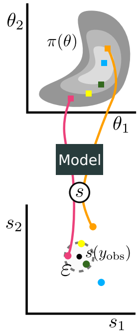
Learning robust summary statistics and cost functions in ABC

INCOME, 2021/03/01-04

Yannik Schaelte[1,2], Jan Hasenauer[1,2,3]

1: Helmholtz Center Munich, Institute of Computational Biology, 2: Technical University Munich, Department of Mathematics, 3: University Bonn, Research Unit Biomathematics

Mini-Intro: ABC

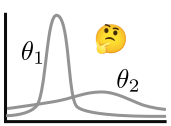


- Approximate Bayesian Computation enables likelihood-free inference
- until N acceptances:
  1. sample parameters $\theta \sim \pi(\theta)$
  2. simulate data $y \sim \pi(y|\theta)$
  3. accept if $d(s(y), s(y_{\text{obs}})) < \varepsilon$
- often combined with an SMC scheme

The problem

- Easily gives bad results if summary statistics and distance are not properly calibrated

ABC Posterior



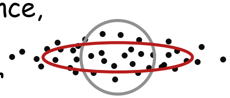
- Posteriors can look very different
- Sufficiency vs Monte-Carlo error
- Can we learn good representations automatically?

A zoo of summary statistics

reliable, different scales, less or uninformative, model error, replicates, complex relations, ... anything can happen

Adaptive distance functions

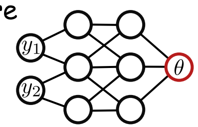
- Prangle 2017: Weight by sample variance,
  $$w_i \propto \text{std}\{s_j\}_i$$
- adaptively adjust weights to posterior



- ▸ robustification: Account for sample bias
  $$w_i \propto \text{std}(\{s_j\})_i + \text{bias}(\{s_j\}, s_{\text{obs}})_i$$
- ▸ actually, what we really want is to account for the "information" of data on parameters $y \mapsto \theta$ ...

Learning statistics via regression

- Fearnhead et al 2012: Good statistics are
  $$s(y) = \mathbb{E}[\theta|y] \approx f(y)$$
  with a linear model f



- Alternatives: Ridge (Blum et al 2013), NNs (Jiang et al 2017), GPs (Borowska et al 2020)
- extract information from high-dimensional raw data
- ▸ feature normalization and adaptivity

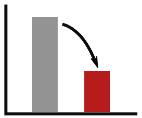
Let's combine the two! 💍

- ▸ weight-adjusted adaptive automatic statistic calculation
- ▸ weights accounting for variance, bias, and sensitivity

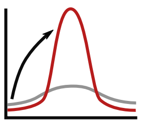
(First) Results

- ✓ basic algorithms implemented in a modular manner
- ✓ first checks: perform robustly on classical test problems
- ✓ robustification identifies model error
- ✓ combination allows to learn problem structure faster

Samples



Posterior θ



Outlook

- combination of various regressors and cost functions
- model selection
- semi-automatic updating in ABC-SMC
- application to agent-based models
- easily usable implementation in pyABC

Looking for:

- model selection and training for DNNs, GPs, ...
- challenging likelihood-free problems

github.com/yannikschaelte
yannik.schaelte@gmail.com
@yannik_schaelte

Fit Multi Cell

HelmholtzZentrum münchen
German Research Center for Environmental Health

TUM

UNIVERSITÄT BONN