Erratum in
Gaussian Process Optimization with Mutual Information

Emile Contal[1], Vianney Perchet[2], and Nicolas Vayatis[1]

[1]CMLA, UMR CNRS 8536, ENS Cachan, France
[2]LPMA, Université Paris Diderot, France

June 8, 2015

Abstract

This paper describes a mistake in the article "Gaussian Process Optimization with Mutual Information", Proceedings of the 31st International Conference on Machine Learning, JMLR, pp.253-261, 2014. It appears that the information given to the algorithm is not sufficient for the main theorem to hold true. The theoretical guarantees would remain valid in a setting where the algorithm observes the instantaneous regret instead of noisy samples of the unknown function.

Introduction. In our paper "Gaussian Process Optimization with Mutual Information" [1], we analyze an algorithm for sequential global optimization using Gaussian processes and we aim at proving upper bounds on the cumulative regret incurred by the algorithm. We found an error in the proof of Lemma 1, which invalidates the main theorem.

Notations. Let $f : \mathcal{X} \to \mathbb{R}$ be the unknown function to be optimized, which is a sample from a Gaussian process. Let's fix $x^\star, x_1, \dots, x_T \in \mathcal{X}$ and the observations $y_t = f(x_t) + \epsilon_t$ where the noise variables ϵ_t are independent Gaussian noise $\mathcal{N}(0, \sigma^2)$. We define the instantaneous regret $r_t = f(x^\star) - f(x_t)$ and,

$$M_T = \sum_{t=1}^{T} \Big(r_t - \mathbb{E}\big[r_t \mid y_1, \dots, y_{t-1} \big] \Big) .$$

Erratum. In Lemma 1, we claimed that M_T is a Gaussian martingale with respect to $\mathbf{Y}_T = y_1, \dots, y_T$. Even if $M_t - M_{t-1}$ is a centered Gaussian conditioned on $\mathbf{Y}_{T-1}$, it is wrong to say that M_T is a martingale since it is not measurable with respect to $\mathbf{Y}_T$.

Correction and consequences. In order to fix Lemma 1, it is possible to modify M_T and use its natural filtration $\mathcal{F}_T = \{r_t\}_{t \leq T}$ instead of $\mathbf{Y}_T$,

$$M_T = \sum_{t=1}^{T} \Big(r_t - \mathbb{E}\big[r_t \mid \mathcal{F}_{t-1}\big] \Big) .$$

Then M_T is a Gaussian martingale with respect to $\mathcal{F}_T$. Now to adapt the algorithm for this new quantity it needs to observe r_t instead of y_t to be able to compute both the posterior expectation and variance for all x in $\mathcal{X}$:

$$\mu_t(x) = \mathbb{E}\big[f(x) \mid \mathcal{F}_{t-1}\big] \text{ and } \sigma_t^2(x) = \mathrm{Var}\big[f(x) \mid \mathcal{F}_{t-1}\big] .$$

Comments on the experiments. We remark that the experiments performed in [1] are remarkably good in spite of Lemma 1 being unproved. After having discovered the mistake we were able to build scenarios were the GP-MI algorithm is overconfident and misses the optimum of f, and therefore incurs a linear cumulative regret.

References

[1] Emile Contal, Vianney Perchet, and Nicolas Vayatis. Gaussian process optimization with mutual information. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 253–261, 2014.