# Curiosity simulation

Franziska Brändle, Charley M. Wu & Eric Schulz

March 2020

## 1 Problem statement

We want to explore a Bayesian optimization problem and compare the sampling patterns of different models. We define target function $f(x)$ on some bounded set $\mathcal{X}$, which we will take to be a subset of $\mathbb{R}^D$. We use a Gaussian Process ($\mathcal{GP}$) as a probabilistic model for $f(x)$ and then exploit this model via an acquisition function to decide where in $\mathcal{X}$ to evaluate the function next. Our goal is to assess how different acquisition functions differ in their mean confidence ranks over their sampled input points.

## 2 Gaussian Process

Let $f(\boldsymbol{x})$ be a function mapping an input $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$ to an output $y$. A $\mathcal{GP}$ defines a distribution $p(f)$ over such functions. A $\mathcal{GP}$ is parameterized by a mean function $m(\boldsymbol{x})$ and a kernel function, $k(\boldsymbol{x}, \boldsymbol{x}')$ :

$$m(\boldsymbol{x}) = \mathbb{E}\left[f(\boldsymbol{x})\right] \tag{1}$$

$$k(\boldsymbol{x}, \boldsymbol{x}') = \mathbb{E}\left[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))\right] \tag{2}$$

At time $t$, we have collected observations $\mathbf{y}_{1:t} = [y_1, y_2, \ldots, y_t]^\top$ at inputs $\boldsymbol{X}_{1:t} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t\}$, which together form the dataset $\mathcal{D}_{1:t} = \{\mathbf{X}_{1:t}, \mathbf{y}_{1:t}\}$. For each outcome $y_t$, we assume $y_t = f(\boldsymbol{x}_t) + \epsilon_t$ with $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Given a $\mathcal{GP}$ prior on the functions $f(\boldsymbol{x}) \sim \mathcal{GP}\left(m(\boldsymbol{x}), k(\boldsymbol{x}, \boldsymbol{x}')\right)$, the posterior over $f$ conditioned on observed data $\mathcal{D}_{1:t}$ is also a $\mathcal{GP}$ with mean and variance defined as:

$$\mathbb{E}\left[f(\mathbf{x})|\mathcal{D}_{1:t}\right] = \mathbf{k}_{1:t}(\boldsymbol{x})^\top(\boldsymbol{K}_{1:t} + \sigma_\epsilon^2 \boldsymbol{I}_t)\boldsymbol{y}_{1:t} \tag{3}$$

$$\mathbb{V}[f(\mathbf{x})|\mathcal{D}_{1:t}] = k(\boldsymbol{x}, \boldsymbol{x}') - \boldsymbol{k}_{1:t}(\boldsymbol{x})^\top(\boldsymbol{K}_{1:t} + \sigma_\epsilon^2 \boldsymbol{I}_t)^{-1}\boldsymbol{k}_{1:t}(\boldsymbol{x}') \tag{4}$$

where $\boldsymbol{k}_{1:t}(\boldsymbol{x}) = [k(\boldsymbol{x}_1, \boldsymbol{x}), \ldots, k(\boldsymbol{x}_t, \boldsymbol{x})]^\top$, $\boldsymbol{K}_{1:t}$ is the positive definite kernel matrix $[k(\boldsymbol{x}_i, \boldsymbol{x}_j)]_{i,j=1,\ldots,t}$, and $\boldsymbol{I}_t$ is a $t$ by $t$ identity matrix.

## 3 Setup details

We use the Radial Basis Function (RBF) kernel as a component of the GP function learning algorithm, which specifies the correlation between inputs.

$$k(\boldsymbol{x}, \boldsymbol{x}') = \exp\left(-\frac{||\boldsymbol{x} - \boldsymbol{x}'||^2}{2\lambda^2}\right) \tag{5}$$

We sample a univariate target function $f^* \sim \mathcal{GP}(0, k(\boldsymbol{x}, \boldsymbol{x}'))$ over a discretized input space from -5 to 5 with equal distance between each points. We corrupt the signal of this function with a noise of $\sigma_\epsilon^2 = 0.1$, such that

$$f^*(\boldsymbol{x}) + \epsilon,\ \epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2) \tag{6}$$

We then use a Gaussian Process to explore the sampled function by sampling options based on different acquisition functions. For all simulations, we set the length-scale of both the generating and the sampling RBF kernel to $\lambda = 1$.

# 4 Acquisition functions

We compare three different acquisition functions: novelty-based sampling, complexity approximations, and upper confidence bound sampling. For simplicity, we redefine the posterior mean as $\mu(\boldsymbol{x}) := \mathbb{E}\left[f(\boldsymbol{x})|\mathcal{D}_{1:t}\right]$ and the posterior standard deviation as $\sigma(\mathbf{x}) := \sqrt{\mathbb{V}[f(\boldsymbol{x})|\mathcal{D}_{1:t}]}$, based on the standard convention for describing a Gaussian distribution.

## 4.1 Uncertainty sampling

Uncertainty sampling (US) sampling picks the next point that currently has the highest predictive uncertainty

$$\mathrm{US}(\boldsymbol{x}) = \sigma(\boldsymbol{x}). \tag{7}$$

NB sampling treats uncertainty as positive and samples the option it is currently the most uncertain about. This mimics novelty-based approaches to curiosity.

## 4.2 Expected model change

Since GPs are nonparametric, our implementation of expected model change (EMC) sampling used the expected reduction in prediction uncertainty over all options as a measure of model change:

$$\mathrm{EMC}(\boldsymbol{x}) = \int_{\mathcal{D}_{1:t}} \Delta\sigma(\boldsymbol{x})\,\mathrm{d}y. \tag{8}$$

In practice, this integral is approximated using Monte Carlo sampling. We first compute the sum of the current posterior standard deviation over all options (i.e., global model uncertainty). Then we use Monte Carlo samples from the posterior predictive distribution to simulate new observations, and recompute the global uncertainty conditioned on these simulated observations. Thus, EMC is the difference in global model uncertainty, averaged over the Monte Carlo samples.

## 4.3 Upper confidence bound sampling

Upper confidence bound (UCB) sampling picks the next point that currently has the highest upper confidence bound

$$\mathrm{UCB}(\boldsymbol{x}) = \mu(\boldsymbol{x}) + \beta\sigma(\boldsymbol{x}). \tag{9}$$

UCB sampling is an optimistic strategy that samples based on an explicit exploration-exploitation trade-off, and has been proven to produce sub-linear regret. We set the exploration parameter to $\beta = 3$ to mimic highly curious agents.

# 5 Simulation details

We let each acquisition function explore 10,000 target functions over 10 trials each. Thus, there are 10,000 simulation runs for each acquisition function. On each trial, a softmax choice rule transforms each model's valuations into a probability distribution over options:

$$p(\mathbf{x}) = \frac{\exp(q(\boldsymbol{x})/\tau)}{\sum_{j=1}^{N}\exp(q(x_j)/\tau)}, \tag{10}$$

where $q(\mathbf{x})$ is the predicted value of each option $\boldsymbol{x}$ for a given acquisition function (e.g., $q(\mathbf{x}) = \text{UCB}(\boldsymbol{x})$ for UCB), and $\tau$ is the temperature parameter. We set $\tau = 0.001$ to mimic minimally noisy sampling.

For every trial per simulation, we rank all options based on their current confidence, measured by $\sigma(gvecx)$, and then track the average confidence rank[1]. After each run, we calculate the average confidence ranks over trials, leading to a collection of 10,000 such mean ranks for each acquisition function.

# 6 Results

We found that the approximations produced the expected patterns of behavior (Fig. 1). Uncertainty sampling showed a linear relation between confidence and sampling probability, resembling the predictions of novelty-based theories of curiosity. Both expected model change sampling and upper confidence bound sampling produced inverse U-shaped curves, resembling the predictions of complexity-based theories of curiosity. Therefore it can be difficult to distinguish between these strategies based on their sampling behaviour.

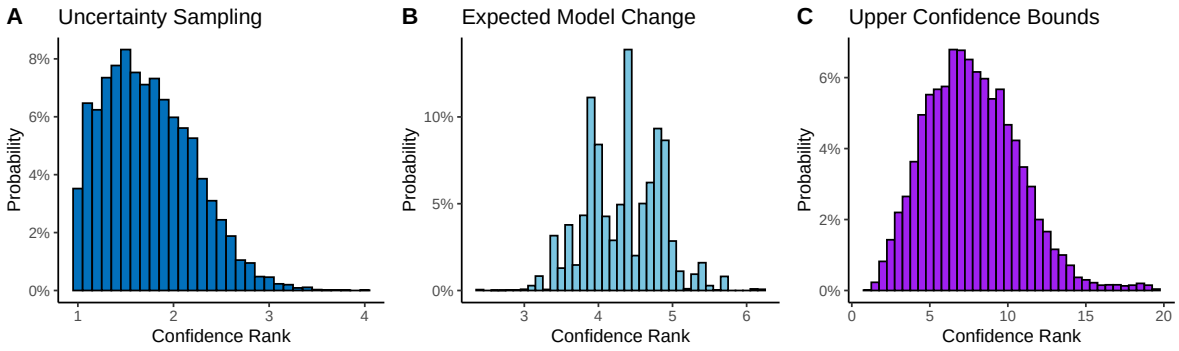


Figure 1: **Simulation Results.** Agents sample options based on their uncertainty, according to three different acquisition functions.

[1] We use rank values, since $\sigma$ changes with every trial.