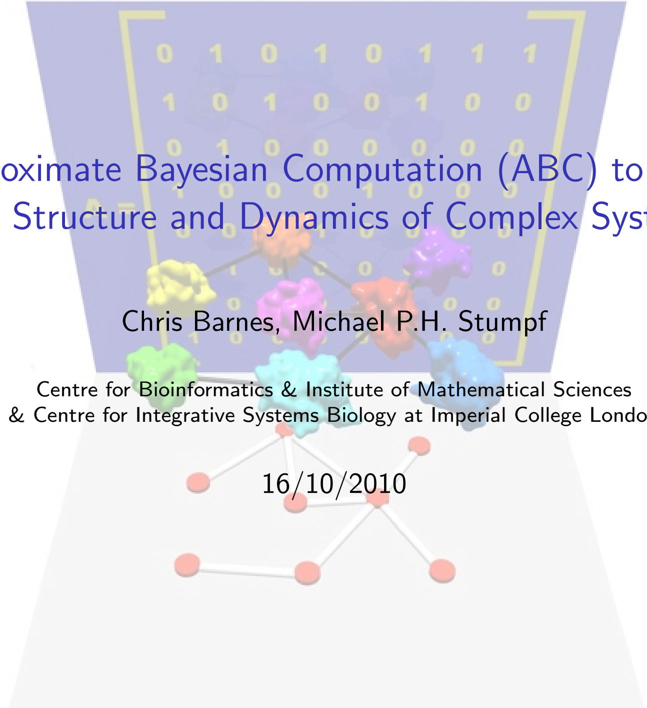


Approximate Bayesian Computation (ABC) to Learn the Structure and Dynamics of Complex Systems

Chris Barnes, Michael P.H. Stumpf

Centre for Bioinformatics & Institute of Mathematical Sciences
& Centre for Integrative Systems Biology at Imperial College London

16/10/2010

Outline

Our Problems

- What can we learn about the structure and dynamics of biological systems from data?
- How do networks evolve and what does their structure tell us about their function?

Outline

Our Problems

- What can we learn about the structure and dynamics of biological systems from data?
- How do networks evolve and what does their structure tell us about their function?

Inference and Models: Approximate vs. Exact

We have observed data, $\mathcal{D}$, that was generated by some system of in general unknown structure that we seek to describe by a mathematical model. In principle we can have a model-set, $\mathcal{M} = \{M_1, \ldots, M_\nu\}$, where each model M_i has an associated parameter $\boldsymbol{\theta}_i$.

We may know the different constituent parts of the system, X_i, and have measurements for some or all of them under some experimental designs, $\mathcal{T}$.

Inference and Models: Approximate vs. Exact

We have observed data, $\mathcal{D}$, that was generated by some system of in general unknown structure that we seek to describe by a mathematical model. In principle we can have a model-set, $\mathcal{M} = \{M_1, \dots, M_\nu\}$, where each model M_i has an associated parameter $\boldsymbol{\theta}_i$.
We may know the different constituent parts of the system, X_i, and have measurements for some or all of them under some experimental designs, $\mathcal{T}$.

$$\overbrace{\Pr(M_i|\mathcal{T},\mathcal{D})}^{\text{Model Posterior}}$$

Inference and Models: Approximate vs. Exact

We have observed data, $\mathcal{D}$, that was generated by some system of in general unknown structure that we seek to describe by a mathematical model. In principle we can have a model-set, $\mathcal{M} = \{M_1, \ldots, M_\nu\}$, where each model M_i has an associated parameter $\boldsymbol{\theta}_i$.

We may know the different constituent parts of the system, X_i, and have measurements for some or all of them under some experimental designs, $\mathcal{T}$.

$$\overbrace{\Pr(M_i|\mathcal{T},\mathcal{D})}^{\text{Model Posterior}} = \frac{\overbrace{\Pr(\mathcal{D}|M_i,\mathcal{T})}^{\text{Likelihood}}\overbrace{\pi(M_i)}^{\text{Prior}}}{\underbrace{\sum_{j=1}^{\nu}\Pr(\mathcal{D}|M_j,\mathcal{T})\pi(M_j)}_{\text{Evidence}}}$$

Inference and Models: Approximate vs. Exact

We have observed data, $\mathcal{D}$, that was generated by some system of in general unknown structure that we seek to describe by a mathematical model. In principle we can have a model-set, $\mathcal{M} = \{M_1, \ldots, M_\nu\}$, where each model M_i has an associated parameter $\boldsymbol{\theta}_i$.

We may know the different constituent parts of the system, X_i, and have measurements for some or all of them under some experimental designs, $\mathcal{T}$.

$$\overbrace{\Pr(M_i|\mathcal{T},\mathcal{D})}^{\text{Model Posterior}} = \frac{\overbrace{\Pr(\mathcal{D}|M_i,\mathcal{T})}^{\text{Likelihood}}\overbrace{\pi(M_i)}^{\text{Prior}}}{\underbrace{\sum_{j=1}^{\nu}\Pr(\mathcal{D}|M_j,\mathcal{T})\pi(M_j)}_{\text{Evidence}}}$$

For complicated models and/or detailed data the likelihood evaluation can become prohibitively expensive.

Inference and Models: Approximate vs. Exact

We have observed data, $\mathcal{D}$, that was generated by some system of in general unknown structure that we seek to describe by a mathematical model. In principle we can have a model-set, $\mathcal{M} = \{M_1, \ldots, M_\nu\}$, where each model M_i has an associated parameter $\boldsymbol{\theta}_i$.

We may know the different constituent parts of the system, X_i, and have measurements for some or all of them under some experimental designs, $\mathcal{T}$.

$$\overbrace{\Pr(M_i|\mathcal{T},\mathcal{D})}^{\text{Model Posterior}}=\frac{\overbrace{\Pr(\mathcal{D}|M_i,\mathcal{T})}^{\text{Likelihood}}\overbrace{\pi(M_i)}^{\text{Prior}}}{\underbrace{\sum_{j=1}^{\nu}\Pr(\mathcal{D}|M_j,\mathcal{T})\pi(M_j)}_{\text{Evidence}}}$$

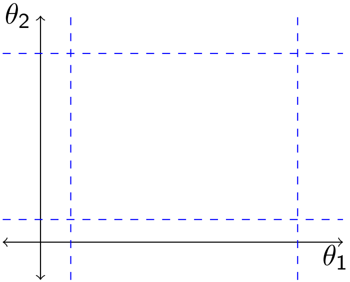
For complicated models and/or detailed data the likelihood evaluation can become prohibitively expensive.

Approximate Inference

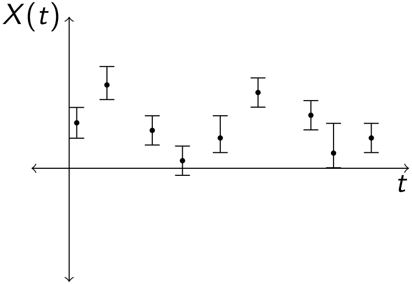
We can approximate the likelihood and/or the models. The “true” model is unlikely to be in $\mathcal{M}$ anyway.

Approximate Bayesian Computation

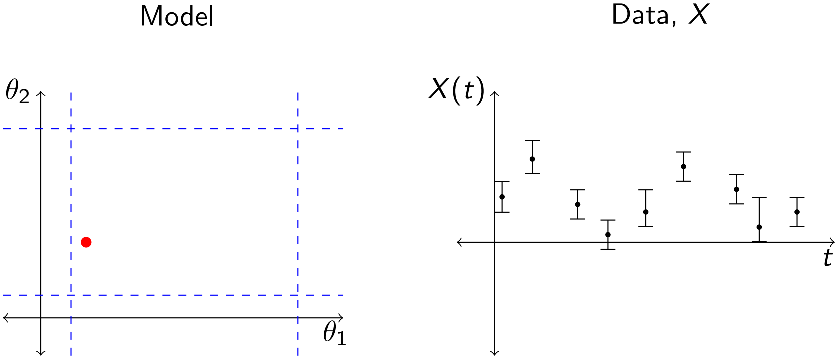
Model



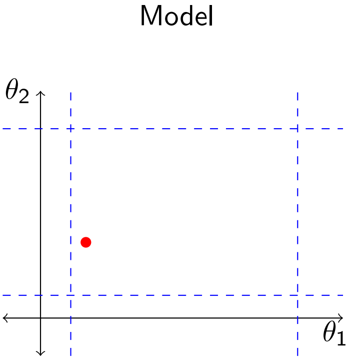
Data, X

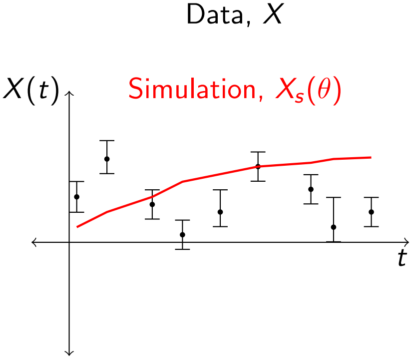


Approximate Bayesian Computation



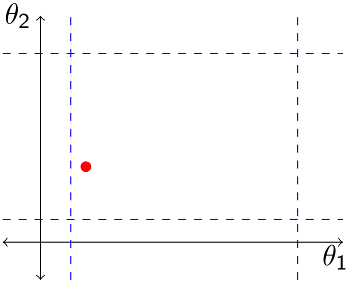
Approximate Bayesian Computation



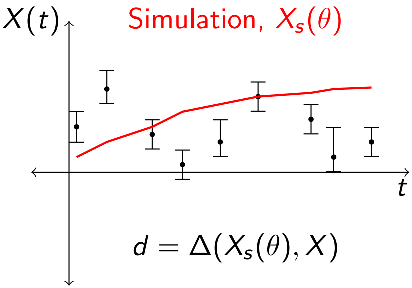


Approximate Bayesian Computation

Model



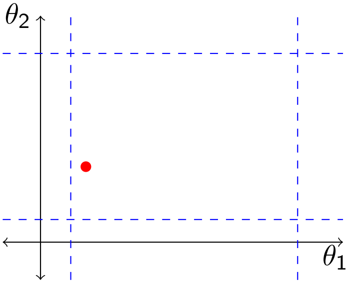
Data, X



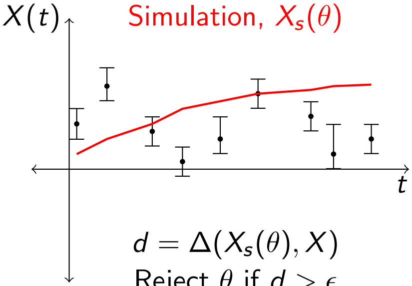
$$d = \Delta(X_s(\theta), X)$$

Approximate Bayesian Computation

Model



Data, X

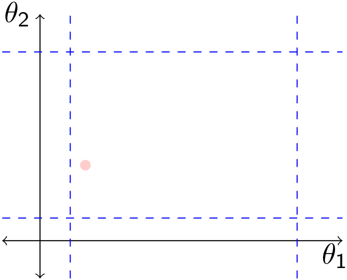


$$d = \Delta(X_s(\theta), X)$$

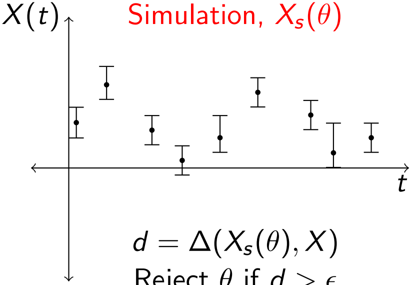
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

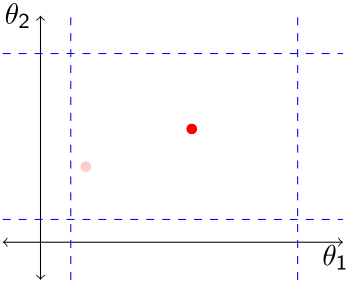


$$d = \Delta(X_s(\theta), X)$$

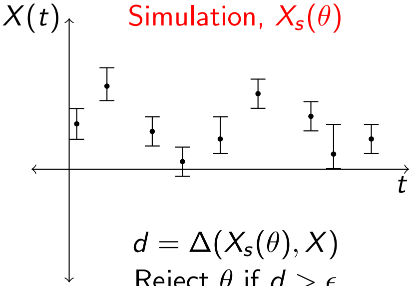
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

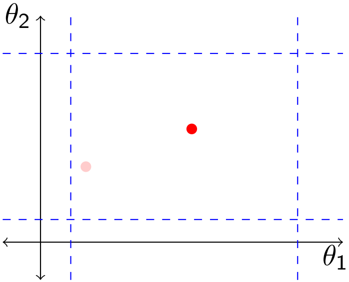


$$d = \Delta(X_s(\theta), X)$$

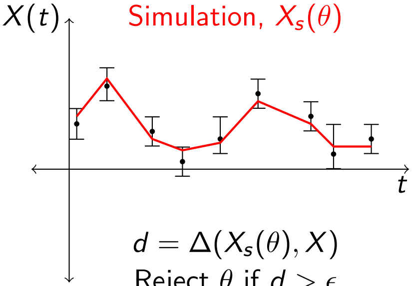
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

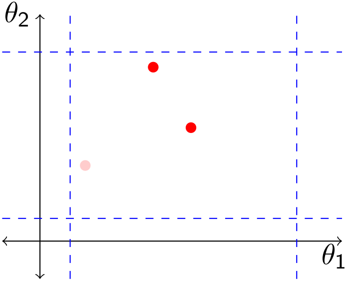


$$d = \Delta(X_s(\theta), X)$$

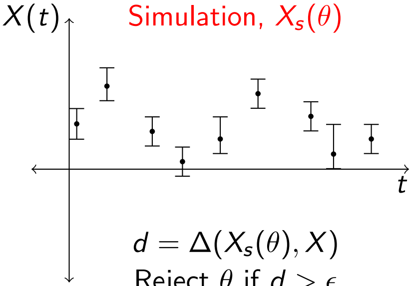
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

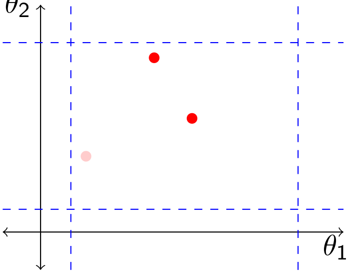


$$d = \Delta(X_s(\theta), X)$$

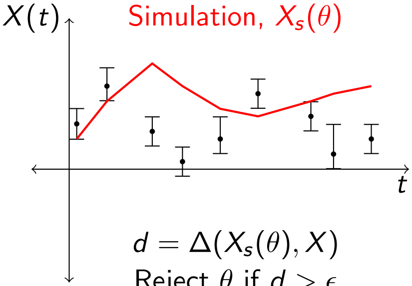
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

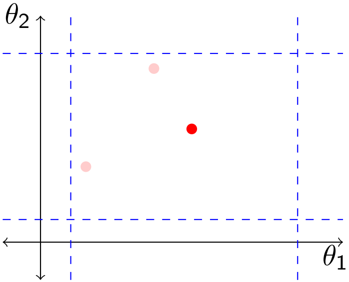


$$d = \Delta(X_s(\theta), X)$$

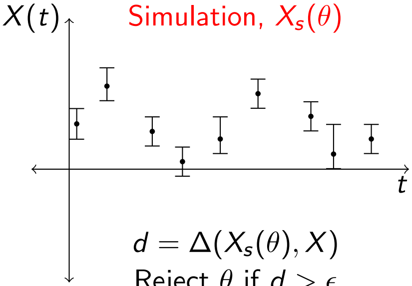
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

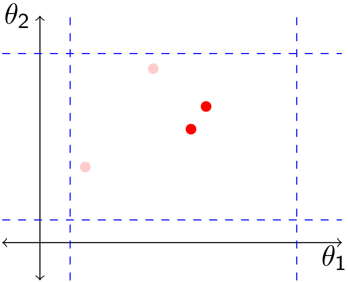


$$d = \Delta(X_s(\theta), X)$$

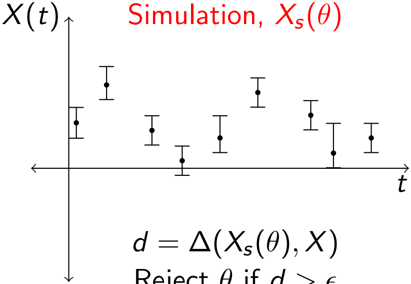
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X



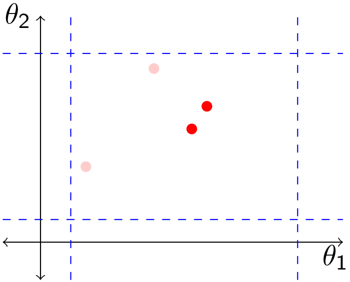
$d = \Delta(X_s(\theta), X)$

Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

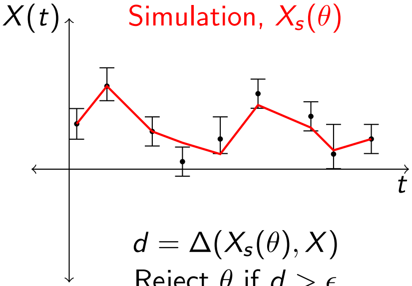
Imperial College London

Approximate Bayesian Computation

Model



Data, X

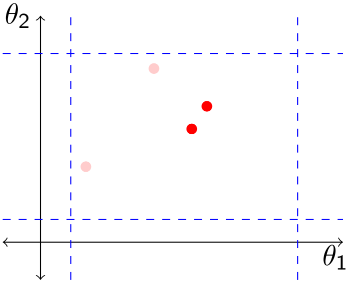


$$d = \Delta(X_s(\theta), X)$$

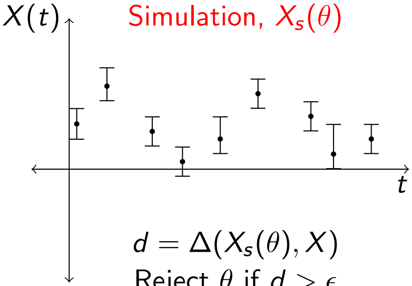
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X



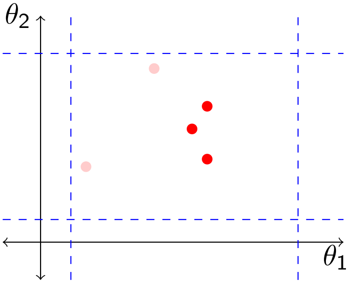
$$d = \Delta(X_s(\theta), X)$$

Reject θ if $d > \epsilon$

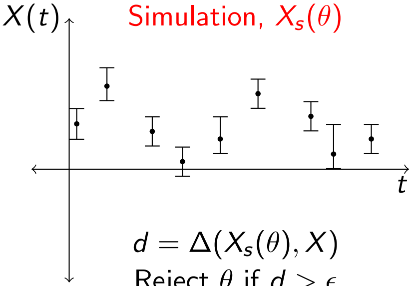
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X



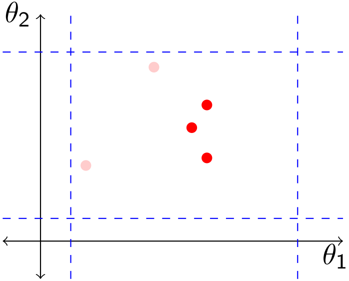
$$d = \Delta(X_s(\theta), X)$$

Reject θ if $d > \epsilon$

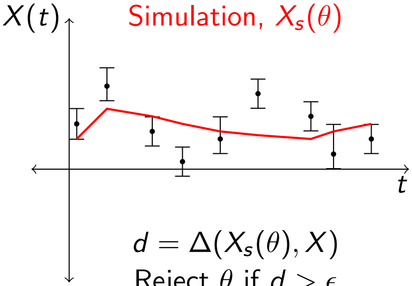
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

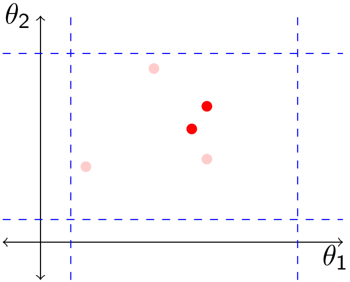


$$d = \Delta(X_s(\theta), X)$$

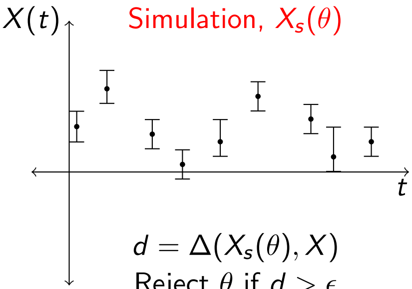
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

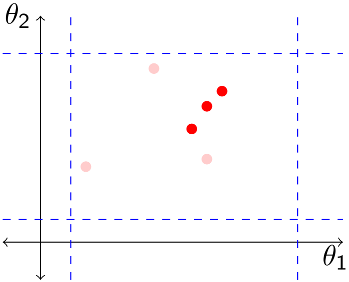


$$d = \Delta(X_s(\theta), X)$$

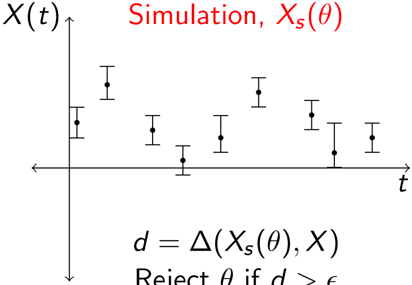
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X



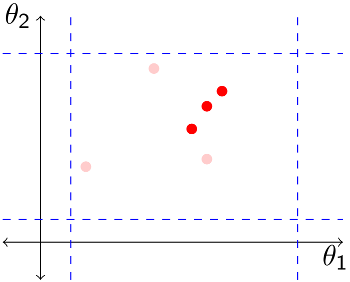
$d = \Delta(X_s(\theta), X)$

Reject θ if $d > \epsilon$

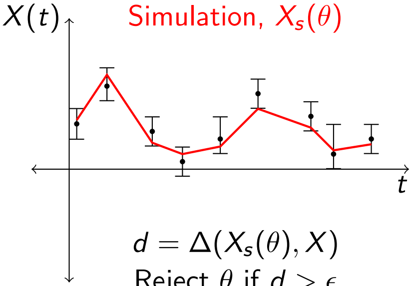
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



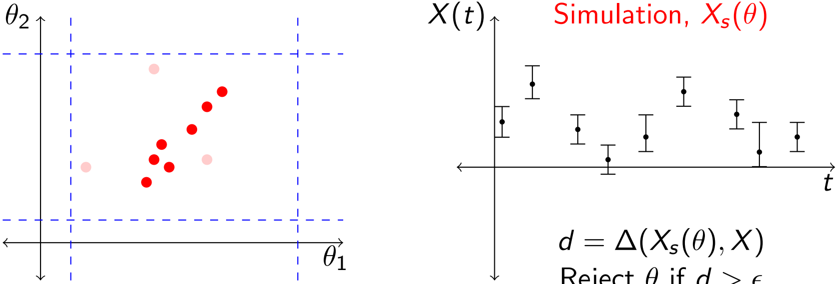
Data, X



$d = \Delta(X_s(\theta), X)$

Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation



Model

Data, X

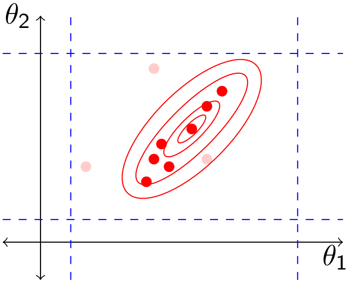
Simulation, $X_s(\theta)$

$$d = \Delta(X_s(\theta), X)$$

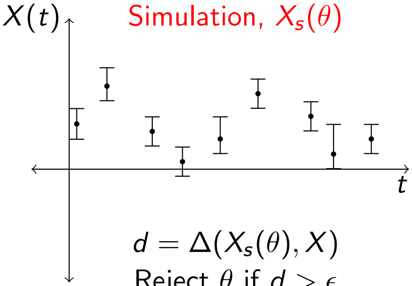
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



Data, X

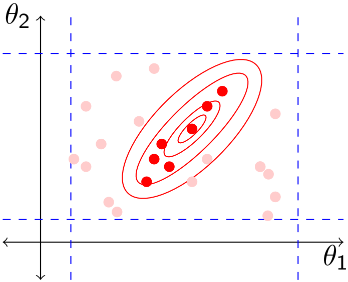


$$d = \Delta(X_s(\theta), X)$$

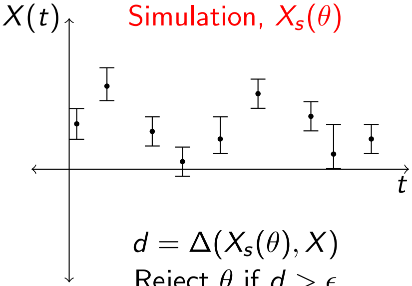
Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

Approximate Bayesian Computation

Model



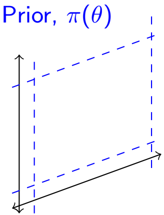
Data, X

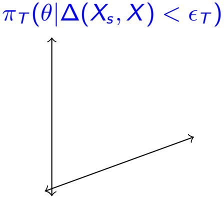


$$d = \Delta(X_s(\theta), X)$$

Reject θ if $d > \epsilon$
Accept θ if $d \leq \epsilon$

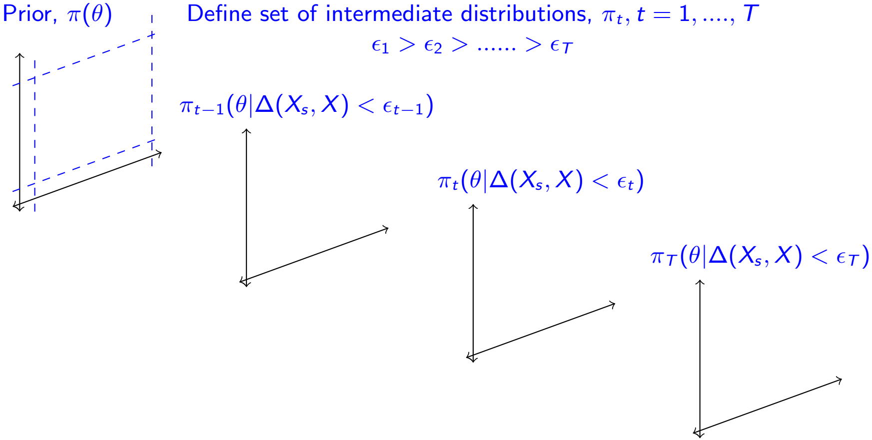
ABC SMC





Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC



Prior, $\pi(\theta)$

Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$

$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$

$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$

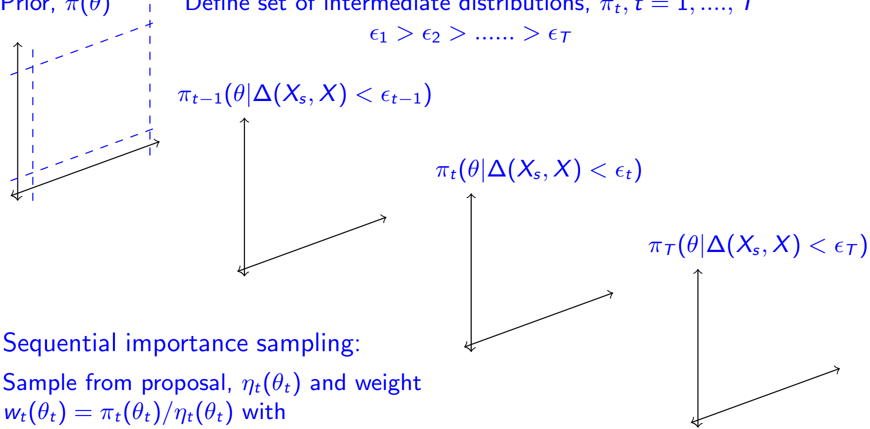
$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$

Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC

Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$



Prior, $\pi(\theta)$

$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$

$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$

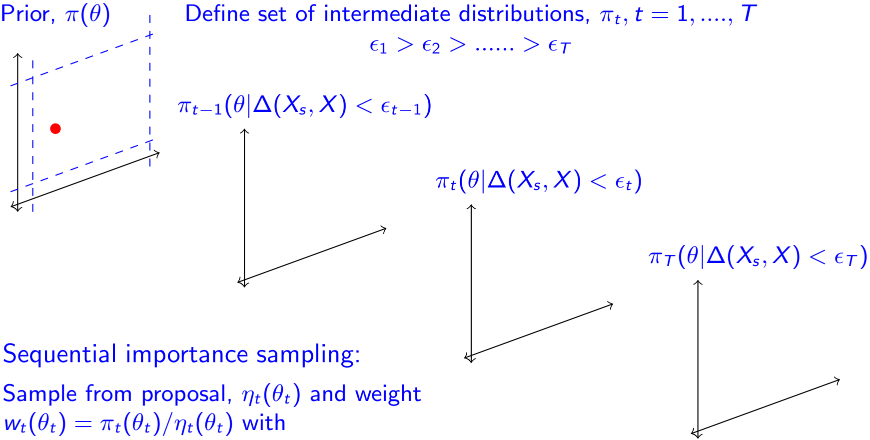
$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$

Sequential importance sampling:

Sample from proposal, $\eta_t(\theta_t)$ and weight
$w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$ with
$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$ where
$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC



Prior, $\pi(\theta)$

Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$

$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$

$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$

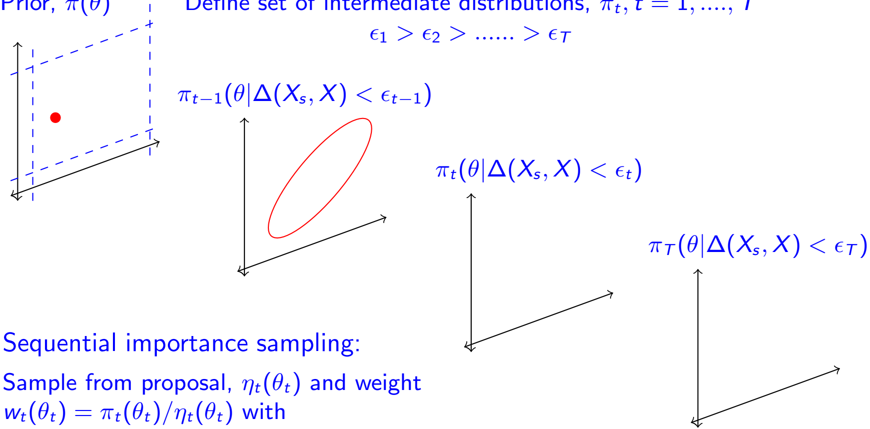
$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$

Sequential importance sampling:

Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$ with $\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$ where $K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC



Prior, $\pi(\theta)$

Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$

$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$

$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$

$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$

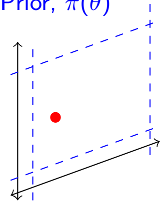
Sequential importance sampling:

Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$ with $\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$ where $K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC

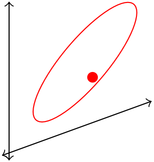
Prior, $\pi(\theta)$



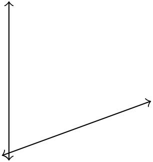
Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$

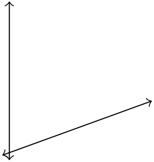
$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$



$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$



$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$

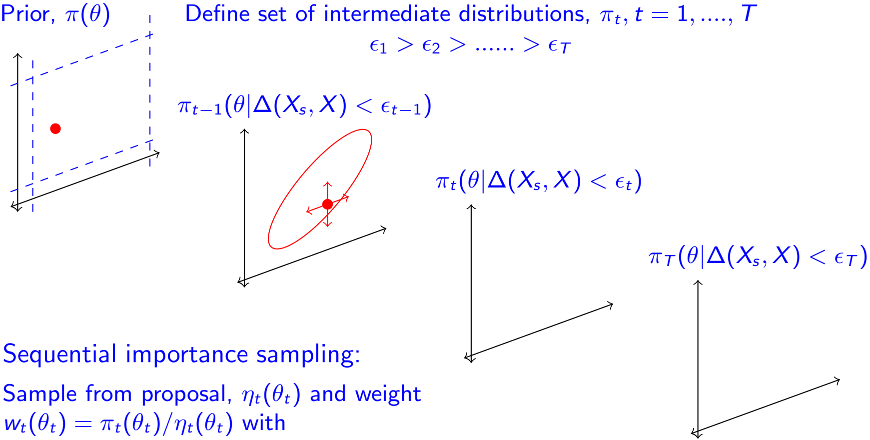


Sequential importance sampling:

Sample from proposal, $\eta_t(\theta_t)$ and weight
$w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$ with
$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$ where
$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC



Prior, $\pi(\theta)$

Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$

$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$

$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$

$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$

Sequential importance sampling:

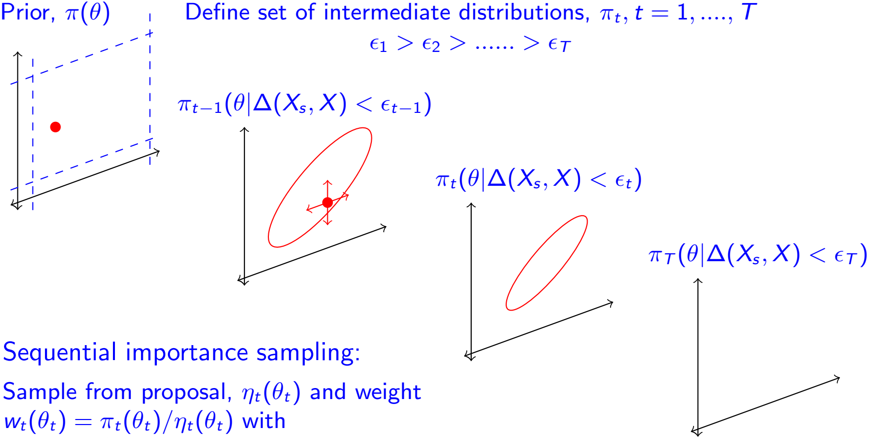
Sample from proposal, $\eta_t(\theta_t)$ and weight
$w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$ with
$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$ where
$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC

Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$



Prior, $\pi(\theta)$

$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$

$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$

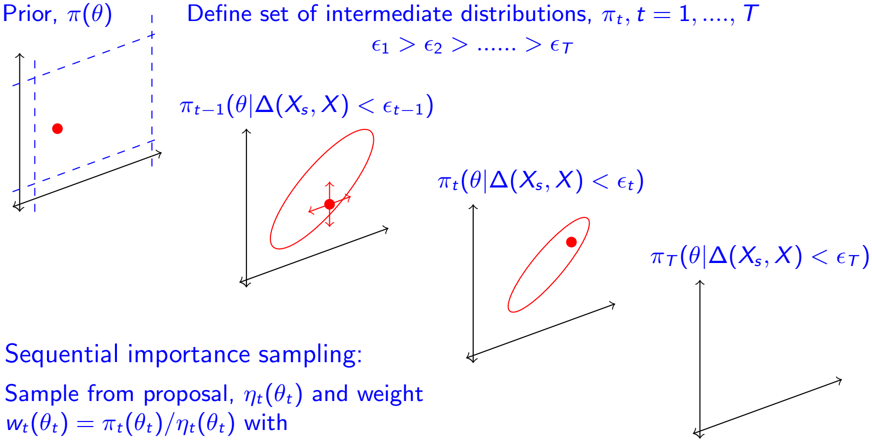
$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$

Sequential importance sampling:

Sample from proposal, $\eta_t(\theta_t)$ and weight
$w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$ with
$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$ where
$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC



Prior, $\pi(\theta)$

Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$

$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$

$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$

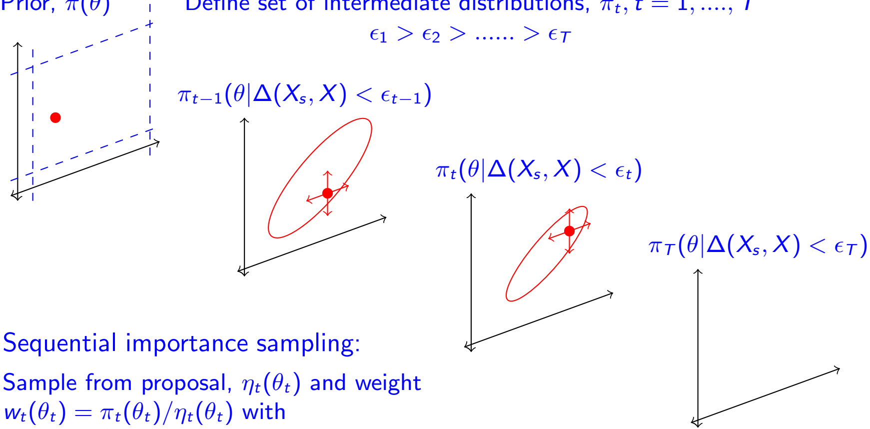
$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$

Sequential importance sampling:

Sample from proposal, $\eta_t(\theta_t)$ and weight
$w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$ with
$\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$ where
$K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC



Prior, $\pi(\theta)$

Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$

$\pi_{t-1}(\theta|\Delta(X_s, X) < \epsilon_{t-1})$

$\pi_t(\theta|\Delta(X_s, X) < \epsilon_t)$

$\pi_T(\theta|\Delta(X_s, X) < \epsilon_T)$

Sequential importance sampling:

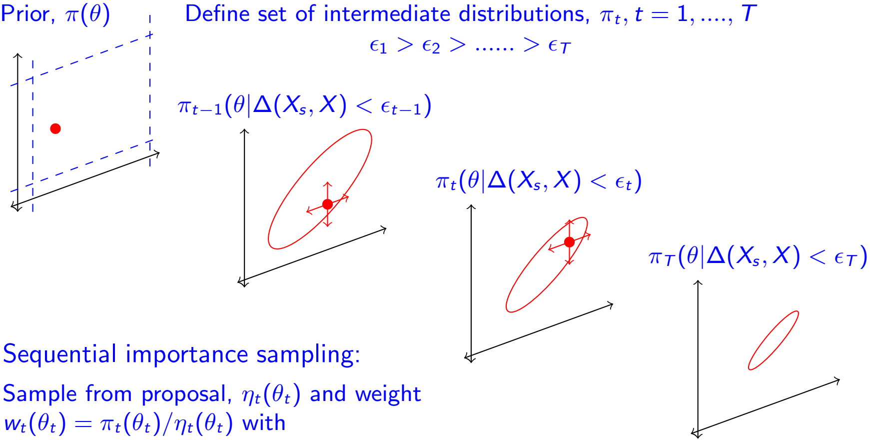
Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$ with $\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$ where $K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

Toni et al., J.Roy.Soc. Interface (2009).

ABC SMC

Define set of intermediate distributions, $\pi_t, t = 1, ..., T$

$$\epsilon_1 > \epsilon_2 > ...... > \epsilon_T$$



Sequential importance sampling:

Sample from proposal, $\eta_t(\theta_t)$ and weight $w_t(\theta_t) = \pi_t(\theta_t)/\eta_t(\theta_t)$ with $\eta_t(\theta_t) = \int \pi_{t-1}(\theta_{t-1})K_t(\theta_{t-1}, \theta_t)d\theta_{t-1}$ where $K_t(\theta_{t-1}, \theta_t)$ is Markov perturbation kernel

Toni et al., J.Roy.Soc. Interface (2009).

Model selection on a joint space

M_1 M_2 M_3 M_4

Toni & Stumpf, Bioinformatics (2010).

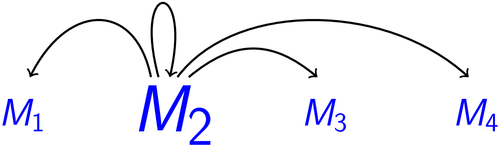
Model selection on a joint space

M^*

M_1 M_2 M_3 M_4

Toni & Stumpf, Bioinformatics (2010).

Model selection on a joint space

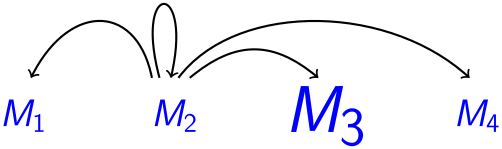


M^*

$M^{**} \sim KM(M|M^*)$

Toni & Stumpf, Bioinformatics (2010).

Model selection on a joint space



$$M^*$$

$$M^{**} \sim KM(M|M^*)$$

Toni & Stumpf, Bioinformatics (2010).

Model selection on a joint space

(M_3, θ_3)

(M_3, θ_7)

(M_3, θ_6)

(M_3, θ_2)

(M_3, θ_8)

(M_3, θ_5)

(M_3, θ_1)

(M_3, θ_4)

(M_3, θ_9)

$$M^*$$

$$M^{**} \sim KM(M|M^*)$$

$$\theta^*$$

Toni & Stumpf, Bioinformatics (2010).

Model selection on a joint space

(M_3, θ_3)

(M_3, θ_7)

(M_3, θ_6) (M_3, θ_2)

(M_3, θ_8) $\large (M_3, \theta_5)$ (M_3, θ_1)

(M_3, θ_4)

(M_3, θ_9)

M^*

$M^{**} \sim KM(M|M^*)$

θ^*

Toni & Stumpf, Bioinformatics (2010).

Model selection on a joint space

$$(M_3, \theta_3)$$

$$(M_3, \theta_7)$$

$$(M_3, \theta_6)$$

$$(M_3, \theta_2)$$

$$(M_3, \theta_8) \quad \boldsymbol{(M_3, \theta_5)} \quad (M_3, \theta_1)$$

$$(M_3, \theta_4)$$

$$(M_3, \theta_9)$$

$$M^*$$

$$M^{**} \sim KM(M|M^*)$$

$$\theta^*$$

$$\theta^{**} \sim KP(\theta|\theta^*)$$

Toni & Stumpf, Bioinformatics (2010).

Model selection on a joint space

$$(M^{**}, \theta^{**})$$

$$M^*$$

$$M^{**} \sim KM(M|M^*)$$

$$\theta^*$$

$$\theta^{**} \sim KP(\theta|\theta^*)$$

accept / reject

Toni & Stumpf, Bioinformatics (2010).

Model selection on a joint space

$$w(M^{**}, \theta^{**})$$

M^*

$M^{**} \sim KM(M|M^*)$

θ^*

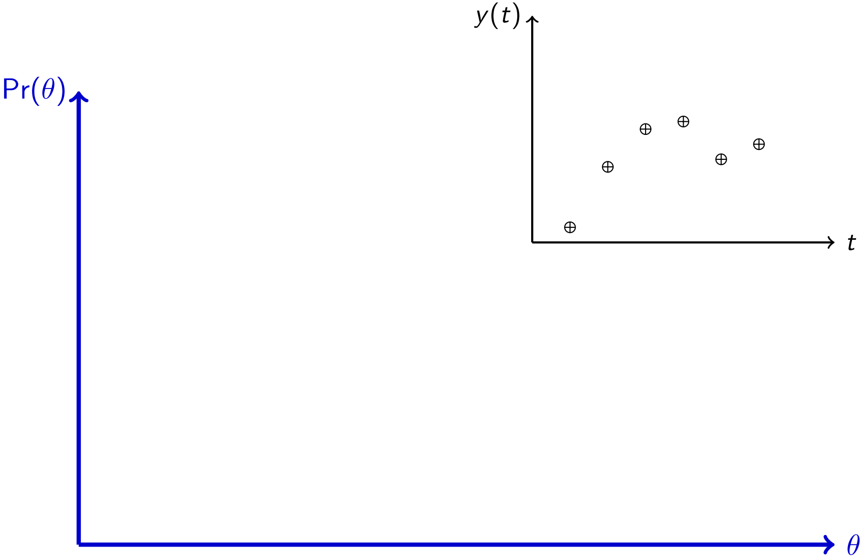
$\theta^{**} \sim KP(\theta|\theta^*)$

accept / reject

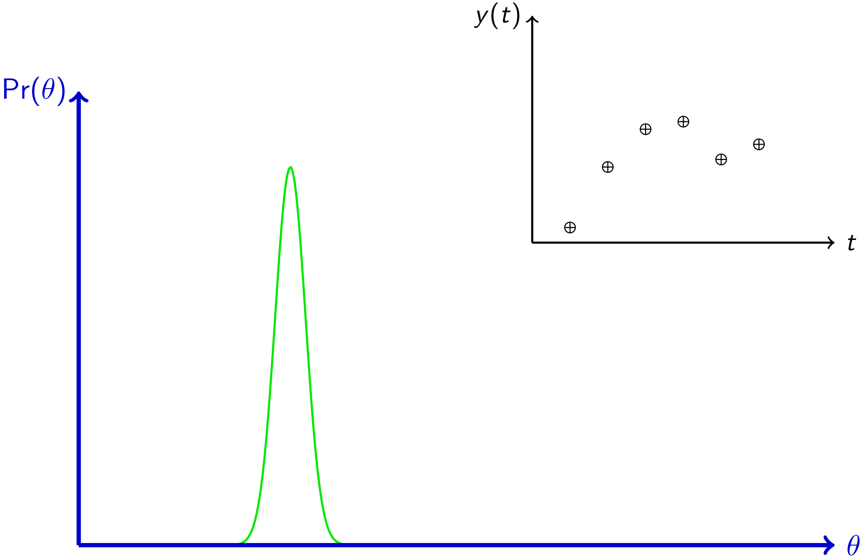
calculate w

Toni & Stumpf, Bioinformatics (2010).

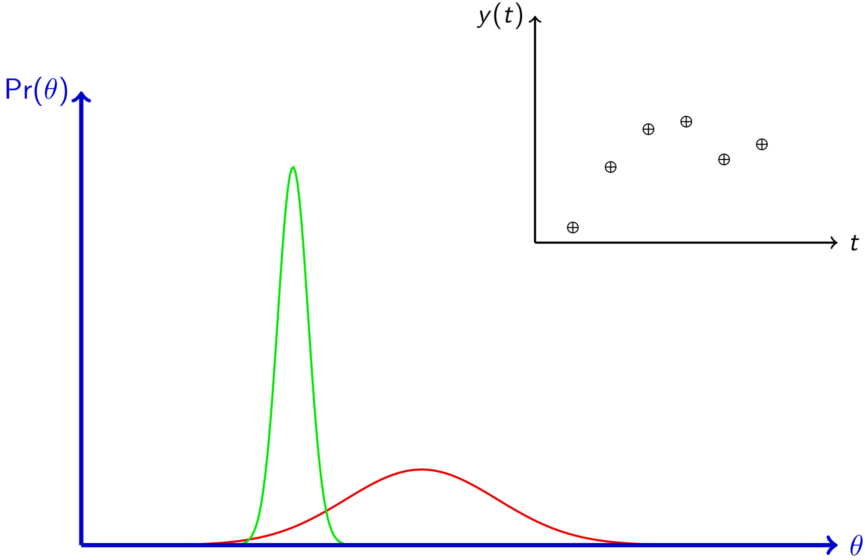
Optimal Models



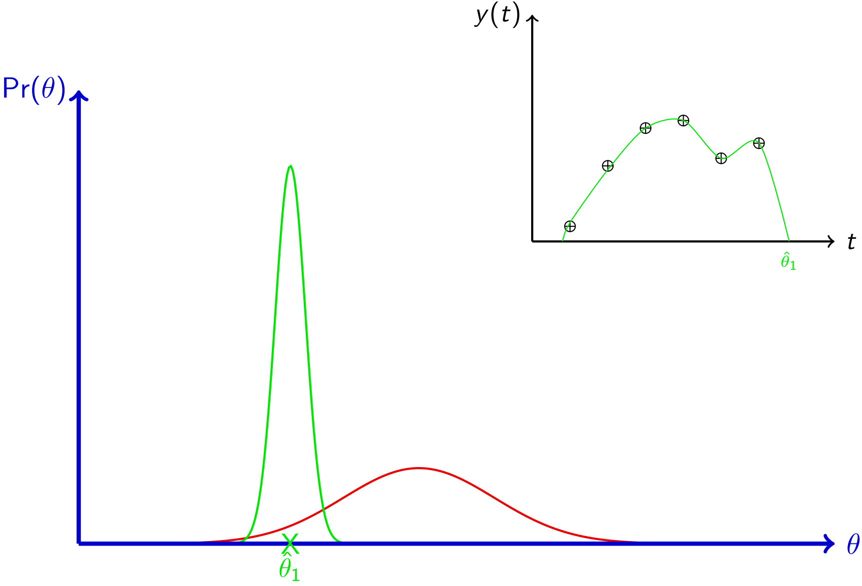
Optimal Models



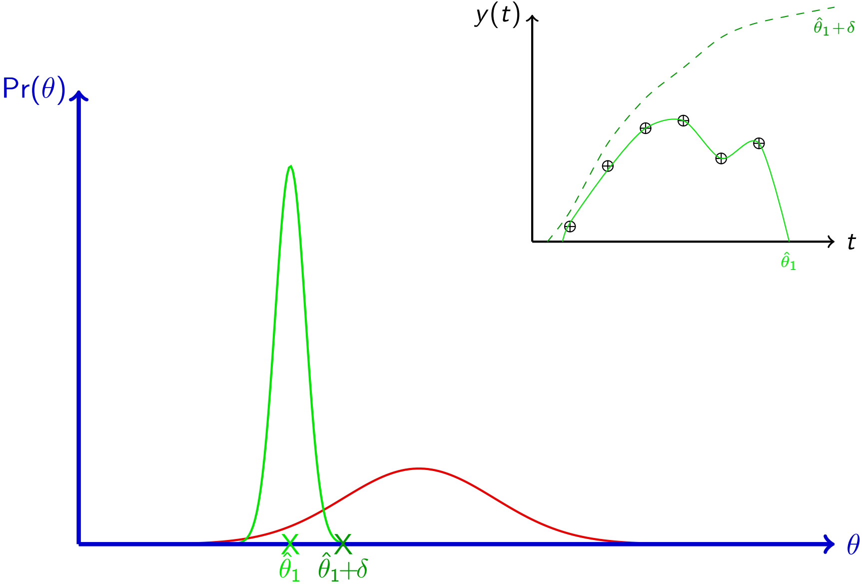
Optimal Models



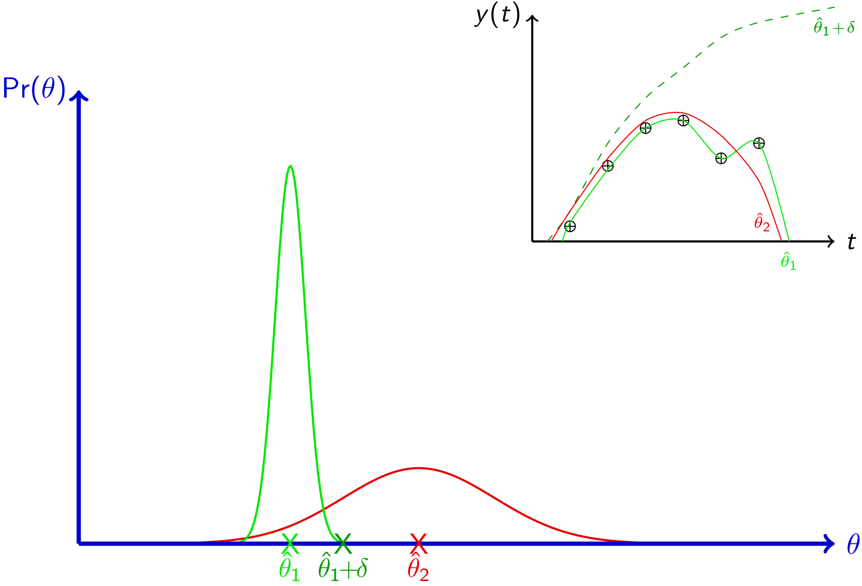
Optimal Models



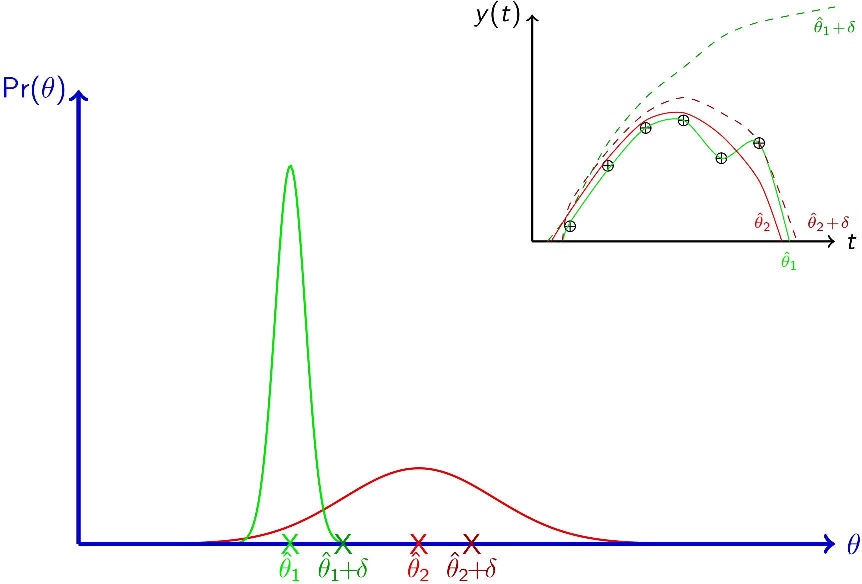
Optimal Models



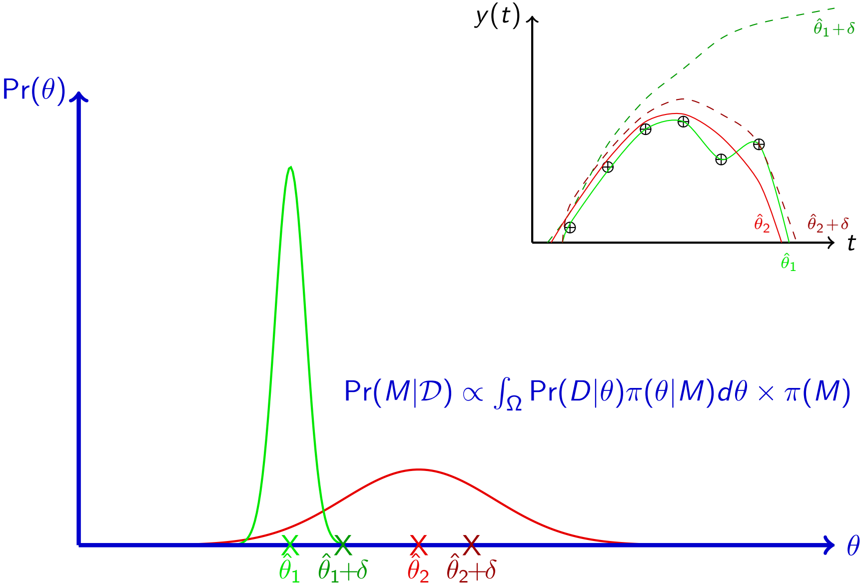
Optimal Models



Optimal Models



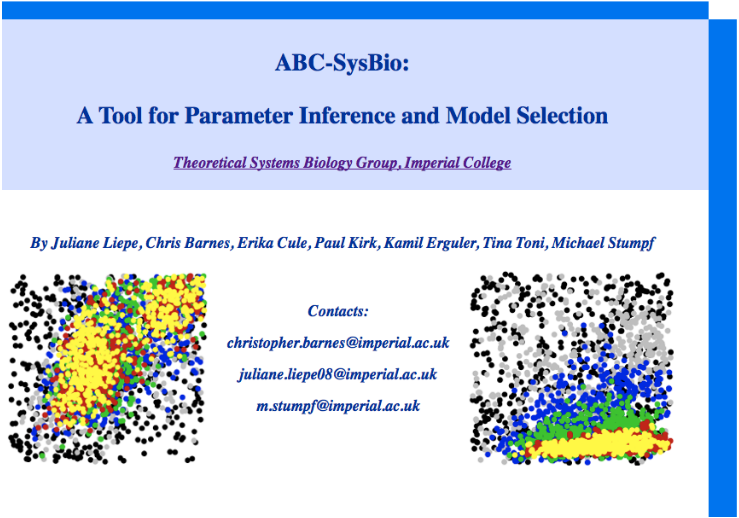
Optimal Models



$$\mathrm{Pr}(M|\mathcal{D}) \propto \int_\Omega \mathrm{Pr}(D|\theta)\pi(\theta|M)d\theta \times \pi(M)$$

ABC-SysBio

`http://abc-sysbio.sourceforge.net/`



ABC-SysBio:

A Tool for Parameter Inference and Model Selection

Theoretical Systems Biology Group, Imperial College

By Juliane Liepe, Chris Barnes, Erika Cule, Paul Kirk, Kamil Erguler, Tina Toni, Michael Stumpf

Contacts:

christopher.barnes@imperial.ac.uk

juliane.liepe08@imperial.ac.uk

m.stumpf@imperial.ac.uk

Liepe et al., Bioinformatics (2010).

What can ABC-SysBio do?

Input models in SBML format or python/CUDA code and supply time series data from which to infer parameters.

Model 1

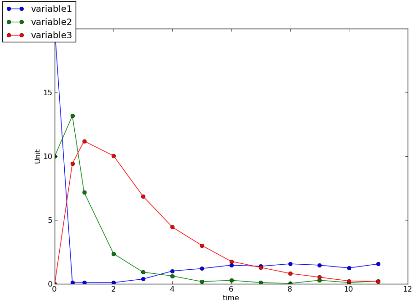
$$\dot{S} = \alpha - \gamma SI - dS$$
$$\dot{I} = \gamma SI - \upsilon I - dI$$
$$\dot{R} = \upsilon I - dR$$

Model 2

$$\dot{S} = \alpha - \gamma SI - dS$$
$$\dot{L} = \gamma SI - \delta L - dL$$
$$\dot{I} = \delta L - \upsilon I - dI$$
$$\dot{R} = \upsilon I - dR$$

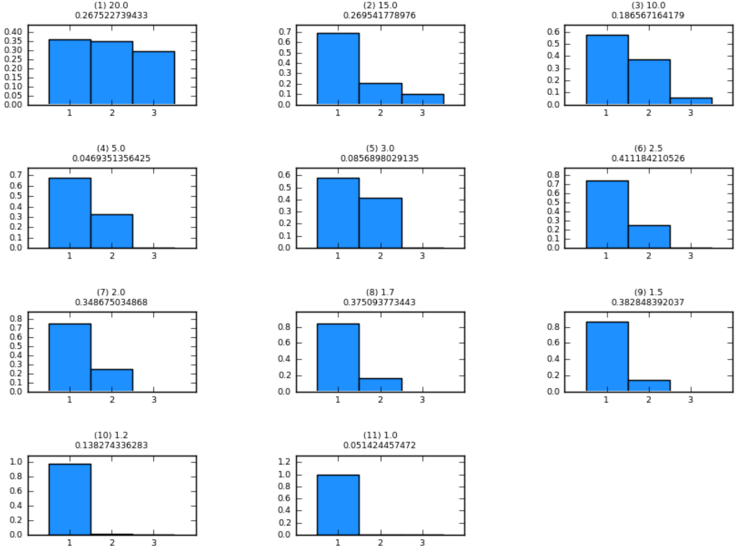
Model 3

$$\dot{S} = \alpha - \gamma SI - dS + eR$$
$$\dot{I} = \gamma SI - \upsilon I - dI$$
$$\dot{R} = \upsilon I - dR - eR$$


Data simulated from Model 1

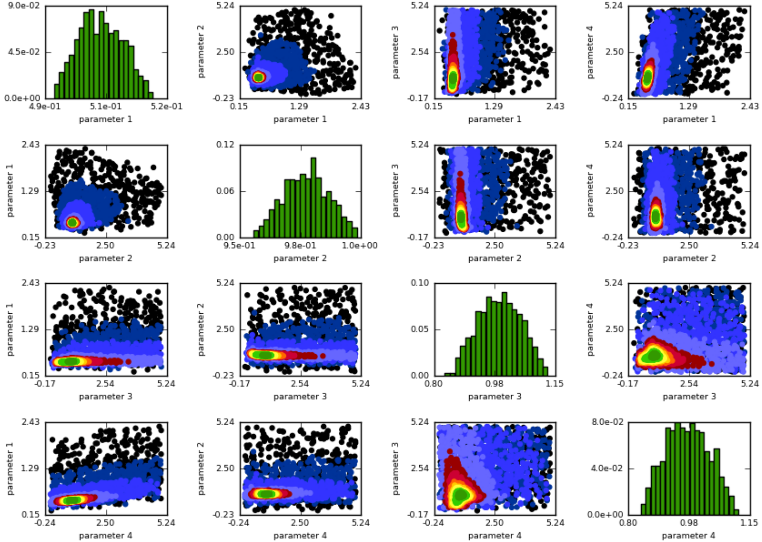
What can ABC-SysBio do?

Model selection



What can ABC-SysBio do?

Parameter inference



Computation on graphics processing units (GPUs)

GPUs are massively multithreaded many-core chips and provide a platform for cheap parallel computation.

Computation on graphics processing units (GPUs)

GPUs are massively multithreaded many-core chips and provide a platform for cheap parallel computation.

Multiple instruction multiple data (MIMD)

- Multiple independent processors execute different instructions on different data
- API, Clusters, GRID computing
- Main drawback: cost

Computation on graphics processing units (GPUs)

GPUs are massively multithreaded many-core chips and provide a platform for cheap parallel computation.

Multiple instruction multiple data (MIMD)

- Multiple independent processors execute different instructions on different data
- API, Clusters, GRID computing
- Main drawback: cost

Single instruction multiple data (SIMD)

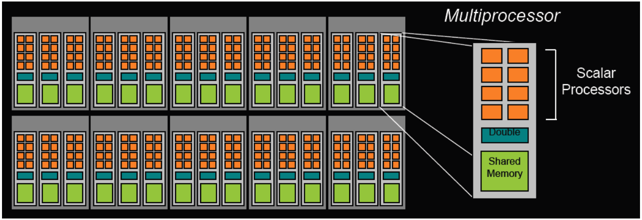
- Multiple processors execute same instruction on different data
- Supercomputers from 70s-80s based on this architecture
- GPUs follow this paradigm and are cheap
- Main drawback: Programming paradigm differs from CPU, not all applications can be accelerated

NVIDIA Compute Unified Device Architecture (CUDA)

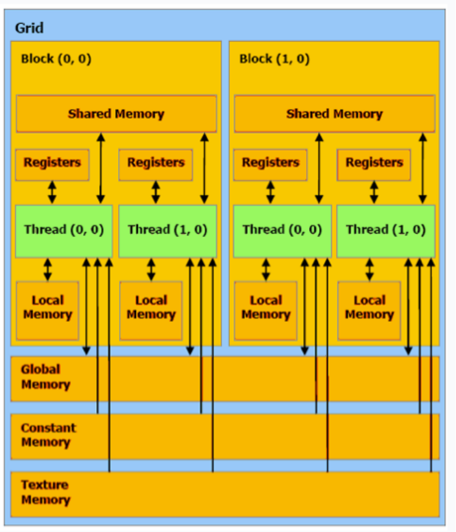
GPGPU: General purpose GPU

- GPUs evolved from dedicated computer graphics to general-purpose parallel processors
- Dedicated computation GPUs manufactured by NVIDIA, ATI

NVIDIA dedicated computational GPUs include the Tesla range.
Tesla C1060 : 30 multi x 8 (floating point) processors = 240 cores

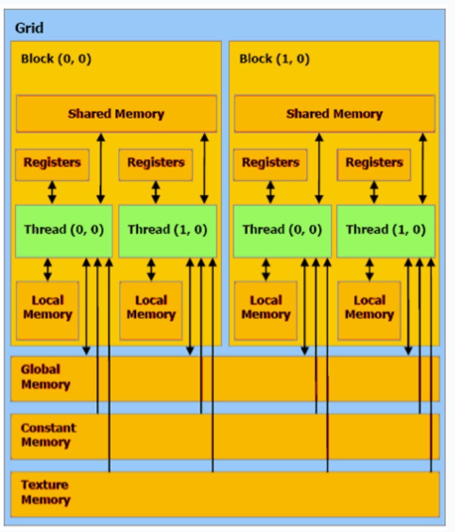


Threads, Blocks, Grids, Memory....



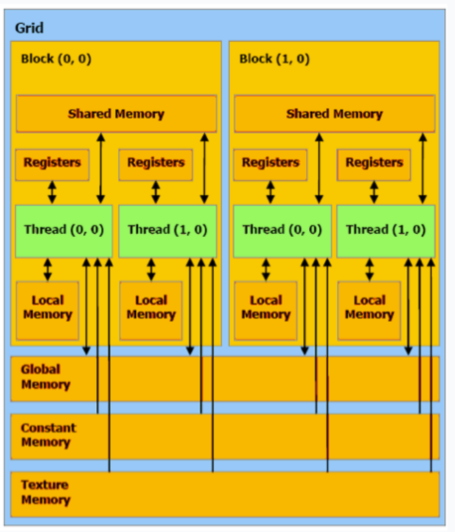
Imperial College London

Threads, Blocks, Grids, Memory....



GPU programs require the efficient use of different memories.
Data transfer between host (CPU) and device (GPU) must be minimized.

Threads, Blocks, Grids, Memory....

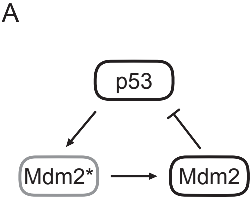


GPU programs require the efficient use of different memories.
Data transfer between host (CPU) and device (GPU) must be minimized.

	speed	scope
global	150x slower	dev, host
local	150x slower	thread
texture	faster (cached)	dev, host
constant	faster (cached)	dev, host
shared	fastest	block
registers	fastest	thread

p53 oscillations : simple negative feedback loop

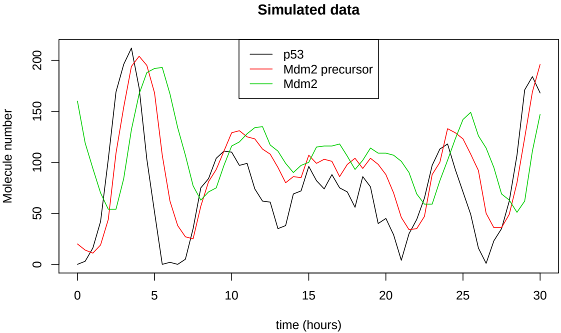
Geva-Zatorsky et al., Mol. Syst. Biol. (2006).



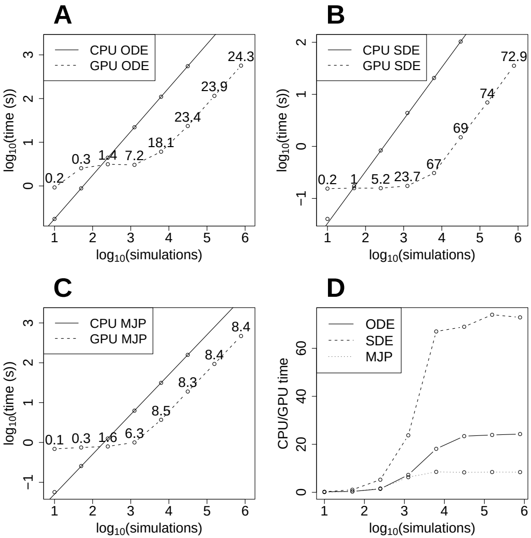
B

$$\dot{x} = \beta_x\Omega - \alpha_x x - \alpha_k y \frac{x}{x+k\Omega}$$
$$\dot{y_0} = \beta_y x - \alpha_0 y_0$$
$$\dot{y} = \alpha_0 y_0 - \alpha_y y$$

x : nuclear p53 (p53)	β_x : 0.9[1]
y_0 : Mdm2 precursor (Mdm2*)	β_y : 1.2[1]
y : nuclear Mdm2 (Mdm)	α_x : 0.3[1]
	α_y : 0.8[1]
$x(0)$: 0	α_0 : 0.9[1]
$y_0(0)$: 20	α_k : 1.7[1]
$y(0)$: 160	k : 0.0001
	Ω : 200



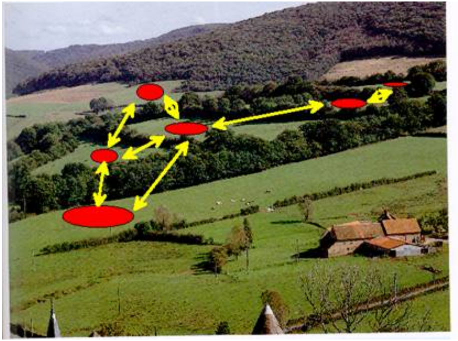
Timing improvements using CUDA + Tesla C1060



`http://cuda-sim.sourceforge.net/`

Mechanistic modelling of metapopulation dynamics

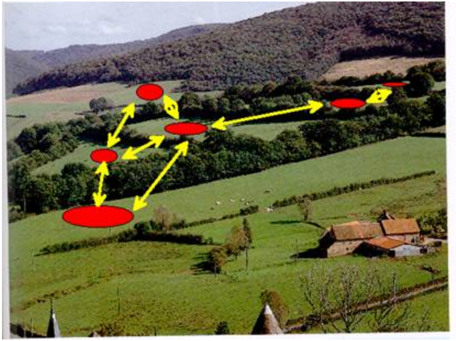
Predator-prey systems are unstable and prone to extinction but one mechanism for promoting stability is spatial heterogeneity.
A metapopulation is a set of linked sub populations or 'patches' and limited dispersal and asynchronous dynamics between patches can increase total persistence.



http://www.bio.uni-potsdam.de

Mechanistic modelling of metapopulation dynamics

Predator-prey systems are unstable and prone to extinction but one mechanism for promoting stability is spatial heterogeneity.
A metapopulation is a set of linked sub populations or 'patches' and limited dispersal and asynchronous dynamics between patches can increase total persistence.



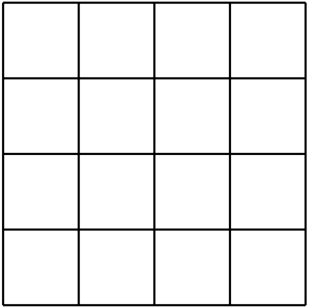
http://www.bio.uni-potsdam.de

Mechanistic modelling is important for understanding how to maximize metapopulation persistence and has obvious applications for conservation.

Experimental setup

Host-parasitoid systems

- Callosobruchus chinensis (bruchid beetle, bean weavils)
- Anisopteromalus calandrae (wasps)

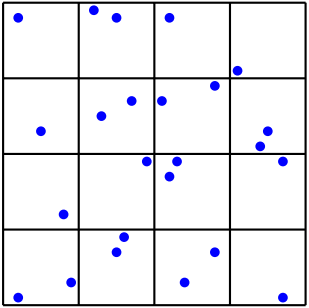


- Laboratory microcosm
- 4 × 4 clear plastic boxes (73×73×30 mm)

Experimental setup

Host-parasitoid systems

- Callosobruchus chinensis (bruchid beetle, bean weavils)
- Anisopteromalus calandrae (wasps)

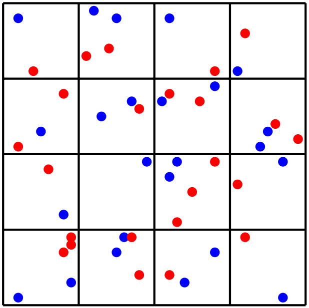


- Laboratory microcosm
- 4 × 4 clear plastic boxes (73×73×30 mm)
- Establish bruchid beetle on black eyed peas

Experimental setup

Host-parasitoid systems

- Callosobruchus chinensis (bruchid beetle, bean weavils)
- Anisopteromalus calandrae (wasps)



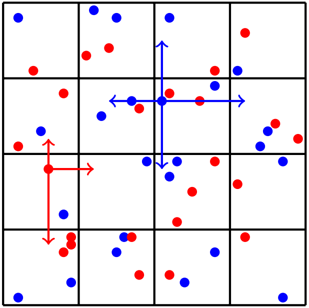
- Laboratory microcosm
- 4 × 4 clear plastic boxes (73×73×30 mm)
- Establish bruchid beetle on black eyed peas
- Introduce wasp populations

Imperial College London

Experimental setup

Host-parasitoid systems

- Callosobruchus chinensis (bruchid beetle, bean weavils)
- Anisopteromalus calandrae (wasps)

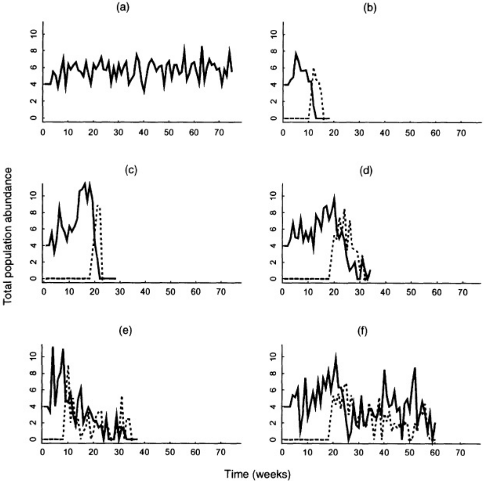


- Laboratory microcosm
- 4 × 4 clear plastic boxes (73×73×30 mm)
- Establish bruchid beetle on black eyed peas
- Introduce wasp populations
- Control inter cell migration using gates
  - Limited dispersal (3 hours per day)
  - Unlimited dispersal

Metapopulation structure affects persistence

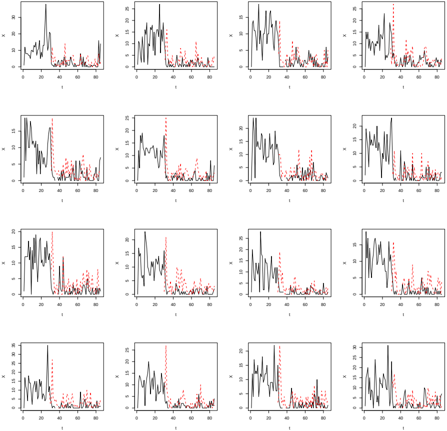
a. Prey in absence of predators
b. Single isolated system
c. Small metapopulation system unlimited dispersal
d. Small metapopulation system limited dispersal
e. Large metapopulation system unlimited dispersal
f. Large metapopulation system limited dispersal

(Bonsall *et al* J. Anim. Ecol. (2002))

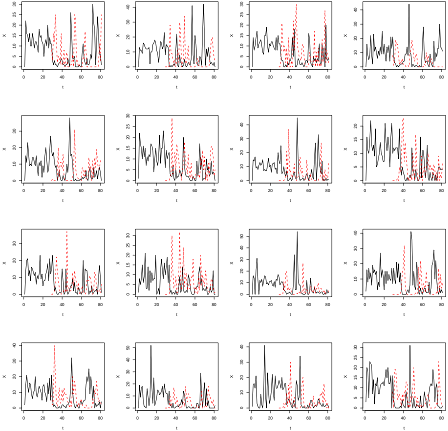


Time series data

Unlimited dispersal



Limited dispersal



Stochastic Model

Logistic growth + Lotka-Volterra interaction with migration

X_i, Y_i are beetles, wasps in cell i

process	hazard
$X_i \to 2X_i$	$b_1 X_i$
$X_i \to \varnothing$	$d_1 X_i^2$
$X_i + Y_i \to 2Y_i$	$pX_i Y_i$
$Y_i \to \varnothing$	$d_2 Y_i$
$X_i \to X_j$	$m_{Xc} X_i$
$X_i + X_i' \to X_i + X_j$	$m_{Xd} X_i^2$
$Y_i \to Y_j$	$m_{Yc} Y_i$
$Y_i + Y_i' \to Y_i + Y_j$	$m_{Yd} Y_i^2$

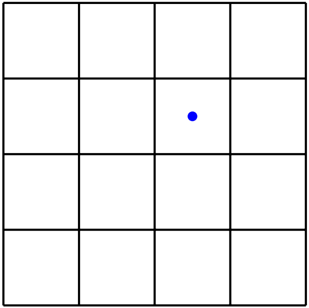
Stochastic Model

Logistic growth + Lotka-Volterra interaction with migration

X_i, Y_i are beetles, wasps in cell i

process	hazard
$X_i \rightarrow 2X_i$	$b_1 X_i$
$X_i \rightarrow \varnothing$	$d_1 X_i^2$
$X_i + Y_i \rightarrow 2Y_i$	$pX_i Y_i$
$Y_i \rightarrow \varnothing$	$d_2 Y_i$
$X_i \rightarrow X_j$	$m_{Xc} X_i$
$X_i + X_i' \rightarrow X_i + X_j$	$m_{Xd} X_i^2$
$Y_i \rightarrow Y_j$	$m_{Yc} Y_i$
$Y_i + Y_i' \rightarrow Y_i + Y_j$	$m_{Yd} Y_i^2$

Q: "Given a migration event occurs, where does the individual move?"



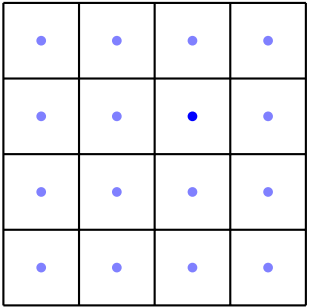
Stochastic Model

Logistic growth + Lotka-Volterra interaction with migration

X_i, Y_i are beetles, wasps in cell i

process	hazard
$X_i \to 2X_i$	$b_1 X_i$
$X_i \to \varnothing$	$d_1 X_i^2$
$X_i + Y_i \to 2Y_i$	$pX_i Y_i$
$Y_i \to \varnothing$	$d_2 Y_i$
$X_i \to X_j$	$m_{Xc} X_i$
$X_i + X_i' \to X_i + X_j$	$m_{Xd} X_i^2$
$Y_i \to Y_j$	$m_{Yc} Y_i$
$Y_i + Y_i' \to Y_i + Y_j$	$m_{Yd} Y_i^2$

Q: "Given a migration event occurs, where does the individual move?"



Global movement: $X_i \to X_j$ where $i \neq j$

Imperial College London

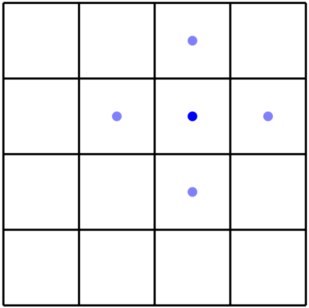
Stochastic Model

Logistic growth + Lotka-Volterra interaction with migration

X_i, Y_i are beetles, wasps in cell i

process	hazard
$X_i \rightarrow 2X_i$	$b_1 X_i$
$X_i \rightarrow \varnothing$	$d_1 X_i^2$
$X_i + Y_i \rightarrow 2Y_i$	$pX_i Y_i$
$Y_i \rightarrow \varnothing$	$d_2 Y_i$
$X_i \rightarrow X_j$	$m_{Xc} X_i$
$X_i + X_i' \rightarrow X_i + X_j$	$m_{Xd} X_i^2$
$Y_i \rightarrow Y_j$	$m_{Yc} Y_i$
$Y_i + Y_i' \rightarrow Y_i + Y_j$	$m_{Yd} Y_i^2$

Q: "Given a migration event occurs, where does the individual move?"

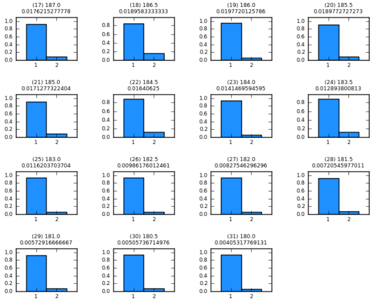


Local movement: $X_i \rightarrow X_j$ where $j \in$ nearest neighbours of i

Movement and migration models

Movement models

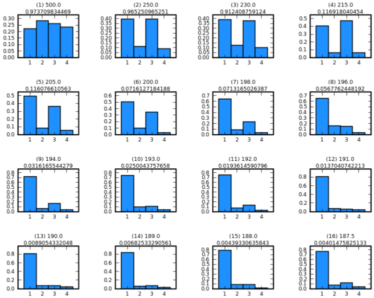
global : local



Global model has most support.

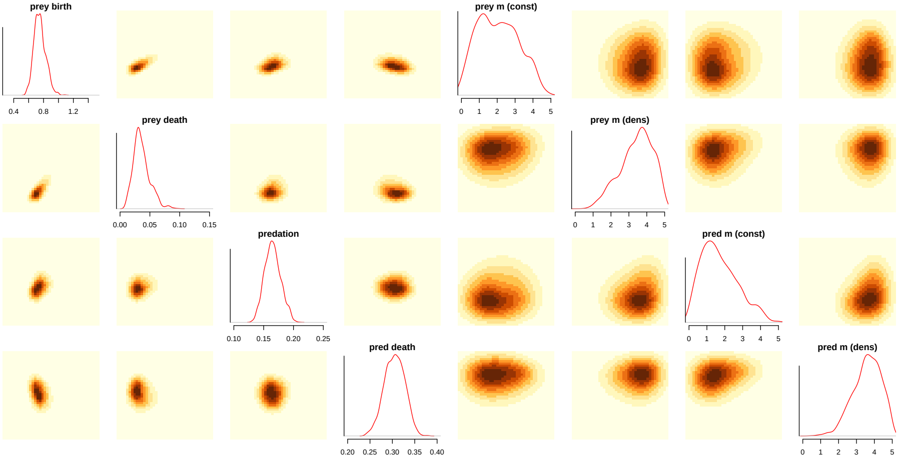
Migration models

d,d : c+d,d : d, c+d : c+d, c+d



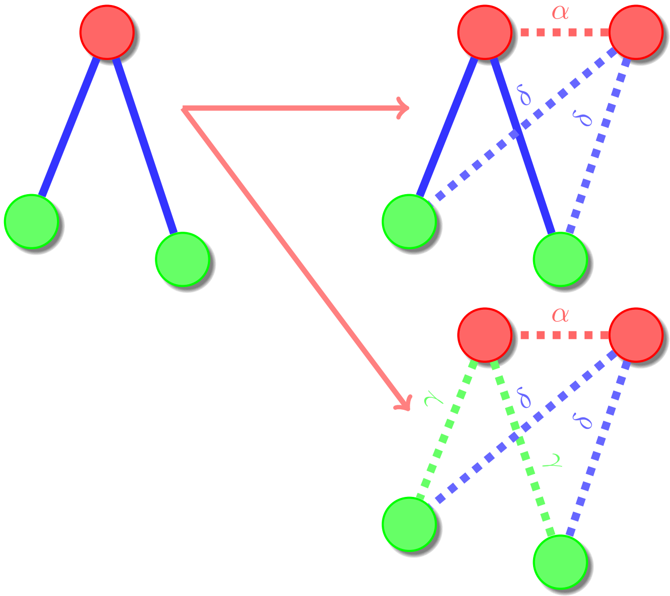
Density dependence dominates.

Inference: Parameters of the global movement model



Imperial College London

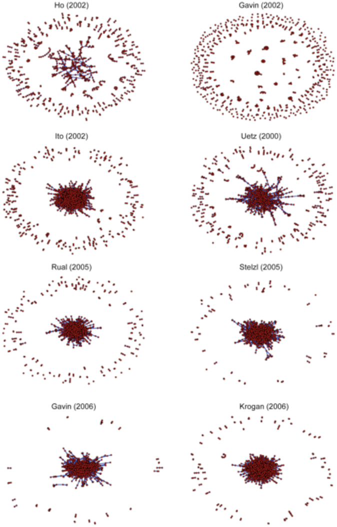
Network growth models



ABC on Networks

Summarizing Networks

- Data are noisy and incomplete.

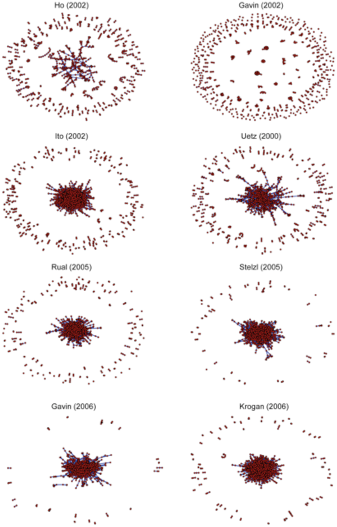


Stumpf & Wiuf, J. Roy. Soc. Interface (2010).

ABC on Networks

Summarizing Networks

- Data are noisy and incomplete.
- We can simulate models of network evolution, but this does not allow us to calculate likelihoods for all but very trivial models.

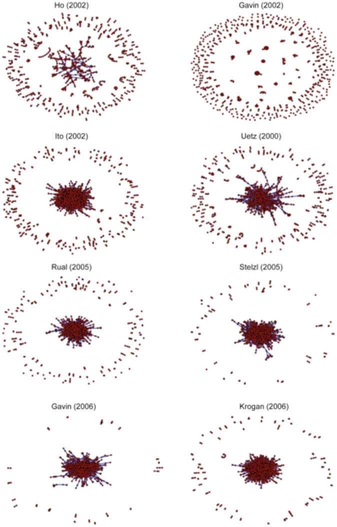


Stumpf & Wiuf, J. Roy. Soc. Interface (2010).

ABC on Networks

Summarizing Networks

- Data are noisy and incomplete.
- We can simulate models of network evolution, but this does not allow us to calculate likelihoods for all but very trivial models.
- There is also no sufficient statistic that would allow us to summarize networks, so ABC approaches require some thought.

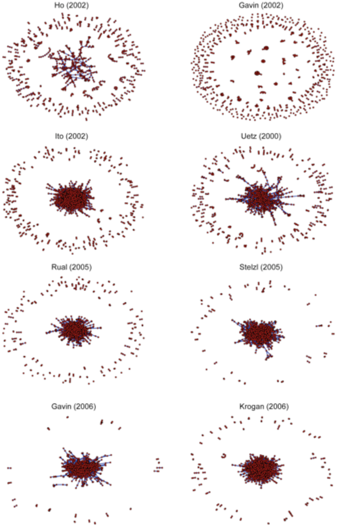


Stumpf & Wiuf, J. Roy. Soc. Interface (2010).

ABC on Networks

Summarizing Networks

- Data are noisy and incomplete.
- We can simulate models of network evolution, but this does not allow us to calculate likelihoods for all but very trivial models.
- There is also no sufficient statistic that would allow us to summarize networks, so ABC approaches require some thought.
- Many possible summary statistics of networks are expensive to calculate.



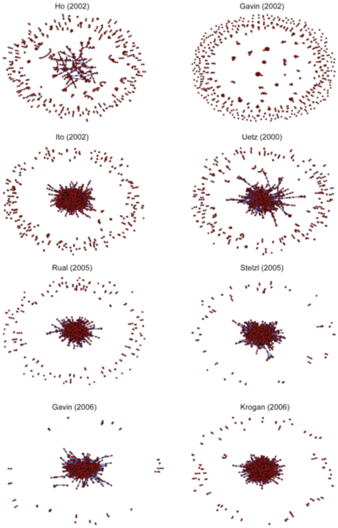
Stumpf & Wiuf, J. Roy. Soc. Interface (2010).

ABC on Networks

Summarizing Networks

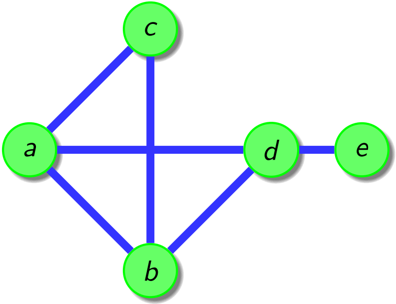
- Data are noisy and incomplete.
- We can simulate models of network evolution, but this does not allow us to calculate likelihoods for all but very trivial models.
- There is also no sufficient statistic that would allow us to summarize networks, so ABC approaches require some thought.
- Many possible summary statistics of networks are expensive to calculate.

Full likelihood: Wiuf et al., PNAS (2006).
ABC: Ratman et al., PLoS Comp.Biol. (2008).



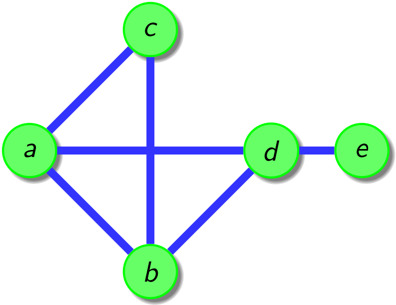
Stumpf & Wiuf, J. Roy. Soc. Interface (2010).

Graph Spectrum



$$A = \begin{array}{cccccc} a & b & c & d & e & \\ \left(\begin{array}{c} 0 \\ 1 \\ 1 \\ 1 \\ 0 \end{array}\right. & \begin{array}{c} 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{array} & \begin{array}{c} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{array} & \begin{array}{c} 1 \\ 1 \\ 0 \\ 0 \\ 1 \end{array} & \left.\begin{array}{c} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{array}\right) & \begin{array}{c} a \\ b \\ c \\ d \\ e \end{array} \end{array}$$

Graph Spectrum



$$A = \begin{array}{c} \\ \\ \end{array}\begin{array}{ccccc|c} a & b & c & d & e & \\ 0 & 1 & 1 & 1 & 0 & a \\ 1 & 0 & 1 & 1 & 0 & b \\ 1 & 1 & 0 & 0 & 0 & c \\ 1 & 1 & 0 & 0 & 1 & d \\ 0 & 0 & 0 & 1 & 0 & e \end{array}$$

Graph Spectra

Given a graph G comprised of a set of nodes N and edges $(i,j) \in E$ with $i,j \in N$, the adjacency matrix, A, of the graph is defined by

$$a_{i,j} = \begin{cases} 1 & \text{if } (i,j) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

The eigenvalues, λ, of this matrix provide one way of defining the graph spectrum.

Spectral Distances

A simple distance measure between graphs having adjacency matrices A and B, known as the edit distance, is to count the number of edges that are not shared by both graphs,

$$D(A, B) = \sum_{i,j} (a_{i,j} - b_{i,j})^2.$$

Spectral Distances

A simple distance measure between graphs having adjacency matrices A and B, known as the edit distance, is to count the number of edges that are not shared by both graphs,

$$D(A, B) = \sum_{i,j} (a_{i,j} - b_{i,j})^2.$$

However for unlabelled graphs we require some mapping h from $i \in N_A$ to $i' \in N_B$ that minimizes the distance

$$D(A, B) \geq D'_h(A, B) = \sum_{i,j} (a_{i,j} - b_{h(i),h(j)})^2,$$

Spectral Distances

A simple distance measure between graphs having adjacency matrices A and B, known as the edit distance, is to count the number of edges that are not shared by both graphs,

$$D(A,B) = \sum_{i,j} (a_{i,j} - b_{i,j})^2.$$

However for unlabelled graphs we require some mapping h from $i \in N_A$ to $i' \in N_B$ that minimizes the distance

$$D(A,B) \geq D'_h(A,B) = \sum_{i,j} (a_{i,j} - b_{h(i),h(j)})^2,$$

Given a spectrum we have

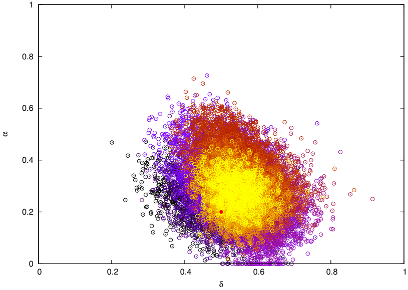
$$D'(A,B) = \sum_{l} (\lambda_l^{(\alpha)} - \lambda_l^{(\beta)})^2$$

Spectrum calculation would be prohibitive without using GPU + CUDA.

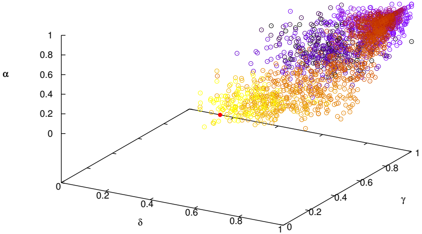
Estimating Parameters of Network Growth Models

We simulate networks, sample a dataset from it (e.g. interactions among 80% of nodes) and then try and infer the parameters of the evolutionary model.

Duplication Attachment (DA)



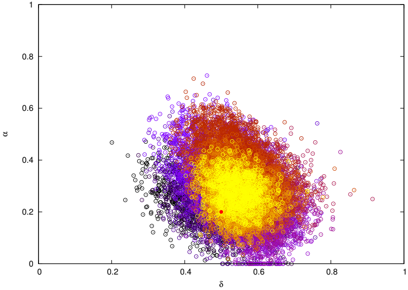
DA + complimentarity (DAC)



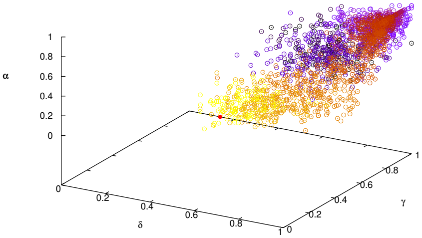
Estimating Parameters of Network Growth Models

We simulate networks, sample a dataset from it (e.g. interactions among 80% of nodes) and then try and infer the parameters of the evolutionary model.

Duplication Attachment (DA)



DA + complimentarity (DAC)

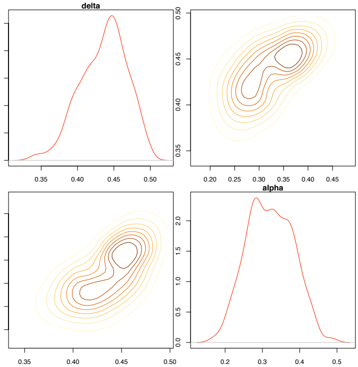


Inference of Evolutionary Parameters

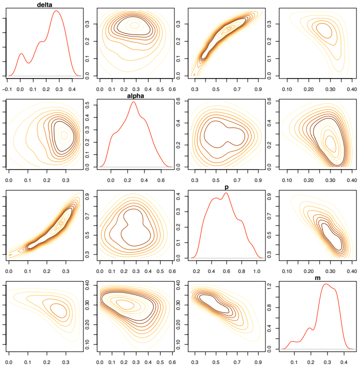
In all simulations the posterior over model parameters was reasonably large — as would be expected for evolutionary models, where the variance tends to overwhelm the mean.

Evolutionary Models of the Yeast PIN

Duplication Attachment (DA)

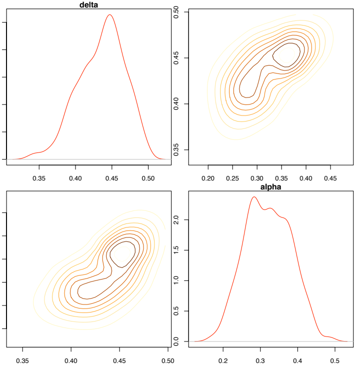


DAC + uniform (DAC + UR)

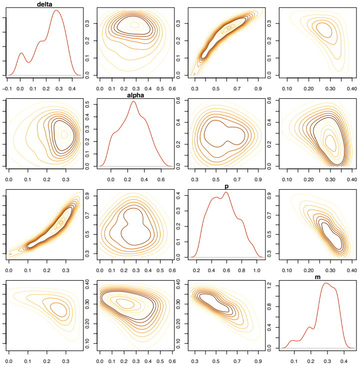


Evolutionary Models of the Yeast PIN

Duplication Attachment (DA)



DAC + uniform (DAC + UR)

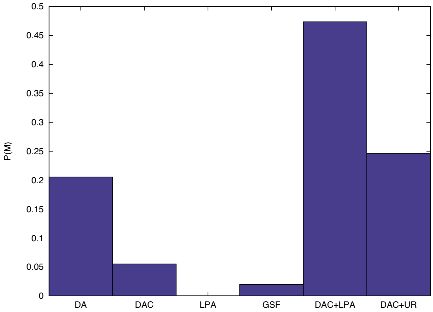


Data Uncertainties

Conditioning on incomplete data and known whole-genome duplication data can radically alter the inferred parameters.

While the spectrum appears to be a better summary statistic than others used so far, our ABC approach does not (yet) condition on other aspects of the data, such as functional annotations.

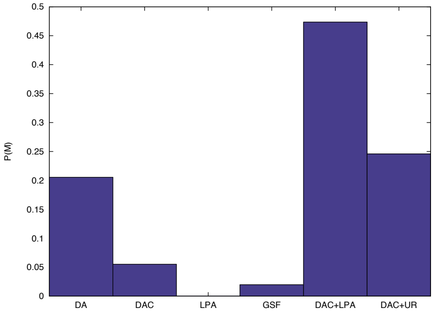
Selecting Models of Network Evolution



Probability of Models. (**DA**: very basic duplication; **DAC**: duplication preserving complementarity; **LPA**: linear preferential attachment; **GSF** - generalised scale free (change scaling coefficient between 2-3); **DAC+LPA**: combination of **DAC** and **LPA**; **DAC+UR**: DAC plus random attachment.

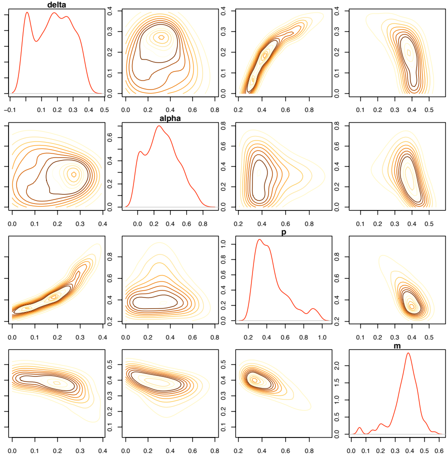
Most models have finite weight and only one, LPA, can be ruled out.

Selecting Models of Network Evolution



Probability of Models. (**DA**: very basic duplication; **DAC**: duplication preserving complementarity; **LPA**: linear preferential attachment; **GSF** - generalised scale free (change scaling coefficient between 2-3); **DAC+LPA**: combination of **DAC** and **LPA**; **DAC+UR**: DAC plus random attachment.

Most models have finite weight and only one, LPA, can be ruled out.



Thanks!

Acknowledgements

Yanxiang Zhou
Tom Thorne
Michael Bonsall
Michael Stumpf

bbsrc
biotechnology and biological sciences
research council

christopher.barnes@imperial.ac.uk

`http://www3.imperial.ac.uk/theoreticalsystemsbiology`

`http://abc-sysbio.sourceforge.net/`

`http://cuda-sim.sourceforge.net/`

Perturbation kernels

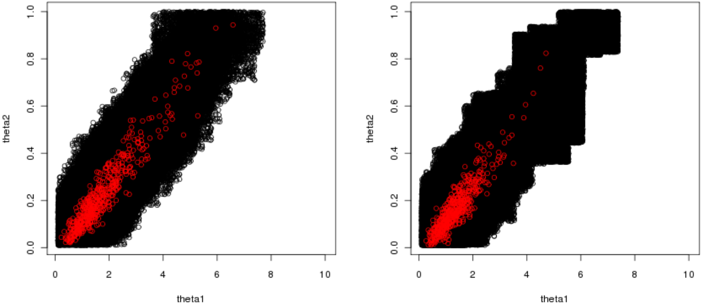
Componentwise independent random walk proposals

Empirical studies have shown, for each component i,

$$K(\theta_t^i|\theta_{t-1}^i) = \theta_{t-1}^i + \delta^i U(-1,1) \text{ where } \delta^i = \frac{1}{2}(\max\{\theta_{t-1}^i\} - \min\{\theta_{t-1}^i\})$$

provides good coverage for most applications.

However this can be very wasteful for correlated posteriors.



Perturbations in varimax rotation space

Rotate into a new basis O'

$$\theta'_{t-1} = V^T_{t-1}\theta_{t-1}$$

where V_{t-1} results from the spectral decomposition of the covariance matrix $\Sigma_{t-1}(\pi_{t-1})$

$$\Sigma_{t-1}(\pi_{t-1}) = V^T_{t-1}\Lambda_{t-1}(\pi_{t-1})V_{t-1}.$$

The importance weight calculation requires $K_t(\theta_{t-1}, \theta_t)$ which is the probability of observing the current particle given the previous population.
In the case of uniform, independent kernels, this reduces to two questions

- What is the volume defined by θ_{t-1}?
- Is θ_t contained within the volume defined by θ_{t-1}?

Perturbations in varimax rotation space

Rotate into a new basis O'

$$\theta'_{t-1} = V_{t-1}^T \theta_{t-1}$$

where V_{t-1} results from the spectral decomposition of the covariance matrix $\Sigma_{t-1}(\pi_{t-1})$

$$\Sigma_{t-1}(\pi_{t-1}) = V_{t-1}^T \Lambda_{t-1}(\pi_{t-1}) V_{t-1}.$$

The importance weight calculation requires $K_t(\theta_{t-1}, \theta_t)$ which is the probability of observing the current particle given the previous population.
In the case of uniform, independent kernels, this reduces to two questions

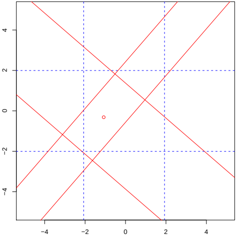
- What is the volume defined by θ_{t-1}?
- Is θ_t contained within the volume defined by θ_{t-1}?

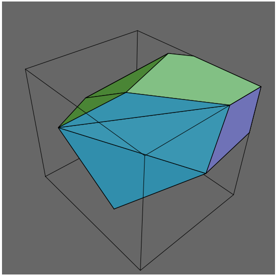
Standard computational geometry problems.

n-polytope volume calculation

Duality between $\mathcal{H}$ and $\mathcal{V}$ representations and typically P hard in one representation.

Using both representations is more efficient but there is a tradeoff between the reduction in samples and volume calculations. Expect volume calculation to become lengthy in high dimensional space.

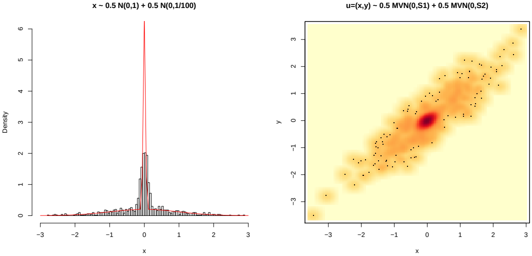


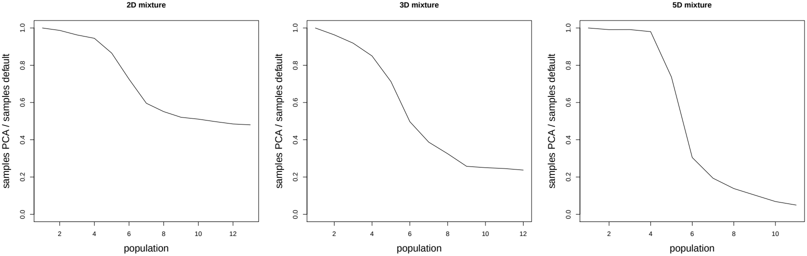


cddlib: `http://www.ifor.math.ethz.ch/~fukuda/cdd_home/`
Vinci: `http://www.math.u-bordeaux1.fr/~enge/Vinci.html`

B Bueler, A Enge, K Fukuda, Combinatorics and Computation (2000)

Toy mixture model testing





What is the Point of Such Simple Models?

- We can gain generic insights into evolutionary processes underlying the architecture of networks.
- We can use Bayesian model averaging in order to make predictions.

$$E_{\mathrm{BMA}}[Q] = \sum_{i=1}^{\nu} \mathrm{Pr}(M_i|\mathcal{D})E_{M_i}[Q]$$

True Models?

Even if the correct model is not included among the ν candidates, we can often obtain reliable predictions.
BMA trades in explanatory for predictive power, but allows us to predict structural and organizational properties of PINs.

p53 oscillations : parameter inference using SDE's