ABC and Model Selection

Christoph Leuenberger

Université de Fribourg, Switzerland

Paris, June 26, 2009

Almost two thirds of a 20-dimensional water melon with a radius of 20 cm consists of rind if the rind is 1 cm thick.

Two ways to overcome the curse of dimensionality in ABC

Speed up the convergence to the observed summary statistics :

- **ABC-MCMC :** Marjoram, Molitor, Plagnol, Tavaré (2003)
- **ABC-SMC :** Sisson, Fan, Tanaka (2007)
- **ABC-PRC :** Beaumont, Cornuet, Marin, Robert (2009)
- **ABC-SMC :** Toni, Welch, Strelkowa, Ipsen, Stumpf (2009)
- . . .

Model the retained parameters and statistics :

- **ABC-REG :** Beaumont, Zhang, Balding (2002)
- **ABC-ANCH :** Blum, François (2009)
- **ABC-GLM :** L., Wegmann, Excoffier (2009)
- . . .

Truncated Model

Likelihood of truncated model $\mathcal{M}_\epsilon(\mathbf{s}_{obs})$ obtained by ABC process is given by

$$f_\epsilon(\mathbf{s}|\boldsymbol{\theta}) = \text{Ind}(\mathbf{s} \in \mathcal{B}_\epsilon(\mathbf{s}_{obs})) \cdot f_{\mathcal{M}}(\mathbf{s}|\boldsymbol{\theta}) \cdot (\int_{\mathcal{B}_\epsilon} f_{\mathcal{M}}(\mathbf{s}|\boldsymbol{\theta}) d\mathbf{s})^{-1}$$

where $\mathcal{B}_\epsilon = \mathcal{B}_\epsilon(\mathbf{s}_{obs}) = \{\mathbf{s} \in \mathbb{R}^n | \text{dist}(\mathbf{s}, \mathbf{s}_{obs}) < \epsilon\}$ is the ϵ-ball in the space of summary statistics and $\text{Ind}(\cdot)$ is the indicator function.

If the parameters are generated from a prior $\pi(\boldsymbol{\theta})$ then the distribution of the parameters retained after ABC are given by

$$\pi_\epsilon(\boldsymbol{\theta}) = \frac{\pi(\boldsymbol{\theta}) \int_{\mathcal{B}_\epsilon} f_{\mathcal{M}}(\mathbf{s}|\boldsymbol{\theta}) d\mathbf{s}}{\int_\Pi \pi(\boldsymbol{\theta}) \int_{\mathcal{B}_\epsilon} f_{\mathcal{M}}(\mathbf{s}|\boldsymbol{\theta}) d\mathbf{s} d\boldsymbol{\theta}}.$$

Combining :

$$\pi(\boldsymbol{\theta}|\mathbf{s}_{obs}) = \frac{f_\epsilon(\mathbf{s}_{obs}|\boldsymbol{\theta})\pi_\epsilon(\boldsymbol{\theta})}{\int_\Pi f_\epsilon(\mathbf{s}_{obs}|\boldsymbol{\theta})\pi_\epsilon(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

$\pi_\epsilon(\boldsymbol{\theta})$: truncated prior $=$ posterior of ABC

$f_\epsilon(\mathbf{s}|\boldsymbol{\theta})$: likelihood of truncated model $\mathcal{M}_\epsilon(\mathbf{s}_{obs})$

Model Assuption

Assume that $\mathcal{M}_\epsilon(\mathbf{s}_{obs})$ is a General Linear Model (GLM) :

$$\mathbf{s}|\boldsymbol{\theta} = \mathbf{C}\boldsymbol{\theta} + \mathbf{c}_0 + \boldsymbol{\epsilon}$$

$\mathbf{s}$: summary statistics (of dimension n)

$\boldsymbol{\theta}$: parameters (of dimension m)

$\mathbf{C}$: $n \times m$ design matrix

$\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_s)$: normal error

- Estimate of $\mathbf{C}$, $\mathbf{c}_0$ and $\mathbf{\Sigma}_s$ with OLS
- Smoothen out the empirical distribution of $\pi_\epsilon(\boldsymbol{\theta})$ using retained parameters $\boldsymbol{\theta}^j$:

$$\pi_\epsilon(\boldsymbol{\theta}) \approx \frac{1}{N}\sum_{j=1}^{N} \phi(\boldsymbol{\theta} - \boldsymbol{\theta}^j, \mathbf{\Sigma}_\theta), \tag{1}$$

$\phi(\boldsymbol{\theta} - \boldsymbol{\theta}^j, \mathbf{\Sigma}_\theta) = \frac{1}{|2\pi\mathbf{\Sigma}_\theta|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}^j)^t \mathbf{\Sigma}_\theta^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}^j)}$: Gaussian peaks

$\mathbf{\Sigma}_\theta = \text{diag}(\sigma_1, \ldots, \sigma_m)$: bandwidths

$$\pi(\theta|\mathbf{s}_{obs}) \propto \sum_{j=1}^{N} c(\boldsymbol{\theta}^j) e^{-\frac{1}{2}(\boldsymbol{\theta}-\mathbf{t}^j)^t \mathbf{T}^{-1}(\boldsymbol{\theta}-\mathbf{t}^j)},$$

where

$$\begin{aligned} c(\boldsymbol{\theta}^j) &= \exp\left[-\frac{1}{2}\left((\boldsymbol{\theta}^j)^t \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{\theta}^j - (\mathbf{v}^j)^t \mathbf{T} \mathbf{v}^j\right)\right] \\ \mathbf{T} &= \left(\mathbf{C}^t \boldsymbol{\Sigma}_s^{-1} \mathbf{C} + \boldsymbol{\Sigma}_\theta^{-1}\right)^{-1} \\ \mathbf{t}^j &= \mathbf{T}\mathbf{v}^j \\ \mathbf{v}^j &= \mathbf{C}^t \boldsymbol{\Sigma}_s^{-1}(\mathbf{s}_{obs} - \mathbf{c}_0) + \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{\theta}^j. \end{aligned}$$

Marginal posterior of parameter θ_k :

$$\pi(\theta_k|\mathbf{s}_{obs}) = a \cdot \sum_{j=1}^{N} c(\boldsymbol{\theta}^j) \exp\left(-\frac{(\theta_k - t_k^j)^2}{2\tau_{k,k}}\right).$$

where $\tau_{k,k}$ is the k-th diagonal element of the matrix $\mathbf{T}$, t_k^j is the k-th component of the vector $\mathbf{t}^j$.

Marginal Density and Bayes Factors

Marginal density of $\mathcal{M}$:

$$f_{\mathcal{M}}(\mathbf{s}_{obs}) = \int_{\Pi} f(\mathbf{s}_{obs}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

We have

$$f_{\mathcal{M}}(\mathbf{s}_{obs}) = \frac{A_{\epsilon}(\mathbf{s}_{obs}, \pi)}{N|2\pi\mathbf{D}|^{1/2}} \sum_{j=1}^{N} e^{-\frac{1}{2}(\mathbf{s}_{obs}-\mathbf{m}^j)^t\mathbf{D}^{-1}(\mathbf{s}_{obs}-\mathbf{m}^j)}$$

$\mathbf{D} = \boldsymbol{\Sigma}_s + \mathbf{C}\boldsymbol{\Sigma}_{\theta}\mathbf{C}^t$
$\mathbf{m}^j = \mathbf{c}_0 + \mathbf{C}\boldsymbol{\theta}^j$
$A_{\epsilon}(\mathbf{s}_{obs}, \pi) := \int_{\Pi} \pi(\boldsymbol{\theta}) \int_{\mathcal{B}_{\epsilon}} f_{\mathcal{M}}(\mathbf{s}|\boldsymbol{\theta})d\mathbf{s}d\boldsymbol{\theta}$: acceptance rate of ABC

Bayes factor for two models :

$$B_{AB} = \frac{f_{\mathcal{M}_A}(\mathbf{s}_{obs})}{f_{\mathcal{M}_B}(\mathbf{s}_{obs})}$$

Independent Loci

Loci assumed to be independent : $\mathbf{s}_{obs} = \{\mathbf{s}_1, \ldots, \mathbf{s}_L\}$

Posterior :

$$\pi(\boldsymbol{\theta}|\mathbf{s}_{obs}) \propto \pi_\epsilon(\boldsymbol{\theta}) \prod_{i=1}^{L} f_\epsilon(\mathbf{s}_i|\boldsymbol{\theta})$$

Assumption can be incorporated directly into GLM (block matrices in diagonals of $\mathbf{C}$ and $\boldsymbol{\Sigma}_s$).

It suffices to simulate a single locus to estimate GLM , see Thalmann *et al.* (preprint).

Assessing Model Fit for GLM

Under null-hypothesis that GLM fits the truncated model $\mathcal{M}_\epsilon(\mathbf{s}_{obs})$, the residuals $\mathbf{r}_j$ of the regression are multivariate normal with zero mean :

$$\mathbf{r}_j^T \mathbf{\Sigma}_s^{-1} \mathbf{r}_j \sim \chi^2_n.$$

Closeness to χ^2-distribution is expressed e.g. by Kolmogorov-Smirnov test statistic.

Total variation distance between two distributions π_0 and π_1 :

$$d_1(\pi_0, \pi_1) = \frac{1}{2} \int |\pi_0(\theta) - \pi_1(\theta)| d\theta$$

$$\mathbf{s}|\boldsymbol{\theta} = \mathbf{C}\boldsymbol{\theta} + \mathbf{c}_0 + \boldsymbol{\epsilon}, \quad \pi(\boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\theta)$$

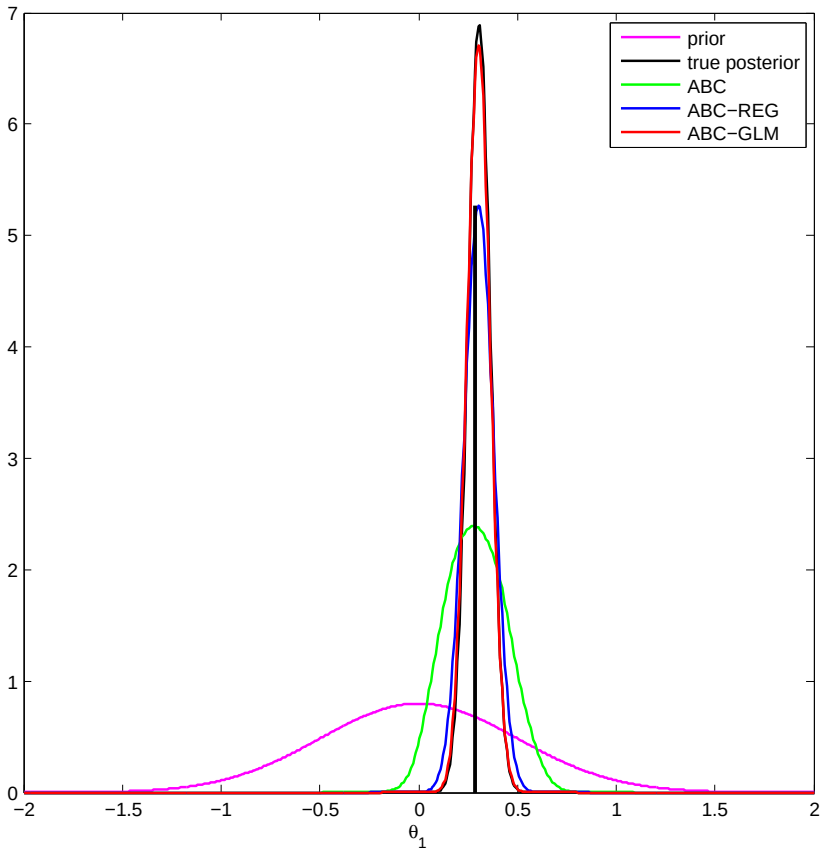

FIGURE: GLM, prior and marginal posteriors, acceptance rate $p = 0.1$

acc. rate p	$d_1(\pi_0, \pi_\epsilon)$	$d_1(\pi_0, \pi_{REG})$	$d_1(\pi_0, \pi_{GLM})$	KS statistics
1.00	0.51 ± 0.22	0.15 ± 0.10	0.01 ± 0.001	0.004 ± 0.001
0.50	0.42 ± 0.19	0.13 ± 0.10	0.02 ± 0.008	0.007 ± 0.003
0.10	0.29 ± 0.18	0.13 ± 0.11	0.03 ± 0.01	0.02 ± 0.01
0.05	0.24 ± 0.16	0.13 ± 0.12	0.03 ± 0.01	0.03 ± 0.01
0.01	0.21 ± 0.17	0.15 ± 0.14	0.05 ± 0.02	0.06 ± 0.02

TABLE: GLM, $m = 3$, $n = 4$, prior $N(0, 0.2^2)$, 100 simulations

Toy Model 2 : Non-linear, uniform errors

$$\mathbf{s}|\theta = \mathbf{1}\theta^3 + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{Unif}([-u, u]^n), \quad \pi(\theta) \sim \mathcal{N}(0, \sigma)$$

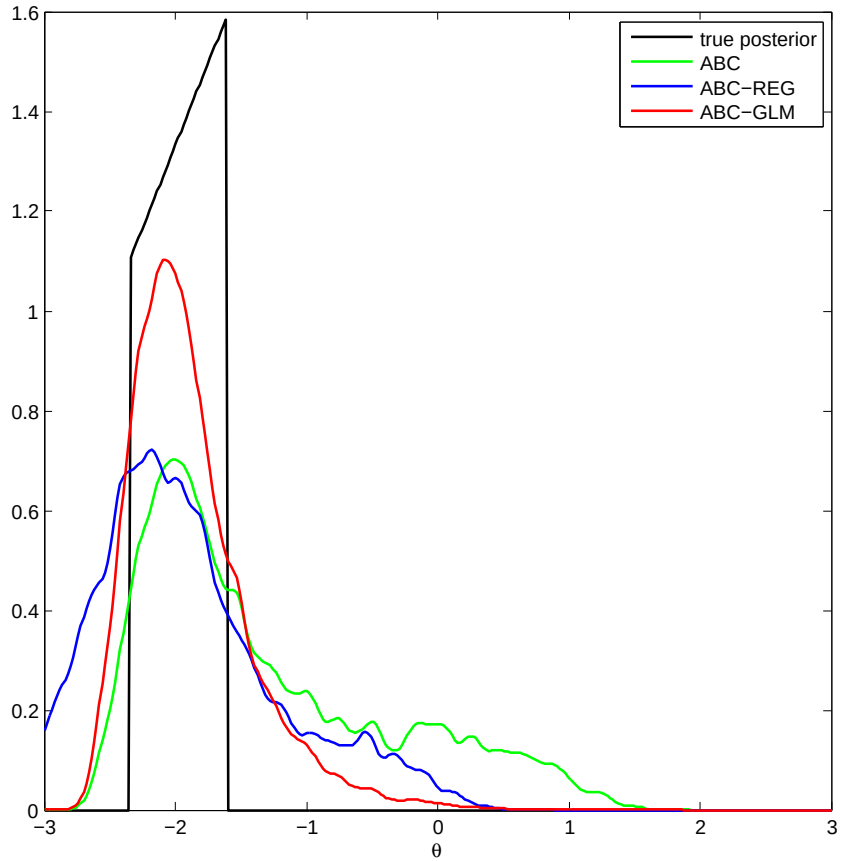

FIGURE: Uniform errors, posteriors, acceptance rate $p = 0.1$

acc. rate p	$d_1(\pi_0, \pi_\epsilon)$	$d_1(\pi_0, \pi_{REG})$	$d_1(\pi_0, \pi_{GLM})$	KS statistics
1.00	0.56 ± 0.24	0.49 ± 0.25	0.46 ± 0.29	0.09 ± 0.01
0.50	0.40 ± 0.30	0.36 ± 0.28	0.37 ± 0.27	0.12 ± 0.01
0.10	0.38 ± 0.28	0.35 ± 0.26	0.34 ± 0.23	0.14 ± 0.03
0.05	0.34 ± 0.29	0.33 ± 0.27	0.32 ± 0.23	0.14 ± 0.02
0.01	0.29 ± 0.23	0.26 ± 0.22	0.26 ± 0.18	0.16 ± 0.03

TABLE: Uniform errors, $m = 1$, $n = 5$, prior $N(0, 2^2)$, error unif$([-10, 10]^n)$,100 simulations

Toy Model 3 : Number of segregating sites (Watterson)

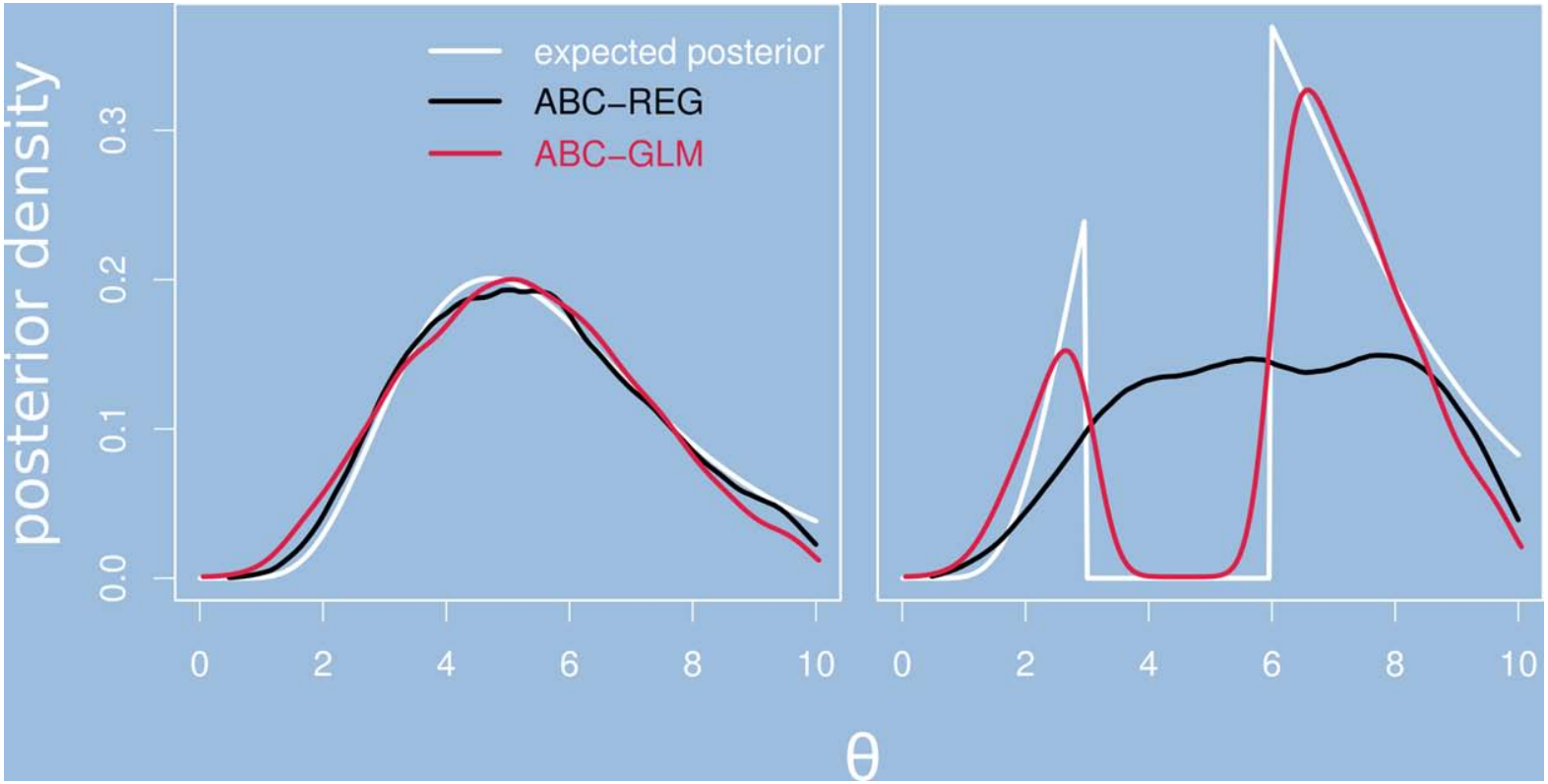

FIGURE: Population mutation parameter $\theta = 4N\mu$

And now for something completely different ...

Definition.
Let the random variable $\mathbf{T}_i = \mathbf{T}_i(\mathbf{S})$ be an m_i-dimensional function of $\mathbf{S}$. We call $\mathbf{T}_i$ *sufficient* for the parameter θ_i if the conditional distribution of $\mathbf{S}$ given $\mathbf{T}_i$ does not depend on θ_i. More precisely, let $\mathbf{t}_{i,obs} = \mathbf{T}_i(\mathbf{s}_{obs})$. Then

$$\begin{aligned}\mathbb{P}(\mathbf{S} = \mathbf{s}_{obs} | \mathbf{T}_i = \mathbf{t}_{i,obs}, \boldsymbol{\theta}) &= \frac{\mathbb{P}(\mathbf{S} = \mathbf{s}_{obs}, \mathbf{T}_i = \mathbf{t}_{i,obs} | \boldsymbol{\theta})}{\mathbb{P}(\mathbf{T}_i = \mathbf{t}_{i,obs} | \boldsymbol{\theta})} \\ &= \frac{\mathbb{P}(\mathbf{S} = \mathbf{s}_{obs} | \boldsymbol{\theta})}{\mathbb{P}(\mathbf{T}_i = \mathbf{t}_{i,obs} | \boldsymbol{\theta})} =: g_i(\boldsymbol{\theta}_{-i}),\end{aligned}$$

where $\boldsymbol{\theta}_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$ is $\boldsymbol{\theta}$ with the i-th component omitted.

MCMC Using Sufficient Statistics : Algorithm

1. Choose an index $i = 1, \ldots, n$ according to a probability distribution $(p_1, \ldots, p_n)$ with $\sum p_i = 1$ and $p_1 > 0$.
2. At $\boldsymbol{\theta} = \boldsymbol{\theta}^{(t)}$ propose $\boldsymbol{\theta}'$ according to the transition kernel $q_i(\boldsymbol{\theta}'|\boldsymbol{\theta})$ where $\boldsymbol{\theta}'$ differs from $\boldsymbol{\theta}$ only in the i-th component :

$$\boldsymbol{\theta}' = (\theta_1, \ldots, \theta_{i-1}, \theta_i', \theta_{i+1}, \ldots, \theta_n).$$

3. Generate $\mathbf{s}'$ using model $\mathcal{M}$ with parameter $\boldsymbol{\theta}'$ and calculate $\mathbf{t}_i' = \mathbf{T}_i(\mathbf{s}')$.
4. If $\mathbf{t}_i' = \mathbf{t}_{i,obs}$ go to step 5 ; otherwise stay at $\boldsymbol{\theta}$ and go to step 6.
5. Calculate the Hastings ratio

$$h(\boldsymbol{\theta}, \boldsymbol{\theta}') = \min\left(1, \frac{\pi(\boldsymbol{\theta}')q_i(\boldsymbol{\theta}|\boldsymbol{\theta}')}{\pi(\boldsymbol{\theta})q_i(\boldsymbol{\theta}'|\boldsymbol{\theta})}\right).$$

   Replace $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}'$ with probability $h(\boldsymbol{\theta}, \boldsymbol{\theta}')$; otherwise stay at $\boldsymbol{\theta}$.
6. Increase t by one unity, save a new parameter value $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}$ and continue with step 1.

Theorem. *The stationary distribution of the Markov chain is* $\pi(\boldsymbol{\theta}|\mathbf{S} = \mathbf{s}_{obs})$.