**Bayesian optimization with prior injection for the calibration of Cost-Effectiveness models**

Cost-effectiveness models are used to evaluate and compare different medical strategies (e.g. strategies to detect cancer), in terms of both health and cost, and to help decision makers determine the optimal allocation of medical resources. These models have several inputs that describe the environment, the disease and the strategies used (inputs like probabilities, sensitivities and specificities for medical procedures, costs, utility values, …) and can have several outputs as needed for the analysis (e.g. average life expectancy, average cost, incidence of the disease, mortality, …).

The stated goal of the research project is to find efficient ways to calibrate these cost-effectiveness models by changing some input parameters so a particular output (usually incidence or mortality) matches a given value found in the scientific literature. In summary, **we can frame the problem as the optimization of a black box function computing (for example) the euclidean distance between the simulated output and a theoretical value**.

In some of our models we find that classical optimization methods need many simulations to converge to a good solution, demanding a lot of time and computing resources. Also, in many cases we are able to provide information about the statistical distribution of some of the inputs that could help in the optimization process. For these reasons I believe Bayesian Optimization (BO) might be a good alternative to adapt to our particular modeling needs.

Some preliminary (ongoing) experiments, using both bayesian and classical methods, are available in a jupyter notebook in <https://github.com/david-gomez-guillen/phd>, using the *hypermapper*[1] python library for BO and *scipy* for classical optimization methods. This first set of tests focus on the performance of different optimization methods on well-known analytical functions that are easy to compute. Once acquainted with BO usage, the next step would be to repeat the tests on the actual cost-effectiveness models.

The conclusions and challenges found in the tests so far include:

1. BO using Gaussian Processes might be too slow for some of our more lightweight simulation models, classical methods converge faster even if they need more function evaluations. It should not be a problem for more computationally expensive simulation models.
2. BO tests performed converge to worse optima compared to classical optimization methods (e.g. Nelder-Mead, BFGS). It might be due to code issues: either implementation problems or an incorrect usage of the library.
3. Gaussian processes regression used in BO becomes more expensive for each iteration due to having to calculate the inverse of a matrix that grows with the number of observations. Regression becomes very slow to compute after a number of evaluations, especially for high-dimensional inputs (depending on the available hardware)[2][3].
4. Regular bayesian methodology (updating iteratively the model after each observation) is more difficult to parallelize than classical methods, though some alternatives exist[4][5].
5. BO in these tests performed with priors for the optimum[6] (implemented in *hypermapper*: <https://github.com/luinardi/hypermapper/wiki/prior-injection>) do not seem to converge faster. It might be due to library issues too: optimum priors implementation is labeled as experimental in the documentation.
6. When optimizing the actual cost-effectiveness models in future tests it would be interesting to add constraints for the inputs[7][8]. E.g: one input parameter being greater than another.

**References**

[1] L. Nardi, A. Souza, D. Koeplinger and K. Olukotun, "HyperMapper: a Practical Design Space Exploration Framework," 2019 IEEE 27th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS), 2019, pp. 425-426, doi: 10.1109/MASCOTS.2019.00053.

[2] Das, Sourish, Sasanka Roy, and Rajiv Sambasivan. “Fast Gaussian Process Regression for Big Data.” *ArXiv:1509.05142 [Cs, Stat]*, August 18, 2017. <http://arxiv.org/abs/1509.05142>.

[3] Terry, Nick, and Youngjun Choe. “Splitting Gaussian Processes for Computationally-Efficient Regression.” *PLOS ONE* 16, no. 8 (August 24, 2021): e0256470. <https://doi.org/10.1371/journal.pone.0256470>.

[4] Daulton, Samuel, Maximilian Balandat, and Eytan Bakshy. “Parallel Bayesian Optimization of Multiple Noisy Objectives with Expected Hypervolume Improvement.” *ArXiv:2105.08195 [Cs, Stat]*, October 26, 2021. <http://arxiv.org/abs/2105.08195>.

[5] Wang, Jialei, Scott C. Clark, Eric Liu, and Peter I. Frazier. “Parallel Bayesian Global Optimization of Expensive Functions.” *ArXiv:1602.05149 [Math, Stat]*, May 5, 2019. <http://arxiv.org/abs/1602.05149>.

[6] Souza, Artur, Luigi Nardi, Leonardo B. Oliveira, Kunle Olukotun, Marius Lindauer, and Frank Hutter. “Bayesian Optimization with a Prior for the Optimum.” *ArXiv:2006.14608 [Cs, Stat]*, April 19, 2021. <http://arxiv.org/abs/2006.14608>.

[7] Ungredda, Juan, and Juergen Branke. “Bayesian Optimisation for Constrained Problems.” *ArXiv:2105.13245 [Cs, Stat]*, May 27, 2021. <http://arxiv.org/abs/2105.13245>.

[8] Gardner, Jacob R, Matt J Kusner, and Gardner Jake. “Bayesian Optimization with Inequality Constraints,” n.d., 9.