Batch simulations and uncertainty quantification in Gaussian process surrogate ABC

Marko Järvenpää[1]* **Aki Vehtari**[1] **Pekka Marttinen**[1]

[1]Aalto University, Finland *Current address: University of Oslo, Norway; m.j.jarvenpaa@medisin.uio.no

Aalto University School of Science

1 Background on ABC inference

- **Approximate Bayesian computation (ABC)** is used for Bayesian inference when the likelihood function $\pi(\mathbf{x}|\boldsymbol{\theta})$ is intractable but simulating pseudo-data $\mathbf{x} \in \mathcal{X}$ is feasible.
- The intractable posterior $\pi(\boldsymbol{\theta}|\mathbf{x}_o)$ is approximated with

$$\pi_{\mathrm{ABC}}(\boldsymbol{\theta}|\mathbf{x}_o) \triangleq \frac{\pi(\boldsymbol{\theta})\int_{\mathcal{X}}\pi_{\varepsilon}(\mathbf{x}_o|\mathbf{x})\pi(\mathbf{x}|\boldsymbol{\theta})\,\mathrm{d}\mathbf{x}}{\int_{\Theta}\pi(\boldsymbol{\theta}')\int_{\mathcal{X}}\pi_{\varepsilon}(\mathbf{x}_o|\mathbf{x}')\pi(\mathbf{x}'|\boldsymbol{\theta}')\,\mathrm{d}\mathbf{x}'\,\mathrm{d}\boldsymbol{\theta}'}, \quad (1)$$

where $\pi_{\varepsilon}(\mathbf{x}_o\,|\,\mathbf{x}) = \mathbb{1}_{\Delta(\mathbf{x}_o,\mathbf{x})\leq\varepsilon}$, $\Delta : \mathcal{X}^2 \to \mathbb{R}_+$ is a discrepancy function and ε is a threshold parameter.
- Common sampling-based ABC methods targeting (1) require huge number of simulations and cannot be used when **simulations are costly**.
- Recently various approaches using neural networks and Gaussian process (GP) **surrogate models** have been proposed to tackle the computational challenges.

2 Contributions

- Reformulation of earlier GP-surrogate based approaches [1, 2, 3] into a coherent framework called **'Bayesian ABC'**.
- **Batch Bayesian experimental design (active learning)** strategies for the parallelisation of the potentially expensive simulations in this framework.
- An approach to **quantify the uncertainty** in the moments and marginals of the ABC posterior.
- Further analysis on the connections between the related problems of **Bayesian ABC**, **Bayesian quadrature** and **Bayesian optimisation**.
- Extensive experiments with several toy and real-world simulation models with intractable likelihoods.

3 Bayesian ABC framework

- **Bayesian ABC** uses another layer of Bayesian inference to estimate the ABC posterior in (1). The previously simulated discrepancy-parameter-pairs $D_t = \{(\Delta_i, \boldsymbol{\theta}_i)\}_{i=1}^t$ are treated as 'data' to learn a **GP surrogate model**, which will predict the discrepancy for a given parameter value. The GP model is used to form an estimator for the ABC posterior (1) and to **adaptively acquire new data**.
- Each discrepancy evaluation Δ_i at corresponding parameter $\boldsymbol{\theta}_i$ is assumed to be generated as

$$\Delta_i = f(\boldsymbol{\theta}_i) + \nu_i, \quad \nu_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_n^2), \quad (2)$$

where $\sigma_n^2 > 0$ is the variance of the discrepancy.
- To encode the assumptions of e.g. smoothness and quadratic shape of $\Delta_{\boldsymbol{\theta}}$, f is given a GP prior.
- If f and σ_n^2 were known, the ABC posterior could be obtained from (1) and (2) as

$$\pi^f_{\mathrm{ABC}}(\boldsymbol{\theta}) \triangleq \frac{\pi(\boldsymbol{\theta})\Phi\left((\varepsilon - f(\boldsymbol{\theta}))/\sigma_n\right)}{\int_{\Theta}\pi(\boldsymbol{\theta}')\Phi\left((\varepsilon - f(\boldsymbol{\theta}'))/\sigma_n\right)\mathrm{d}\boldsymbol{\theta}'}, \quad (3)$$

- Computing the distribution of π^f_{ABC} is difficult due to its non-linear dependence on infinite-dimensional quantity f. However, various statistics of $\tilde{\pi}^f_{\mathrm{ABC}}(\boldsymbol{\theta}) \triangleq \pi(\boldsymbol{\theta})\Phi((\varepsilon - f(\boldsymbol{\theta}))/\sigma_n)$, the numerator of (3), can be computed analytically [2] and combined with MCMC.
- Given D_t, our knowledge about f is $f \sim \Pi^f_{D_t} \triangleq \mathcal{GP}(f; m_t, c_t)$. The posterior of π^f_{ABC} in (3) describes the amount of uncertainty in π^f_{ABC} due to the limited t simulations and is obtained as the push-forward measure through the mapping $f \mapsto \pi^f_{\mathrm{ABC}}$.
- We developed a numerical method to characterise this uncertainty based on normalised importance sampling and GP sample paths. For illustration, see Fig. 1c-d and 5.

Bayesian ABC vs. Bayesian quadrature vs. Bayesian optimisation

- **Bayesian ABC**: 'estimate' π^f_{ABC} in (3) (or $\tilde{\pi}^f_{\mathrm{ABC}}$). Data: $D_t = \{(\Delta_i, \boldsymbol{\theta}_i)\}_{i=1}^t$. Model: $\Delta_i \sim \mathcal{N}(f(\boldsymbol{\theta}_i), \sigma_n^2)$, $f \sim \mathcal{GP}$, typically $\sigma_n \gg 0$.
- **Bayesian quadrature**: 'estimate' $I \triangleq \int g[f(\boldsymbol{\theta})]\pi(\boldsymbol{\theta})\,\mathrm{d}\boldsymbol{\theta}$ where usually $g[f(\boldsymbol{\theta})] = f(\boldsymbol{\theta})$. Data: $\{(y_i, \boldsymbol{\theta}_i)\}_{i=1}^t$. Model: $y_i \sim \mathcal{N}(f(\boldsymbol{\theta}_i), \sigma_n^2)$, $f \sim \mathcal{GP}$, typically $\sigma_n \simeq 0$.
- **BayesOpt**: 'estimate' $\boldsymbol{\theta}^\star \triangleq \arg\min_{\boldsymbol{\theta}} f(\boldsymbol{\theta})$. Data: $\{(y_i, \boldsymbol{\theta}_i)\}_{i=1}^t$. Model: $y_i \sim \mathcal{N}(f(\boldsymbol{\theta}_i), \sigma_n^2)$, $f \sim \mathcal{GP}$, typically $\sigma_n \simeq 0$.

References

[1] M. U. Gutmann and J. Corander. Bayesian optimization for likelihood-free inference of simulator-based statistical models. *Journal of Machine Learning Research*, 17(125):1–47, 2016.

[2] M. Järvenpää, M. U. Gutmann, A. Pleska, A. Vehtari, and P. Marttinen. Efficient acquisition rules for model-based approximate Bayesian computation. *Bayesian Analysis*, 14(2):595–622, 2019.

[3] M. Järvenpää, M. U. Gutmann, A. Vehtari, and P. Marttinen. Parallel Gaussian process surrogate Bayesian inference with noisy likelihood evaluations. Bayesian analysis, to appear, 2020.

[4] O. Thomas, R. Dutta, J. Corander, S. Kaski, and M. U. Gutmann. Likelihood-free inference by ratio estimation. arXiv1611.10242, 2018.

[5] E. Numminen, L. Cheng, M. Gyllenberg, and J. Corander. Estimating the transmission dynamics of streptococcus pneumoniae from strain prevalence data. *Biometrics*, 69(3):748–757, 2013.

4 Batch acquisition functions

- Need to find **informative simulation locations** for estimating π^f_{ABC} in (3) given the postulated GP model.
- Loss function $l(\pi_{\mathrm{ABC}}, d)$ quantifies the penalty of reporting $d \in \mathscr{D}$ as our ABC posterior when the true one is $\pi_{\mathrm{ABC}} \in \mathscr{D}$. Given D_t, the one-batch-ahead **Bayes-optimal** selection of the **next batch** $\boldsymbol{\theta}^{\mathrm{opt}} = [\boldsymbol{\theta}^{\mathrm{opt}}_1, \ldots, \boldsymbol{\theta}^{\mathrm{opt}}_b]$ is $\boldsymbol{\theta}^{\mathrm{opt}} = \arg\min_{\boldsymbol{\theta}^* \in \Theta^b} L_t(\boldsymbol{\theta}^*)$, where

$$L_t(\boldsymbol{\theta}^*) = \mathbb{E}_{\boldsymbol{\Delta}^*|\boldsymbol{\theta}^*, D_t} \min_{d \in \mathscr{D}} \mathbb{E}_{f\,|\,D_t \cup D^*} l(\pi^f_{\mathrm{ABC}}, d). \quad (4)$$

- If we use L^2 loss function $\tilde{l}_2 \triangleq \int_{\Theta}(\tilde{\pi}^f_{\mathrm{ABC}}(\boldsymbol{\theta}) - \tilde{d}(\boldsymbol{\theta}))^2\,\mathrm{d}\boldsymbol{\theta}$, then the optimal estimator for $\tilde{\pi}^f_{\mathrm{ABC}}$ is the mean and the resulting **expected integrated variance (EIV) acquisition function**, denoted as $L^{\mathrm{v}}_t(\boldsymbol{\theta}^*)$, is

$$L^{\mathrm{v}}_t(\boldsymbol{\theta}^*) = 2\int_{\Theta}\pi^2(\boldsymbol{\theta})\left[T\left(a_t(\boldsymbol{\theta}), \frac{\sqrt{\sigma_n^2 + s_t^2(\boldsymbol{\theta}) - \tau_t^2(\boldsymbol{\theta};\boldsymbol{\theta}^*)}}{\sqrt{\sigma_n^2 + s_t^2(\boldsymbol{\theta}) + \tau_t^2(\boldsymbol{\theta};\boldsymbol{\theta}^*)}}\right) - T\left(a_t(\boldsymbol{\theta}), \sigma_n/\sqrt{\sigma_n^2 + 2s_t^2(\boldsymbol{\theta})}\right)\right]\mathrm{d}\boldsymbol{\theta}, \quad (5)$$

where $a_t(\boldsymbol{\theta}) \triangleq (\varepsilon - m_t(\boldsymbol{\theta}))/\sqrt{\sigma_n^2 + s_t^2(\boldsymbol{\theta})}$, $s_t^2(\boldsymbol{\theta}) = c_t(\boldsymbol{\theta}, \boldsymbol{\theta})$, $\tau_t^2(\boldsymbol{\theta};\boldsymbol{\theta}^*) = c_t(\boldsymbol{\theta}, \boldsymbol{\theta}^*)[c_t(\boldsymbol{\theta}^*, \boldsymbol{\theta}^*) + \sigma_n^2\mathbf{I}]^{-1}c_t(\boldsymbol{\theta}^*, \boldsymbol{\theta})$ and T is Owen's T function.
- Similarly, L^1 loss function produces the marginal median as the optimal estimator for $\tilde{\pi}^f_{\mathrm{ABC}}$ and results **expected integrated MAD (EIMAD) acquisition function**.
- We use **greedy optimisation** and the integral over Θ is approximated using **importance sampling**.
- Heuristically-motivated batch methods (MAXV, MAXMAD) based on **uncertainty sampling** and random strategy RAND are used as baselines.

Illustration of the main idea

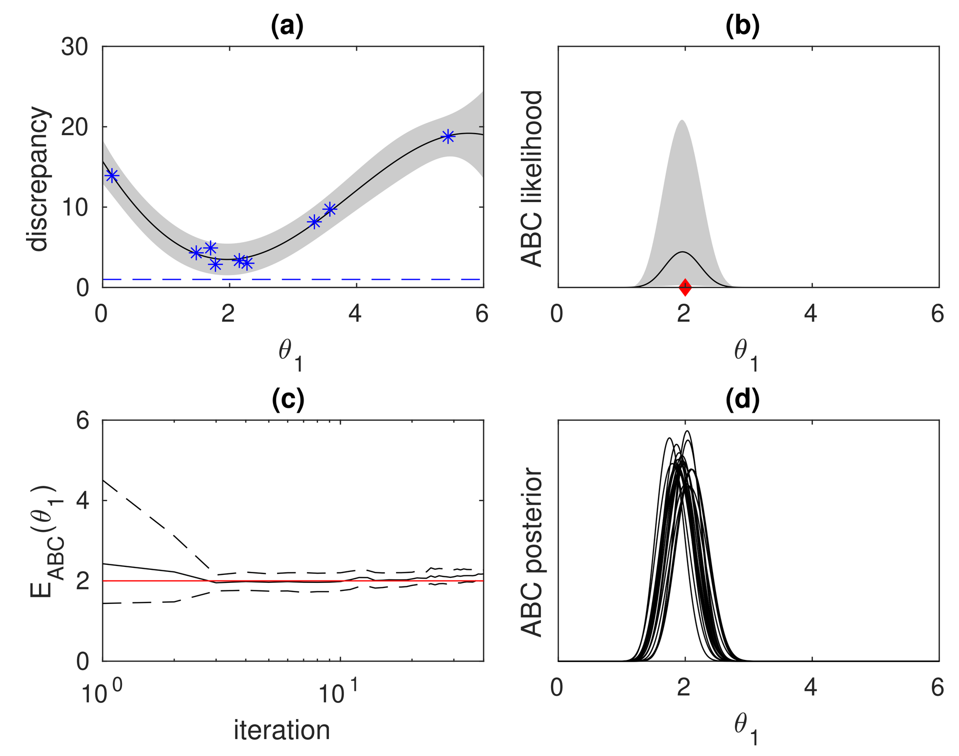

Figure 1: ABC posterior uncertainty quantification using Lorenz model with parameter θ_2 fixed. (a) GP for Δ_{θ_1} (blue dashed line ε, blue stars 9 discrepancy evaluations), (b) uncertainty of unnormalised ABC posterior $\tilde{\pi}^f_{\mathrm{ABC}}$, (c) evolution of GP-based ABC posterior expectation (black line) and its 95% CI (dashed black) for 40 iterations, (d) uncertainty of ABC posterior π^f_{ABC}.

5 Main results

- Experiments show that ABC posterior uncertainty quantification approach is useful in low dimensions and that **Bayesian ABC framework is well-suited for parallel simulations**.
- EIV and EIMAD, both based on the same Bayesian decision theoretic framework but different loss functions, performed similarly and better than the baselines.
- For full details and experiments, see **our full paper**.

Results with 2D synthetic simulation models

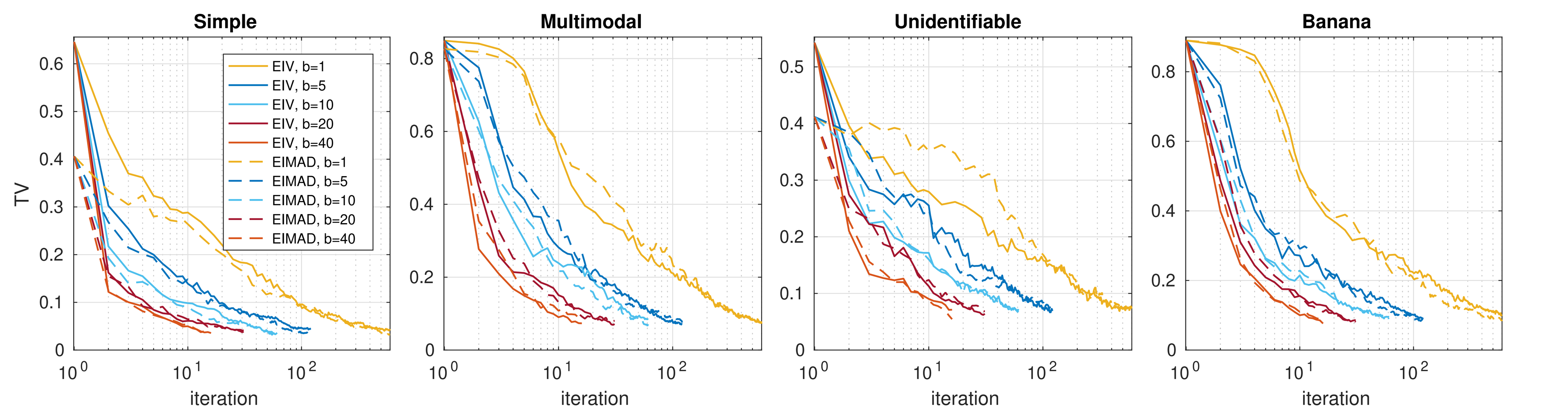

Figure 2: Four 2D toy simulation models using greedy EIV and EIMAD acq. functions and various batch sizes b.

Real-world test cases: Lorenz and bacterial infections models

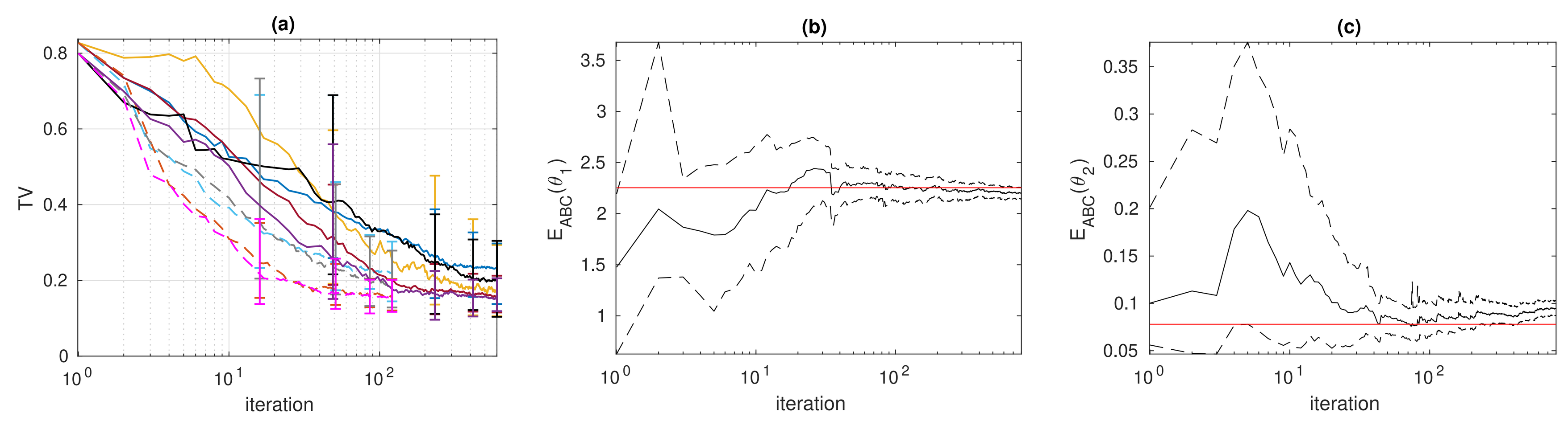

Figure 3: (a) Lorenz model [4, 3]. The intervals show the 90% variability. See Fig. 4a for the legend. (b-c) Black line is the mean and dashed black the 95% CI of the ABC posterior expectations. Red line shows the true value.

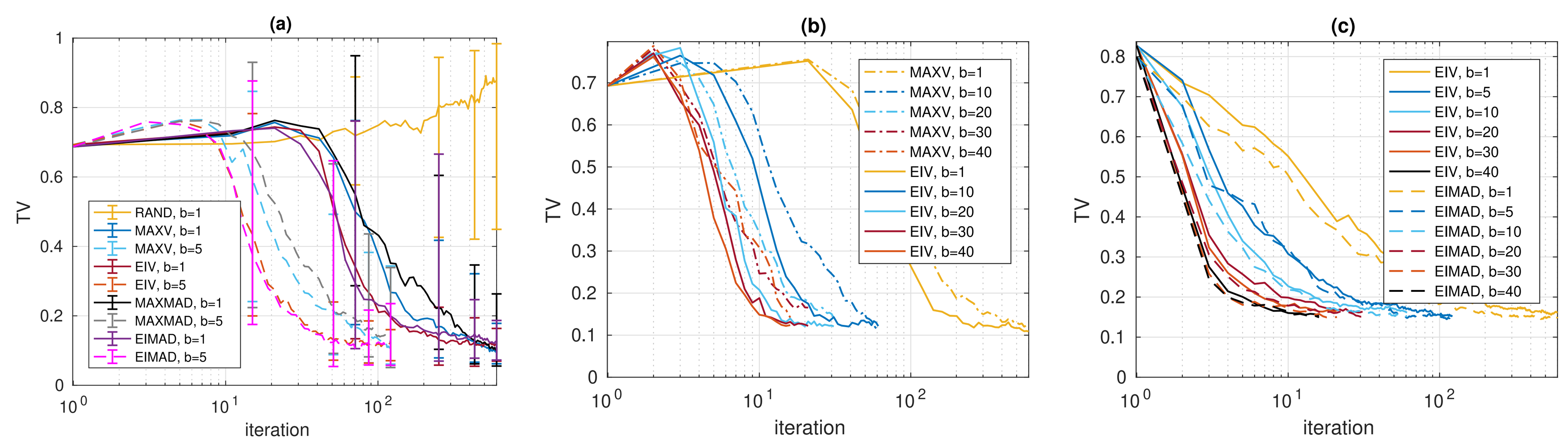

Figure 4: (a) Bacterial infections model [5]. The intervals show the 90% variability. (b) Bacterial infections model with different batch sizes and two chosen acquisition methods. (c) Additional experiments with Lorenz model.

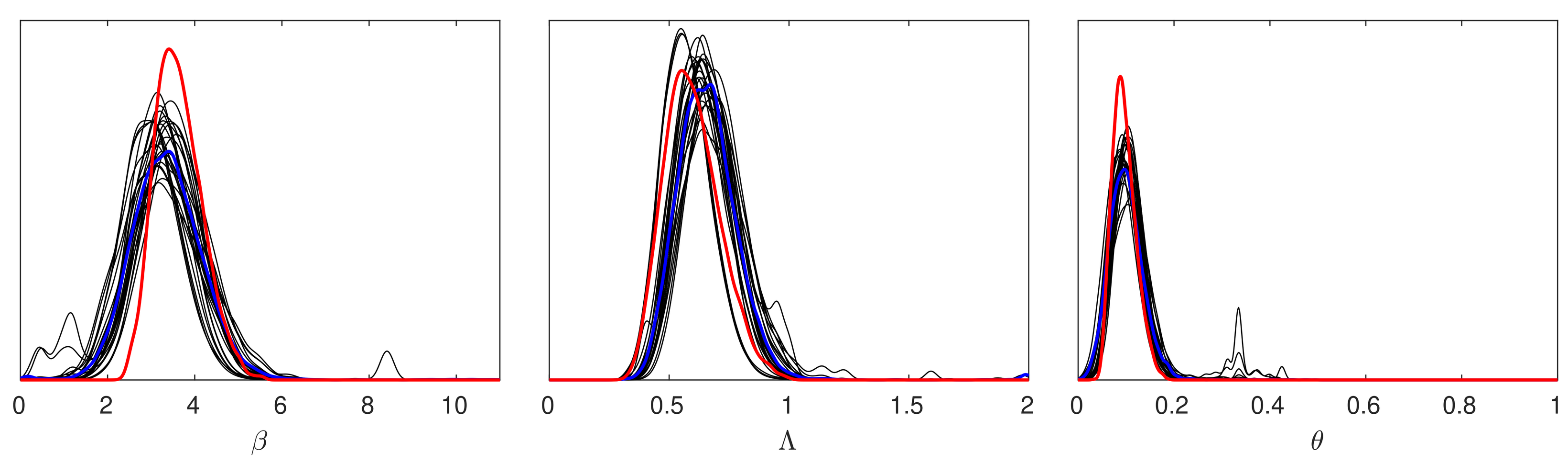

Figure 5: Uncertainty quantification for the ABC posterior marginals of the bacterial infections model after $t = 120$ simulations. Red line: true ABC posterior, blue line: the mean-based point estimate, black lines: sampled ABC marginal posteriors to (approximately) represent the uncertainty due to the limited number of simulations t.