Approximate Bayesian Computation

PGDH19

Outline

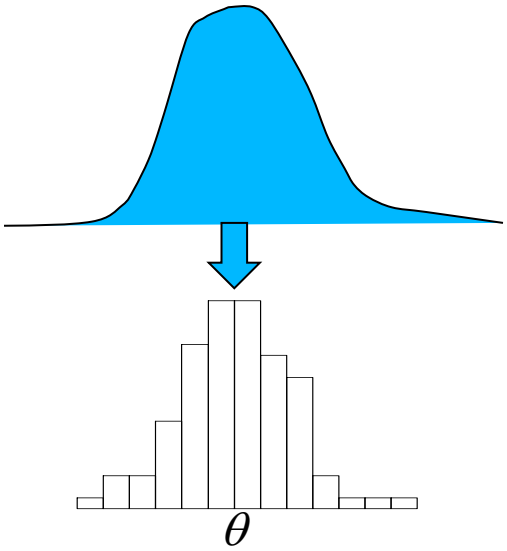
1. Introduction to basic idea of ABC
2. ABC model choice
3. Curse of dimensionality
4. Regression Adjustment
5. Semi-automatic ABC
6. Conclusions
7. Introduction to the practical.

Bayesian Inference and Monte Carlo

- **Bayesian methods** aim to compute the probability distribution of parameter values θ given data x (posterior distribution):

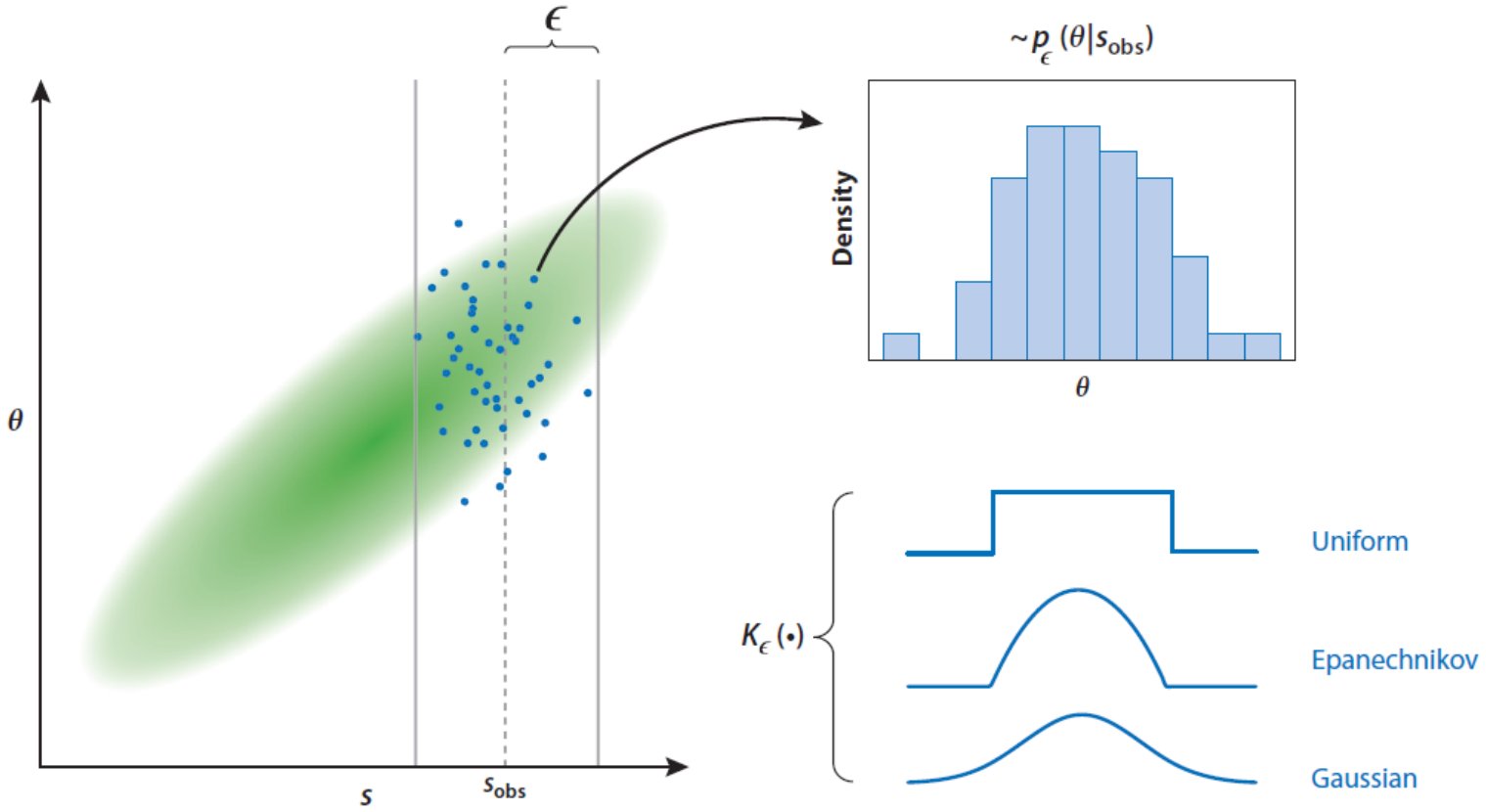
$$p(\theta \mid x)$$

- **Monte Carlo methods** aim to approximate this by drawing **random** values of θ from the posterior distribution $p(\theta \mid x)$.



Approximate Bayesian Computation (ABC)

•ABC is a very specific type of Monte Carlo method, in which the data x are transformed into summary statistics s, and instead of drawing θ from $p(\theta \mid x)$ we draw θ and s from $p_\varepsilon(\theta, s \mid s_{\text{obs}})$ for a bandwidth ε with a kernel function $K_\varepsilon()$



Key Features of ABC

- The method depends on simulations from the 'generative model' that explains the data. It does not require a likelihood function.
- 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.
- Our tolerance (level of approximation) is measured by the bandwidth ε.

ABC model choice

- It is also straightforward in the ABC framework to compare the posterior probabilities of different models
- We can give each model an indicator (e.g. 1, 2, 3, etc) and treat this as a discrete random variable that is sampled from its prior (corresponding to the prior probability of the relevant model).
- The ABC algorithm will then sample values of the model indicator approximately according to its posterior probability.
- A warning: ABC model choice can be sensitive to the summary statistics used (Robert et al, 2011), particularly when only a few are used.

The Curse of Dimensionality

- We do not expect to capture all the information in the data with summary statistics.

- Intuitively, the more summary statistics we use the more information we capture.

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

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

The Curse of Dimensionality

A simple illustration:

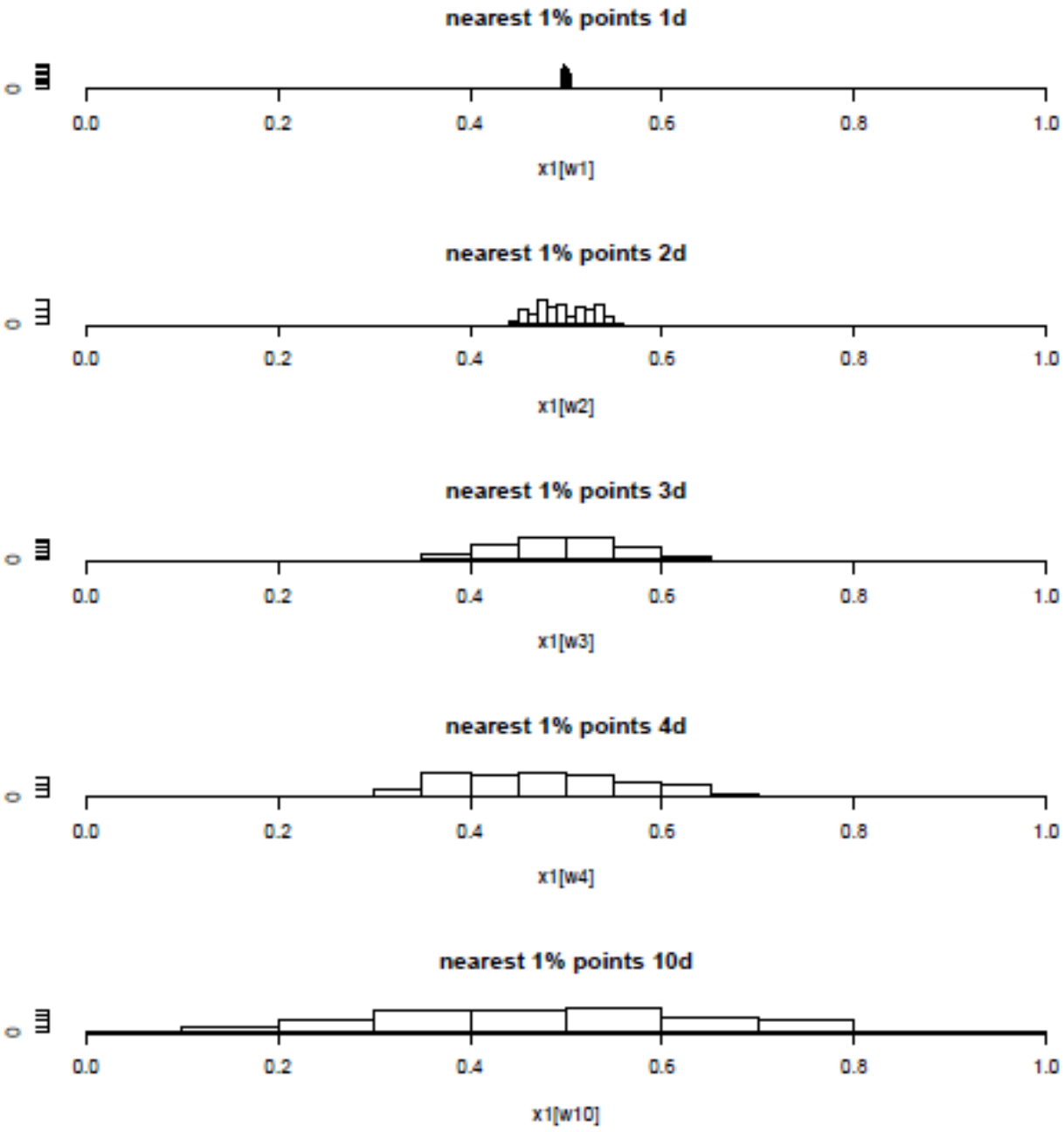
- Let us assume we can choose any number of summary statistics
- Assume they are uncorrelated
- Assume they are uniformly distributed under the priors on the parameters
- For points that are the closest 1% to the target, what does this tell us about the range of values among the accepted summary statistic values? How does this scale with dimension?

Toy example

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

compute $\| X_i - 0{\cdot}5 \|$;

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



How to Address the Curse of Dimensionality

These results show that as the dimensionality increases we may accept any possible value of an **individual** summary statistic, however far it is from the data.

One way to help is to try to develop methods that reduce the bandwidth – the tolerance interval – e.g. MCMC or sequential methods.

However, substantial improvement comes from the use of methods that try to model the relationship between the summary statistics and parameter values.

MCMC-ABC and SMC-ABC

Note:

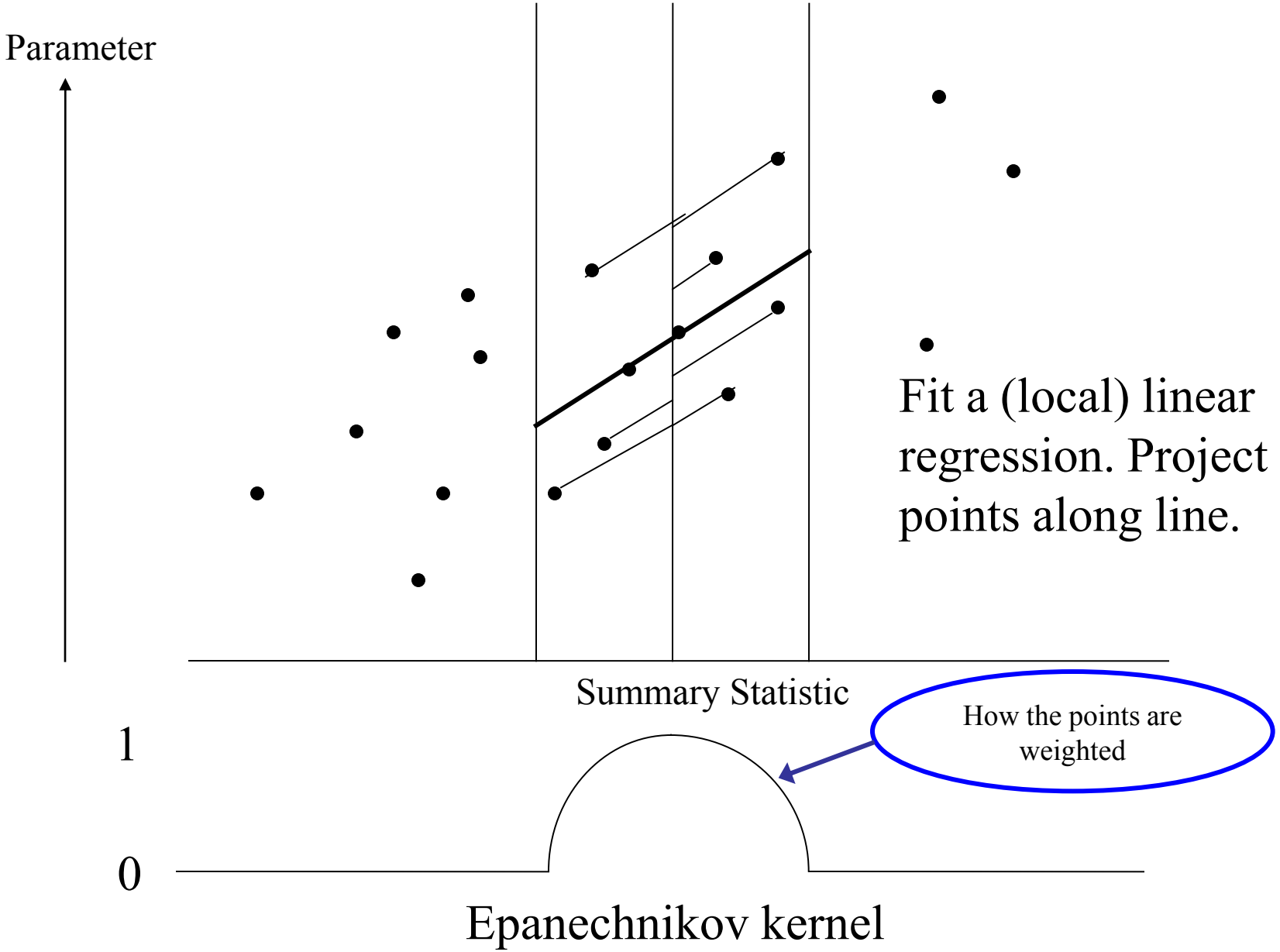
this lecture will not describe MCMC and sequential approaches to ABC, although they are also useful methods that aim to work with smaller tolerance intervals.

Good methods that use these approaches are:

ABCtoolbox (http://cmpg.unibe.ch/software/ABCtoolbox/)

EasyABC (https://cran.r-project.org/web/packages/EasyABC/index.html)

ABC regression-adjustment method



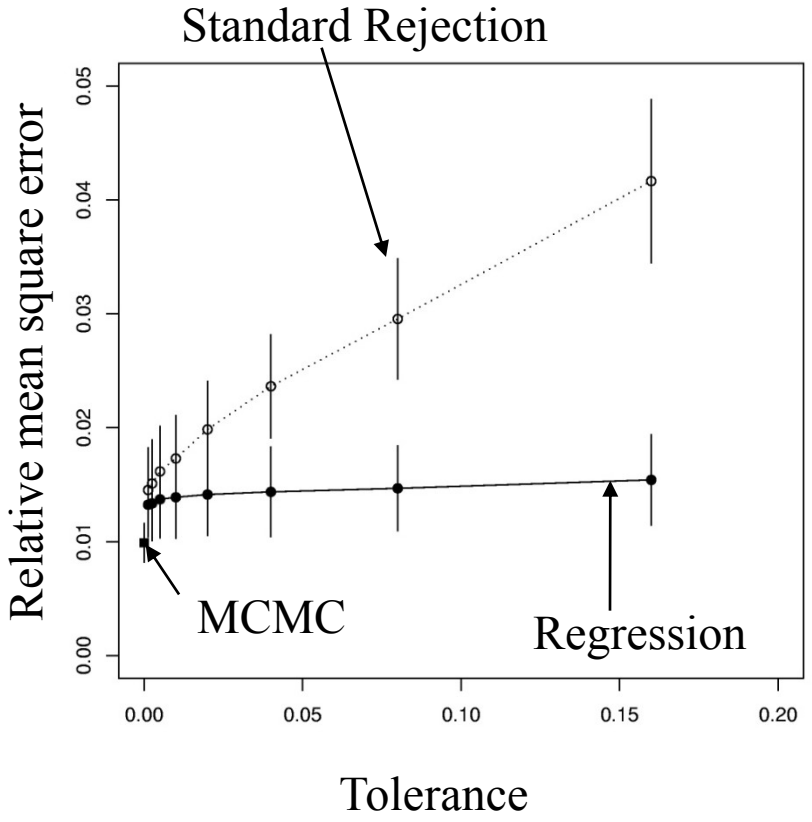
Accuracy in the estimation of scaled mutation rate $\theta = 2N\mu$

Data:-

- linked microsat loci

Summary statistics:-

- mean variance in length
- mean heterozygosity
- number of haplotypes



Alternative Regression approaches to ABC

The basic regression approach has been improved by Blum, François and colleagues in two ways:

- Allow for non constant variance in the regression adjustment
- Use non-linear regression with the nnet package in R
- This is implement in the R package abc (Csillery et al, 2012), which we will use today

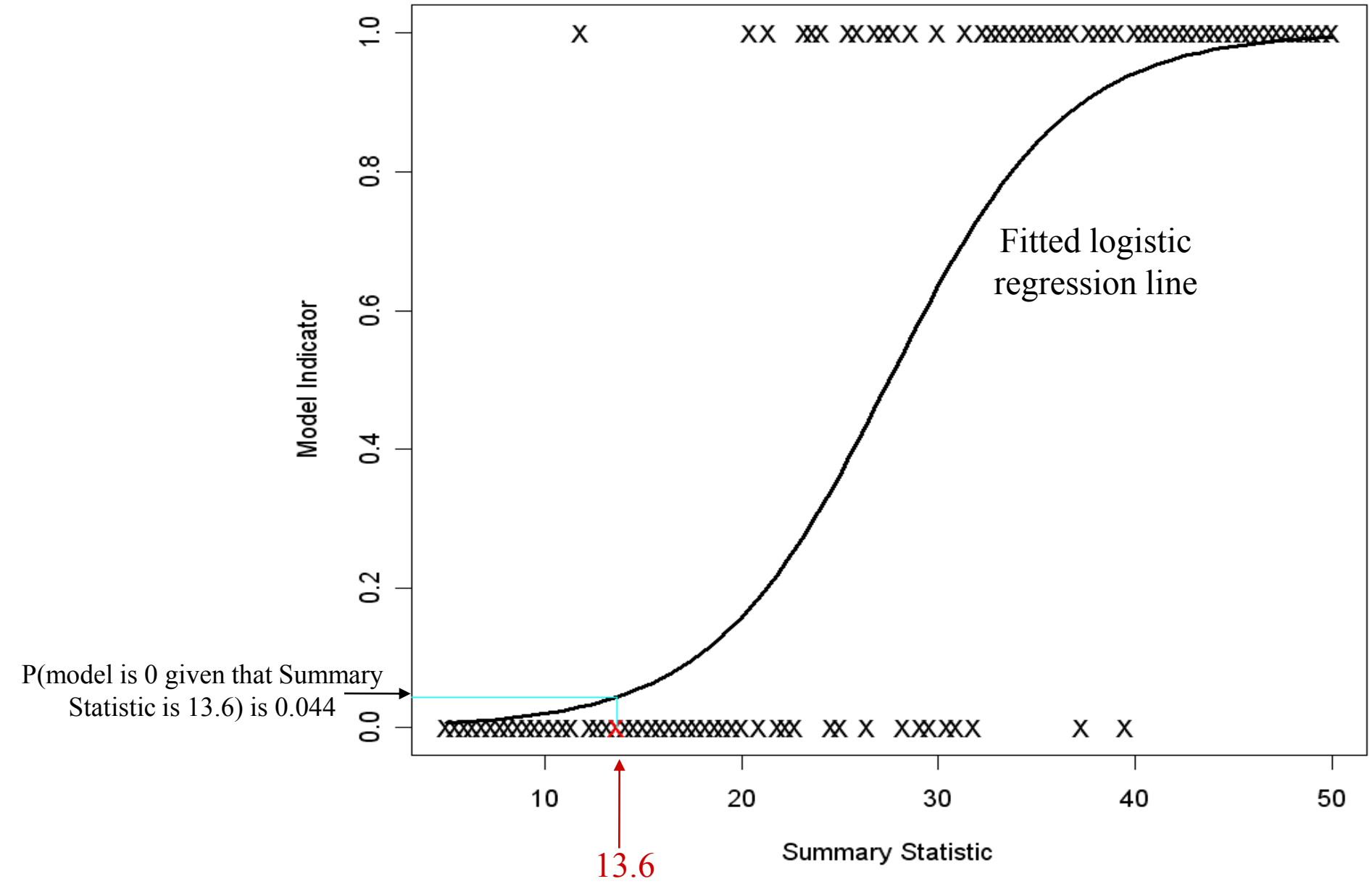
- An alternative regression adjustment approach has been developed by Wegmann et al (2009), in which the summary statistics are modelled in terms of the parameters (i.e. switching the order of regression).
- This is implemented in the package ABCtoolbox (Wegmann et al, 2010)

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 $P(Y=j \mid s = s_{obs})$. Use weighted regression, as before.


Model Indicator
0.0
0.2
0.4
0.6
0.8
1.0
10
20
30
40
50
Summary Statistic
Fitted logistic
regression line
P(model is 0 given that Summary
Statistic is 13.6) is 0.044
13.6

Alternative regression methods for model choice

- The `abc` package uses the same approach, but with the `nnet` package for non-linear model-fitting.
- `ABCtoolbox` aims to get an estimate of the marginal likelihood directly – i.e. it aims to compute the probability of the observed data, $p(s_{obs})$.

Semi-automatic ABC

One approach to address the curse of dimensionality is that of Fearnhead and Prangle (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 (i.e. the average squared deviation from the true parameter value).
- Regression gives an estimate of the posterior mean, which we write as
- $\hat{E}(\theta \mid S(x))$.
- So we can use the linear predictor (i.e. the regression equation) as a projection to map the vector of summary statistics to a 1d (scalar) quantity
- $\hat{E}(\theta_j \mid S(x))$
- for each parameter (*j*th component of the parameter vector) and use this projection in place of the original summary statistics.

This is implemented in the `abctools` package (Nunes and Prangle, 2015)

https://cran.r-project.org/web/packages/abctools/index.html

Extensions of the sa-ABC approach

- A number of research groups have taken further the idea of projecting high-dimensional summary statistics to a lower dimension.
- The use of deep-learning methods to predict the parameter values from summary statistics has been quite widely used (not just in an ABC context: e.g. work of the Schrider group).
- A recent genetic paper that uses deep-learning to project the summary statistics is

Mondal, M., Bertranpetit, J., & Lao, O. (2019). Approximate Bayesian computation with deep learning supports a third archaic introgression in Asia and Oceania. *Nature communications*, *10*(1), 246.

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.

Main Packages

R packages:

abc – Package from Michael Blum's lab, originated from MAB code. Variety of regression-adjustment techniques

abctools – Package from Dennis Prangle and Matt Nunes. Implements semi-automatic abc, and a variety of other methods.

easyabc – Package from Franck Jabot's lab. Broad set of sequential ABC algorithms.

abcrf – Package from group of Arnaud Estoup and Jean-Michel Marin. Implements a machine-learning method for model choice.

Standalone:

DIY-ABC – Package from Jean-Marie Cornuet/Arnaud Estoup's group. 'point and click' interface. Regression-adjustment methods.

ABCtoolbox – Daniel Wegmann and co-workers. Extensive MCMC-ABC implementation, their own regression approach, PLS.

Interlude to Generate Summary Statistics for Exercise

Overview of the abc package (based on the vignette)

- The logic of the vignette is that you originally have a number of models that you wish to compare
- So you first perform model selection and cross-validation for model selection
- **Then you examine goodness of fit – for example use a PCA of the summary statistics generated under the prior and compare the first two PCs with the PCs of the target summary statistics.
- The abc package also shows how posterior predictive checks can be performed (this requires some work, including additional simulations).
- **It allows cross-validation for parameter estimation
- **Finally the vignette discuss the use of `abc` for parameter estimation.

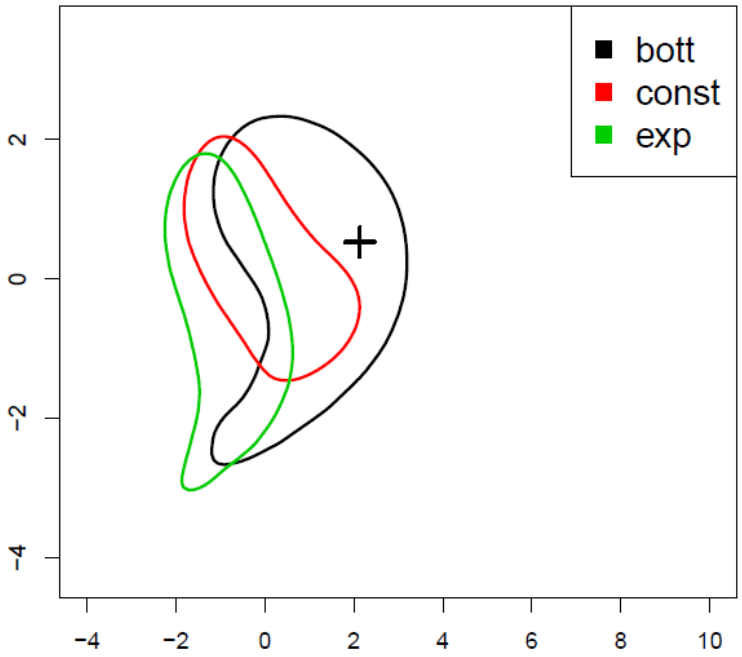
(**discussed in this presentation.)

Goodness of Fit

• The distribution of simulated summary statistics under parameters drawn from the prior is know as the prior predictive distribution.

•ABC only behaves well when the target summary statistics can be regarded as drawn from this distribution.

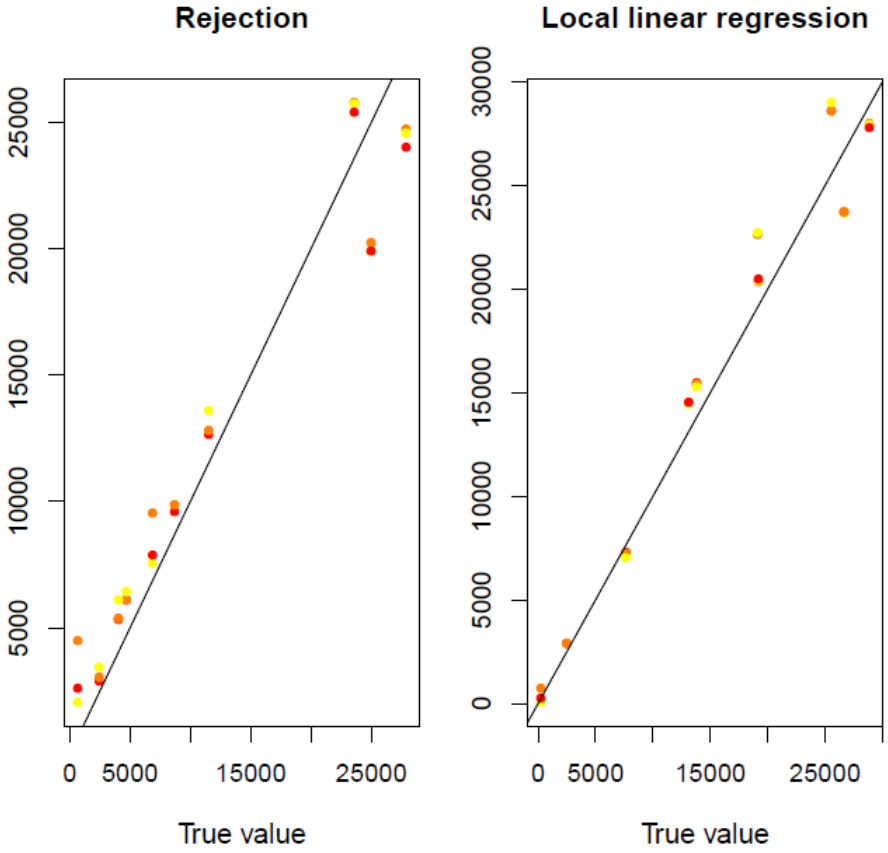
•One way to illustrate the behaviour is to plot the position of the simulated and target summary statistics on the first two PCs



Cross-Validation

• The package provides a function cv4abc() which uses a small subset of simulated parameters and summary statistics as test values to check whether the ABC estimates from these match the true values.

•The resulting object can be plotted:



Parameter Inference

- The `abc` function allows for parameter inference.
- You provide the function with the target summary statistics, the matrix of parameter values simulated from the prior, and the matrix of resulting summary statistics
- You also need to specify a tolerance (the proportion of points closest to the target).
- You can specify whether data are first transformed if you are using rejection adjustment (to ensure the adjusted values stay within the prior).
- You can choose a method: `neuralnet`, `loclinear`, or `rejection`.

Displaying the results

- The output can be displayed in a variety of ways.
- You can print it directly, or use `summary()`, or `hist()`
- I suggest also using `str()` to see what components of the object are stored.
- Additionally you can use `plot()`, which provides a visual summary of the prior, the distribution of distances, the distribution of the posterior against the prior, and a q-q plot of residuals (i.e. showing how well it fits)