A Nonparametric Analysis of ABC

Arnaud Guyader

Journées MAS de la SMAI
27-29 août 2014, Toulouse

Framework and Objective [Marin *et al.* (2012)]

- Parameter: $\theta \in \mathbb{R}^p$ generated from the prior $\pi(\theta)$.
- Observations: $y \in \mathbb{R}^m$ generated from the likelihood $f(y|\theta)$.
- Goal: given a **fixed** observation y_0, estimate the posterior

$$\pi(\theta|y_0) = \frac{f(y_0|\theta)\pi(\theta)}{f(y_0)} \propto f(y_0|\theta)\pi(\theta).$$

- Classical Tool: MCMC methods (e.g. Metropolis algorithm), but sometimes computationally intractable...

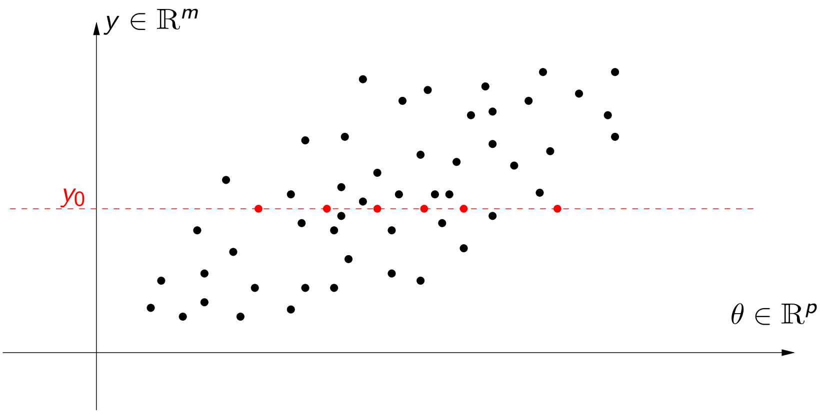
$\Rightarrow$ Another Strategy: Approximate Bayesian Computation (ABC), a family of likelihood-free computational techniques.

The Original ABC Algorithm [Rubin (1984), Tavaré *et al.* (1997)]

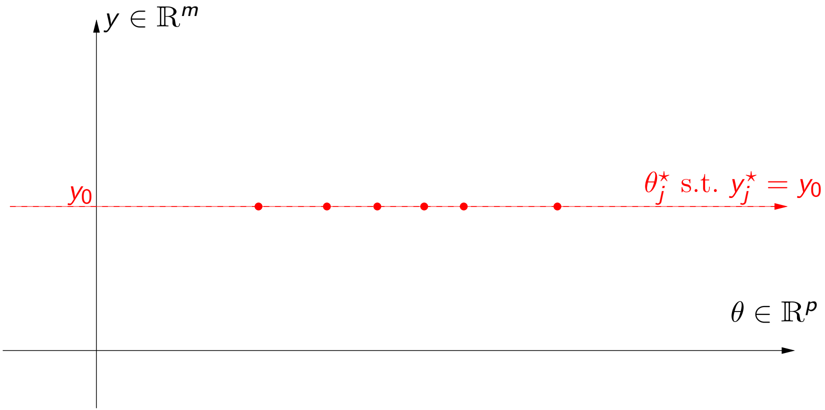
Require: An integer N
for $i = 1$ to N **do**
Generate θ_i from the prior $\pi(\theta)$
Generate y_i from the likelihood $f(.|\theta_i)$
end for
return The values $\theta_j^\star$ such that $y_j^\star = y_0$.

- Conclusion: the $\theta_j^\star$'s are i.i.d. with law $\pi(\theta|Y = y_0)$.
- Drawback: unrealistic unless the support of Y is countable.

Illustration



Illustration

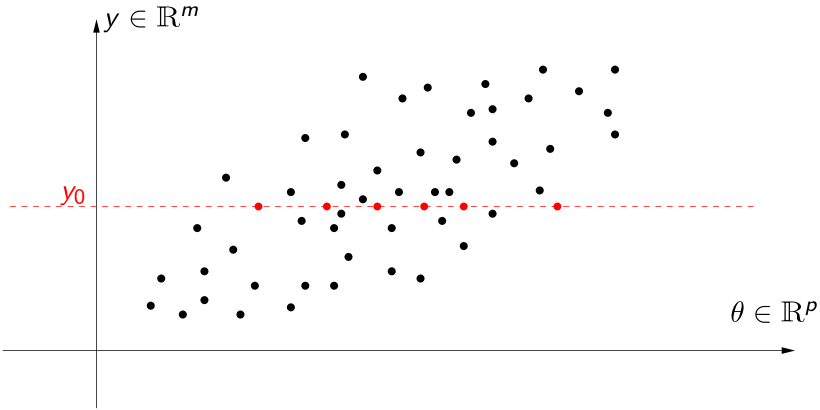


Extension of ABC [Pritchard *et al.* (1999)]

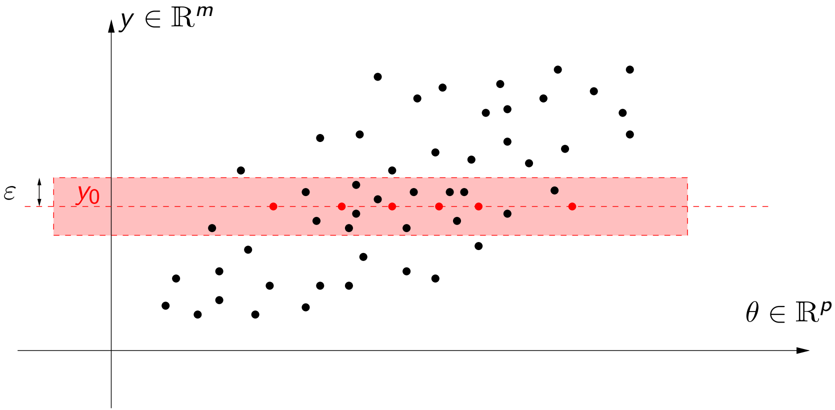
Require: An integer N, a tolerance level ε, a distance d on $\mathbb{R}^m$
for $i = 1$ to N **do**
Generate θ_i from the prior $\pi(\theta)$
Generate y_i from the likelihood $f(.|\theta_i)$
end for
return The couples $(\theta_j^\star, y_j^\star)$ such that $d(y_j^\star, y_0) \leq \varepsilon$.

- Practical Issue: use a low-dimensional summary statistic $s(y)$ and a distance $\rho(s(y), s(y_0))$ instead of $d(y, y_0)$.
- Question: how to tune ε?

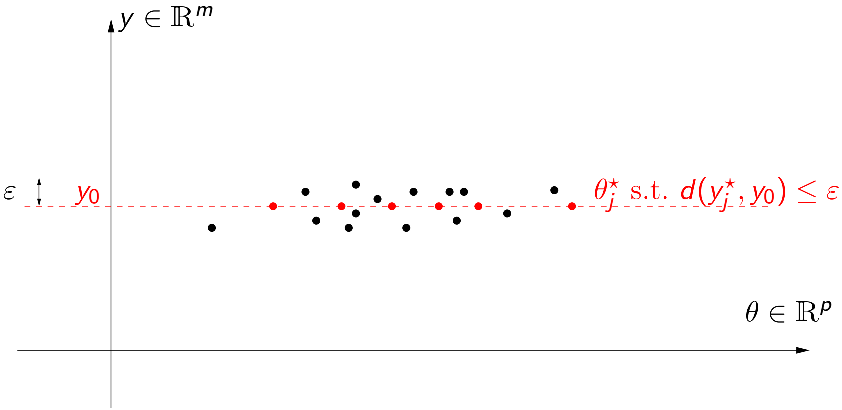
Illustration



Illustration



Illustration



ABC in Practice

Require: Integers N and k, a distance d on $\mathbb{R}^m$

for $i = 1$ to N **do**

Generate θ_i from the prior $\pi(\theta)$

Generate y_i from the likelihood $f(.|\theta_i)$

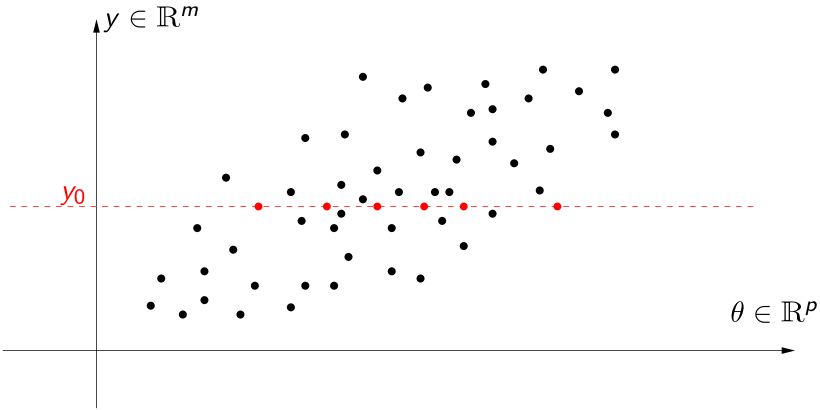
end for

return The k pairs $(\theta_j^\star, y_j^\star)$ such that $y_j^\star$ belongs to the k nearest neighbors of y_0, i.e. such that

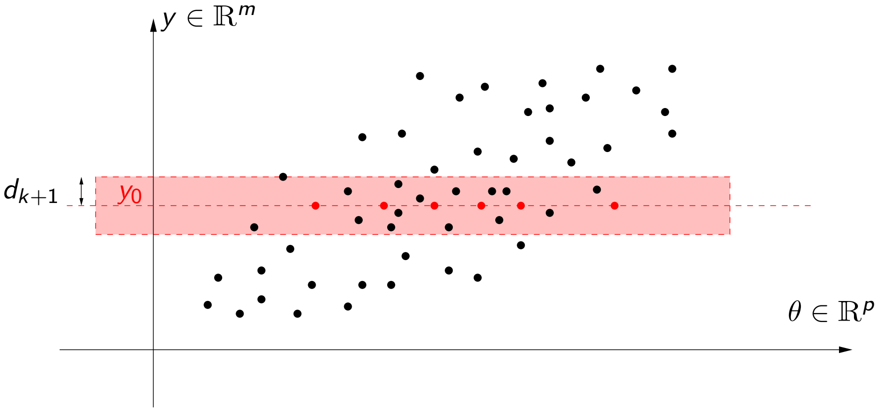
$$d(y_j^\star, y_0) < d(y_{(k+1)}, y_0) =: d_{k+1}.$$

Remark: in practice, $k = k_N$ is most commonly expressed as a percentile of N, e.g. $N = 10^6$ and $k_N/N = 0.1\%$.

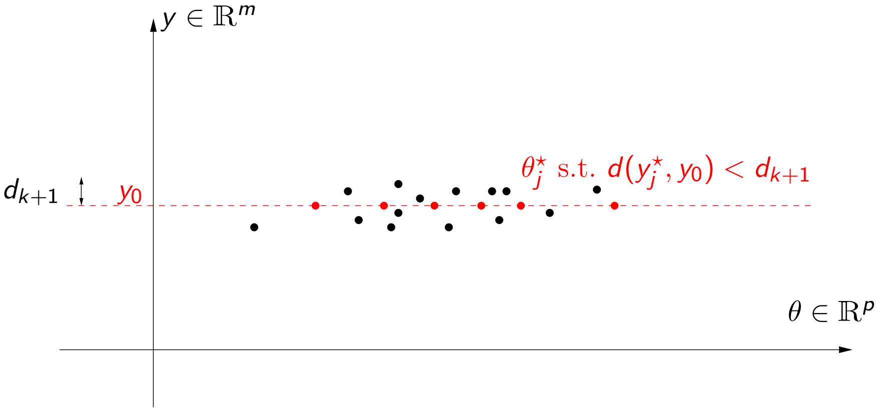
Illustration



Illustration



Illustration



Why Does It Work?

Proposition (Conditional Distribution)

Given d_{k+1}, *the* $(\Theta_j^\star, Y_j^\star)_{1\leq j\leq k}$ *are i.i.d. according to*

$$\frac{f(\theta, y)\mathbb{1}_{\mathcal{B}(y_0,d_{k+1})}}{C_{k+1}} = \frac{f(\theta, y)\mathbb{1}_{\mathcal{B}(y_0,d_{k+1})}}{\int_{\mathcal{B}(y_0,d_{k+1})} f(y)dy}$$

that is, the law $\mathcal{L}((\Theta, Y)|d(Y, y_0) < d_{k+1})$.

Corollary (Strong Law of Large Numbers)

Assume that $k_N \to +\infty$, $k_N/N \to 0$, *and* $k_N/\log\log N \to +\infty$. *Then, for any bounded function* φ, *one has*

$$\frac{1}{k_N}\sum_{j=1}^{k_N} \varphi(\Theta_j^\star) \xrightarrow[N\to+\infty]{a.s.} \mathbb{E}[\varphi(\Theta)|Y = y_0].$$

Kernel Density Estimate

- Density Estimator:

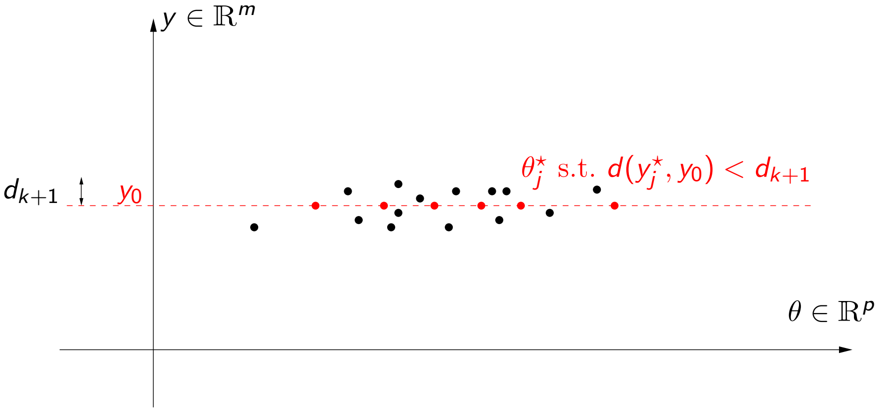
$$\hat{\pi}_N(\theta_0|y_0) = \frac{1}{k_N h_N^p} \sum_{j=1}^{k_N} K\left(\frac{\Theta_j^\star - \theta_0}{h_N}\right).$$

- This is a hybrid between a k-nearest neighbor and a kernel density estimation procedure.
- Remark: an additional degree of smoothing [Blum (2010)]

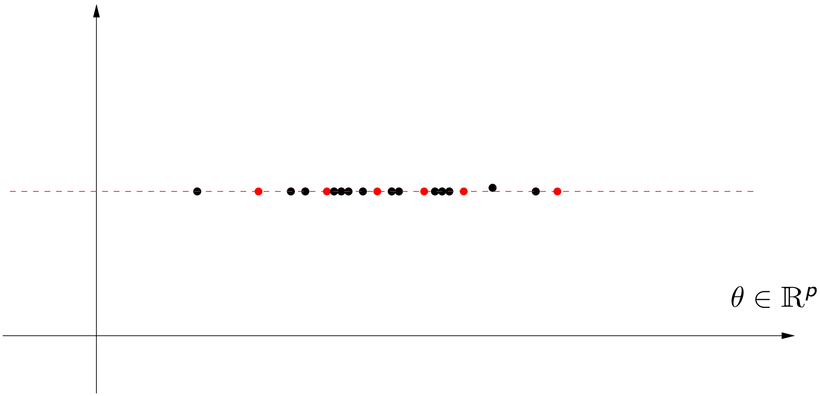
$$\tilde{\pi}_N(\theta_0|y_0) = \frac{\sum_{i=1}^N L\left(\frac{y_i - y_0}{\delta_N}\right) K\left(\frac{\Theta_i - \theta_0}{h_N}\right)}{h_N^p \sum_{i=1}^N L\left(\frac{y_i - y_0}{\delta_N}\right)}.$$

$\Rightarrow$ Questions: Consistency? Rates of convergence?

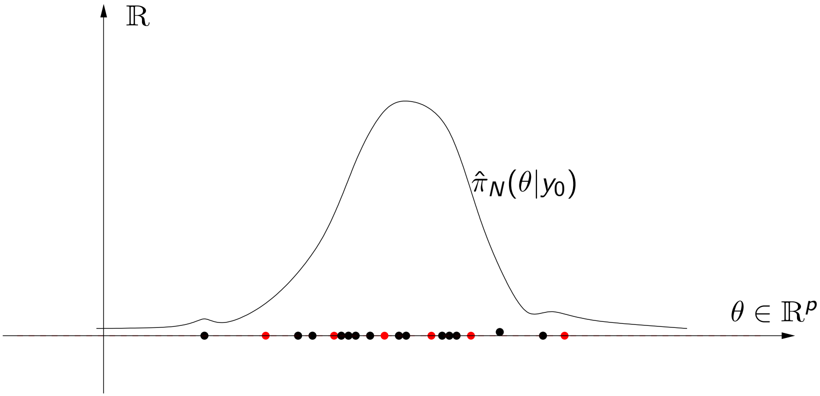
Illustration



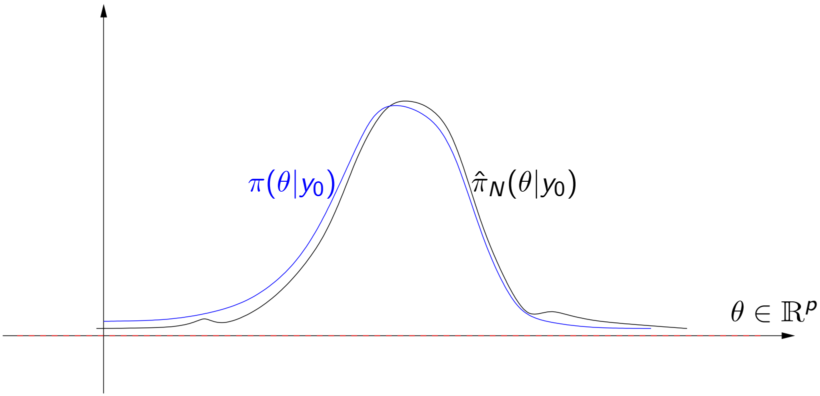
Illustration



Illustration



Illustration



Pointwise Mean Square Error Consistency

Theorem

Assume that the joint probability density f is such that

$$\int_{\mathbb{R}^p}\int_{\mathbb{R}^m} f(\theta, y)\log^+ f(\theta, y) d\theta dy < \infty.$$

If $k_N \to \infty$, $k_N/N \to 0$, $h_N \to 0$ *and* $k_N h_N^p \to \infty$, *then*

$$\mathbb{E}\left[\left(\hat{\pi}_N(\theta_0|y_0) - \pi(\theta_0|y_0)\right)^2\right] \xrightarrow[N\to\infty]{\lambda_p \otimes \lambda_m \ a.e.} 0.$$

Remark: the assumption on f is not very restrictive...

Bias-Variance Decomposition

Conditioning on $d_{k+1} = d_{k_N+1}$ yields

$$\mathbb{E}\left[\left(\hat{\pi}_N(\theta_0|y_0) - \pi(\theta_0|y_0)\right)^2\right] = \mathbb{E}\left[B(d_{k+1})^2\right] + \mathbb{E}\left[V(d_{k+1})\right],$$

where

$$B(d_{k+1}) = \mathbb{E}[\hat{\pi}_N(\theta_0|y_0) \,|\, d_{k+1}] - \pi(\theta_0|y_0),$$

and

$$V(d_{k+1}) = \mathbb{E}\left[\left(\hat{\pi}_N(\theta_0|y_0) - \mathbb{E}[\hat{\pi}_N(\theta_0|y_0) \,|\, d_{k+1}]\right)^2 \,\Big|\, d_{k+1}\right].$$

The Bias Term

Recall: We have to prove that $\mathbb{E}[B(d_{k+1})^2] \to 0$, with

$$B(d_{k+1}) = \mathbb{E}[\hat{\pi}_N(\theta_0|y_0) \big| d_{k+1}] - \pi(\theta_0|y_0),$$

where $\pi(\theta_0|y_0) = f(\theta_0, y_0)/f(y_0)$, and

$$\mathbb{E}[\hat{\pi}_N(\theta_0|y_0) \mid d_{k+1}] = \left(\frac{1}{V_m\, d_{k+1}^m} \int_{\mathcal{B}(\mathbf{y_0}, d_{k+\mathbf{1}})} f(y) \mathrm{d}y \right)^{-1}$$
$$\times \left(\frac{1}{V_m\, d_{k+1}^m} \int_{\mathbb{R}^p} \int_{\mathcal{B}(\mathbf{y_0}, d_{k+\mathbf{1}})} K_h(\theta - \theta_0) f(\theta, y) \mathrm{d}\theta \mathrm{d}y \right)$$

$\Rightarrow$ Tools: Extensions of Lebesgue's differentiation theorem, and of Jessen-Marcinkiewicz-Zygmund theorem.

The Variance Term

Recall that

$$\mathbb{E}\left[V(d_{k+1})\right] = \mathbb{E}\left[\mathbb{E}\left[\left(\hat{\pi}_N(\theta_0|y_0) - \mathbb{E}[\hat{\pi}_N(\theta_0|y_0) \mid d_{k+1}]\right)^2 \mid d_{k+1}\right]\right].$$

Thus, assuming that $\|K\|_\infty = \sup K(\theta) < \infty$, we are led to

$$\mathbb{E}\left[V(d_{k+1})\right] \leq \frac{C(\theta_0, y_0)\|K\|_\infty}{k_N h_N^p},$$

and everything is OK, provided that

$$k_N h_N^p \xrightarrow[N\to\infty]{} \infty.$$

Rates of Convergence

Theorem (MISE in the case $m > 4$)

Assume that Y has a bounded support. Then, under some regularity assumptions on $f(\theta, y)$ and $f(y)$, we have

$$\mathbb{E}\left[\int_{\mathbb{R}^p} \left[\hat{\pi}_N(\theta_0|y_0) - \pi(\theta_0|y_0)\right]^2 d\theta_0\right] \leq \frac{\int_{\mathbb{R}^p} K^2(\theta) d\theta}{k_N h_N^p}$$

$$+ A(y_0)\left(\frac{k_N}{N}\right)^{\frac{4}{m}} + B(y_0)\left(\frac{k_N}{N}\right)^{\frac{2}{m}} h_N^2 + C(y_0) h_N^4 + o\left(\left(\frac{k_N}{N}\right)^{\frac{4}{m}} + h_N^4\right)$$

$\Rightarrow$ For $k_N \propto N^{\frac{p+4}{m+p+4}}$ and $h_N \propto N^{\frac{-1}{m+p+4}}$, this leads to

$$\mathbb{E}\left[\int_{\mathbb{R}^p} \left[\hat{\pi}_N(\theta_0|y_0) - \pi(\theta_0|y_0)\right]^2 \mathrm{d}\theta_0\right] \leq D(y_0)\ N^{\frac{-4}{m+p+4}}.$$