)})^2}{w_d^2 + \sigma_d^2})\prod_{d=1}^{D}(\frac{w_d^2 + \sigma_d^2}{w_d^2})^{-1/2}ds$$

Bayesian Quadrature - Closed Form

$$\int k(s, x)p(s)ds = r$$

$$r_i = \int exp(-0.5 \sum_{d=1}^{D} \frac{(s_d - x_d^{(i)})^2}{w_d^2})\ exp(-0.5 \sum_{d=1}^{S} \frac{s_d^2}{\sigma_d^2}) \prod_{d=1}^{S} 2\pi\sigma_d ds$$

$$r_i = exp(-0.5 \sum_{d=S+1}^{D} \frac{(s_d - x_d^{(i)})^2}{w_d^2})\ z_i$$

Bayesian Optimization

$$x^* = \operatorname*{argmax}_x f(x)$$

Model f and use GP posterior mean and variance to select queries.

$$\mu(x) + \beta\sigma(x)$$

For Max Integral Optimization, we apply UCB to selecting a.

μ and σ are from estimating the integral.

Method

Our method resembles Bayesian Optimization and Bayesian Quadrature, because we sample to both reduce uncertainty of the inner integral and find the maximum value a.

However, by considering both these objectives and using an acquisition function that accurately captures them, we can converge with fewer function evaluations.

Method

Algorithm 1 Max Integral Optimization

function OPTMIO($\boldsymbol{f}, p, n$)
    $GP = \text{INIT}(f, p)$
    for n iterations **do**
        $a = \text{ActionAquisition}(GP)$
        $s = \text{StateAcquisition}(GP, p, a)$
        $y = f((s, a))$
        $GP = \text{AddObervation}(GP, ((s, a), y))$
    end for
    return $argmax_a \int m_{GP}(s, a) ds$
end function

Experiments

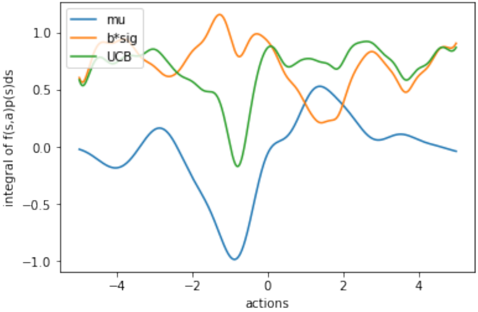
- Synthetic Functions
- Reinforcement Learning

Synthetic Functions

We first demonstrate our method on random functions drawn from a Gaussian Process.

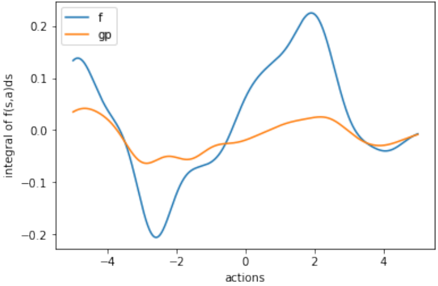
We evaluate settings with 1 or 2 dimensional action space and up to 3 dimensions state spaces.

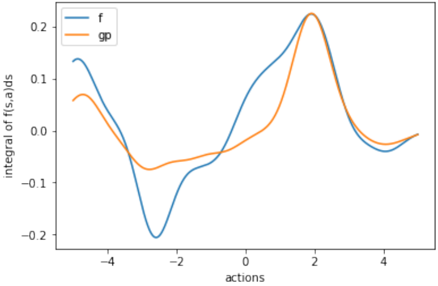
Synthetic Functions



UCB on actions represents a confidence bound on the integral estimate (along that action dimension).

Synthetic Functions





Reinforcement Learning

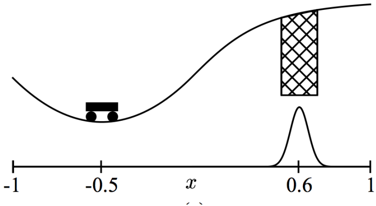
We can apply these methods to the Bellman Update:

$$V^{\pi}(x) = \max_{u \in A} \int_{x' \in S} p(x'|u)\big(R(x'|x,u) + \gamma^{\Delta t} V^{\pi}(x')\big)\ \mathrm{d}x'$$

This setting differs from the synthetic functions because now p depends on the action.

We are working on using this for Value Iteration and RTDP.

Reinforcement Learning



Rasmussen and Kuss (2004) use GP to model:

- System Dynamics
- Value Function

And this defines an implicit policy:

$$\pi(\mathbf{s}) \leftarrow \operatorname*{argmax}_{\mathbf{a}\in\mathcal{A}(\mathbf{s})} \int \mathcal{P}^{\mathbf{a}}_{\mathbf{s},\mathbf{s}'}\left[\mathcal{R}^{\mathbf{a}}_{\mathbf{s},\mathbf{s}'} + \gamma V(\mathbf{s}')\right] d\mathbf{s}'$$

Reinforcement Learning

$$\pi(\mathbf{s}) \leftarrow \underset{\mathbf{a}\in\mathcal{A}(\mathbf{s})}{\operatorname{argmax}} \int \mathcal{P}^{\mathbf{a}}_{\mathbf{s},\mathbf{s}'} \left[\mathcal{R}^{\mathbf{a}}_{\mathbf{s},\mathbf{s}'} + \gamma V(\mathbf{s}')\right] d\mathbf{s}'.$$

Even if we use the closed form integration, it depends on all support points of V, which may be large. And we still have to evaluate this for each action we wish to test.

Using our method, we can selectively query V and still find the optimal a.

Next Steps

- Value Iteration and RTDP
- Pushing Experiment

Other Prior Distributions

Mixture of Gaussians

$$\int f(x)p(x)dx = \alpha \int f(x)q_1(x)dx + (1-\alpha) \int f(x)q_2(x)dx$$

Importance Sampling

$$\int f(x)p(x)dx = \int \frac{f(x)p(x)}{q(x)} q(x)dx$$

Questions?

Bibliography

[1] O'Hagan, A. (1991). Bayes–Hermite quadrature. Journal of Statistical Planning and Inference, 29(3), 245–260.
[2] M. P. Deisenroth, J. Peters, and C. E. Rasmussen, "Approximate dynamic programming with Gaussian processes," in Proc. of the IEEE American Control Conference (ACC), 2008, pp. 4480–4485.
[3] Rasmussen, C.E., Ghahramani, Z.: Bayesian Monte Carlo. In Becker, S., Obermayer, K., eds.: Advances in Neural Information Processing Systems. Volume 15. MIT Press, Cambridge, MA (2003)