Likelihood-free inference with deep Gaussian processes

Alexander Aushev, Henri Pesonen, Markus Heinonen, Jukka Corander, Samuel Kaski

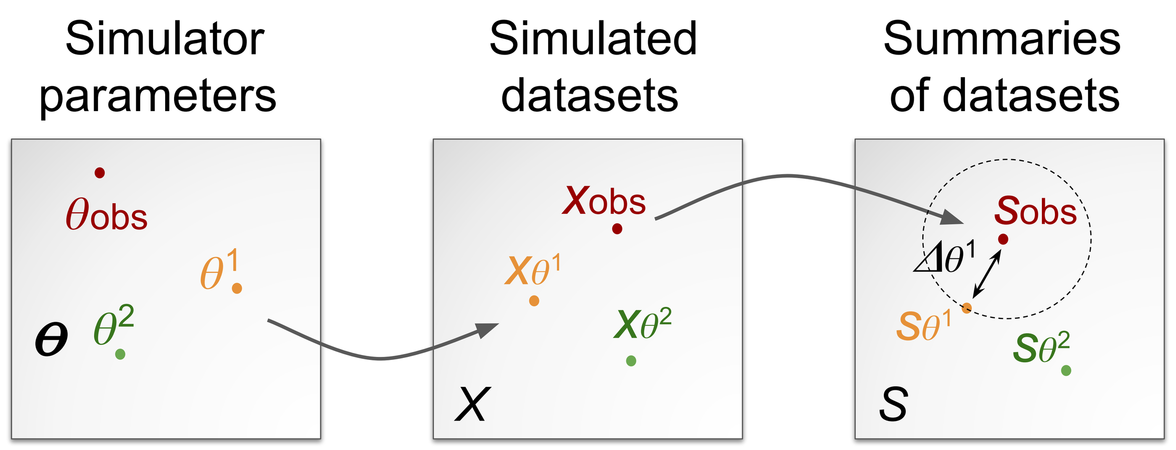
NeurIPS 2020 Workshop on Deep Learning and Inverse Problems, December 11

Aalto University

The University of Manchester

Introduction

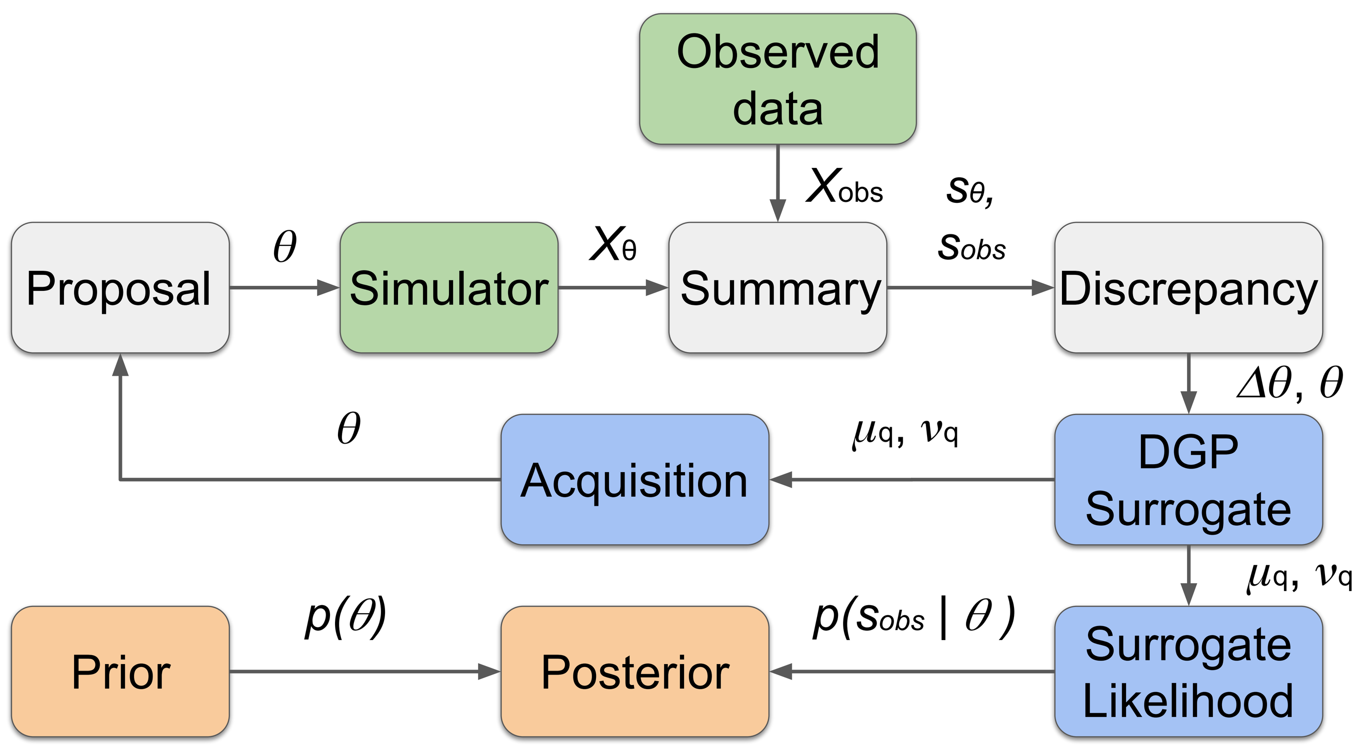
Given observed data X_{obs} and a simulator $g(\theta) = X$, we need to determine the parameter of the simulator θ that generated X_{obs}. Current likelihood-free inference (LFI) methods either can not approximate multimodal target distributions or require thousands of samples, which are rarely available when dealing with computationally expensive simulators.

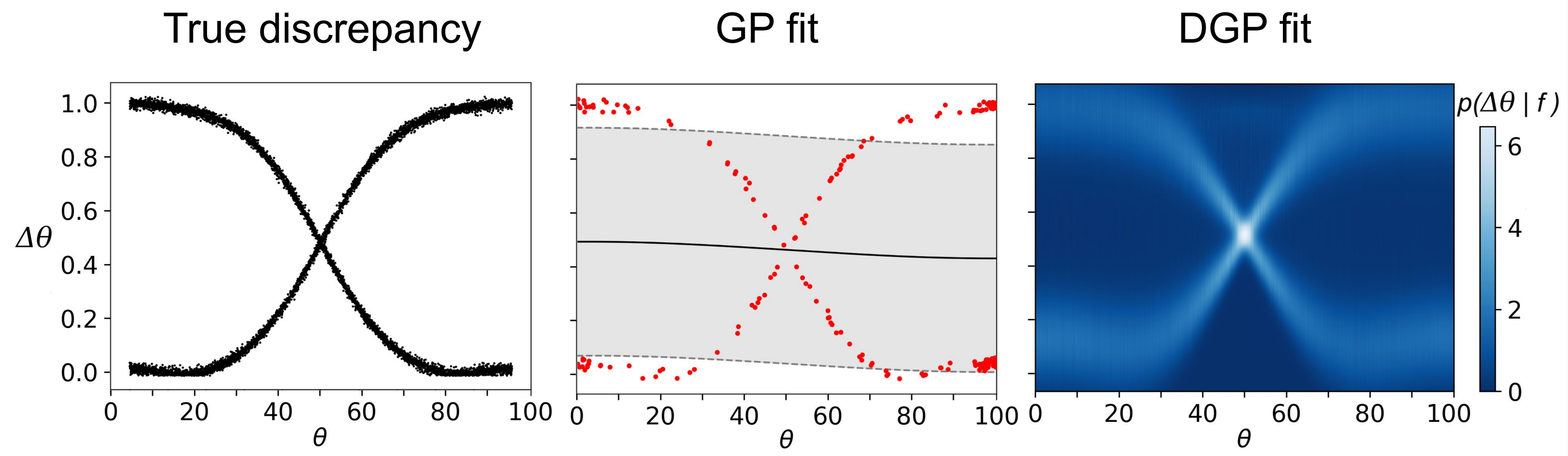


Assuming p(X|θ) ≈ p(S|θ)

Methods

We find $p(\theta|X_{obs})$ by minimizing the discrepancy Δ (e.g Euclidean distance) between summaries (e.g. number of clusters, variance) of the observed data s_{obs} and synthetic data s_θ. We use Bayesian Optimization (BO) for LFI [1], replacing a Gaussian Process (GP) surrogate with deep GPs. We use importance-weighted variational inference [2] for DGPs. The general overview of the approach:





We use quantile conditioning Q(.) on DGP samples:

$$\mu_q(\theta) = E[\Delta\theta_i : \Delta\theta_i \leq Q(q)]$$
$$\nu_q(\theta) = \mathrm{var}[\Delta\theta_i : \Delta\theta_i \leq Q(q)]$$

We use them in BO acquisition (η is a user-defined const)

$$A^t(\theta) = \mu_q(\theta) - \sqrt{\eta_t^2 \cdot \nu_q(\theta)}$$
$$\theta^{t+1} = \mathrm{argmin}_\theta \{A^t(\theta)\}$$

And likelihood approximation (ϵ is a discrepancy threshold)

$$p(s_{obs}|\theta) \propto F\left(\frac{\epsilon - \mu_q(\theta)}{\sqrt{\nu_q(\theta) + \sigma^2}}\right)$$

where F is a normal cdf with mean 0 and variance 1, and σ is a Gaussian likelihood noise. Quantile-conditioning filters DGPs samples, retaining only those that correspond to the low-valued discrepancy regions, where the discrepancy model is the most accurate.

Experiments

DGPs outperform GPs, masked autoregressive flows (MAFs) and mixture density networks (MDNs) in terms of scaled Wasserstein distance between the surrogate posterior and the true posterior for multimodal simulator TE2 and NW (the lower distance is the better).

TE2 simulator is a 1d demonstration of multimodality, shown in the figure above, **BDM** is a unimodal Gaussian-like simulator of tuberculosis spread [3] and **NW** is a multimodal simulation of a reinforcement learning agent in the grid-world planning environment.

The 95%-confidence intervals of the Wasserstein distance for each model are shown below:

Model	TE2	BDM	NW
LV-GP	**(1.6, 1.64)**	(1.51, 1.61)	**(1.24, 1.29)**
LV-3GP	(1.7, 1.74)	(1.5, 1.6)	(1.26, 1.29)
GP	(2.65, 2.68)	**(1.23, 1.25)**	(1.67, 1.7)
MAF	(1.99, 2.02)	(2.03, 2.16)	(2.37, 2.5)
MDN	(15.63, 18.16)	(1.38, 1.4)	(1.8, 1.83)

Conclusions

- Non-Gaussian uncertainties pose a problem for likelihood-free inference with computationally expensive simulators;
- DGPs are able to model irregular distributions, allowing accurate representation of the uncertainty in BO;
- We introduce quantile conditioning on DGP samples to handle the acquisition and likelihood approximation in multimodal cases;
- Our experiments show that DGPs outperform GPs on multimodal cases and retain comparable performance and sample-efficiency on the rest of the cases;
- DGPs in BO outperform deep learning alternatives on tasks with only hundreds simulator calls available.

References

1. Gutmann MU, Corander J. Bayesian optimization for likelihood-free inference of simulator-based statistical models. The Journal of Machine Learning Research. 2016 Jan 1;17(1):4256-302.
2. Salimbeni H, Dutordoir V, Hensman J, Deisenroth MP. Deep gaussian processes with importance-weighted variational inference. arXiv preprint arXiv:1905.05435. 2019 May 14.
3. Lintusaari J, Blomstedt P, Rose B, Sivula T, Gutmann MU, Kaski S, Corander J. Resolving outbreak dynamics using approximate Bayesian computation for stochastic birth–death models. Wellcome Open Research. 2019 Aug 30;4(14):14.