Approximate Bayesian Computation for Parameter Estimation of Complex Stochastic Models

Yannik Schälte[1,2] Jan Hasenauer[1,2]

[1] Helmholtz Center Munich, Institute of Computational Biology

[2] Technical University Munich, Department of Mathematics

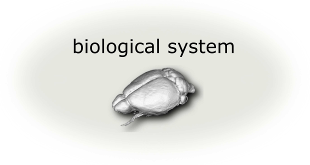
CRM SNS Pisa, 2018-10-03

HelmholtzZentrum münchen
German Research Center for Environmental Health

Table of Contents

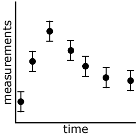
Table of Contents

Parameter estimation



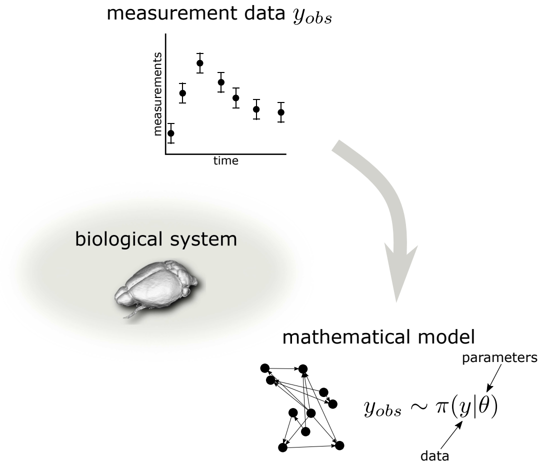
Parameter estimation

measurement data y_{obs}



biological system

Parameter estimation



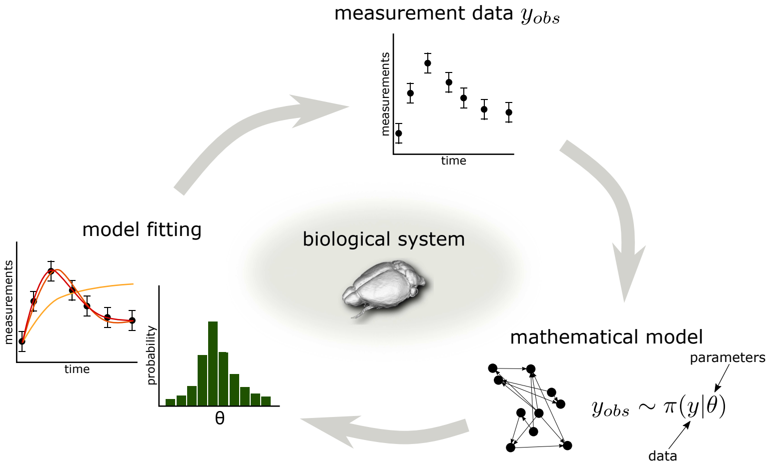
measurement data y_{obs}

biological system

mathematical model

$$y_{obs} \sim \pi(y|\theta)$$

Parameter estimation



measurement data y_{obs}

biological system

mathematical model

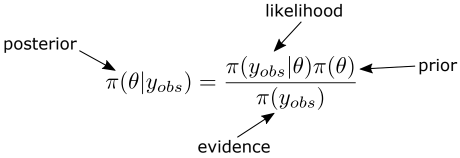
parameters

$$y_{obs} \sim \pi(y|\theta)$$

data

model fitting

Bayesian inference



likelihood

posterior

$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

prior

evidence

- goal: infer parameters θ given data y_{obs}, i.e. analyze the posterior distribution

Bayesian inference

likelihood

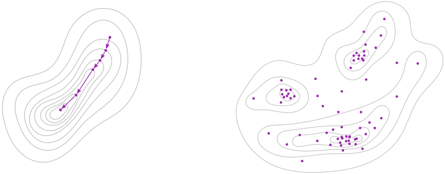
posterior

prior

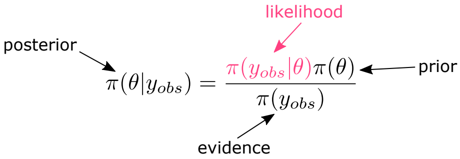
$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

evidence

- goal: infer parameters θ given data y_{obs}, i.e. analyze the posterior distribution
- optimization and sampling methods like MCMC commonly require evaluating the (unnormalized) likelihood



Likelihood-free Bayesian inference



likelihood

posterior

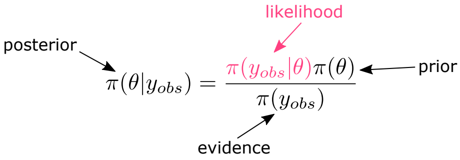
$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

prior

evidence

- can happen: numerical evaluation of likelihood infeasible

Likelihood-free Bayesian inference



likelihood

posterior

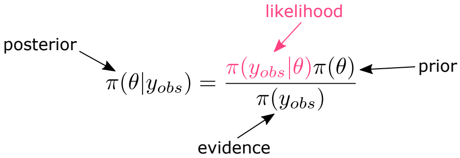
$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

prior

evidence

- can happen: numerical evaluation of likelihood infeasible
- but possible to simulate data $y \sim \pi(y|\theta)$

Likelihood-free Bayesian inference



likelihood

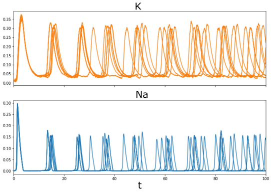
posterior

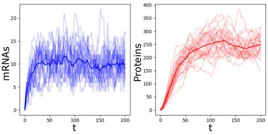
$$\pi(\theta|y_{obs}) = \frac{\pi(y_{obs}|\theta)\pi(\theta)}{\pi(y_{obs})}$$

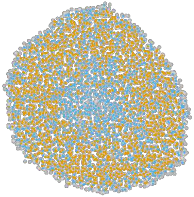
prior

evidence

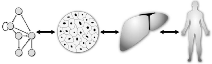
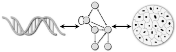
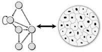
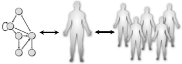
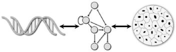
- can happen: numerical evaluation of likelihood infeasible
- but possible to simulate data $y \sim \pi(y|\theta)$
- often the case for complex stochastic models





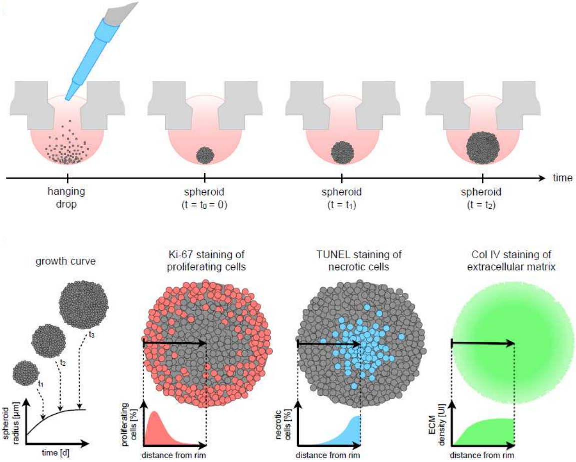


Multi-scale models

		Boolean / Petri net models	Constraint-based models	Markov chain models	SDE models	ODE models	PDE models	Agent-based models	
whole-heart (tissue and fluid mechanics, signaling)		■				■	■		Hunter and Borg, *Nature Reviews Molecular Cell Biology*, 4(3):237-243, 2003
cancer growth (signaling, cell division and death, angiogenesis, tissue remodeling)					■		■	■	Anderson and Quaranta, *Nature Reviews Cancer*, 8(3):227-234. 2008
liver lobule (cell division and cell death, tissue mechanics)				■	■		■	■	Hoehme et al., *PNAS*, 107(23):10371-10376, 2010
glucose-insulin-glucagon regulation (blood and interstitial flow, organ uptake, signaling)						■	■	■	Schaller et al., *CPT: Pharmacometrics and Systems Pharmacology*, 2:e65, 2013
whole-cell (transcription, translation, DNA replication, metabolism, replication)		■	■	■		■		■	Karr et al., *Cell*, 150(2):389-401, 2012

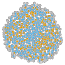
Hasenauer; Data-driven modeling of biological multi-scale processes; J. Coup. Sys. and Mult. Dyn.; 2015

Example: Multi-scale model of tumor growth



Jagiella et al.; Parallelization and high-performance computing enables automated statistical inference of multi-scale models; Cell Systems; 2017

Example: Multi-scale model of tumor growth



proliferating cells

- hybrid discrete-continuous model
- cells modeled as stochastically interacting agents, dynamics of extracellular substances by reaction-diffusion equations
- simulate up to 10^6 cancer cells
- 10s - 1h for one forward simulation
- 7 - 18 parameters

Jagiella et al.; Parallelization and high-performance computing enables automated statistical inference of multi-scale models; Cell Systems; 2017

Example: Multi-scale model of tumor growth

What we tried:

- multi-start local optimization
  - deterministic gradient descent
    - Levenberg-Marquardt
    - trust-region
    - interior-point
  - stochastic gradient descent
  - Bayesian optimization
- global optimization
  - simulated annealing
  - > 20 particle methods
  - enhanced scatter search

Example: Multi-scale model of tumor growth

What we tried:

- multi-start local optimization
  - deterministic gradient descent
    - Levenberg-Marquardt
    - trust-region
    - interior-point
  - stochastic gradient descent
  - Bayesian optimization
- global optimization
  - simulated annealing
  - $>$ 20 particle methods
  - enhanced scatter search

worked — failed

Example: Multi-scale model of tumor growth

What we tried:

- multi-start local optimization
  - deterministic gradient descent
    - Levenberg-Marquardt
    - trust-region
    - interior-point
  - stochastic gradient descent
  - Bayesian optimization
- global optimization
  - simulated annealing
  - $>$ 20 particle methods
  - enhanced scatter search

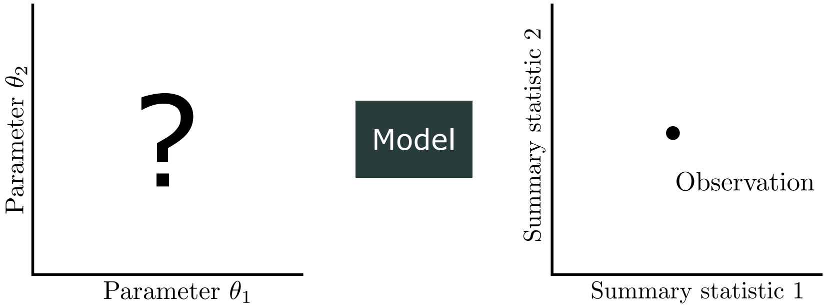
worked — failed

Key problem: Objective function cannot be evaluated, but only stochastically approximated.

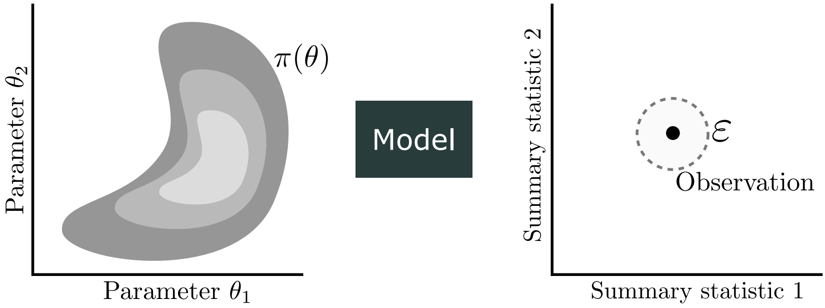
How to do parameter estimation for complex stochastic models?

Table of Contents

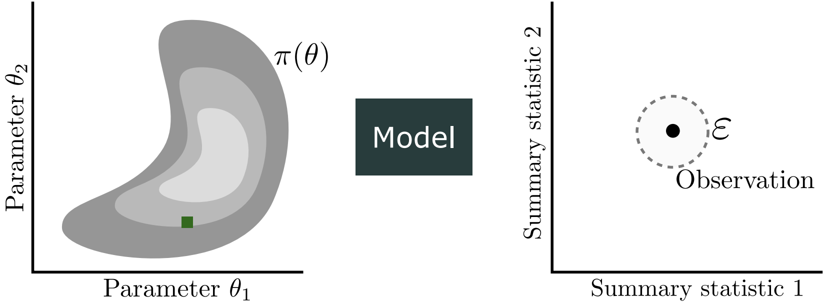
ABC-Rejection



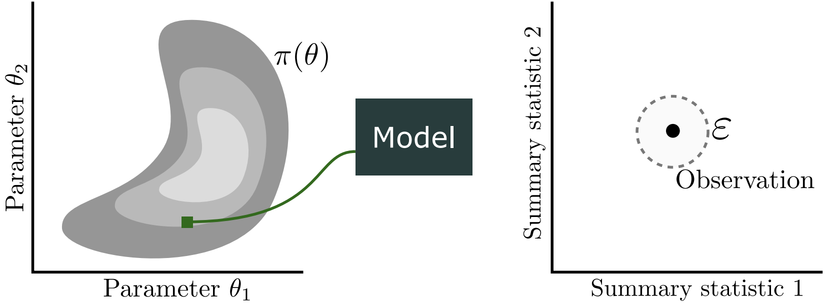
ABC-Rejection



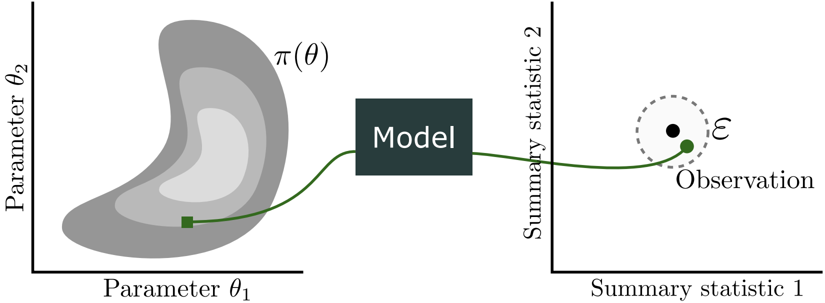
ABC-Rejection



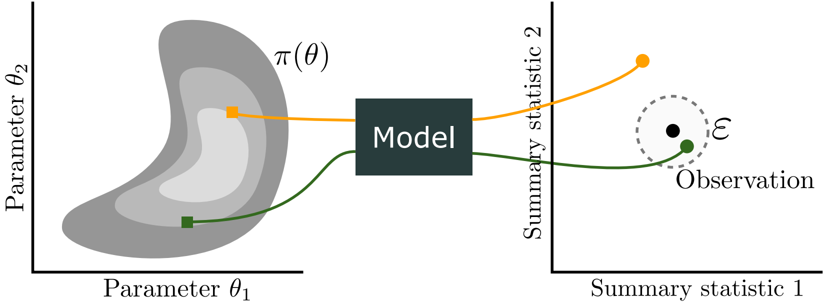
ABC-Rejection



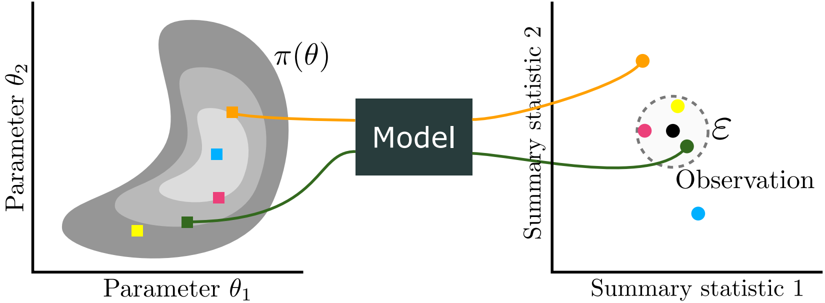
ABC-Rejection



ABC-Rejection



ABC-Rejection



Idea: Rejection sampling

Background: Want to sample from f, but can only sample from g with $g \gg f$.

until N acceptances:

1. sample $\theta^* \sim g(\theta)$
2. accept θ^* with probability $\propto \frac{g(\theta^*)}{f(\theta^*)}$

Accepted θ^* are independent samples from $f(\theta)$.

Idea: Rejection sampling

Background: Want to sample from f, but can only sample from g with $g \gg f$.

until N acceptances:

1. sample $\theta^* \sim g(\theta)$
2. accept θ^* with probability $\propto \frac{g(\theta^*)}{f(\theta^*)}$

Accepted θ^* are independent samples from $f(\theta)$.

Here: $f = \pi(\theta|y_{obs})$, $g = \pi(\theta)$, so that $\frac{\pi(\theta^*|y_{obs})}{\pi(\theta^*)} \propto \pi(y_{obs}|\theta^*)$

Idea: Rejection sampling

Background: Want to sample from f, but can only sample from g with $g \gg f$.

until N acceptances:

1. sample $\theta^* \sim g(\theta)$
2. accept θ^* with probability $\propto \frac{g(\theta^*)}{f(\theta^*)}$

Accepted θ^* are independent samples from $f(\theta)$.

Here: $f = \pi(\theta|y_{obs})$, $g = \pi(\theta)$, so that $\frac{\pi(\theta^*|y_{obs})}{\pi(\theta^*)} \propto \pi(y_{obs}|\theta^*)$

- not available for us
- idea: **circumvent likelihood evaluation** by **simulating data** and matching them to the observed data

Likelihood-free rejection sampling

until N acceptances:

1. sample parameter $\theta^* \sim \pi(\theta)$
2. simulate data $y^* \sim \pi(y|\theta^*)$
3. accept θ^* if $y^* = y_{obs}$

Likelihood-free rejection sampling

```
until $N$ acceptances:
    1. sample parameter $\theta^* \sim \pi(\theta)$
    2. simulate data $y^* \sim \pi(y|\theta^*)$
    3. accept $\theta^*$ if $y^* = y_{obs}$
```

Acceptance probability $\mathbb{P}[y_{obs}]$

- can be small, in particular 0 for continuous data
- idea: accept simulations that are "**similar**" to y_{obs}

ABC-Rejection

With distance d, and treshold $\varepsilon > 0$:

until N acceptances:

1. sample parameter $\theta^* \sim \pi(\theta)$
2. simulate data $y^* \sim \pi(y|\theta^*)$
3. accept θ^* if $d(y^*, y_{obs}) \leq \varepsilon$

ABC-Rejection

With distance d, and treshold $\varepsilon > 0$:

until N acceptances:

1. sample parameter $\theta^* \sim \pi(\theta)$
2. simulate data $y^* \sim \pi(y|\theta^*)$
3. accept θ^* if $d(y^*, y_{obs}) \leq \varepsilon$

- **curse of dimensionality**: if the data are too high-dimensional, the probability of simulating similar data sets is small
- reduce the dimension using **summary statistics**

ABC-Rejection

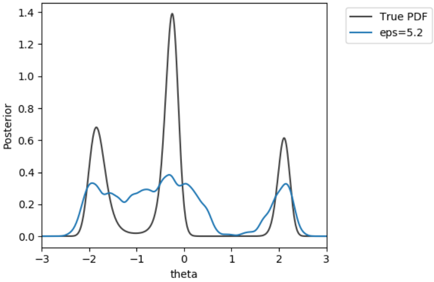
With distance d, threshold $\varepsilon > 0$, and summary statistics s:

until N acceptances:

1. sample parameter $\theta^* \sim \pi(\theta)$
2. simulate data $y^* \sim \pi(y|\theta^*)$
3. accept θ^* if $d(s(y^*), s(y_{obs})) \leq \varepsilon$

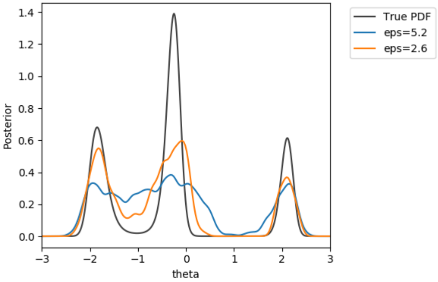
Example

$y \sim \mathcal{N}(2(\theta-2)\theta(\theta+2), 1+\theta^2)$,
$\theta \sim U[-3,3]$,
$d = \|\cdot\|_1$,
$y_{obs} = 2$,
$N = 1000$ acceptances



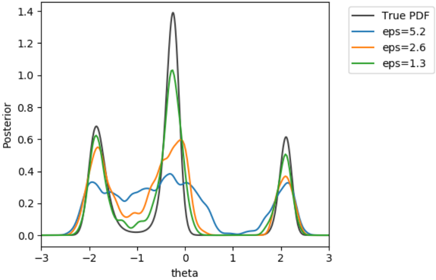
Example

$y \sim \mathcal{N}(2(\theta-2)\theta(\theta+2), 1+\theta^2)$,
$\theta \sim U[-3,3]$,
$d = \|\cdot\|_1$,
$y_{obs} = 2$,
$N = 1000$ acceptances



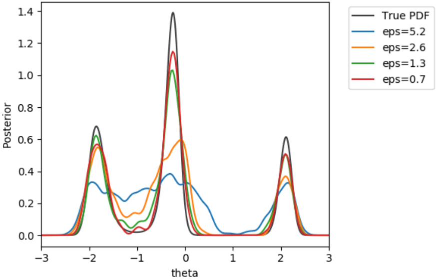
Example

$y \sim \mathcal{N}(2(\theta-2)\theta(\theta+2), 1+\theta^2)$,
$\theta \sim U[-3,3]$,
$d = \|\cdot\|_1$,
$y_{obs} = 2$,
$N = 1000$ acceptances



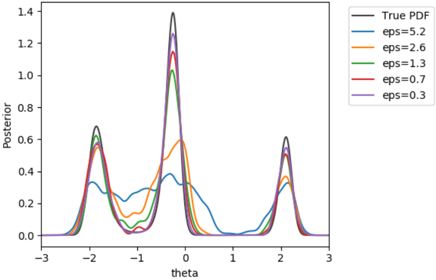
Example

$y \sim \mathcal{N}(2(\theta - 2)\theta(\theta + 2), 1 + \theta^2)$,
$\theta \sim U[-3, 3]$,
$d = \|\cdot\|_1$,
$y_{obs} = 2$,
$N = 1000$ acceptances



Example

$y \sim \mathcal{N}(2(\theta-2)\theta(\theta+2), 1+\theta^2),$
$\theta \sim U[-3, 3],$
$d = \|\cdot\|_1,$
$y_{obs} = 2,$
$N = 1000$ acceptances



Will the approximation always converge to the true posterior?

Formally

We want:

$$\pi(\theta|y_{obs}) \propto \pi(y_{obs}|\theta)\pi(\theta)$$

We get:

$$\pi_{ABC}(\theta|s_{obs}) \propto \int I(\{d(s,s_{obs}) \leq \varepsilon\})\pi(s|\theta)\pi(\theta)ds \overset{\propto}{\approx} \frac{1}{N}\sum_{i=1}^{N}\delta_{\theta^{(i)}}(\theta)$$

Theorem

Under certain assumptions it holds that

- $\frac{1}{N}\sum_{i=1}^{N}\delta_{\theta^{(i)}}(\theta) \xrightarrow{w} \pi_{ABC}(\theta|s_{obs})$ for $N \to \infty$,
- $\pi_{ABC}(\theta|s_{obs}) \xrightarrow{w} \pi(\theta|s_{obs})$ for $\varepsilon \to 0$.

Sources of approximation errors in ABC

- model error (as for every model of reality)
- Monte-Carlo error (as for sampling in general)
- summary statistics
- epsilon threshold

John Tukey 1962

"Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise."

Gelman and Rubin 1996

"[...] one of the great scientific advantages of simulation analysis of Bayesian methods is the freedom it gives the researcher to formulate appropriate models rather than be overly interested in analytically neat but scientifically inappropriate models."

from U. Picchini; Lund University; PhD course Statistical inference for partially observed stochastic processes

Efficient samplers

- **ABC-Rejection**, the basic ABC algorithm, can be **inefficient** due to repeatedly sampling from the prior

Efficient samplers

- **ABC-Rejection**, the basic ABC algorithm, can be **inefficient** due to repeatedly sampling from the prior
- smaller ε leads to lower acceptance rates

Efficient samplers

- **ABC-Rejection**, the basic ABC algorithm, can be **inefficient** due to repeatedly sampling from the prior
- smaller ε leads to lower acceptance rates
- many (likelihood-based) Monte-Carlo sampling algorithms like IS, MCMC, Gibbs, SMC, today have **ABC-fied** versions

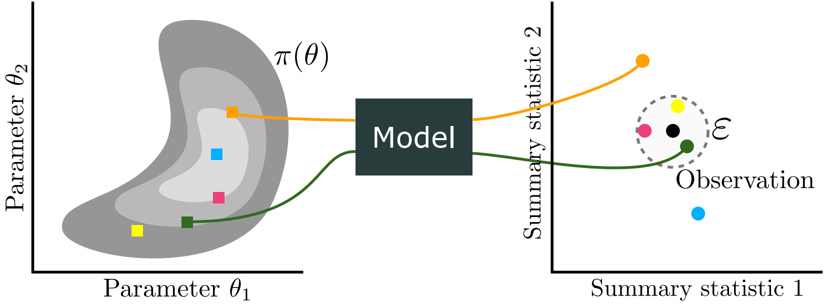
Efficient samplers

- **ABC-Rejection**, the basic ABC algorithm, can be **inefficient** due to repeatedly sampling from the prior
- smaller ε leads to lower acceptance rates
- many (likelihood-based) Monte-Carlo sampling algorithms like IS, MCMC, Gibbs, SMC, today have **ABC-fied** versions

- here: focus on **ABC-SMC**

ABC-SMC

Combine with a Sequential Monte-Carlo Scheme

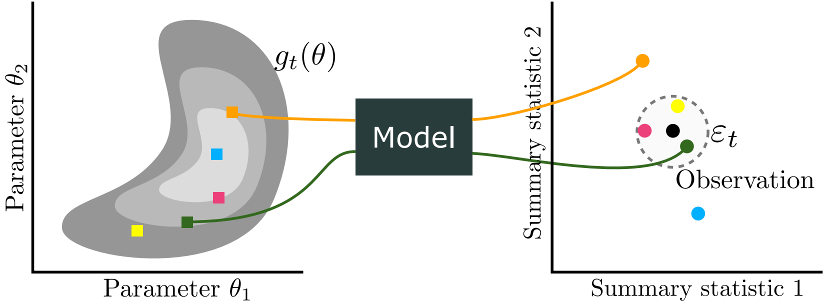


- idea: decrease $\varepsilon = \varepsilon_t$ while sampling from an increasingly better approximation of the posterior

Sisson et al. 2007, Toni et al. 2008, Beaumont et al. 2009

ABC-SMC

Combine with a Sequential Monte-Carlo Scheme

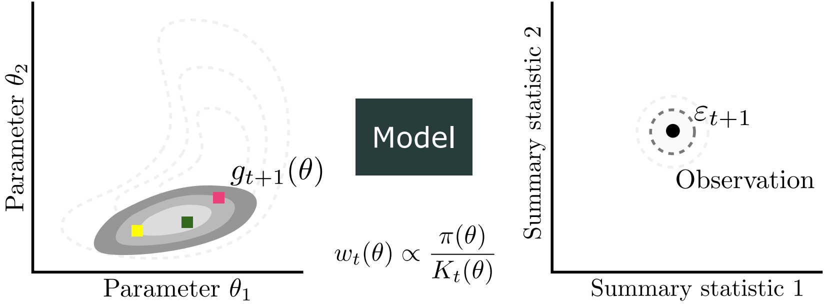


- idea: decrease $\varepsilon = \varepsilon_t$ while sampling from an increasingly better approximation of the posterior

Sisson et al. 2007, Toni et al. 2008, Beaumont et al. 2009

ABC-SMC

Combine with a Sequential Monte-Carlo Scheme



- idea: decrease $\varepsilon = \varepsilon_t$ while sampling from an increasingly better approximation of the posterior

Sisson et al. 2007, Toni et al. 2008, Beaumont et al. 2009

ABC-SMC

```
for $t = 1, \dots, t_{max}$
    1. until $N$ acceptances
        1. sample parameter $\theta^* \sim g_t(\theta)$
        2. simulate data $y^* \sim \pi(y|\theta^*)$
        3. accept $\theta^*$ if $d(y^*, y_{obs}) \leq \varepsilon_t$
       denote by $\theta_1^t, \dots, \theta_N^t$ the accepted parameters
    2. compute weights $w_i^t = \frac{\pi(\theta_i^t)}{g_t(\theta_i^t)}$
```

Here, the proposal distribution is

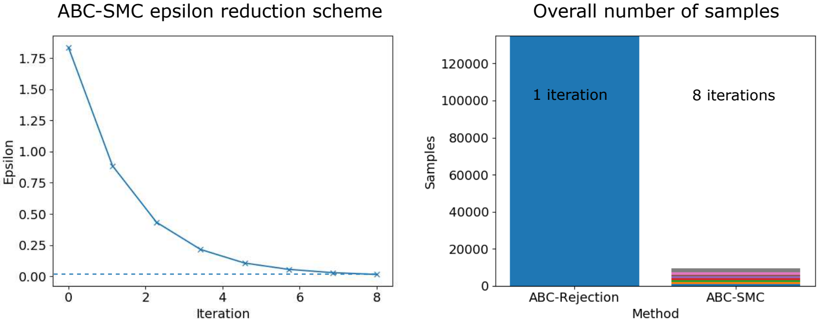
$$g_t(\theta) = \begin{cases} \pi(\theta) & , t = 1 \\ \sum_{i=1}^N w_i^{t-1} K_t(\theta|\theta_i^{t-1}) / \sum_{i=1}^N w_i^{t-1} & , \text{otherwise} \end{cases}.$$

Then, $\pi_{ABC}(\theta|s_{obs}) \sim \{\theta_i^{t_{max}}, w_i^{t_{max}}\}_{1 \leq i \leq N}$ (importance sampling).

ABC-SMC

Example

$y \sim U[-0.05, 0.05]$, $\theta \sim U[-4, 4]$, $d = \|\cdot\|_2$, $y_{obs} = 0$, same ε threshold



The SMC scheme significantly reduces the needed number of samples.

Challenges in ABC-SMC

- summary statistics
- distance functions
- epsilon thresholds
- population sizes

Challenges in ABC-SMC

- **summary statistics**
- distance functions
- epsilon thresholds
- population sizes

Summary statistics

$$y \mapsto s(y)$$

- **low-dimensional** representation of data to **extract relevant features** and **increase acceptance rates**

Summary statistics

$$y \mapsto s(y)$$

- **low-dimensional** representation of data to **extract relevant features** and **increase acceptance rates**
- ideally: **minimal sufficient statistics**, i.e. s is minimal s.t. $\pi(\theta|y) = \pi(\theta|s(y))$ for almost every y

Summary statistics

$$y \mapsto s(y)$$

- **low-dimensional** representation of data to **extract relevant features** and **increase acceptance rates**
- ideally: **minimal sufficient statistics**, i.e. s is minimal s.t. $\pi(\theta|y) = \pi(\theta|s(y))$ for almost every y
- practically: usually not available (essentially only for exponential family models), therefore **trade-off between information loss and performance**

Summary statistics

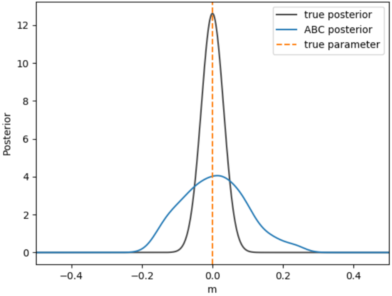
$$y \mapsto s(y)$$

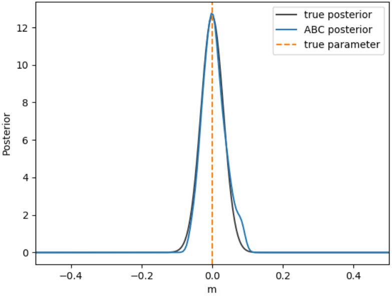
- **low-dimensional** representation of data to **extract relevant features** and **increase acceptance rates**
- ideally: **minimal sufficient statistics**, i.e. s is minimal s.t. $\pi(\theta|y) = \pi(\theta|s(y))$ for almost every y
- practically: usually not available (essentially only for exponential family models), therefore **trade-off between information loss and performance**
- **(semi-)automatic selection of summary statistics** (Fearnhead and Prangle 2012, and the review Blum et al. 2013)

Summary statistics

Example

$y \sim \bigotimes_{i=1}^{R} \mathcal{N}(\theta, 1)$ for $R = 10000$, $\theta \sim U[-4, 4]$, $d = \|\cdot\|_2$,
$y_{obs} = [0, \ldots, 0]$





$s = y$

$s = \bar{y} = \frac{1}{R}\sum_{i=1}^{R} y_i$

time: 107s, samples: 5.2e3

time: 32s, samples: 1.8e3

Summary statistics

Example 2: gk distribution

- gk distribution $\mathrm{gk}(y|\theta)$ with $\theta = (A, B, g, k)$ given via quantile transform

$$Q(q|A, B, g, k) = A + B\left[1 + c\frac{1 - \exp(-gz(q))}{1 + \exp(-gz(q))}\right](1 + z(q)^2)^k z(q)$$

  for $B > 0, k > -\frac{1}{2}$, $c = 0.8$, where $z(q) = \Phi^{-1}(q)$ is the quantile transform of $\mathcal{N}(0, 1)$
- density function has no closed form

Summary statistics

Example 2: gk distribution

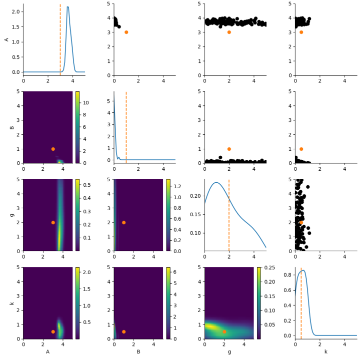
$y \sim \bigotimes_{i=1}^{R} \mathrm{gk}(y|\theta)$ for $R = 1000$, $\theta \sim U[0,5]$, $d = \|\|_2$, y_{obs} sampled from the likelihood with $\theta = (3.0, 1.0, 2.0, 0.5)$

summary statistics:

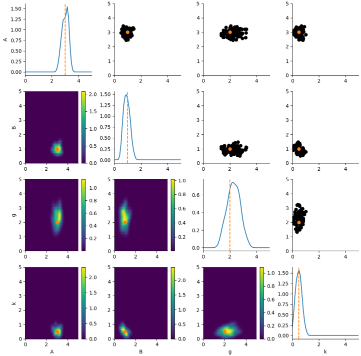
1. $s = y$
2. $s = (s_A, s_B, s_g, s_k)$ where $s_A = E_4$, $s_B = E_6 - E_2$, $s_g = (E_6 + E_2 - 2E_4)/s_B$, $s_k = (E_7 - E_5 + E_3 - E_1)/s_B$ where $E_1 \leq \ldots \leq E_8$ are the octiles of y (Drovandi and Pettitt 2011)

Summary statistics

Example 2: gk distribution



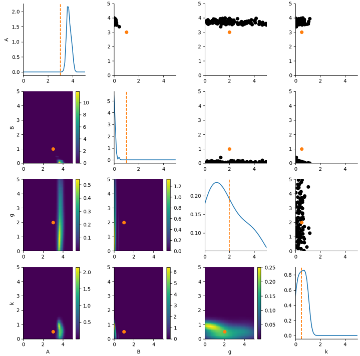
based on full data



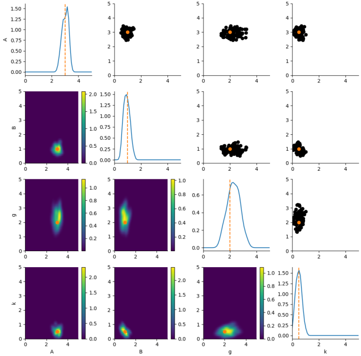
based on order statistics

Summary statistics

Example 2: gk distribution



based on full data



based on order statistics

Only the use of proper summary statistics ensures convergence in a reasonable computation time.

Challenges in ABC-SMC

- summary statistics
- **distance functions**
- epsilon thresholds
- population sizes

Distance functions

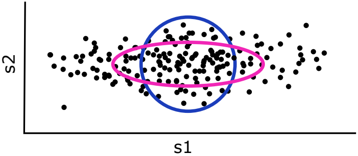
$$d(s(y^*), s(y_{obs}))$$

- in practice often simply p-norm distance (e.g. Euclidean distance $p = 2$) used, i.e $d(x, y) = \left(\sum_{i=1}^{n_s} |x_i - y_i|^p\right)^{1/p}$ where n_s is the summary statistics dimension
- many other distances possible (McKinley 2009)

Distance functions

Weighted distances

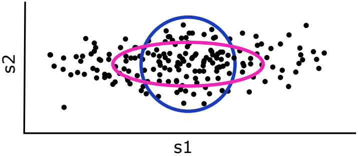
- problem: summary statistics can vary on **different scales**



Distance functions

Weighted distances

- problem: summary statistics can vary on **different scales**

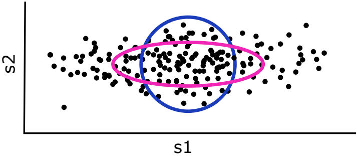


- therefore: **weighted distance** $d(x,y) = \left(\sum_{i=1}^{n_s} \omega_i |x_i - y_i|^p\right)^{1/p}$
- usually: pre-calibrate weights

Distance functions

Weighted distances

- problem: summary statistics can vary on **different scales**

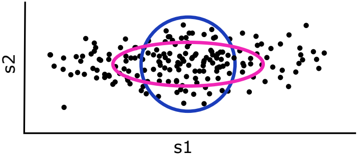


- therefore: **weighted distance** $d(x,y) = \left(\sum_{i=1}^{n_s} \omega_i |x_i - y_i|^p\right)^{1/p}$
- usually: pre-calibrate weights
- requires additional effort, and in iterative methods the proposal distributions can vary over time

Distance functions

Weighted distances

- problem: summary statistics can vary on **different scales**

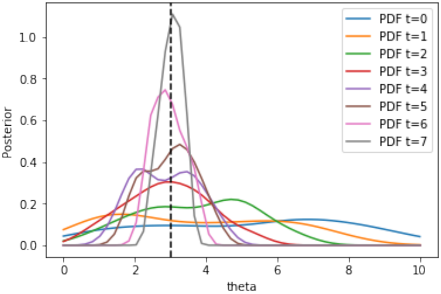


- therefore: **weighted distance** $d(x,y) = \left(\sum_{i=1}^{n_s} \omega_i |x_i - y_i|^p\right)^{1/p}$
- usually: pre-calibrate weights
- requires additional effort, and in iterative methods the proposal distributions can vary over time
- Prangle 2015: **adapt weights iteratively** based on samples from previous iteration
- note: assumes equally informative summary statistics

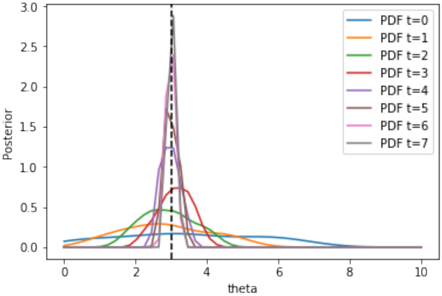
Distance functions

Adaptive weights: Example

$s_1 \sim \mathcal{N}(1+\theta, 0.01)$, $s_2 \sim \mathcal{N}(2, 100)$, $s_{obs} = [4, 2]$ corresponding to $\theta_{true} = 3$, $\theta \sim U[0, 10]$



$d = \|\cdot\|_2$



$d =$ adaptively weighted $\|\cdot\|_2$

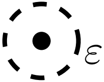
Accounting for data heterogeneity improves convergence.

Challenges in ABC-SMC

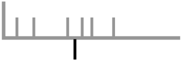
- summary statistics
- distance functions
- **epsilon thresholds**
- population sizes

Epsilon thresholds

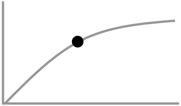
How to choose epsilon?



pre-specified list



quantiles



predict threshold-acceptance rate

Silk, Filippi, Stumpf 2013

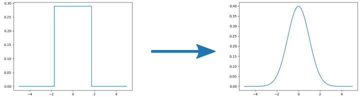
Epsilon thresholds

Generalization and re-interpretation

- generalize $I(\{d(s(y), s(y_{obs})) \leq \varepsilon\}) \rightsquigarrow K_\varepsilon(s(y) - s(y_{obs}))$ for some **kernel** K_ε, i.e.

  3.' accept with probability $\frac{K_\varepsilon(s(y)-s(y_{obs}))}{K_\varepsilon(0)}$

  (can represent the previous $0, 1$-cutoff by a $U[-\varepsilon, \varepsilon]$) kernel)



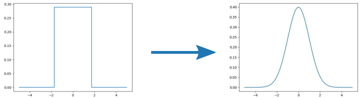
Epsilon thresholds

Generalization and re-interpretation

- generalize $I(\{d(s(y), s(y_{obs})) \leq \varepsilon\}) \rightsquigarrow K_\varepsilon(s(y) - s(y_{obs}))$ for some **kernel** K_ε, i.e.

  3.' accept with probability $\frac{K_\varepsilon(s(y)-s(y_{obs}))}{K_\varepsilon(0)}$

  (can represent the previous $0, 1$-cutoff by a $U[-\varepsilon, \varepsilon]$) kernel)



- Wilkinson 2013: ABC gives exact inference under the assumption of measurement noise: it samples from the model $s \sim \pi(s|\theta) + \delta$ where $\delta \sim K_\varepsilon$ is an independent error term

Epsilon thresholds

Assessing measurement noise

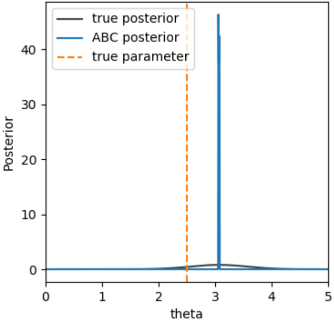
- if there is **measurement noise**, it **must be accounted for** in ABC
- ignoring leads to wrong parameter estimates

Epsilon thresholds

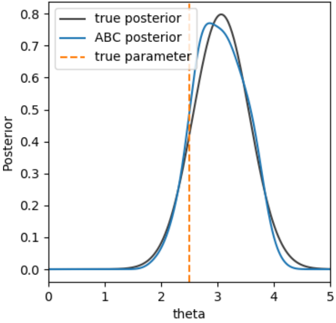
Assessing measurement noise

- if there is **measurement noise**, it **must be accounted for** in ABC
- ignoring leads to wrong parameter estimates

$\theta \sim U[0,5]$, y_{obs} sampled from $\mathcal{N}(2.5, 0.5^2)$



$y = \theta$



$y \sim \mathcal{N}(\theta, 0.5^2)$

Epsilon thresholds

Assessing measurement noise

- if there is **measurement noise**, it **must be accounted for** in ABC
- ignoring leads to wrong parameter estimates
- idea: we can use Wilkinson's insight to encode actual measurement noise not in the simulation, but in the **acceptance step**, and perform **exact Bayesian inference**

Challenges in ABC-SMC

- summary statistics
- distance functions
- epsilon thresholds
- **population sizes**

Population sizes

How to choose the population sizes N_t in ABC-SMC?

Population sizes

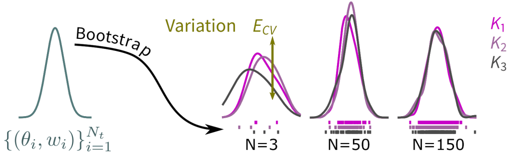
How to choose the population sizes N_t in ABC-SMC?

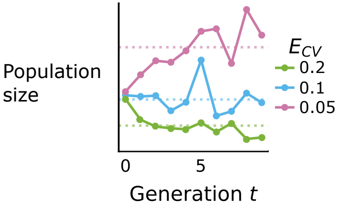
- trade-off accuracy – computational effort

Population sizes

How to choose the population sizes N_t in ABC-SMC?

- trade-off accuracy – computational effort
- idea: **adapt population sizes** trying to **match a specified target accuracy**
- expressed in terms of the variation associated with kernel density estimates





Klinger et al.; A Scheme for Adaptive Selection of Population Sizes in ABC-SMC; CMSB; 2017

Further notes

- there is a lot more to discuss
- including adequate proposal distributions, automatic summary statistics selection, threshold schedules, ABC-MCMC, regression ABC, approximate ABC, and many variations of the presented algorithms, …

- also model selection possible in ABC by augmenting the parameter space

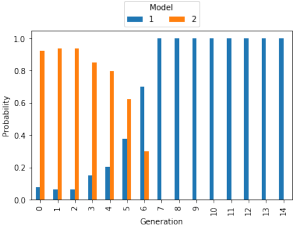


Table of Contents

pyABC

https://github.com/icb-dcm/pyabc

- implements an ABC-SMC algorithm
- HPC scalable using dynamic scheduling
- analysis, visualization and easy customization
- adaptive local/global transition kernels, distances, acceptance threshold schedules, population sizes, early rejection, ...

user-friendly

scalable

flexible

Klinger, Rickert, Hasenauer; pyABC: distributed, likelihood-free inference; Bioinformatics; 2018

pyABC

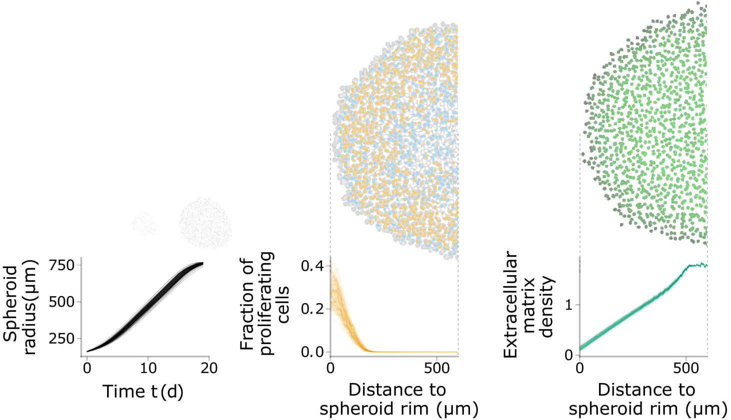
Three lines get you started

```
# pass model data
abc = pyabc.ABCSMC(model, prior, distance)

# pass observations
abc.new("sqlite:///database.db", observations)

# run it
abc.run(minimum_epsilon=1e-2, max_nr_populations=30)
```

Klinger, Rickert, Hasenauer; pyABC: distributed, likelihood-free inference; Bioinformatics; 2018

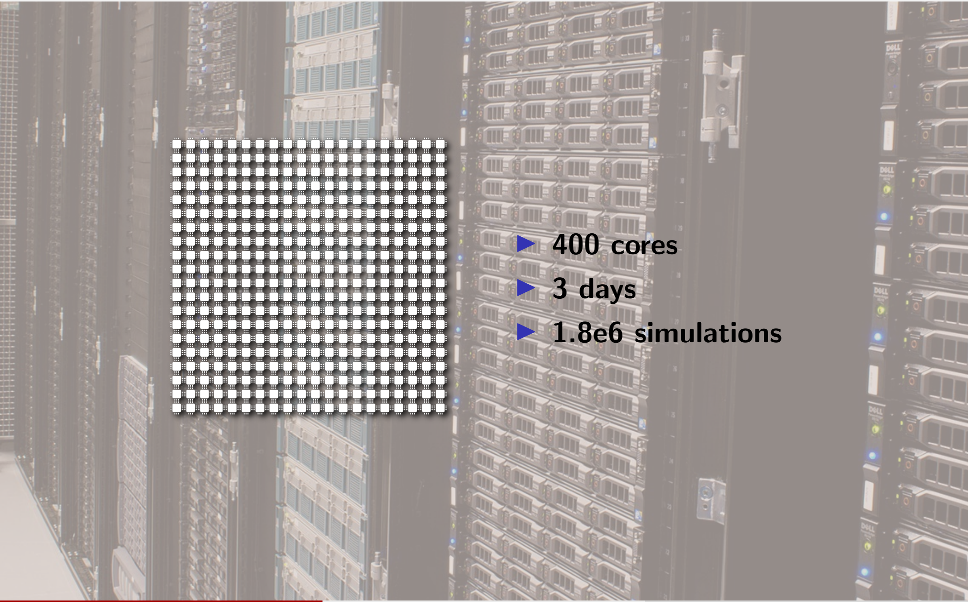
Multi-scale model of tumor growth

Summary statistics



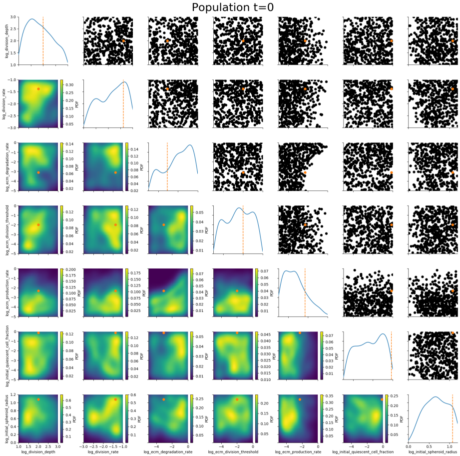
Jagiella et al.; Parallelization and high-performance computing enables automated statistical inference of multi-scale models; Cell Systems; 2017

Multi-scale model of tumor growth

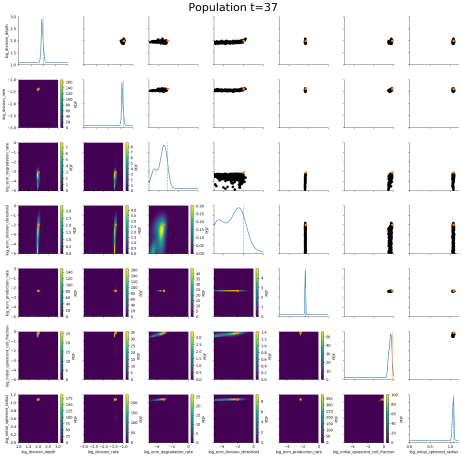


- **400 cores**
- **3 days**
- **1.8e6 simulations**

Multi-scale model of tumor growth

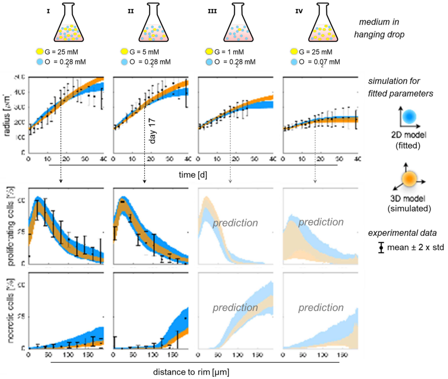


Multi-scale model of tumor growth



ABC worked where many other methods had failed.

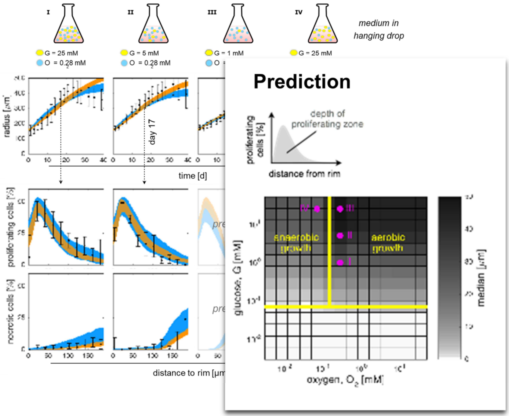
Multi-scale model of tumor growth



ABC enables automatic multi-experiment data integration.

Jagiella et al.; Parallelization and high-performance computing enables automated statistical inference of multi-scale models; Cell Systems; 2017

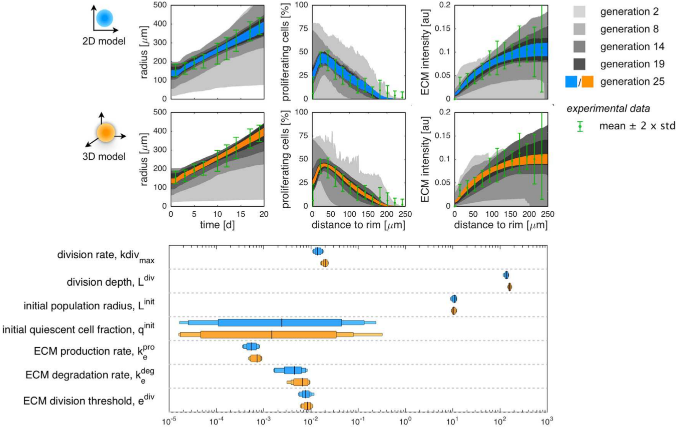
Multi-scale model of tumor growth



ABC enables uncertainty-aware predictions.

Jagiella et al.; Parallelization and high-performance computing enables automated statistical inference of multi-scale models; Cell Systems; 2017

Multi-scale model of tumor growth

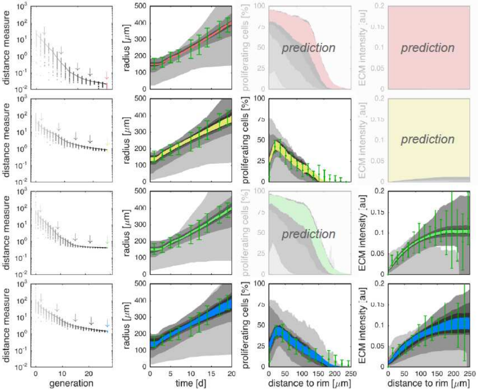


ABC enables hypothesis testing.

Jagiella et al.; Parallelization and high-performance computing enables automated statistical inference of multi-scale models; Cell Systems; 2017

Multi-scale model of tumor growth

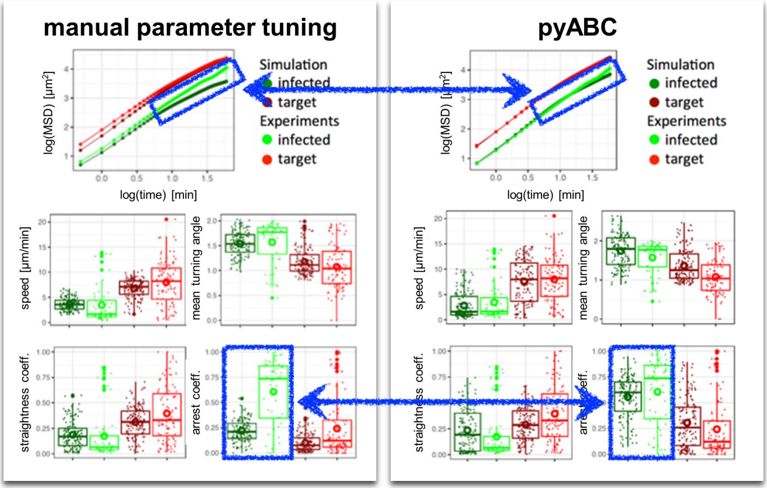
What data do I need?



ABC enables experimental design.

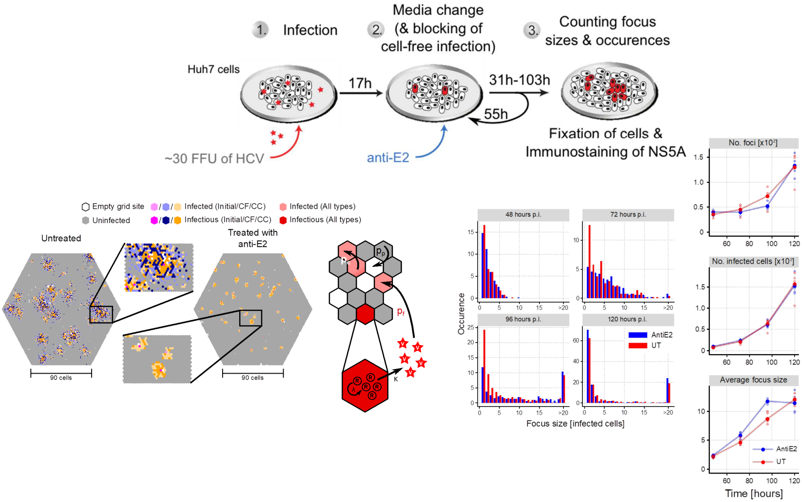
Jagiella et al.; Parallelization and high-performance computing enables automated statistical inference of multi-scale models; Cell Systems; 2017

Analysis of HIV infection dynamics



Jana Fehr, Emmanuel Klinger, Frederik Graw, Jan Hasenauer

Comparing HCV transmission modes



with Elba Raimúndez-Álvarez, Peter Kumberger

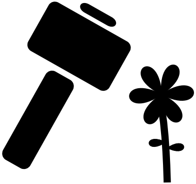
Table of Contents

Conclusion

- parameter estimation when we cannot evaluate the likelihood is challenging
- ABC enables reliable statistical inference with uncertainty information
- samples from an approximation of the true posterior
- broadly applicable
- scalable
- increasingly popular in many research areas
- not a silver bullet – if possible, use (sufficiently good approximations of) likelihoods

Conclusion

- parameter estimation when we cannot evaluate the likelihood is challenging
- ABC enables reliable statistical inference with uncertainty information
- samples from an approximation of the true posterior
- broadly applicable
- scalable
- increasingly popular in many research areas
- not a silver bullet – if possible, use (sufficiently good approximations of) likelihoods

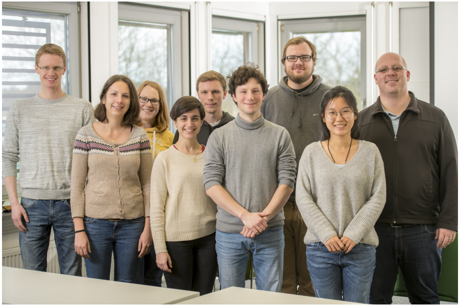


Not everything is a nail.

Acknowledgments

Thanks to:

- Elba Raimúndez Álvarez
- Emmanuel Klinger
- Dennis Rickert
- Jan Hasenauer
- Rest of the ICB-DCM group



HelmholtzZentrum münchen
German Research Center for Environmental Health

Bundesministerium
für Bildung
und Forschung

e:Med
SYSTEMS MEDICINE

Further reading

- Sisson, Scott A. and Fan, Yanan and Beaumont, Mark
  Handbook of Approximate Bayesian Computation.
  CRC Press, 2018.
- Beaumont, Mark A.
  Approximate Bayesian Computation in Evolution and Ecology.
  Annual Review of Ecology, Evolution, and Systematics, 41(1):379–406, 2010.
- Blum M. G.
  Choosing the Summary Statistics and the Acceptance Rate in Approximate Bayesian Computation.
  Proceedings of COMPSTAT, Physica, 2010.

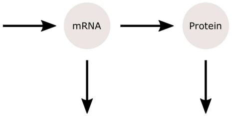
ABC-MCMC

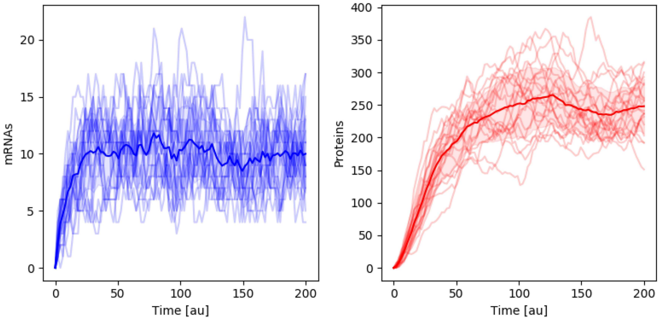
Combine with a Markov-Chain Monte-Carlo Scheme

initialize some θ_0 and simulate $y_0 \sim \pi(y|\theta_0)$
until enough acceptances

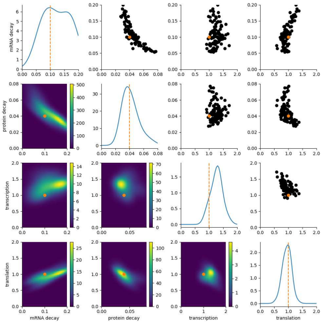
1. sample $\theta^* \sim g(\theta|\theta_{n-1})$
2. simulate $y^* \sim \pi(y|\theta^*)$
3. calculate $\alpha = \min\left[1, \frac{\pi(\theta^*)g(\theta^*|\theta_{n-1})I(\{d(s(y^*),s(y_{obs}))\leq\varepsilon\})}{\pi(\theta_{n-1})g(\theta_{n-1}|\theta^*)I(\{d(s(y_{n-1}),s(y_{obs}))\leq\varepsilon\})}\right]$
4. accept with probability α and update $\theta_n = \theta^*, y_n = y^*$

Example: Gene expression

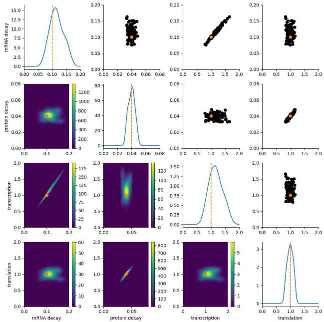




Example: Gene expression



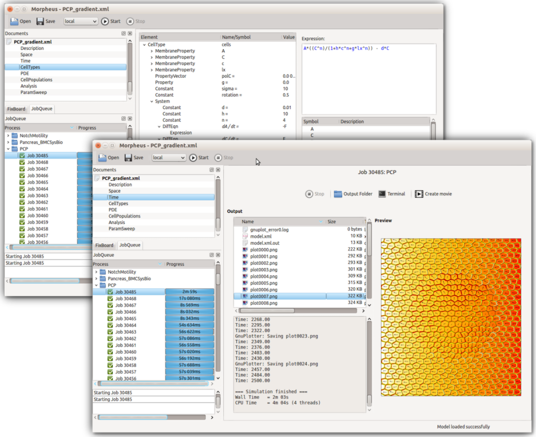
only protein counts



mRNA and protein counts

Model construction using Morpheus

https://morpheus.gitlab.io



Element	Name/Symbol	Value
CellType	cells	
MembraneProperty	A	
MembraneProperty	C	
MembraneProperty	c	
MembraneProperty	lx	
PropertyVector	polC =	0.0 0..
Property	g =	0.0
Constant	sigma =	10
Constant	rotation =	0.5
System		
Constant	d =	0.01
Constant	h =	10
Constant	n =	4
DiffEqn	*dA / dt =*	-F
Expression		

```
A*((C^n)/(1+h*c^n+g*lx^n)) - d*C
```

Process	Progress
NotchMotility	
Pancreas_BMCSysBio	
PCP	
☑ Job 30485	2m 59s
☑ Job 30468	17s 080ms
☑ Job 30467	8s 569ms
☑ Job 30466	8s 032ms
☑ Job 30465	8s 343ms
☑ Job 30464	54s 634ms
☑ Job 30463	56s 622ms
☑ Job 30462	57s 086ms
☑ Job 30461	56s 558ms
☑ Job 30460	57s 020ms
☑ Job 30459	56s 192ms
☑ Job 30458	57s 688ms
☑ Job 30457	57s 039ms
☑ Job 30456	57s 301ms

```
Starting Job 30485
Starting Job 30485
```

Name	Size
gnuplot_error0.log	0 bytes
model.xml	10 KB
model.xml.out	13 KB
plot0000.png	222 KB
plot0001.png	292 KB
plot0002.png	293 KB
plot0003.png	301 KB
plot0004.png	309 KB
plot0005.png	315 KB
plot0006.png	320 KB
plot0007.png	322 KB
plot0008.png	324 KB

```
Time: 2268.00
Time: 2295.00
Time: 2322.00
GnuPlotter: Saving plot0023.png
Time: 2349.00
Time: 2376.00
Time: 2403.00
Time: 2430.00
GnuPlotter: Saving plot0024.png
Time: 2457.00
Time: 2484.00
Time: 2500.00

=== Simulation finished ===
Wall Time    = 2m 03s
CPU Time     = 4m 04s (4 threads)
```

Staruss et al.; Morpheus: a user-friendly modeling environment for multiscale and multicellular systems biology; Bioinformatics; 2014