Predictive Entropy Search for Constrained Bayesian Optimization

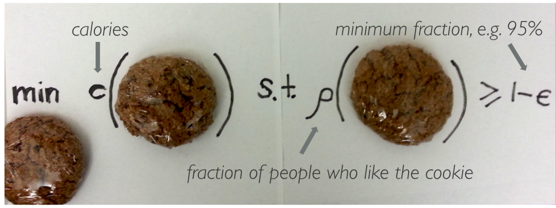
José Miguel Hernández-Lobato*, Michael Gelbart*, Matthew Hoffman†, Ryan Adams*, Zoubin Ghahramani†

*Harvard University, †Cambridge University

HARVARD INTELLIGENT PROBABILISTIC SYSTEMS

Introduction

Motivation: A cookie company wants to create a low-calorie cookie that is just as tasty as the original. This is a **constrained optimization** problem over the parameterized space of cookie recipes. **Bayesian Optimization** (BO) methods based on **Gaussian Processes** (GP) can find the optimal recipe using a small number of function evaluations.



In general, we want to solve: max $f(\mathbf{x})$ s.t. $\mathcal{C}_1(\mathbf{x}) \geq 0, \ldots, \mathcal{C}_K(\mathbf{x}) \geq 0$, where the evaluations of f and $\mathcal{C}_1, \ldots, \mathcal{C}_K$ may be corrupted by noise.

Problem: Existing BO solutions work by combining the **Expected Improvement** (EI) heuristic with feasibility indicators [1, 2]. These methods may run into problems because the EI framework does not naturally extend to the constrained case.

Solution: approaches that maximize the amount of **information** on the location of the global solution do not have the aforementioned problems. We extend an information-based method called **Predictive Entropy Search** (PES) [3] to handle constraints. The resulting method is called **Predictive Entropy Search with Constraints** (PESC).

Expected Improvement with Constraints (EIC)

EIC [2, 1] extends the expected improvement (EI) heuristic

$$\text{EI}(\mathbf{x}|y^\star) = \int \max(0, y^\star - y) p(y|\mathbf{x}) \text{d}y\,, \quad y^\star \equiv \text{best result}\,, \quad (1)$$

to consider the expected feasible improvement, i.e., the EI subject to $\mathcal{C}_k(\mathbf{x}) \geq 0$, $k = 1, \ldots, K$. The EIC acquisition function is then

$$\alpha(\mathbf{x}) = \text{EI}(\mathbf{x}|y^\star) \textstyle\prod_{k=1}^K \Pr(\mathcal{C}_k(\mathbf{x}) \geq 0)\,, \quad y^\star \equiv \text{best } \textbf{feasible} \text{ result}\,. \quad (2)$$

When the evaluation of the constraints is corrupted by noise, EIC uses probabilistic constraints: $\mathbf{x}$ is feasible if $\Pr(\mathcal{C}_k(\mathbf{x})|\mathcal{D}) \geq 1 - \delta_k$, for $k = 1, \ldots, K$, where the δ_k are small positive numbers.

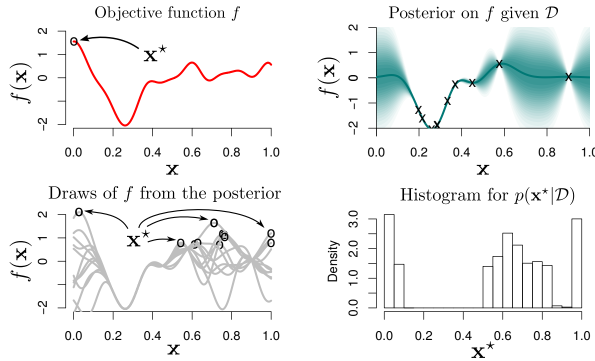
Problems:

- $y^\star$ is not defined if no feasible solution has been found yet.
- EIC breaks down in a **decoupled scenario** where we can choose to evaluate only one of the functions (objective or constraints):

  For $\mathbf{x}$ to improve over the current best feasible solution, **both** the objective and the constraints must produce high values at $\mathbf{x}$. But this is not possible if we only evaluate one of the functions.

Predictive Entropy Search (PES)

In the unconstrained case, PES [3] collects data actively to find the global maximizer $\mathbf{x}_\star$ of f:



PES maximizes the expected reduction in the entropy of $p(\mathbf{x}_\star|\mathcal{D})$

$$\alpha(\mathbf{x}) = \text{H}(p(\mathbf{x}_\star|\mathcal{D})) - \mathbb{E}_{p(y|\mathcal{D},\mathbf{x})}\left[\text{H}(p(\mathbf{x}_\star|\mathcal{D} \cup \{\mathbf{x}, y\}))\right] \equiv I(\mathbf{x}_\star, y)\,, \quad (3)$$

where $H[\cdot]$ computes the entropy of a distribution. (3) is also used by Entropy Search [4]. Because $I(\mathbf{x}_\star, y) = I(y, \mathbf{x}_\star)$ PES rewrites $\alpha(\mathbf{x})$ as

$$\alpha(\mathbf{x}) = \text{H}(p(y|\mathcal{D},\mathbf{x})) - \mathbb{E}_{p(\mathbf{x}_\star|\mathcal{D})}\left[\text{H}(p(y|\mathcal{D},\mathbf{x},\mathbf{x}_\star))\right] \equiv I(y, \mathbf{x}_\star)\,. \quad (4)$$

PES approximates the expectation with respect to $p(\mathbf{x}_\star|\mathcal{D})$ in (4) using an approximate **Monte Carlo** method. **Expectation propagation** (EP) [5] is used to approximate $p(y|\mathcal{D},\mathbf{x},\mathbf{x}_\star)$ in the acquisition function.

PES with Constraints (PESC)

PESC collects data to find the global maximizer $\mathbf{x}_\star$ of f that is feasible. The PESC acquisition function is

$$\alpha(\mathbf{x}) = \text{H}\left[p(y^f, y^1, \ldots, y^K|\mathcal{D},\mathbf{x})\right] - \mathbf{E}\left\{\text{H}\left[p(y^f, y^1, \ldots, y^K|\mathcal{D},\mathbf{x},\mathbf{x}_\star)\right]\right\}\,,$$

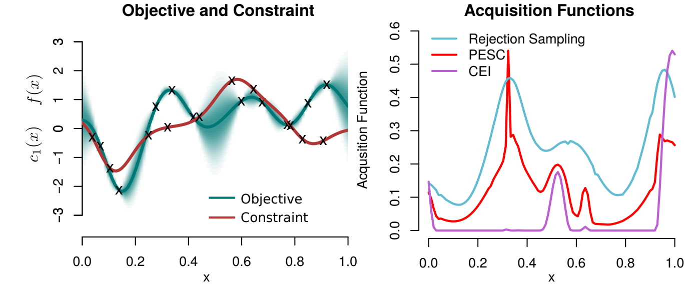
- The expectation with respect to $p(\mathbf{x}_\star|\mathcal{D})$ is approximated as in PES.
- The distribution conditioned to $\mathbf{x}_\star$ is intuitively given by

$$p(y^f, y^1, \ldots, y^K|\mathcal{D},\mathbf{x},\mathbf{x}_\star) \propto \int \delta[\bar{f} - f(\mathbf{x})]\mathcal{N}(y^f|\bar{f}, \sigma_f^2)p(f|\mathcal{D}_n^f)$$
$$\left[\textstyle\prod_{\mathbf{x}' \neq \mathbf{x}_\star}\left(\left\{\prod_{k=1}^K \Theta[c_k(\mathbf{x}')]\right\}\Theta[f(\mathbf{x}_\star) - f(\mathbf{x}')] + \left\{1 - \prod_{k=1}^K \Theta[c_k(\mathbf{x}')]\right\}\right)\right]$$
$$\left[\textstyle\prod_{j=1}^k \Theta[c_j(\mathbf{x}_\star)]\right]\left[\prod_{k=1}^K \left\{\delta[\bar{c}_k - c_k(\mathbf{x})]\mathcal{N}(y^k|\bar{c}_k, \sigma_k^2)p(c_k|\mathcal{D}_n^k)\right\}\right] df\, dc_1 \cdots dc_K\,,$$

where Θ is the Heaviside step function, $\delta(\cdot)$ is Dirac's delta and f, $c_1, \ldots, c_K$ are infinite dimensional vectors encoding the objective and the constraints.

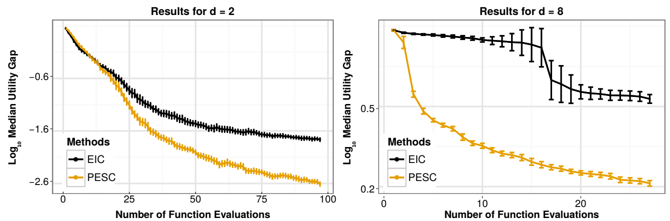
Approximation: We first approximate the infinite product with a **finite dimensional** one only over the locations at which we have data. We then find a Gaussian approximation using EP [5].

Quality of the PESC Approximation



We compare the acquisition functions (right plot) given by PESC, EIC and and a ground truth for PESC given by **rejection sampling** on synthetic data (right plot). The ground truth and PESC obtain the highest values at similar locations.

Experiments: PESC vs. EIC



Objective and **one** constraint are sampled from a zero-mean GP with a squared exponential covariance function with unit amplitude and length scale $\ell = 0.1$ in each dimension. Evaluations are corrupted with Gaussian noise with variance 0.01. We use **probabilistic constraints** and fix $\delta_1 = 0.05$. The y-axis shows median utility gap over 500 runs between the recommended solution and the true solution, where utility is set to the worst possible value if the recommended solution violates the constraint.

References

[1] Matthias Schonlau, William J Welch, and Donald R Jones. Global versus local search in constrained optimization of computer models. *Lecture Notes-Monograph Series*, pages 11–25, 1998.

[2] Michael A. Gelbart, Jasper Snoek, and Ryan P. Adams. Bayesian optimization with unknown constraints. In *UAI*, 2014.

[3] J. M Hernández-Lobato, M. W. Hoffman, and Z. Ghahramani. Predictive entropy search for efficient global optimization of black-box functions. In *NIPS*. 2014.

[4] Philipp Hennig and Christian J. Schuler. Entropy search for information-efficient global optimization. *JMLR*, 2012.

[5] Thomas P Minka. *A family of algorithms for approximate Bayesian inference.* PhD thesis, Massachusetts Institute of Technology, 2001.