A fast, efficient algorithm for quantification of rare events in dynamical systems

Antoine Blanchard*, Themistoklis P. Sapsis*
*Department of Mechanical Engineering, Massachusetts Institute of Technology, USA

Summary. We introduce a fast, efficient algorithm for quantification of rare events in dynamical systems. The algorithm iteratively learns the "black-box" relationship between parameters (the inputs) and extreme observations (the outputs). To keep the number of black-box evaluations to a minimum, the algorithm cleverly selects the "next-best point" that it should sample based on how much that point will improve our estimate of the output statistics. "Improvement" is quantified by the so-called Q-criterion whose computation (and that of its gradient) can be done analytically, which allows for the possibility of the input space being high-dimensional. We show that the proposed sampling algorithm outperforms other standard sampling approaches.

Background

Extreme events may be thought of as short-lived episodes during which an observable (e.g., drag force, temperature, or stock price) significantly deviates from its mean value. They often have disastrous consequences, as with rogue waves, avalanches, wildfires or extreme weather conditions (see figure 1 and [1]). Quantification of extreme events is particularly difficult because they often arise in complex, high-dimensional, nonlinear systems, and give rise to heavy-tail probability density functions (pdf). The fact that extreme events occur with very low probability means that reliable quantification of the statistics requires an unfathomable amount of data. If the data is collected from a large-scale experiment or a computer simulation with millions of degrees of freedom, the task of extreme-event quantification is virtually hopeless. Thus, the challenge is to come up with a way to estimate the heavy-tail statistics of the system using as few samples as possible.

(a) Rogue waves (b) Avalanches (c) Wildfires (d) Hurricanes

Figure 1: Examples of extreme events arising in nature.

Mohamad & Sapsis [2] recently proposed a sampling strategy for quantification of extreme events in dynamical systems. The basis for their approach was to approximate the unknown "black box" $y = F(x)$ as the realization of a Gaussian process. Specifically, starting from an initially small dataset, their algorithm uses Gaussian process (GP) regression [4] to construct a surrogate map, and then determines the next-best sample for which evaluation of the black box will most improve knowledge of the statistics. At each iteration, the sought-after pdf can be approximated using the posterior distribution of the surrogate map. In their work, the criterion used to determine the next-best point is problematic for two reasons. First, it takes the form of an integral over the input space, which is intractable in high dimensions. Second, there is no closed-form expression for the gradient of their criterion, which forces them to use non-gradient-based optimizers.

Overview of the algorithm

Our algorithm is inspired by that of Mohamad & Sapsis [2], but considerably improves its scope and efficiency. Assuming that we have available a dataset of input–output pairs $\mathcal{D}_n = \{x_i, F(x_i)\}_{i=1}^n$, we proceed iteratively as follows:

Step 1. Train a GP regressor on $\mathcal{D}_n$, and compute surrogate mean f_n and covariance σ_n^2.
Step 2. Determine the next-best point by solving the minimization problem

$$x^* = \operatorname*{argmin}_x Q(x; f_n, \mathcal{D}_n). \tag{1}$$

Step 3. Evaluate the expensive map F at the new point, and append $\{x^*, F(x^*)\}$ to $\mathcal{D}_n$. Then, go back to Step 1.

To compute the criterion $Q(\tilde{x}; f_n, \mathcal{D}_n)$ at some candidate point $\tilde{x}$, we proceed as follows:

Step Q1. Construct $\tilde{\mathcal{D}}_n$ by appending $\{\tilde{x}, f_n(\tilde{x})\}$ to $\mathcal{D}_n$. (Note that we use the surrogate map f_n rather than F.)
Step Q2. Predict the covariance $\tilde{\sigma}_n^2$ using the GP regressor trained on $\mathcal{D}_n$, and return

$$Q(\tilde{x}; f_n, \mathcal{D}_n) = \int \frac{\tilde{\sigma}_n^2(x) p_x(x)}{p_n(f_n(x))} \, \mathrm{d}x. \tag{2}$$

In the above, $p_x(x)$ and $p_n(f_n(x))$ respectively denote the joint pdf of the inputs (which is assumed to be known) and the approximate pdf of the output as predicted by the surrogate map. Criterion (2) is already an improvement over the

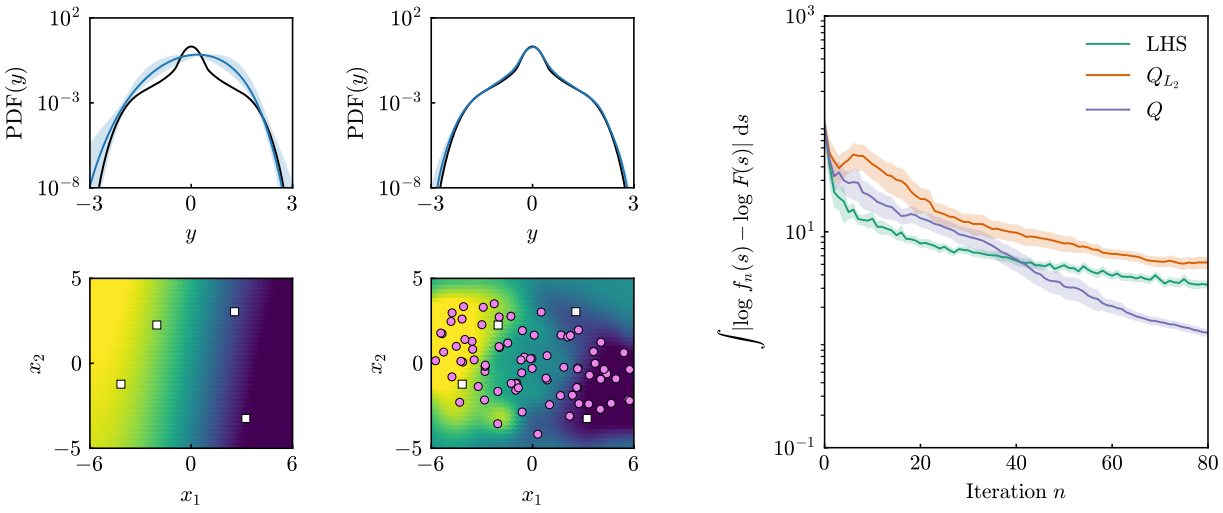
approach of Mohamad & Sapsis [2], for two reasons. First, only the term $\tilde{\sigma}_n^2(x)$ depends on the candidate new point, so computation of the gradient of Q with respect to $\tilde{x}$ is actually tractable, which allows use of gradient-based optimizers to solve (1). Second, as discussed by Sapsis [3], the Q-criterion critically depends on the ratio $p_x(x)/p_n(f_n(x))$ which accounts for the relative importance of the output with respect to the inputs. This is crucial because other approaches (e.g., based on minimization of the Kullback–Leibler divergence) do not take into account the output values of the input–output pairs collected so far, which adversely affects the efficiency of the algorithm.
We go one step further and recognize that for high-dimensional input spaces, evaluating the integral in (2) is daunting. For our algorithm to be efficient and tractable in high-dimensions, we fit a Gaussian Mixture Model (GMM) to

$$\frac{p_x(x)}{p_n(f_n(x))} \approx \sum_{i=1}^{n_{GMM}} \alpha_i \mathcal{N}_x(\mu_i, \Sigma_i) \quad \Longrightarrow \quad Q(\tilde{x}; f_n, \mathcal{D}_n) \approx \sum_{i=1}^{n_{GMM}} \alpha_i \left[\mu_{k_0} - \mathrm{tr}(S_{k_0 k_0}^{-1} C_{k_0 k_0, i})\right], \tag{3}$$

where μ_{k_0}, $S_{k_0 k_0}$ and $C_{k_0 k_0, i}$ are matrices whose dependence on the candidate point $\tilde{x}$ is only through the GP kernel k_0. Therefore, for a range of kernels (in this work, the RBF kernel), the Q-criterion and its gradient can be computed analytically, thereby drastically improving the efficiency of the sampling algorithm.

Results and discussion

We apply our smart sampling algorithm to the stochastically forced oscillator of Mohamad & Sapsis [2] with the same parameters and $m = 2$. Figure 2a shows that the true pdf of the observable (solid black) has heavy tails, which makes this example a good test case for our sampling algorithm. Figures 2a,b show that the pdf of the surrogate map (solid blue) converges to the true pdf after only a few iterations. One advantage of the GP framework is that the estimated pdf comes with uncertainty bounds (shaded blue areas in figures 2a,b). Another advantage of the algorithm is that it provides, in addition to the pdf of the output, a surrogate model for the expensive map F (bottom panel in figures 2a,b). To establish the superiority of our algorithm over existing approaches, we compare it to Latin Hypercube Sampling (LHS) of the input space, a non-iterative method which needs to be bootstrapped every time a new point is added to the dataset; and the Q-criterion (2) in which the term in the denominator is replaced by one, so that the importance of the output relative to the inputs is ignored (Q_{L_2}). Figure 2c demonstrates unequivocal superiority of our algorithm.



(a) Iteration 0 (b) Iteration 80 (c) Error plot

Figure 2: Convergence of the algorithm as more samples (pink circles) are added to the initial dataset (open squares).

It is important to note that our algorithm goes far beyond a simple estimation of the pdf for specific input values. The pdf is in fact a by-product of the sampling process, because what the algorithm learns is the actual black-box map F. This means that the surrogate map can be used to make various kinds of prediction and quantification. Our algorithm thus has significance implications from the standpoint of optimal experimental design, adaptive sampling and active search in complex environments, as well as uncertainty quantification and real-time prediction of extreme events.

References

[1] Lucarini V., Faranda D., Freitas A.C., Freitas J.M., Holland M., Kuna T., Nicol M., Todd M., Vaienti S. (2016) Extremes and Recurrence in Dynamical Systems. Wiley, New York, NY.

[2] Mohamad M. A., Sapsis T. P. (2018) Sequential Sampling Strategy for Extreme Event Statistics in Nonlinear Dynamical Systems. *Proc. Natl. Acad. Sci.* **115**:11138-11143.

[3] Sapsis T.P. (2019) Output-Weighted Optimal Sampling for Bayesian Regression and Rare Event Statistics Using Few Samples. (Submitted.)

[4] Rasmussen C.E, Williams C.K.I. (2006) Gaussian Processes for Machine Learning. The MIT Press, Cambridge, MA.