

# Gaussian Process Surrogate Methods for Sample-Efficient Approximate Bayesian Computation

Gaussin prosessi -surrogaattimenetelmiä likimääräiseen  
Bayesilaiseen päättelyyn

Marko Järvenpää  
Department of Computer Science,  
Aalto University

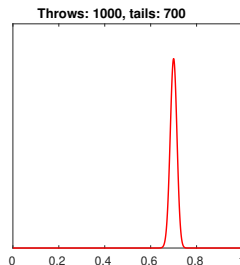
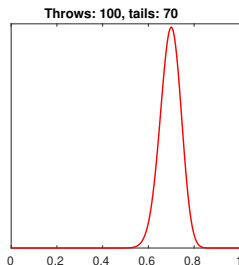
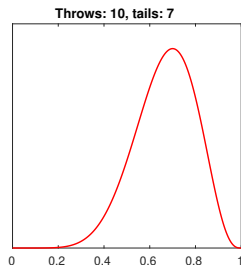
Lectio praecursoria, 25.9.2020

# Statistical inference

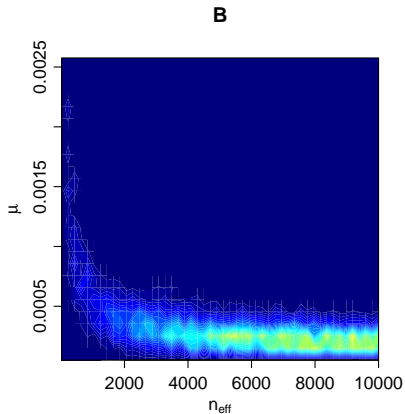
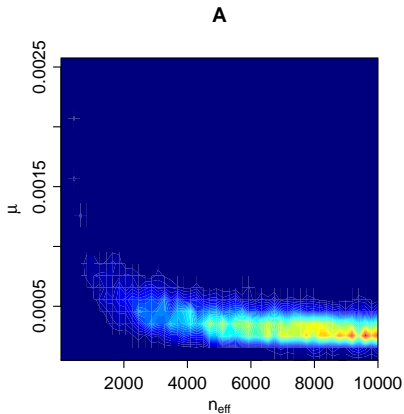
- ▶ **Probability:** e.g. what is the probability to obtain 7 tails if a coin is thrown 10 times?
- ▶ **Statistical inference:** e.g. if we obtain 7 tails and 3 heads, what can we say about the coin?  
What if we instead obtain 70 tails and 30 heads?

# Bayesian inference 1/2

- ▶ **Posterior distribution** tells what we know about some unknown parameters based on prior information and measurements
- ▶ Posterior distribution  $\propto$  prior distribution  $\times$  likelihood function



# Bayesian inference 2/2

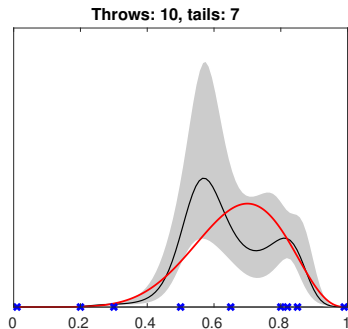
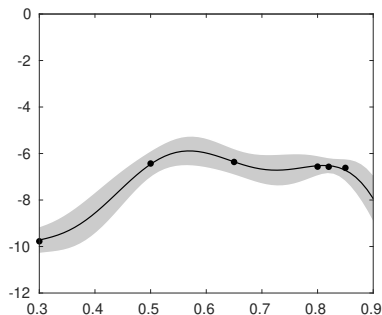


# Bayesian inference - computational challenges

- ▶ Computing the posterior distribution for complicated models can be slow
- ▶ The mathematical formula of the posterior distribution might be too difficult to derive → **simulation-based Bayesian inference**
  - ▶ One might need e.g. 100 000 simulations. If each simulation takes e.g. one hour, the full computation requires **11.4 years**.
  - ▶ Also other challenges (e.g. only approximate solution)

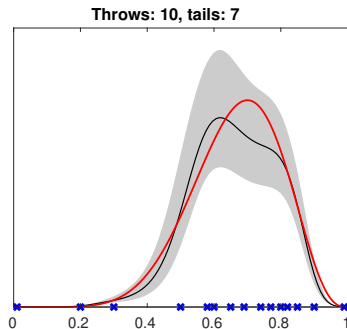
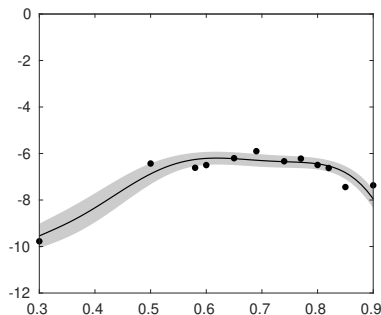
# Surrogate modelling

- ▶ So-called surrogate modelling can be used to **reduce the required posterior evaluations or the amount of simulations**
- ▶ An exact solution is substituted with an approximate one



# Surrogate modelling

- ▶ So-called surrogate modelling can be used to **reduce the required posterior evaluations or the amount of simulations**
- ▶ An exact solution is substituted with an approximate one



# Research questions (in a slightly simplified form)

- ▶ The effect of the surrogate modelling to the accuracy of the resulting approximate posterior distribution
- ▶ How to choose the parameters for simulating the model to minimise the computational cost?
- ▶ How to parallelise the simulations in a best possible way?
- ▶ Applications, especially understanding the colonisation dynamics of MRSA bacteria



# Results: Parallel computation

