Approximate Bayesian Computation

Modelling Complex Systems

What do we want to do?

- ▶ Understand the structure of a model
- ▶ Compare models and choose between them
- ▶ Validate and criticise models
- ▶ Make predictions
- ▶ Make decisions

Likelihood

- In a likelihood-based analysis we aim to maximise the use of the information in the data (likelihood principle)

- Bayesian inference extends likelihood-based modelling to explicitly include a prior probability distribution on parameters.

- Amenable to robust model criticism (*e.g.* posterior predictive checking).

Motivation for ABC

▶ We would like to make use of the likelihood function and Bayesian framework.

▶ For many problems it is easy to simulate data that 'looks like' real data (*e.g.* agent-based models), but impossible to write down the likelihood function in practice.

▶ However it is still possible to carry out likelihood-based and Bayesian inference via simulations (if only approximately).

Approximate Bayesian Computation

•Uses less information than full data-based methods.

•Easier to implement and program (don't need to know the likelihood for a given genealogical history).

Replace the data with summary statistics

Key Points:

•For most problems, we cannot hope to simulate data that are identical to the 'true' data.

•But similar data may have similar posterior distributions.

•If we replace the data with summary statistics, then it is easier to decide how 'similar' data sets are to each other.

Example Applications of ABC

- Fitting and comparing models of cultural evolution using archaeological data (Crema *et al*, *J. Archaeol. Sci*, 2014).

- Modelling epidemiology of bovine tuberculosis in lion populations in Kruger National Park (Kosmala *et al*, *Ecological Applications*, 2015).

- Inferring cosmological parameters from Type Ia Supernova data (Weyant *et al*, *Astrophysical Journal*, 2013).

- Parameterising emulators to model climate sensitivity (effect of doubling CO_2 on equilibrium mean global temperature) (Holden *et al*, *Climate Dynamics*, 2010)

Posterior Distributions Without Likelihoods

If the data are discrete and of low dimension then it is possible to sample from the posterior density of the parameter without an explicit likelihood function, and without approximation, by the following algorithm (Rubin, 1984):

Monte Carlo Algorithm

Given observation y, repeat the following until N points have been accepted:

1. Draw $\theta_i \sim \pi(\theta)$
2. Simulate $x_i \sim p(x|\theta_i)$
3. Reject θ_i if $x_i \neq y$

These are sampled from $p(\theta|y)$.

Approximate Bayesian Computation

With high-dimensional data two approximations are useful:

- ▶ Define a distance function $\rho(\cdot)$
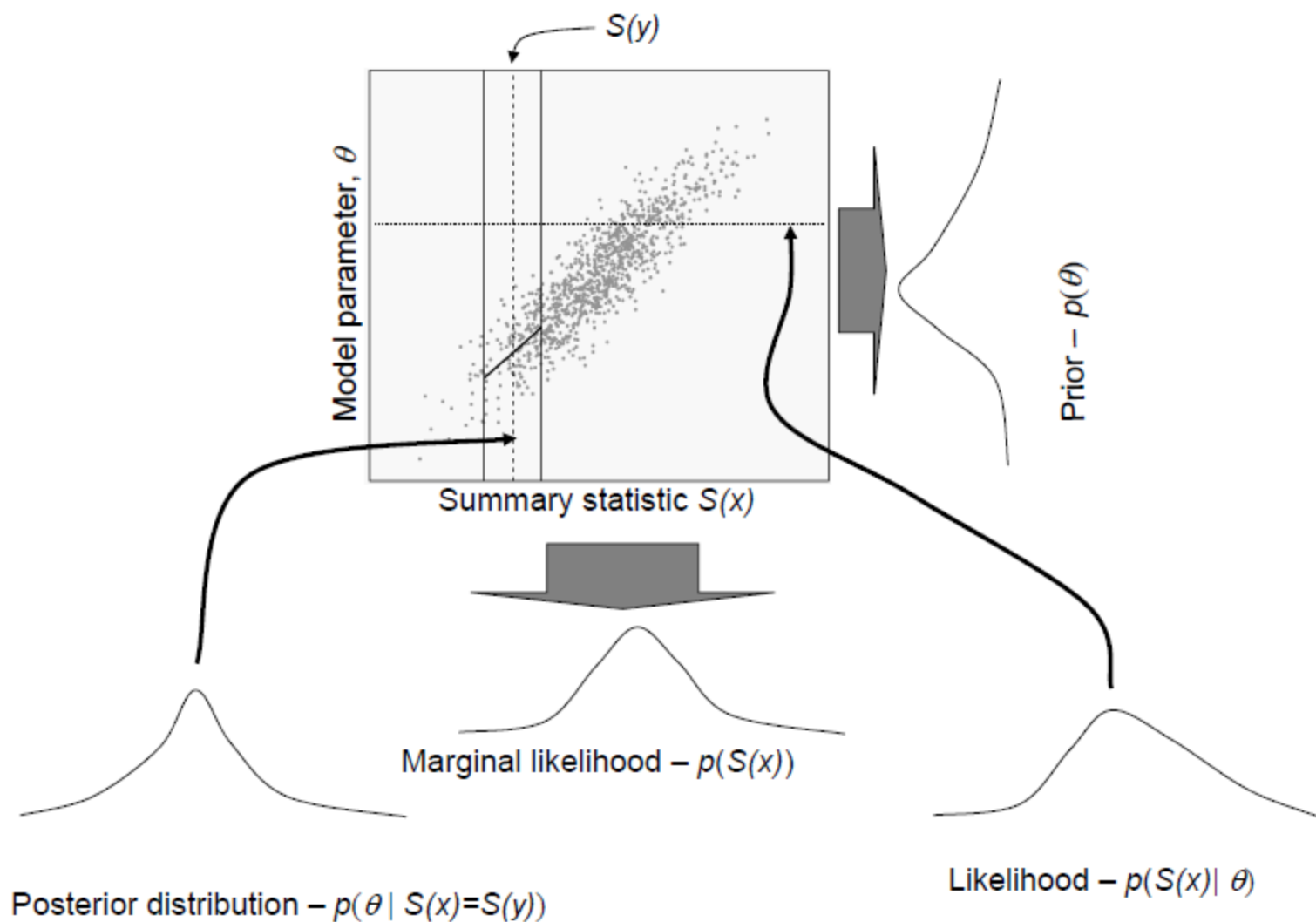- ▶ Define a summary statistic function $S(\cdot)$

ABC Algorithm

Given observation y, repeat the following until N points have been accepted:

1. Draw $\theta_i \sim \pi(\theta)$
2. Simulate $x_i \sim p(x|\theta_i)$
3. Reject θ_i if $\rho(S(x_i), S(y)) > \epsilon$.

Here, as with x, y, and θ, the function $S(\cdot)$ may be vector-valued.

ABC as conditional density estimation

ABC methodology

There are three main computational approaches for ABC:

- Rejection and regression adjustment (Beaumont *et al*, *Genetics*, 2002):
  - Use a local-linear regression model to obtain a conditional-density estimate, given the observed data.
  - Example application: inference of great ape evolutionary history from whole-genome data (Prado-Martinez, J. *et al.* *Nature*, 2013).
- Markov chain Monte Carlo (MCMC-ABC) (Marjoram *et al*, PNAS, 2003)
  - Replace accept-reject Metropolis step in MCMC with accept-reject ABC step.
  - Example application: agent-based model of social structure dynamics in *Drosophila melanogaster* (Foley *et al*, *Am. Nat.*, 2015)
- Sequential Monte Carlo (SMC-ABC) (Sisson *et al*, PNAS, 2007)
  - Iterative refinement where parameter values are initially sampled from the prior, and finally sampled from the approximate posterior.
  - Example application: Modelling bovine tuberculosis spread among farms in the UK (Brooks-Pollock *et al*, *Nature*, 2014).

ABC model choice

It is also straightforward in the ABC framework to compare the posterior probabilities of different models.

1. Simulate n samples from joint distribution $p(m, S(x))$:

    1.1 Simulate a model index $m_i \sim \pi(\mathcal{M} = m)$ $m = 1, \ldots, M$

    1.2 Simulate parameters $\theta_{m_i} \sim \pi_m(\theta_{m_i})$

    1.3 Simulate $x_i \sim p_{m_i}(x|\theta_{m_i})$

    1.4 Reject m_i if $\rho(S(x_i), S(y)) > \epsilon$.

2. Accepted m_i are sampled approximately from $p(\mathcal{M} = m|S(y))$

The above basic rejection approach has counterparts in regression ·ABC, MCMC-ABC, and SMC-ABC.

Summary Statistics and the Curse of Dimensionality

▶ According to the Pitman-Koopman-Darmois theorem only in the case of probability distributions in the Exponential Family can the number of sufficient statistics reach an upper bound with increasing sample size.

▶ *I.e.* generally we do not expect to capture all the information in the data with any predefined set of summary statistics.

▶ Yet ideally we would like to be as close to sufficiency as possible.

▶ But the more statistics we have the less likely we will be able to closely match them for any given tolerance interval ϵ

▶ This is an example of the 'curse of dimensionality'.

Toy example

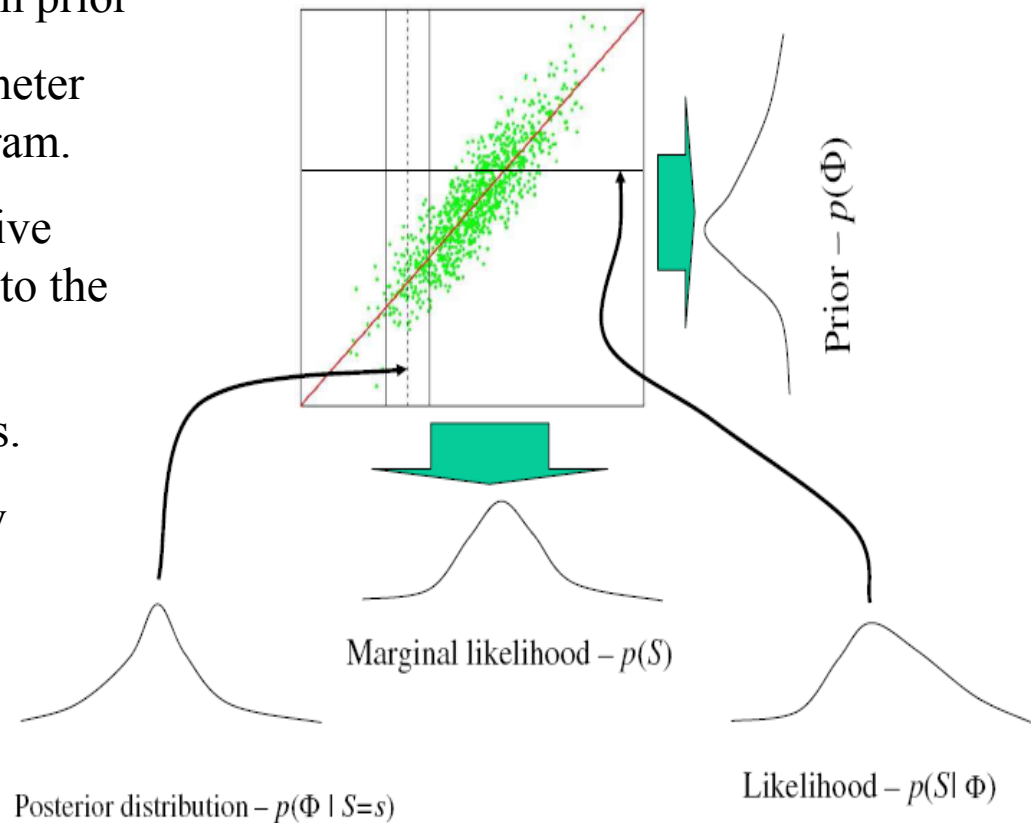Sample $X_i \sim U^d(0,1)$
$i = 1, \ldots, 10,000$;

compute $\| X_i - 0.5 \|$;

for closest 1% of points plot marginal distribution for $X[1]_{i=1,\ldots,10000}$.
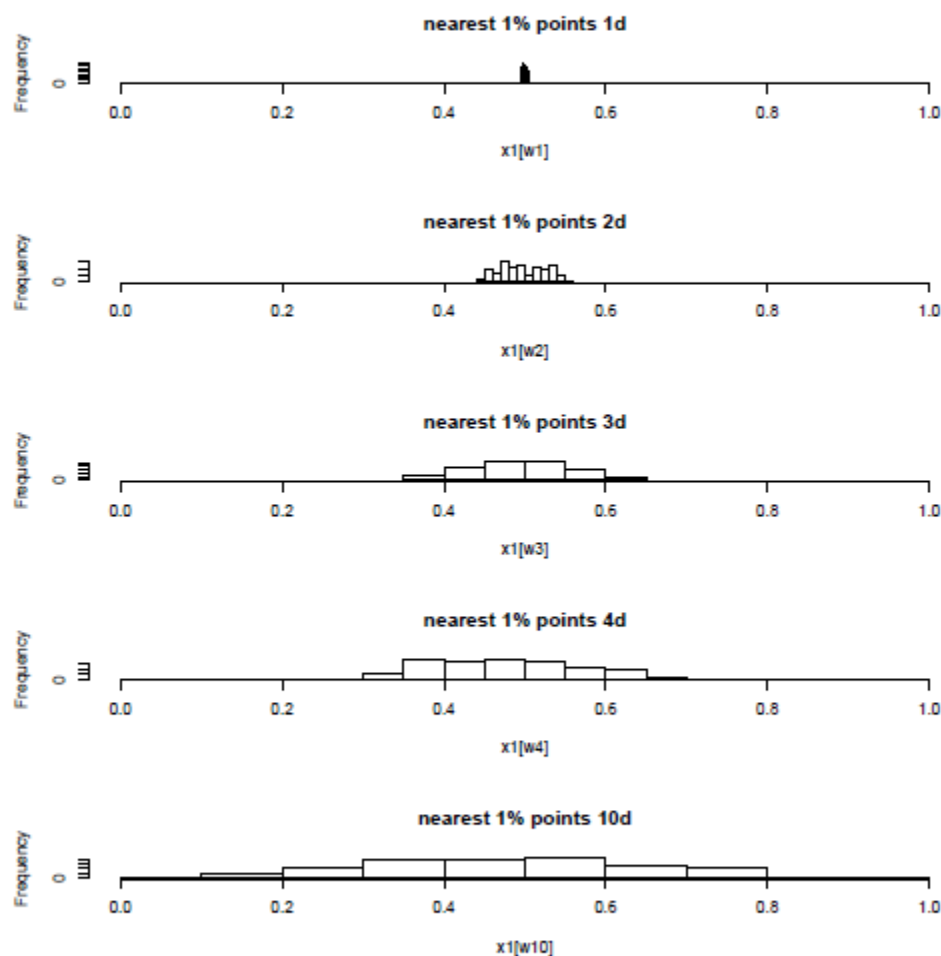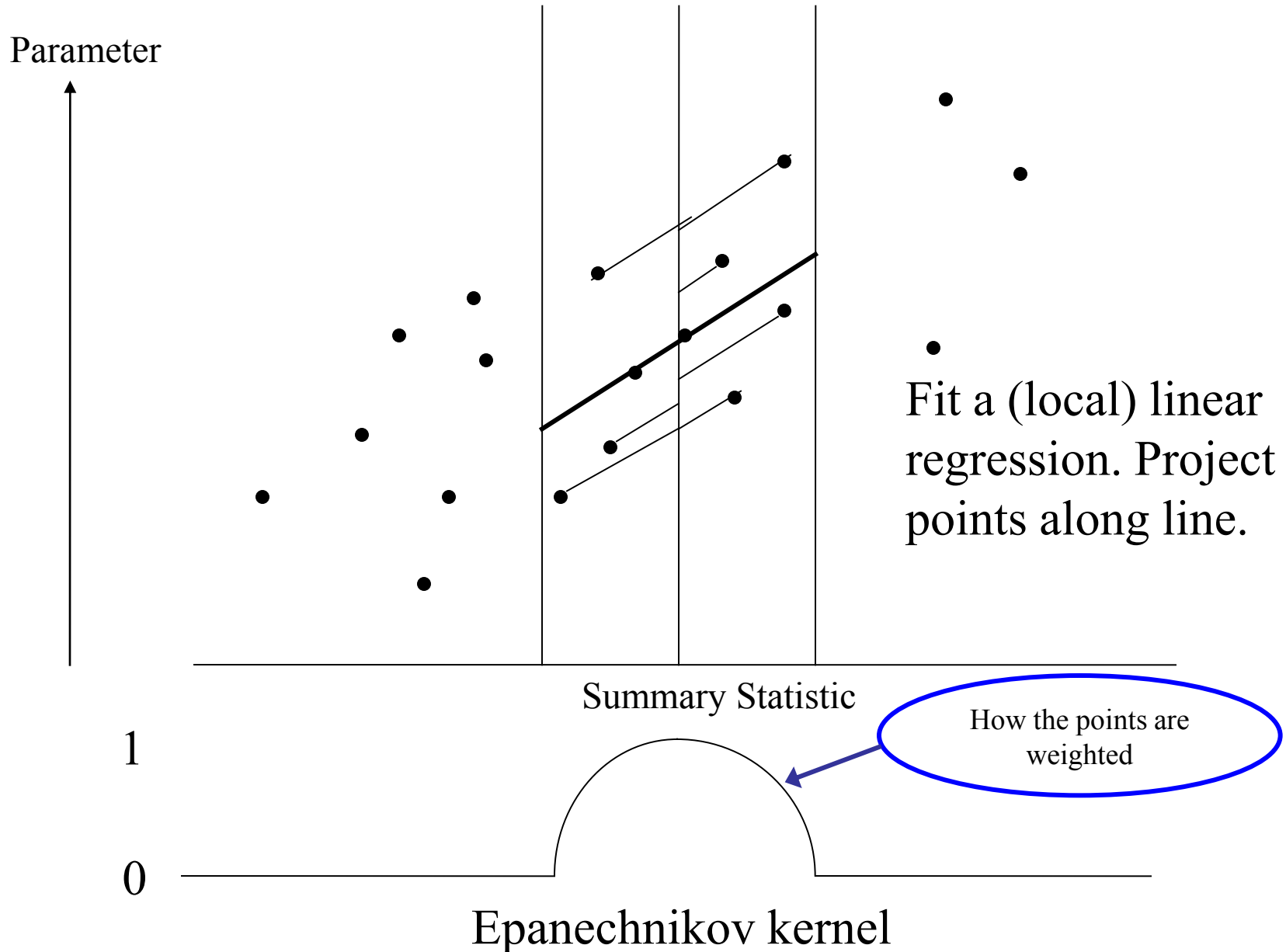
Regression Adjusted ABC

1. Simulate parameter values from prior

2. Simulate data with these parameter values from a simulation program.

3. Retain parameter values that give simulated data that are similar to the 'real' data.

4. Repeat a large number of times.

5. Adjust the parameter values by weighted regression.

Prior – $p(\Phi)$

Marginal likelihood – $p(S)$

Posterior distribution – $p(\Phi \mid S=s)$

Likelihood – $p(S \mid \Phi)$

ABC regression-adjustment method

Parameter

Summary Statistic

Fit a (local) linear regression. Project points along line.

How the points are weighted
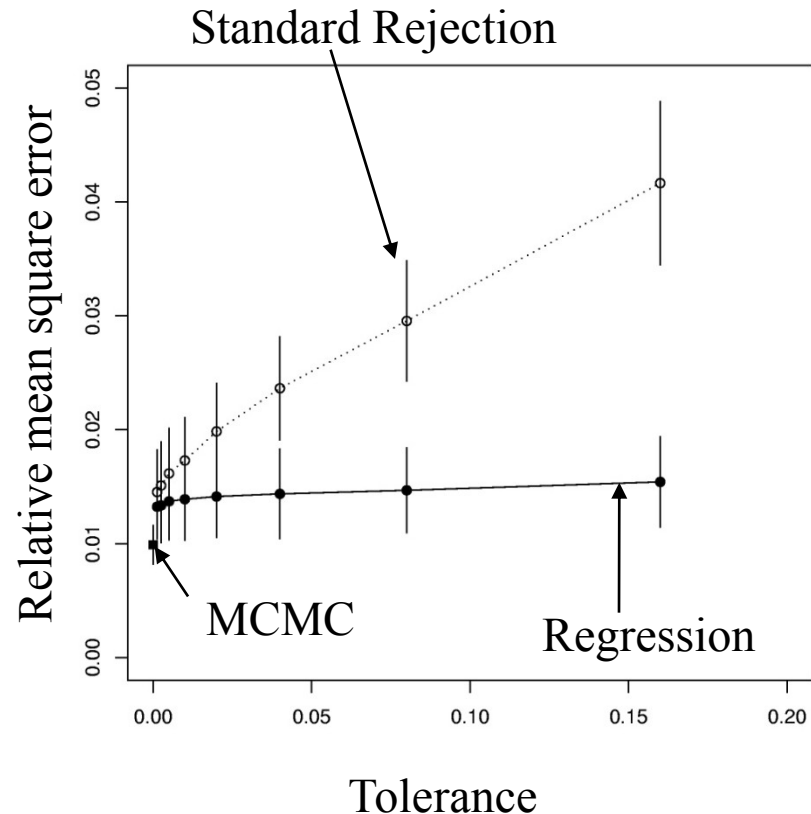
1

0

Epanechnikov kernel

Accuracy in the estimation of scaled mutation rate $\theta = 2N\mu$

Data:-

• linked microsat loci

Summary statistics:-

• mean variance in length

• mean heterozygosity

• number of haplotypes

Improved Regression ABC

The basic regression approach has been improved by Blum and François (*Comput. Stat.*, 2009) in two ways:

- Relax the assumption that the distribution of errors is constant within the tolerance region by performing an extra regression on the residuals.

- Perform multivariate non-linear regression using a neural network.

Regression Approach to Model Selection

Beaumont, M.A. (2008). Joint determination of topology, divergence time, and immigration in population trees, pp 134-154 . In *Simulation, Genetics, and Human Prehistory*, eds. S. Matsumura, P. Forster, & C. Renfrew. (McDonald Institute Monographs.) Cambridge: McDonald Institute for Archaeological Research.

- Use regression framework. Treat model indicator as a categorical variable Y that can take values from $(1,\ldots, n_M)$.

- Then get an estimate of $P(Y=j \mid S = s)$. Use weighted regression, as before.
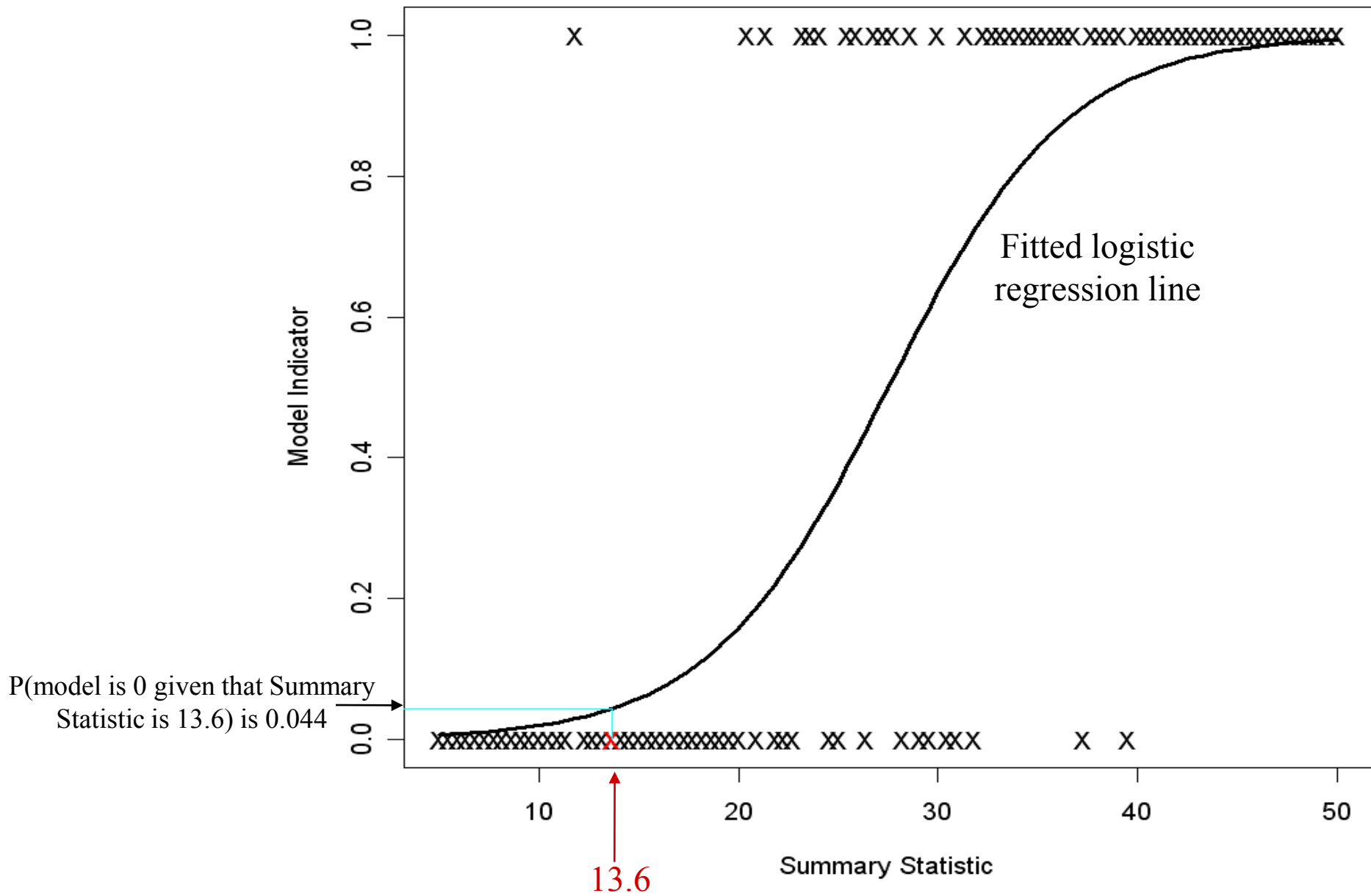
Semi-automatic ABC

One approach to address the curse of dimensionality is that of
Fearnhead and Prangle (RSSB, 2012)

- ▸ Motivation for the method is a proof that if the posterior
  mean for a parameter is used as a summary statistic, this
  minimizes mean square error.

- ▸ Regression gives an estimate of $E(\theta|S(x))$.

- ▸ So we can use the linear predictor from the regression as a
  projection to map the vector of summary statistics to a scalar
  $\hat{E}(\theta|S(x))$ for each parameter and use this projection in place
  of the original summary statistics.

The Advantages and Disadvantages of ABC

Advantages:

- ▶ A likelihood function does not need to be computed.
- ▶ Easily obtain marginals over latent variables.
- ▶ (In simplest versions:) No MCMC convergence problems;
- ▶ Prior/posterior predictive model checking (almost) comes for free.

Disadvantages:

- ▶ Sensitivity to choice of (multivariate) bandwidth.
- ▶ Sensitivity to choice of summary statistics.
- ▶ Uncertainty in how sufficient the summary statistics are.
- ▶ It is still a Monte Carlo method (a method of last resort. . .)
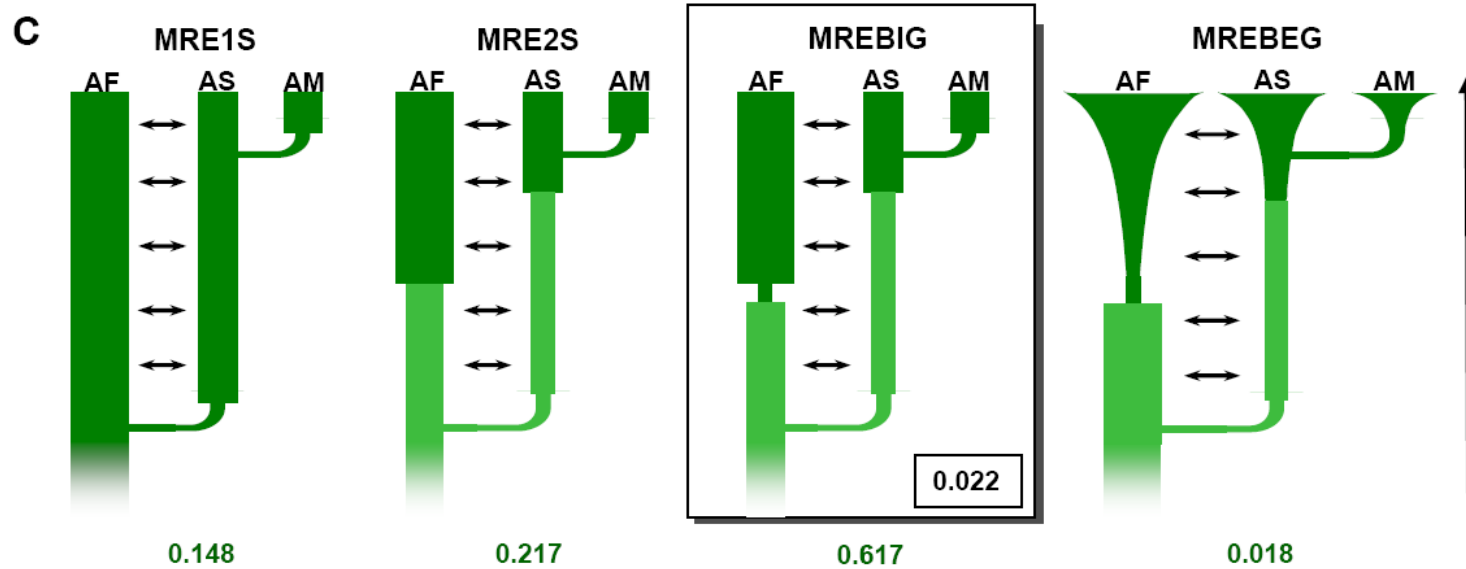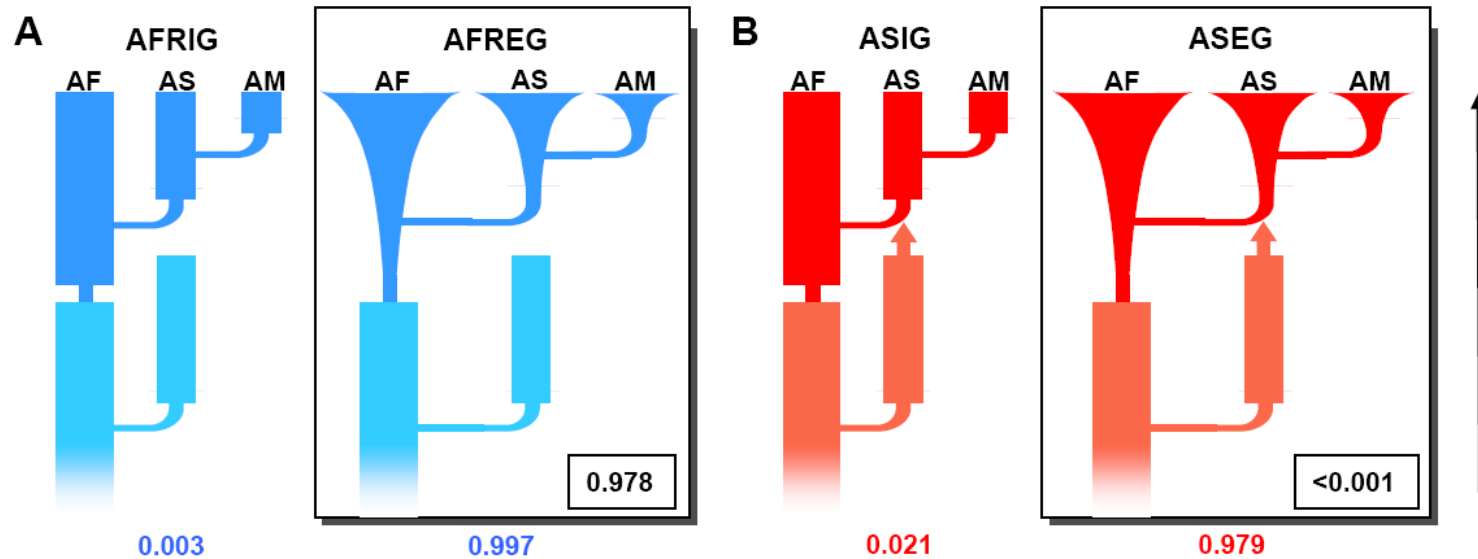
Application of ABC: Human Demographic History

Nelson J. R. Fagundes[†‡§], Nicolas Ray[§], Mark Beaumont[¶], Samuel Neuenschwander[§∥], Francisco M. Salzano[‡††], Sandro L. Bonatto[†,††], and Laurent Excoffier[§††]

[†]Laboratório de Biologia Genômica e Molecular, Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), 90619-900 Porto Alegre, RS, Brazil; [‡]Departamento de Genética, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil; [§]Computational and Molecular Population Genetics (CMPG), Zoological Institute, University of Bern, CH-3012 Bern, Switzerland; [¶]School of Animal and Microbial Sciences, University of Reading, Reading RG6 6AJ, United Kingdom; and [∥]Department of Ecology and Evolution, University of Lausanne, Biophore, CH-1015 Lausanne, Switzerland

- Most genetic data suggest humans originated recently in Africa
- Still unclear if modern humans completely replaced former members of the Homo genus, or if some interbreeding occurred during its range expansion.
- Analysed DNA data from 50 nuclear loci sequenced in African, Asian and American samples.
- Results indicate that a simple African Replacement model with exponential growth has a much higher probability (98%) than alternative multiregional evolution or assimilation scenarios.
- Best supported model points to an origin of our species $\sim$ 145 thousands years ago (Kya), an exit out-of-Africa $\sim$ 54 Kya, and a recent colonization of the Americas 9.5 Kya.

A

AFRIG

AF   AS   AM

AFREG

AF   AS   AM

0.978

0.003   0.997

B

ASIG

AF   AS   AM

ASEG

AF   AS   AM

<0.001

0.021   0.979

C

MRE1S

AF   AS   AM

MRE2S

AF   AS   AM

MREBIG

AF   AS   AM

0.022

MREBEG

AF   AS   AM

0.148   0.217   0.617   0.018

Markov chain Monte Carlo and ABC

- Proposed by Marjoram *et al.* (*PNAS*, 2003).
- Potentially has the advantage of more efficient proposals
- Markov chain is simulated by:

  - Simulate $\theta' \sim K(\theta'|\theta^{(t)})$
  - Simulate $x \sim f(x|\theta')$
  - if $\rho(x, y) < \epsilon$

    then
    $$\theta^{(t+1)} = \theta'$$

    if
    $$u \sim \mathcal{U}(0,1) \leq \pi(\theta')/\pi(\theta^{(t)}) \times K(\theta^{(t)}|\theta')/K(\theta'|\theta^{(t)})$$

    else
    $$\theta^{(t+1)} = \theta^{(t)}$$

$\rho(x,y)$ measures distance between simulated statistic *x* and observed statistic *y*

MCMC-ABC

The MCMC scheme of Marjoram et al. has been widely used and adapted:

- Becquet and Przeworski (*Genome Research*, 2007) used the approach to model population divergence in Chimpanzees

- Wegmann *et al* (*Genetics*, 2009) also analysed chimpanzee data (and introduced Partial Least Squares for 'summarising' summary statistics).
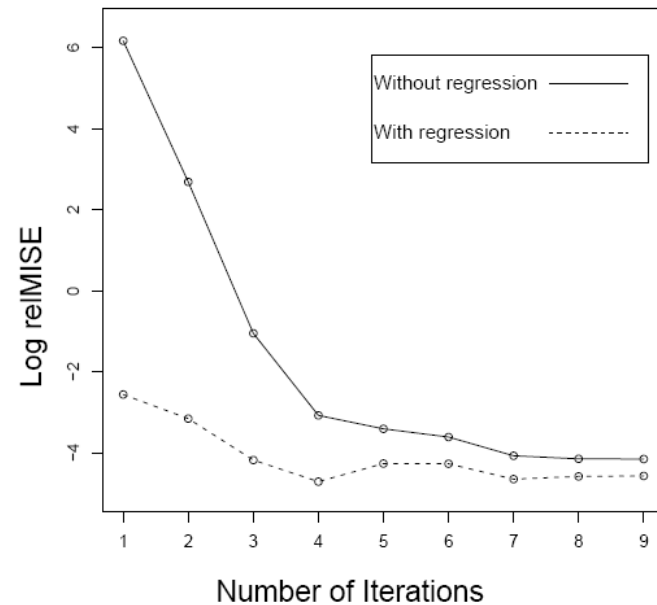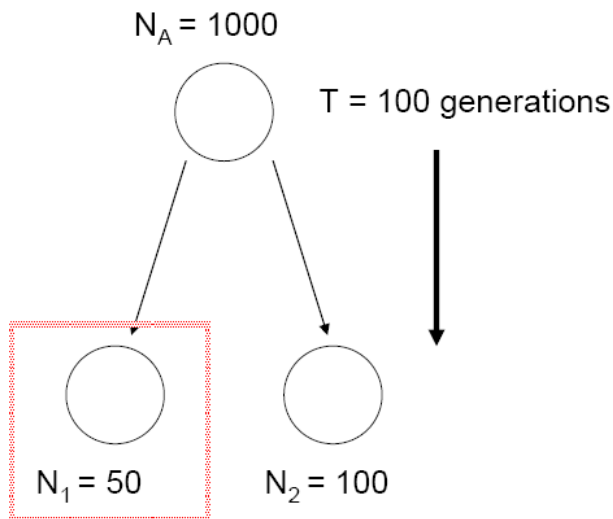
Sequential Monte Carlo and ABC

• An SMC sampler for ABC was introduced by Sisson, Fan and Tanaka (PNAS, 2007)

• Subsequent papers:

  • Toni *et al* (Roy. Soc. Interface, 2009)

  • Beaumont *et al* (Biometrika, 2009)

  • Sisson *et al* (Correction, PNAS, 2009)

(these papers correct a potential bias in the original Sisson *et al* paper)

SMC-ABC

- First step:

    - simulate parameter values from prior

    - Simulate data and summary statistics

    - Retain closest N points

- This set of points can be regarded as drawn from an approximation to posterior

- Then fit a kernel density around the points

- Reduce the tolerance

- Data are simulated from this kernel density, summary statistics computed; closest N points retained.

- Each point is given an importance weight that takes into account that the points are not sampled from the prior.

- This weighted set of points gives an improved approximation because the tolerance interval is smaller.

- The process is continued for further iterations.

SMC-ABC and Regression

- Regression adjustment can lead to improved accuracy of SMC-ABC.
- Inference of population size N_1 from 10 unlinked loci (DNA sequences) in model of population divergence.

Lopes and Beaumont (2010)