Bern Data Science Day 2021 - April 23

Contribution - Sample-based estimation of probability density fields: a spatial extension of the logistic Gaussian process

Topic - DS applications and challenges in Medicine, Natural Sciences, and Engineering
DS algorithms with a view towards Machine Learning and Artifical Intelligence

Author - Athénaïs Gautier,
PhD student in statistics (supervisor: Pr. D. Ginsbourger; SNSF project number 178858),
athenais.gautier@stat.unibe.ch

When studying complex systems, it is common for the response of interest to not be fully determined by the system parameters $\mathbf{x}$, but rather to be random and to follow a probability distribution $\{\mu_{\mathbf{x}},\ \mathbf{x} \in D\}$ that depends on $\mathbf{x}$.

We use the letter t to denote outputs. Our aim here is to estimate the field $\{\mu_{\mathbf{x}},\ \mathbf{x} \in D\}$ based only on a finite number of observations $(\mathbf{x_i}, t_i)_{1 \leq i \leq n}$ where the t_i's were independently sampled from the $\mu_{\mathbf{x}_i}$'s, respectively. Such settings are notably inspired by stochastic optimization and inversion problems, for which estimates of x and associated uncertainty quantification could be instrumental.

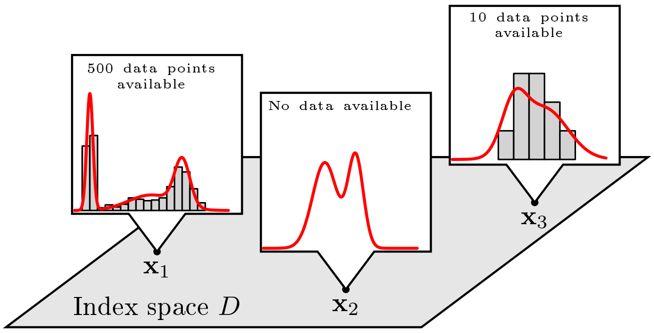
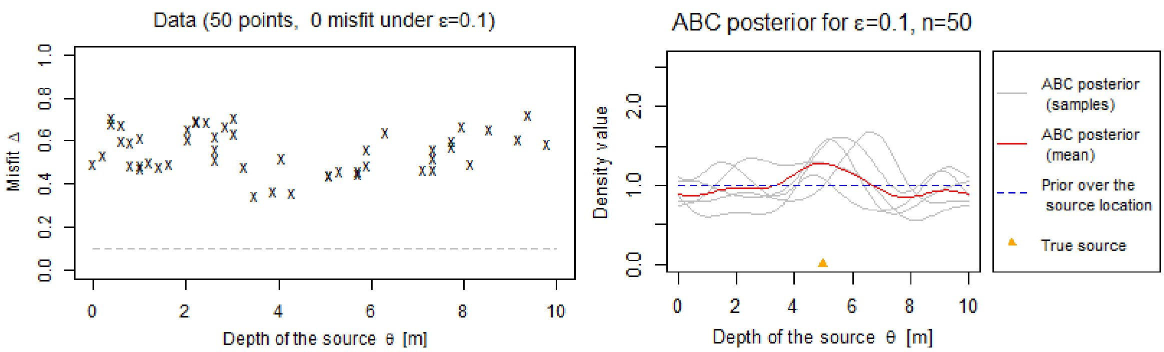


Figure 1: Typical setting: probability density of $\mu_{\mathbf{x}}$ (red curve) versus collected data (histogram).

The approach that we investigate here generalizes to spatial contexts a class of non-parametric Bayesian density models based on logistic Gaussian processes, and allows modelling density-valued fields with complex dependences of $\mu_{\mathbf{x}}$ on $\mathbf{x}$ while accommodating heterogeneous sample sizes. The induced prior on the space of density fields is called Spatial Logistic Gaussian Process (SLGP).

Our contributions build upon bayesian non parametric inferences on fields of probability density functions. The considered models allow for instance performing (approximate) posterior simulations of probability density functions as well as jointly predicting multiple moments or other functionals of target distributions.

We propose an implementation of the SLGP and investigate ways of using the proposed class of model to speed up Approximate Bayesian Computing (ABC) methods and further iterative algorithms involving decisions on points $\mathbf{x}$ where to run new stochastic simulations, be it for optimization or for inversion goals.



(a) Misfits between observed and simulated data. (b) Estimated ABC-posterior of the source depth.

Figure 2: SLGP in stochastic inverse problem: using 50 simulations to infer a contaminant source depth under uncertain geological structure (collab. with G. Pirot, Univ. of Western Australia).