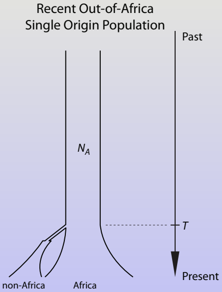
Choosing the Summary Statistics and the Acceptance Rate in Approximate Bayesian Computation (ABC)

Michael G.B. Blum

Laboratoire TIMC-IMAG, CNRS, Grenoble

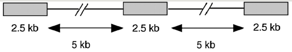
COMPSTAT 2010; Thursday, August 26

A typical application of ABC in population genetics Estimating the time T since the out-of-Africa migration



(a) Model of human origins

a

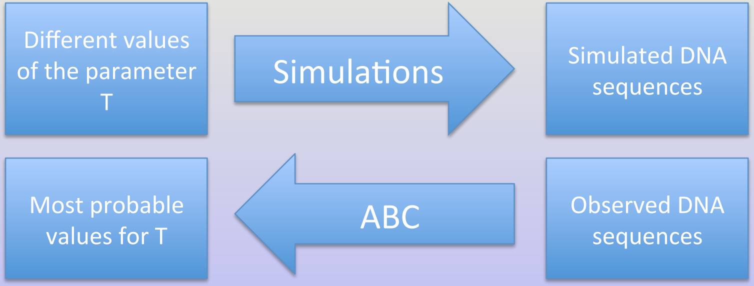


b

	nucleotide position																								frequency		
Haplotype	1555	2137	4330	7388	10369	11290	13106	8491	8502	9274	9316	9404	9495	9638	15452	15463	15548	15660	16080	16426	16635	16641	17348	17365			
Chimp	A	A	C	G	C	A	G	A	A	C	A	G	A	A	T	G	G	A	C	T	T	C	A	T	Af	nA	total
A	.	.	G	.	T	C	.	.	G	.	.	.	.	G	.	.	.	.	T	.	.	.	T	C	2	0	2
B	.	.	.	T	.	.	A	C	.	A	.	.	.	.	.	A	A	.	.	.	.	A	.	.	4	9	13
C	.	C	.	T	.	.	.	.	.	A	.	.	.	.	C	A	A	.	.	.	.	A	.	.	1	0	1
D	.	.	.	T	.	.	.	.	.	A	.	.	.	.	.	A	A	.	.	.	.	A	.	.	1	0	1
E	.	.	.	T	.	.	.	C	.	A	.	.	.	.	.	A	A	.	.	.	.	A	T	C	1	0	1
F	.	.	.	T	.	.	.	.	.	A	.	.	.	.	.	A	A	.	.	.	.	A	T	C	1	0	1
G	C	.	.	T	.	.	.	.	.	A	.	.	.	G	.	A	A	.	.	.	G	A	T	C	0	1	1
H	C	.	.	T	.	.	.	.	.	A	.	T	G	.	.	A	A	.	.	.	G	A	T	C	0	1	1
I	C	.	.	T	.	.	.	.	.	A	.	.	G	.	.	A	A	.	.	.	G	A	T	C	0	11	11
J	C	.	.	T	.	.	.	.	.	A	.	.	.	.	.	A	A	.	.	.	G	A	T	C	0	1	1
K	C	.	.	T	.	.	.	.	.	A	.	.	G	.	.	A	A	.	.	.	.	A	T	C	0	1	1
L	C	.	.	T	.	.	.	.	.	A	.	.	G	.	.	A	A	G	.	.	G	A	T	C	0	1	1
M	.	.	.	T	.	.	A	C	.	A	.	.	.	.	.	A	A	.	.	C	.	A	.	.	0	1	1
N	.	.	.	T	.	.	.	C	.	A	G	.	.	.	.	A	A	.	.	.	.	A	.	.	0	3	3
O	.	.	.	T	.	.	.	.	.	A	G	.	.	.	.	A	A	.	.	.	.	A	.	.	0	1	1
P	.	.	.	T	.	.	.	C	.	A	.	.	.	.	.	A	A	.	.	.	.	A	.	C	0	1	1
Q	C	.	.	T	.	.	.	.	.	A	G	.	G	.	.	A	A	.	.	.	G	A	T	C	0	1	1
			+	+	+	+			+	+						+	+		+			+					

(b) Data

Flowchart of ABC



Rejection algorithm for targeting $p(\phi|S)$

1. Generate a parameter ϕ according to the prior distribution π ;
2. Simulate data $\mathcal{D}'$ according to the model $p(\mathcal{D}'|\phi)$;
3. Compute the summary statistic S' from $\mathcal{D}'$ and accept the simulation if $d(S, S') < \delta$.

Potential problem : the curse of dimensionality limits the number of statistics that rejection-ABC can handle.

Regression-adjustment for ABC

Beaumont, Zhang and Balding; Genetics 2002

- Local linear regression

$$\phi_i | S_i = m(S_i) + \epsilon_i,$$

with a linear function for m.

- Adjustment

$$\phi_i^* = \hat{m}(S) + \tilde{\epsilon}_i,$$

$\hat{m}$ is found with weighted least-squares.

Regression-adjustment for ABC

- Weighted least-squares

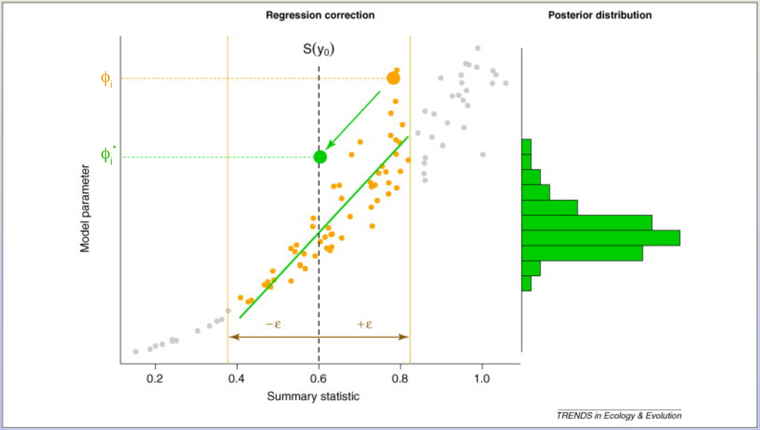
$$\sum_{i=1}^{n}\{\phi_i - (\beta_0 + (S_i - S)^T\beta_1)\}^2 W_i,$$

where $W_i \propto K(||S - S_i||/\delta)$.

- Adjustment

$$\phi_i^* = \hat{\beta}_{\mathrm{LS}}^0 + \tilde{\epsilon}_i = \phi_i - (S_i - S)^T\hat{\beta}_{\mathrm{LS}}^1.$$

Regression-adjustment for ABC



Csilléry, Blum, Gaggiotti and François; TREE 2010

Asymptotic theorem for ABC

Blum; JASA 2010

1. If there is a local homoscedastic relationship between ϕ and S,
   Bias with regression adjustment < Bias with rejection only

2. But
   Rate of convergence of the MSE $= \theta(n^{-4/(d+5)})$
   $d =$ dimension of the summary statistics
   $n =$ number of simulations

A Gaussian example to illustrate potential pitfalls with ABC

Toy example 1 : Estimation of σ^2

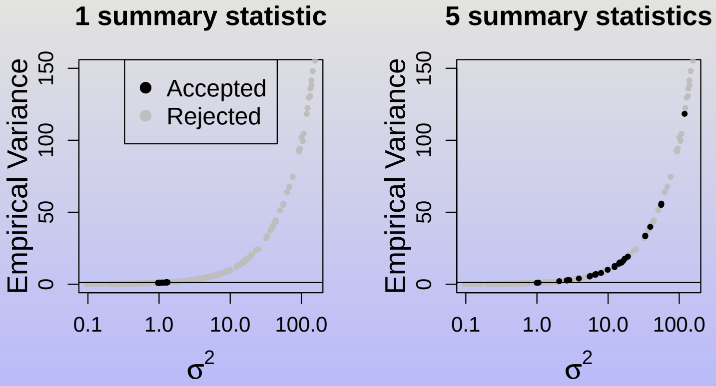
$$\begin{aligned} \sigma^2 &\sim \mathrm{Inv}\chi^2(\mathrm{d.f.} = 1) \\ \mu &\sim \mathcal{N}(0, \sigma^2) \\ N &= 50 \end{aligned}$$

Summary statistics

$$(S^1, \ldots, S^5) = (\bar{x}_N, s_N^2, u_1, u_2, u_3)$$

$$u_j \sim \mathcal{N}(0, 1),\ j = 1, 2, 3$$

A Gaussian example to illustrate potential pitfalls with ABC



Local Bayesian linear regression

Hjort; Book chapter 2003

Prior for the regression coefficients β

$$\beta \sim \mathcal{N}(0, \alpha^{-1} I_{p+1})$$

The *Maximum a posteriori* minimizes the regularized weighted least-squares problem

$$E(\beta) = \frac{1}{2\tau^2} \sum_{i=1}^{n} (\phi_i - (S_i - S)^T \beta)^2 W_i + \frac{\alpha}{2} \beta^T \beta.$$

Local Bayesian linear regression

Posterior distribution of the regression coefficients

$$\beta \sim \mathcal{N}(\beta_{\mathrm{MAP}}, V),$$

$$\begin{aligned} \beta_{\mathrm{MAP}} &= \tau^{-2} V X^T W_\delta \phi \\ V^{-1} &= (\alpha I_{p+1} + \tau^{-2} X^T W_\delta X). \end{aligned}$$

Regression-adjustment for ABC

$$\phi_i^* = \phi_i - (S_i - S)^T \hat{\beta}^1_{\mathrm{MAP}}.$$

The evidence function as an omnibus criterion

Empirical Bayes /Evidence approximation

$$p(\phi|\tau^2, \alpha, p_\delta) = \int \left(\Pi_{i=1}^{n} p(\phi_i|\beta, \tau^2)^{W_i} \right) p(\beta|\alpha)\, d\beta,$$

α is the precision hyperparameter
τ is the variance of the residuals
p_δ is the percentage of accepted simulations.

Maximizing the evidence for

1. choosing p_δ
2. choosing the set of summary statistics

The evidence function as an omnibus criterion

A closed-formed formula

$$\begin{aligned} \log p(\phi|\tau^2, \alpha, p_\delta) &= \frac{p+1}{2}\log\alpha - \frac{N_W}{2}\log\tau^2 - E(\beta_{\mathrm{MAP}}) \\ &\quad -\frac{1}{2}\log|V^{-1}| - \frac{N_W}{2}\log 2\pi, \end{aligned}$$

where $N_W = \sum W_i$.

The evidence function as an omnibus criterion

The evidence as a function of the tolerance rate

$$\log p(\phi|p_\delta) = \max_{(\alpha,\tau)} \log p(\phi|\tau^2, \alpha, p_\delta).$$

The evidence as a function of the set of summary statistics

$$\log p(\phi|S) = \max_{(\alpha,\tau,p_\delta)} \log p(\phi|\tau^2, \alpha, p_\delta).$$

Iterative algorithm for maximizing the evidence w.r.t. α and τ

Updating the value of the hyperparameter

$$\alpha = \frac{\gamma}{\beta_{\mathrm{MAP}}^T \beta_{\mathrm{MAP}}},$$

where γ is the effective number of summary statistics.

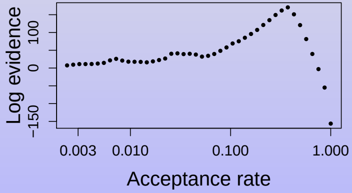
$$\gamma = (p+1) - \alpha \mathrm{Tr}(V).$$

$$\tau^2 = \frac{\sum_{i=1}^n (\phi_i - (S_i - S)^T \beta)^2 W_i}{N_W - \gamma}.$$

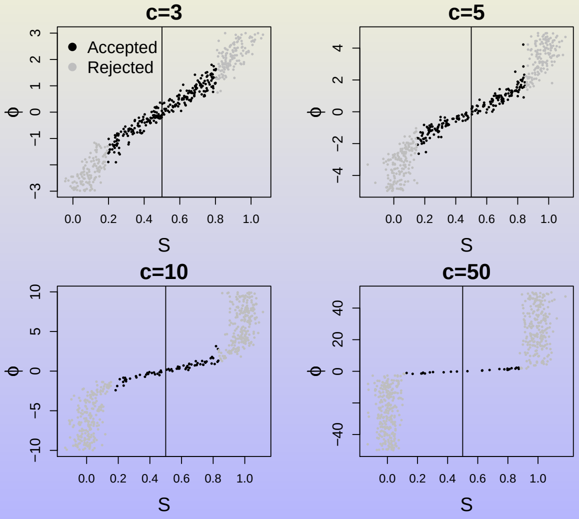
Using the evidence for choosing p_δ

Toy example 2

$$\phi \sim \mathcal{U}_{-c,c}, \quad c \in \mathbb{R},$$
$$S \sim \mathcal{N}\left(\frac{e^{\phi}}{1+e^{\phi}}, \sigma^2 = (.05)^2\right),$$

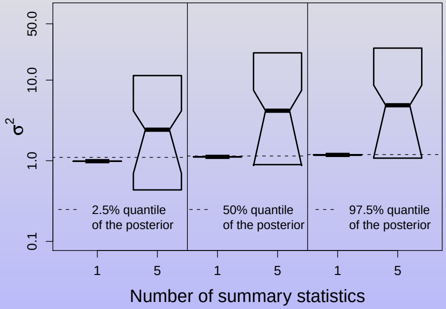


Using the evidence for choosing p_δ



Using the evidence for choosing the summary statistics

Toy example 1 : $(S^1, \ldots, S^5) = (\bar{x}_N, s_N^2, u_1, u_2, u_3)$



Choice of a set of statistics with the evidence

Variance only	Other
100	0

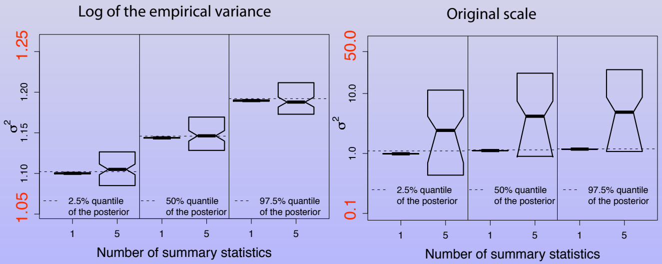
Transformation of the statistics can matter

Left Panel

$$S^1 = \log s_N^2 \text{ or } (S^1, \ldots, S^5) = (\bar{x}_N, \log s^2{}_N, u_1, u_2, u_3)$$

Right Panel

$$S^1 = s_N^2 \text{ or } (S^1, \ldots, S^5) = (\bar{x}_N, s_N^2, u_1, u_2, u_3)$$



Pros and cons

Cons

- Quite complicated
- Model (variable) selection for regression but not for density estimation

Pros

- Similar methodology without regression adjustment
- Omnibus criterion (Choice of the summary statistics, of the tolerance rate p_δ)
- Shrinkage of regression coefficients

Thanks all for your attention