HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Efficient and accurate approximate Bayesian computation

Marko Järvenpää, Michael Gutmann, Arijus Pleska, Aki Vehtari, Pekka Marttinen

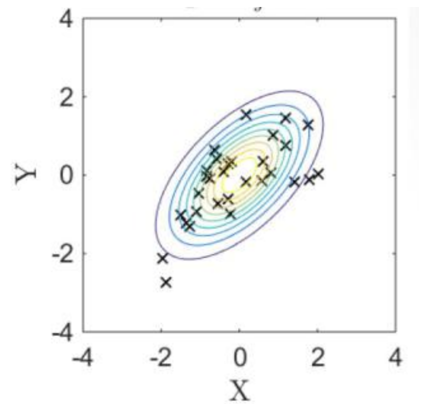
Aalto University

UNIVERSITY OF HELSINKI

Outline

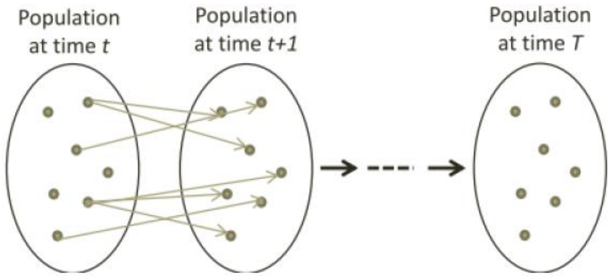
- Simulation-based modeling
- Model-based approximate Bayesian computation (ABC)
- Efficient ABC using Bayesian sequential experimental design
- Results
- Conclusion

Simulation-based modeling



- Statistical inference, the common way
  - Assume some likelihood: *p(data|parameters)*
  - Learn *parameters* that fit the *data*

- Sometimes the likelihood can not be computed, but simulating data from the model is possible
  - Example: population genetics

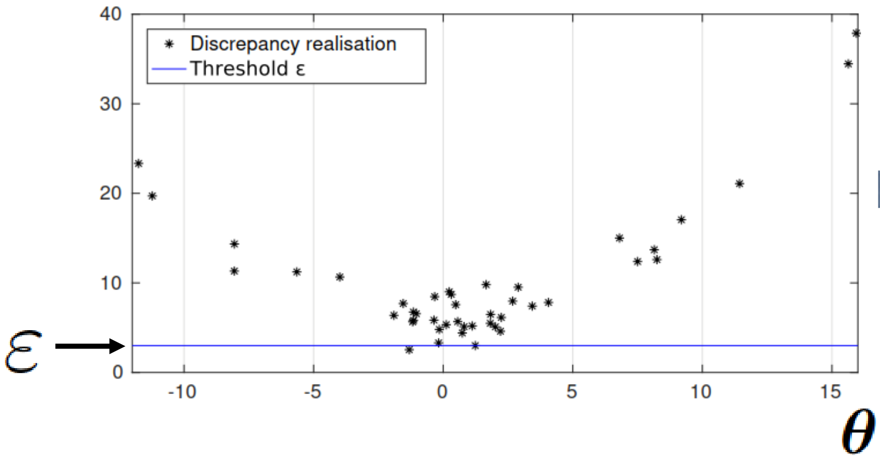


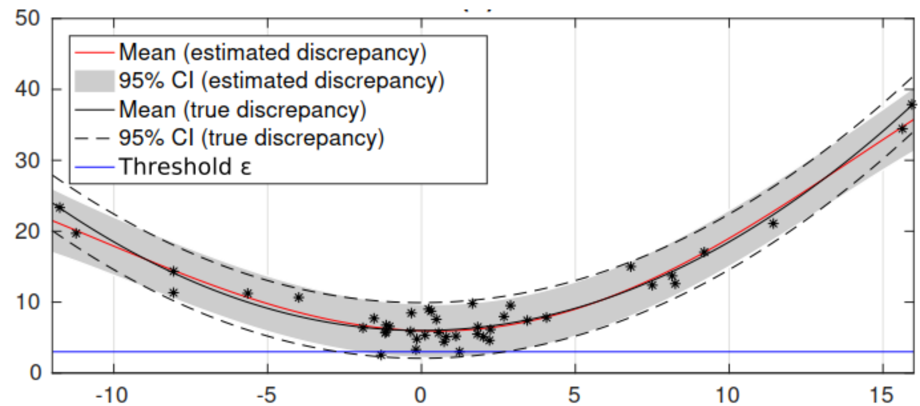
- Applications: economics, material physics, biology, UI design, ...

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Model-based ABC

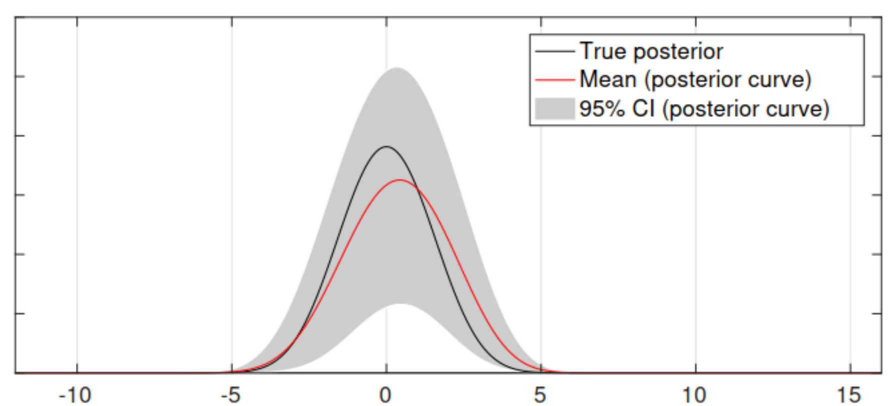
$$\Delta_{\boldsymbol{\theta}} = \Delta(\mathbf{x}_{obs}, \mathbf{x}_{\boldsymbol{\theta}})$$





$$\tilde{\pi}_{\mathrm{ABC}}(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})p_a(\boldsymbol{\theta})$$

$$p_a(\boldsymbol{\theta}) = \hat{\mathbb{P}}(\Delta_{\boldsymbol{\theta}} \leq \varepsilon)$$



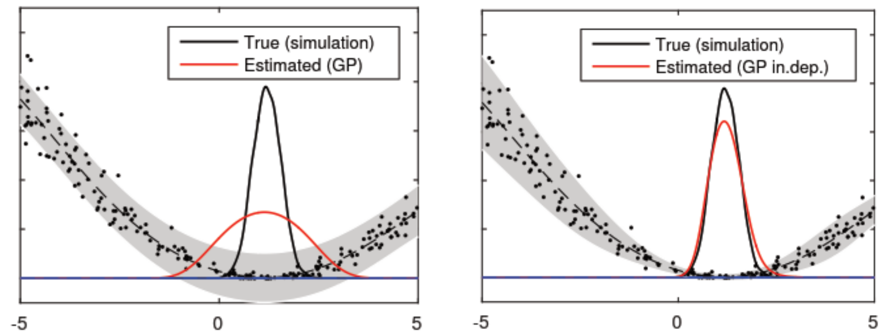
see e.g. Wilkinson (2014); Gutmann and Corander (2016);Järvenpää et al. (2017a)

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Efficient and accurate ABC

- How to model the discrepancies

  Järvenpää et al. (2017a), submitted.

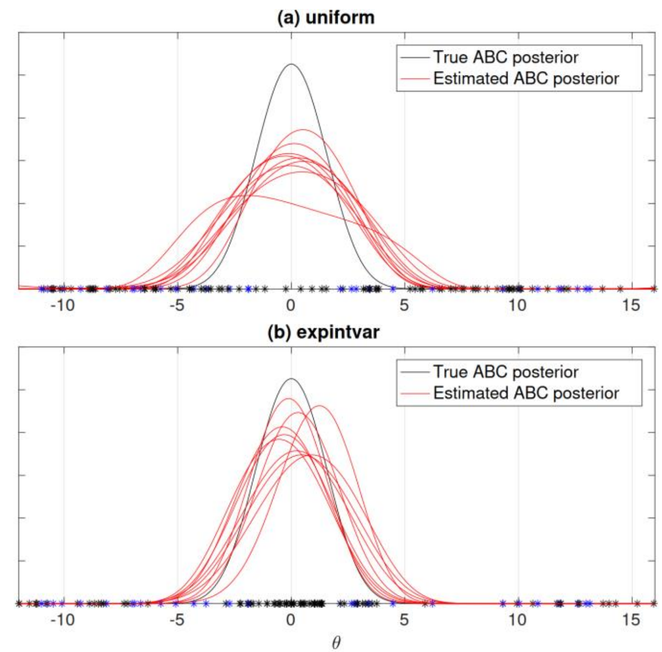


- Where to simulate the model

  Järvenpää et al. (2017a), submitted.

- Implementation: ‘ELFI’

  Lintusaari, et al. (2017), submitted.



HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Where to simulate next?

- Bayesian experimental design
- Simulate with θ^* that minimizes the expected loss

Data simulated with θ^*

Next parameter to simulate

$$\mathbb{E}_{\mathbf{x}^* \mid \boldsymbol{\theta}^*, D_{1:t}}(\mathcal{L}(\pi(\pi_{\mathrm{ABC}} \mid \mathbf{x}^*, \boldsymbol{\theta}^*, D_{1:t})))$$

Estimated ABC posterior

Training data: $\{(\mathbf{x}_i, \boldsymbol{\theta}_i)\}_{i=1}^t$

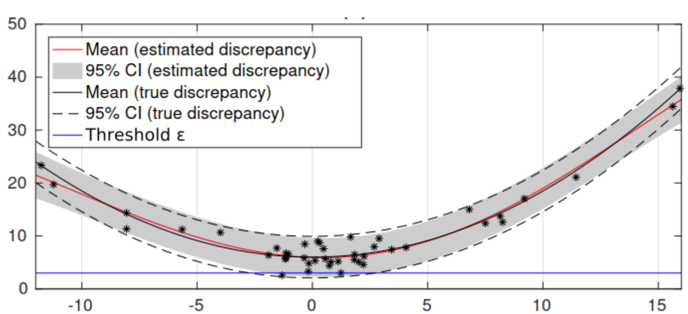
HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

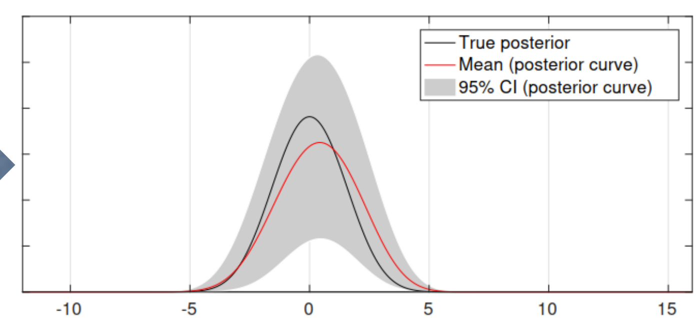
Expected loss

Model for posterior pdf, follows from: $\Delta_{\boldsymbol{\theta}} \sim \mathcal{N}(f(\boldsymbol{\theta}), \sigma_n^2)$
$f \sim \mathcal{GP}$

$$\mathbb{E}_{\mathbf{x}^* \mid \boldsymbol{\theta}^*, D_{1:t}}(\mathcal{L}(\pi(\pi_{\mathrm{ABC}} \mid \mathbf{x}^*, \boldsymbol{\theta}^*, D_{1:t})))$$

Integrated variance: $\int_{\Theta} \mathbb{V}(\tilde{\pi}_{\mathrm{ABC}}(\boldsymbol{\theta}) \mid D_{1:t})\, \mathrm{d}\boldsymbol{\theta}$





HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

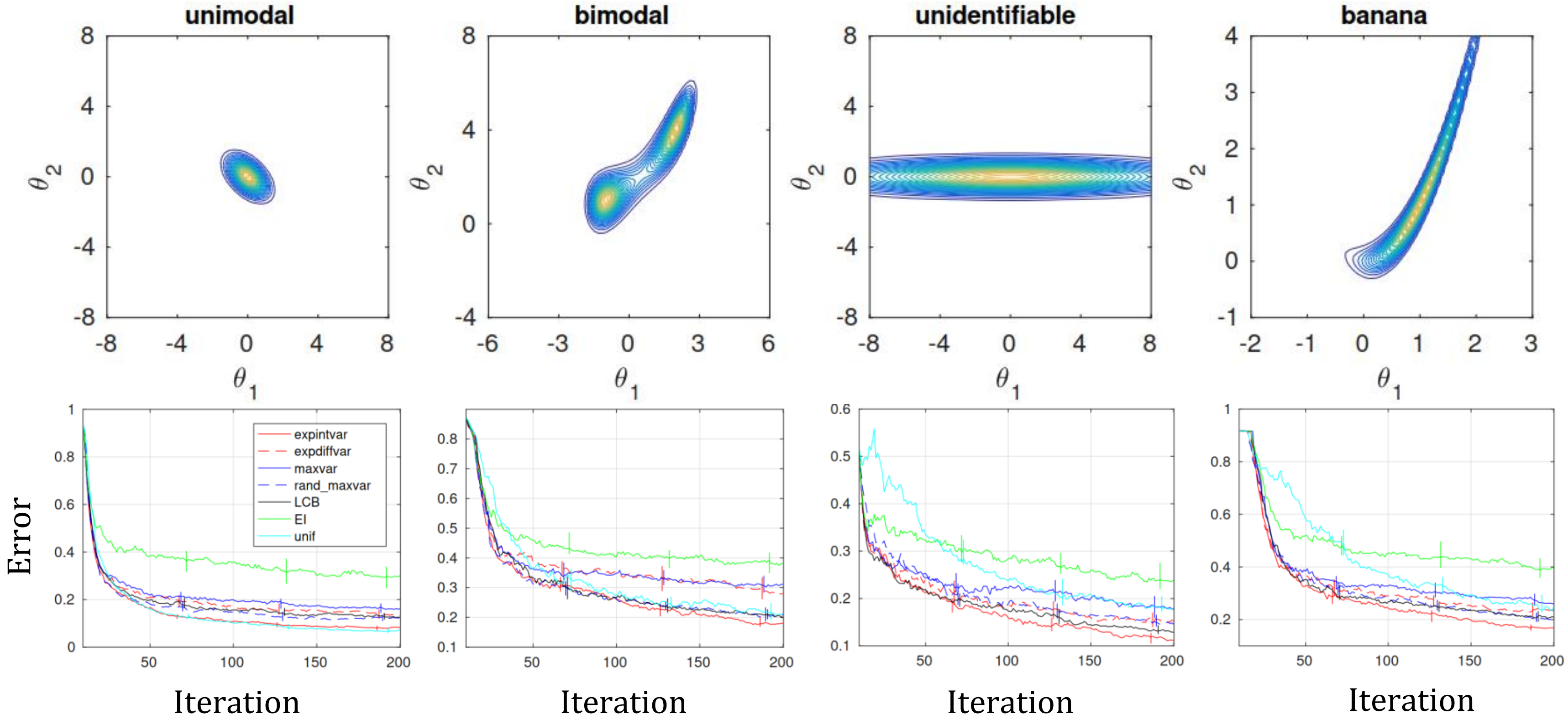
Loss optimization

$$L_{1:t}(\boldsymbol{\theta}^*) = 2\int_{\Theta} \pi^2(\boldsymbol{\theta}) \left[T\left(\frac{\varepsilon - m_{1:t}(\boldsymbol{\theta})}{\sqrt{\sigma_n^2 + v_{1:t}^2(\boldsymbol{\theta})}}, \sqrt{\frac{\sigma_n^2 + v_{1:t}^2(\boldsymbol{\theta}) - \tau_{1:t}^2(\boldsymbol{\theta}, \boldsymbol{\theta}^*)}{\sigma_n^2 + v_{1:t}^2(\boldsymbol{\theta}) + \tau_{1:t}^2(\boldsymbol{\theta}, \boldsymbol{\theta}^*)}} \right) - T\left(\frac{\varepsilon - m_{1:t}(\boldsymbol{\theta})}{\sqrt{\sigma_n^2 + v_{1:t}^2(\boldsymbol{\theta})}}, \frac{\sigma_n}{\sqrt{\sigma_n^2 + 2v_{1:t}^2(\boldsymbol{\theta}))}} \right) \right] \mathrm{d}\boldsymbol{\theta},$$

- For integration -> Importance sampling
- For optimization -> Gradient descent

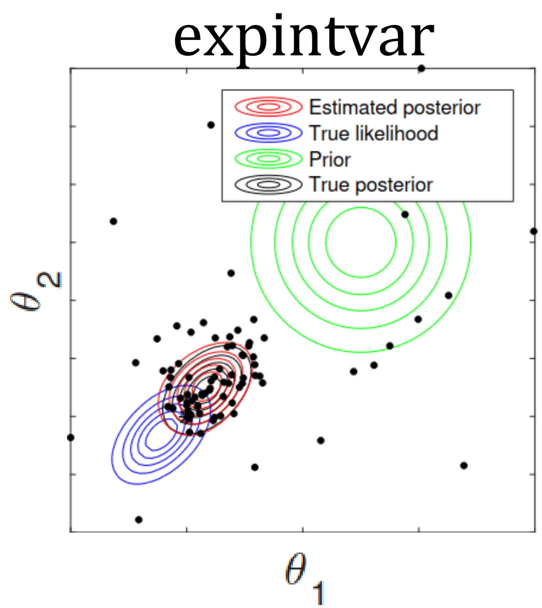
HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

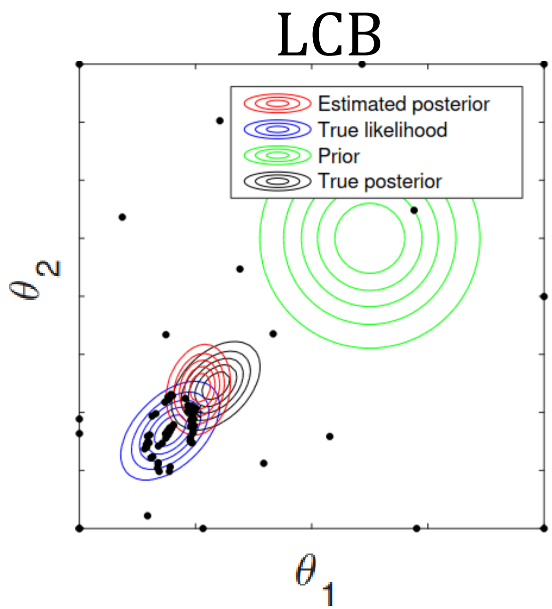
Simulations

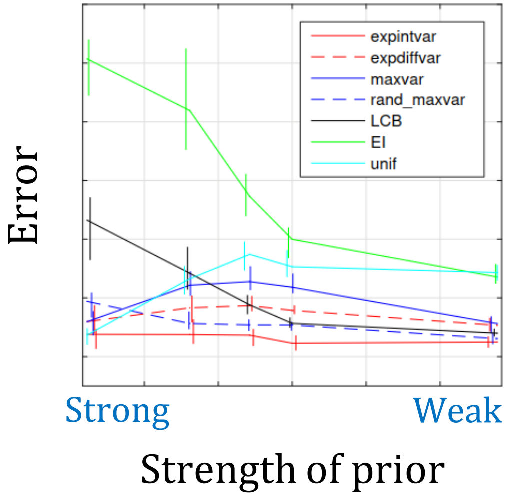


HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

expintvar vs. LCB

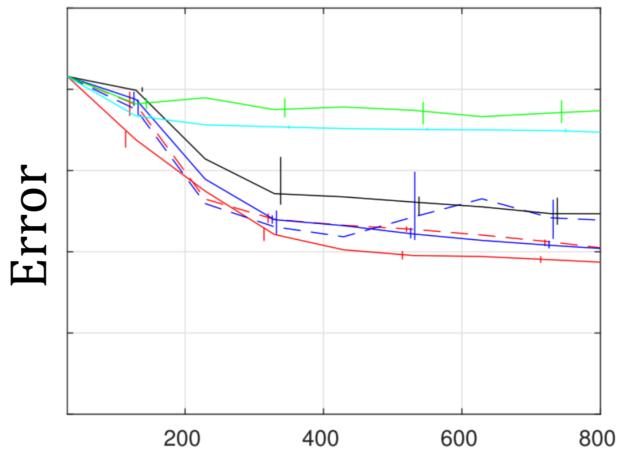


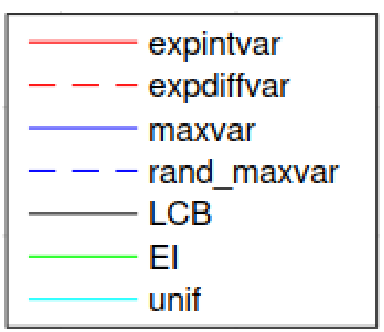




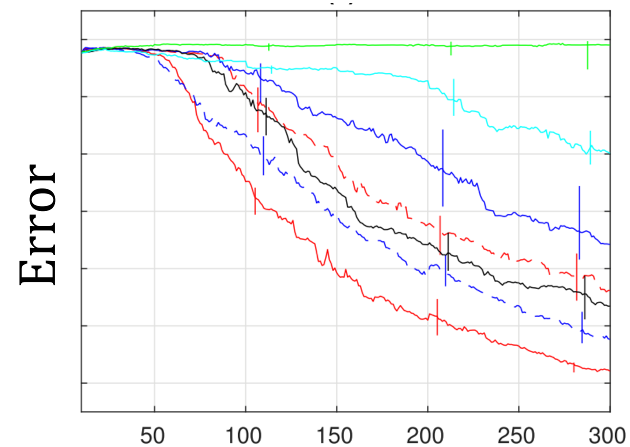
Other examples

10D Gaussian

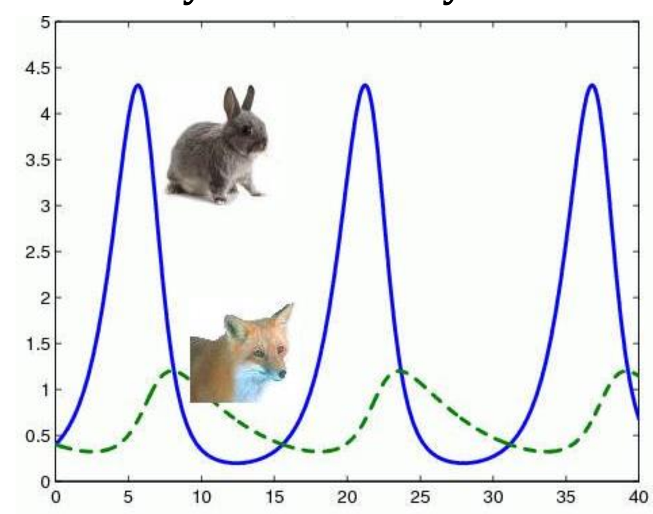




Lotka-Volterra



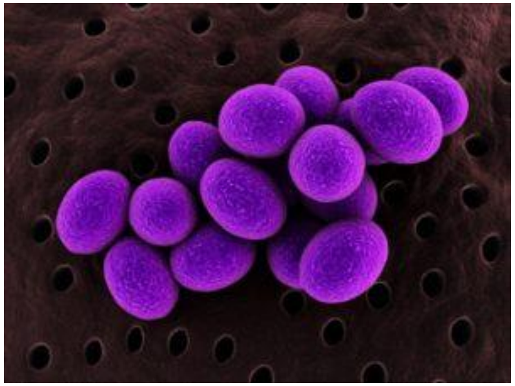
Prey-Predator Cycles



HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

Conclusion

- New principled approach for efficient model-based ABC
- Overall best performance in the experiments, especially when
  - dimension is high
  - there is strong prior information available
- Future:
  - Multi-point proposals for parallelization
  - Models tailored for novel applications



HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY

References

Gutmann and Corander (2016). Bayesian optimization for likelihood-free inference of simulator-based statistical models, JMLR.

Järvenpää, Gutmann, Vehtari, Marttinen (2017a). Gaussian process modeling in approximate Bayesian computation to estimate horizontal gene transfer in bacteria , submitted. (arxiv.org/pdf/1610.06462.pdf)

Järvenpää, Gutmann, Pleska, Vehtari, Marttinen (2017b). Efficient acquisition rules for approximate Bayesian computation, submitted. (arxiv.org/pdf/1704.00520.pdf)

Lintusaari, Vuollekoski, Kangasrääsiö, Skytén, Järvenpää, et al. (2017). ELFI: Engine for likelihood-free inference, submitted. (arxiv.org/pdf/1708.00707.pdf)

Wilkinson (2014). Accelerating ABC methods using Gaussian processes. In AISTATS.

HELSINKI INSTITUTE FOR INFORMATION TECHNOLOGY