Approximate Bayesian Computation under model uncertainty (ABCμ)

Oliver Ratmann[1] C Andrieu[2] C Wiuf[3] S Richardson [1]

[1]Centre for Biostatistics, Imperial College, London, UK
[2]Dep of Mathematics, University of Bristol, Bristol, UK
[3]Bioinformatics Research Centre, University of Aarhus, Aarhus, DK

These slides attempt to present the material of the first section of our recently published work in PNAS [Ratmann O, PNAS 2009]. We expand on the motivation behind ABCμ, and provide a toy example that is easy to re-implement.

These slides do not include material on the second part of the PNAS paper that discusses a new ABCμ algorithm. However, one purpose here is to exemplify with Std-ABC that any ABC method can be modified for the purpose of model criticism.

When use ABC / ABCμ ?

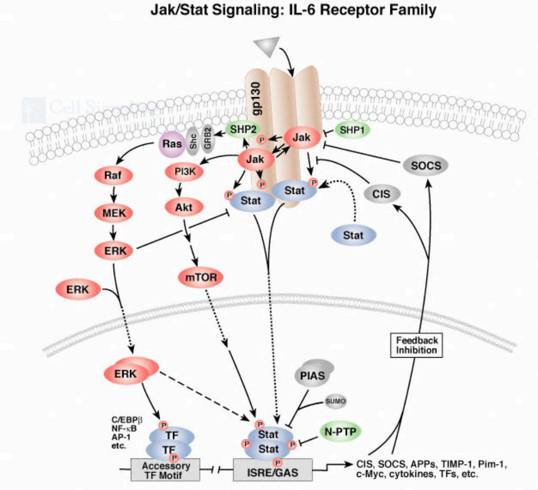
Inference on data-generating stochastic processes

[Diggle PJ, RSSB 1984]

... whose likelihood cannot be readily evaluated

... within Bayesian framework to
(a) infer model parameters,
(b) error terms to see if model adequate

Signalling pathways, [Toni T RoySocInt 2008]



"All models are wrong but some are useful [Box GEP, JASA1976]"

Statistical reasoning in science

1. perform experiment
2. match model to data, assuming that model is correct
3. explore if model is adequate to explain the data in one or several aspects:
   - formally test null hypothesis that model is "adequate"
   - if not, use diagnostics to guide model developments
4. revise model, motivating new experiment

From ABC to ABCμ: outline

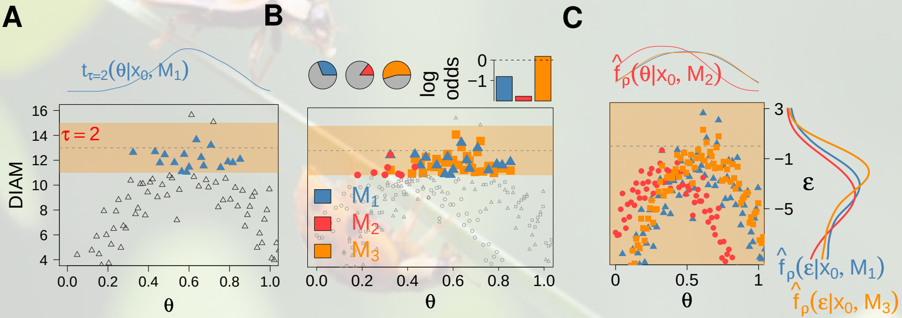
ABC

- exploit model predictions for sampling from approximate posterior densities of θ

ABCμ

- Typically, model predictions form the basis for model criticism [Jeffreys, Th Probability 1961]
- use the data already generated in ABC for this purpose too
- dual use of model predictions reflected in an extension of the state space: we are interested in θ and $\varepsilon_{1:K}$
- to maintain Bayesian framework, crux is to derive a likelihood on the augmented space $(\theta, \varepsilon_{1:K})$.

From ABC to ABCμ: outline



From ABC to ABCμ: notation

- θ random variable, x_0 observed data
- M data-generating stochastic process
- $x \sim f(\cdot|\theta, M)$ simulated data under M given θ

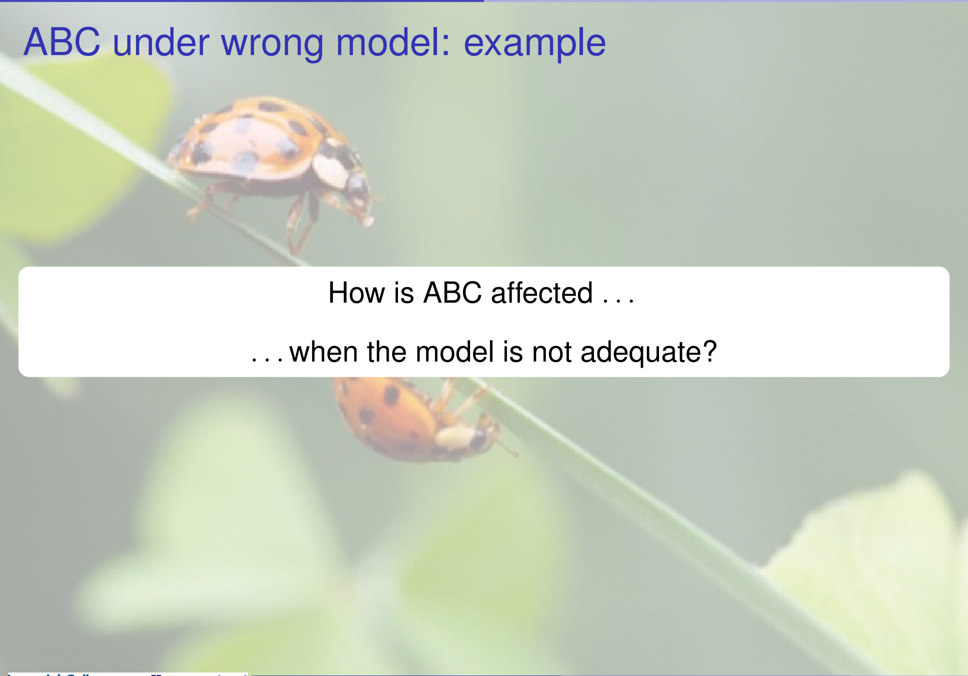
- lower dimensional summaries $\mathbb{S} = \{S_1, \dots, S_K\}$
- introduce mismatch thresholds τ_k
- approximate likelihood here with

$$t_{\rho,\tau}(x_0|\theta, M) = \frac{1}{\prod_k \tau_k} \int \mathbf{1}\Big\{\bigcap_{k=1}^{K} |S_k(x) - S_k(x_0)| \leq \tau_k/2\Big\} f(x|\theta, M)dx$$

- apply Bayes' Theorem

$$t_{\rho,\tau}(\theta|x_0, M) = t_{\rho,\tau}(x_0|\theta, M)\pi_\theta(\theta|M) \,/\, f_{\rho,\tau}(x_0|M)$$

ABC under wrong model: example



How is ABC affected . . .

. . . when the model is not adequate?

ABC under wrong model: example

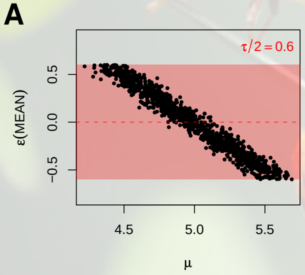
Simulation study

- observe dataset x_0 of 100 sample points, $\overline{x_0} = 5$, 0.25 Quantile $Q_0 = 1.48$
- we believe $x_i \sim \mathcal{N}(\theta, 1)$ whereas in reality $x_i \sim \text{Exp}(0.2)$
- summarize data with 0.25 Quantile Q and $\overline{x}$
- consider ABC inference for separate as well as for joint summaries:

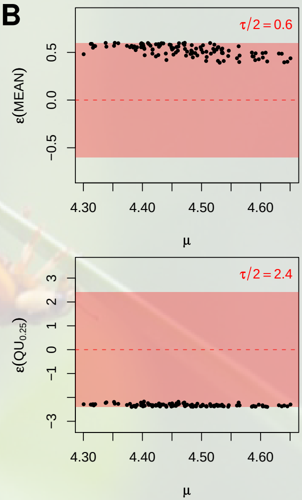
$$\rho\big(\overline{x_0}, Q(x_0)\,;\,\overline{x}, Q(x)\big) = \frac{1}{\tau_1}\mathbf{1}\Big\{\big|\overline{x_0} - \overline{x}\big| \leq \tau_1/2\Big\} \times \frac{1}{\tau_2}\mathbf{1}\Big\{\big|Q(x_0) - Q(x)\big| \leq \tau_2/2\Big\},$$

- reconstruct $t_{\rho,\tau}(\theta|x_0, M)$ with Std-ABC

ABC under wrong model: example (cont'd)

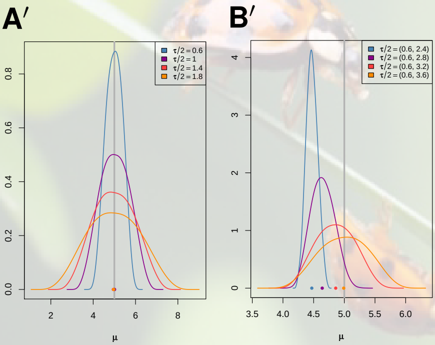


$$\mathbb{S} = \overline{x}$$



$$\mathbb{S} = \{\overline{x}, Q\}$$

ABC under wrong model: example (cont'd)



$\mathbb{S} = \overline{x}$ $\qquad$ $\mathbb{S} = \{\overline{x}, Q\}$

- But what is the meaning of posterior estimates of θ?

make use of the data already generated

Intuitively, want to make information in **B** (computed with ABC) available as job output & interpret; e.g. want to interpret posterior $\varepsilon_{\overline{x}}$ as retained points $\overline{x}_0 - \overline{x}$. Two steps:

1. Take ABC kernel as prior $\pi_{\varepsilon_{1:K}}$.
2. Propose new likelihood definition that provides desired interpretation.

ABCμ: Derivation of augmented likelihood

Predictive error of summaries given θ

- Given θ, consider the "predictive error" ε with probability distribution

$$\mathbb{P}_{\theta,x_0}(\varepsilon \leq e) = \int_{\mathcal{X}} \mathbf{1}\Big\{\rho\big(\mathbb{S}(x), \mathbb{S}(x_0)\big) \leq e\Big\} f(x|\theta, M) dx$$

- $\mathcal{X}$ often finite $\rightarrow$ integral replaced by sum
- not difficult to generalize to multiple errors

ABCμ: Derivation of augmented likelihood

Conditional predictive error density

- Assume $\mathbb{P}_{\theta,x_0}(\varepsilon \leq e)$ has density ξ_{θ,x_0} w.r.t. appropriate measure
- Write ξ_{θ,x_0} as elementary derivative (a. e.)

$$\begin{aligned}
&\xi_{\theta,x_0}(\varepsilon) \\
&\quad = \lim_{h\to 0} \frac{1}{2h} \int \mathbf{1}\Big\{\rho(\mathbb{S}(x),\mathbb{S}(x_0)) - h \;<\; \varepsilon \;\leq \\
&\qquad\qquad \rho(\mathbb{S}(x),\mathbb{S}(x_0)) + h\Big\}\; f(x|\theta,M)dx \\
&\quad = \lim_{h\to 0} \int \delta_h\Big(\rho(\mathbb{S}(x),\mathbb{S}(x_0)) - \varepsilon\Big)\; f(x|\theta,M)dx \\
&\quad \overset{\text{def}}{=} \int \delta\{\rho(\mathbb{S}(x),\mathbb{S}(x_0)) = \varepsilon\} f(x|\theta,M)dx
\end{aligned}$$

ABCμ: Derivation of augmented likelihood

Augmented likelihood on (θ, ε)

- set

$$f_\rho(x_0|\theta,\varepsilon,M) \stackrel{\text{def}}{=} \quad \xi_{\theta,x_0}(\varepsilon) = \int \delta\Big\{\rho(\mathbb{S}(x),\mathbb{S}(x_0)) = \varepsilon\Big\} f(x|\theta,M)dx$$

Interpretation

By construction:

$\rho \neq 0 \;\sim\;$ discrepancies between data and model

- "expect" mode of $\xi_{\theta,x_0}(\varepsilon) \approx 0$ for some θ
  only if model matches data
- if mode "far away" from 0 for all θ,
  detect model mismatch under that discrepancy

Joint posterior density

Joint posterior density of θ and summary errors

Given prior $\pi(\theta, \varepsilon|M)$, apply Bayes' Theorem

$$f_\rho(\theta, \varepsilon|x_0, M) = \xi_{\theta,x_0}(\varepsilon)\, \pi(\theta, \varepsilon|M) \,/\, f_\rho(x_0|M)$$

- take $\pi(\theta, \varepsilon|M_i) = \pi_\varepsilon(\varepsilon|M)\pi_\theta(\theta|M)$
- choose $\pi(\varepsilon|M)$ with mode at zero

"Standard" ABCμ algorithm

"Std"-ABCμ [Ratmann, PNAS 2009]

Std-ABCμ1 Sample $\theta \sim \pi(\theta|M)$, simulate $x \sim f(\cdot|\theta, M)$ and compute $\varepsilon = \rho\big(\mathbb{S}(x), \mathbb{S}(x_0)\big)$.

Std-ABCμ2 Accept (θ, ε, x) with probability proportional to $\pi(\varepsilon|M)$, and go to Std-ABCμ1.

Cmp'd to Standard-ABC almost unchanged (only record realized errors) & no additional cost

Marginally in (θ, ε) we obtain samples from $f_\rho(\theta, \varepsilon|x_0, M)$ by construction. Exploit expression in terms of predictive density to interpret $f_\rho(\theta, \varepsilon|x_0, M)$.

ABCμ: parameter inference

Approximation of true posterior $f(\theta|x_0, M)$

Under regularity assumptions on ξ_{θ,x_0}

$$f_\rho(\theta|x_0, M) \propto \pi_\theta(\theta|M) \int \pi_\varepsilon\Big(\rho\big(\mathbb{S}(x), \mathbb{S}(x_0)\big)\Big) f(x|\theta, M) dx$$

- $\pi_\varepsilon(\varepsilon|M) = \mathbf{1}\{|\varepsilon| \leq \tau/2\}/\tau \;\Rightarrow$ "standard" ABC
- may interpret ABC kernel $\frac{1}{\tau}K\big(\rho(\mathbb{S}(x), \mathbb{S}(x_0)/\tau)\big)$ as particular prior belief on M [Wilkinson R,2008]

We suggest not to stop at prior interpretation of error . . .

ABCμ: model criticism

... **Posterior ε is a compound variable that reflects stochastic fluctuations and systematic biases between the model and x_0.**

K real-valued summary errrors

In contrast to ABC where error $\rho\big(\mathbb{S}(x), \mathbb{S}(x_0)\big)$ is scalar & positive,

consider K real-valued summary errors ε_k
that correspond to $\rho_k\big(S_k(x), S_k(x_0)\big)$.

$\rightarrow$ gain of interpretability and diagnostic value

ABCμ: model criticism

Formally . . .

- define multiple, real-valued prior error terms with density

$$\pi_{\varepsilon_{1:K}}(\varepsilon_{1:K}|M) \stackrel{\text{def}}{=} \prod_{k=1}^{K} 1/\tau_k \; K(\varepsilon_k/\tau_k)$$

think: $K(\varepsilon_k/\tau_k) = \mathbf{1}\{|\varepsilon_k| \leq \tau_k/2\}$

- set augmented likelihood to joint predictive density

$$f_{\rho,\tau}(x_0|\theta, \varepsilon_{1:K}, M) \stackrel{\text{def}}{=} \xi_{\theta,x_0}(\varepsilon_{1:K}).$$

ABCμ: model criticism

Interpretation of posterior $\varepsilon_{1:K}$

Under regularity assumptions on ξ_{θ,x_0}

$$f_{\rho,\tau}(\varepsilon_{1:K}|x_0,M) = \int f_{\rho,\tau}(\theta,\varepsilon_{1:K}|x_0,M)d\theta$$
$$= \pi_{\varepsilon_{1:K}}(\varepsilon_{1:K}|M)L_\rho(\varepsilon_{1:K}|M) \;/\; f_{\rho,\tau}(x_0|M)$$

where (prior) predictive density [Box, RSSA 1980] of summary errors is

$$L_\rho(\varepsilon_{1:K}|M) \stackrel{\text{def}}{=} \lim_{h\to 0}\int_{\mathcal{X}} \delta_h\Big(\big(\rho_k(S_k(x),S_k(x_0)) - \varepsilon_k\big)_{1:K}\Big)\,\pi(x|M)dx$$

and (prior) predictive density of data is

$$\pi(x|M) \stackrel{\text{def}}{=} \int f(x|\theta,M)\pi(\theta|M)d\theta$$

ABCμ: model criticism

$$f_{\rho,\tau}(\varepsilon_{1:K}|x_0, M) = \int f_{\rho,\tau}(\theta, \varepsilon_{1:K}|x_0, M)d\theta$$
$$\propto L_\rho(\varepsilon_{1:K}|M)\pi_{\varepsilon_{1:K}}(\varepsilon_{1:K}|M)$$

- towards more accurate model criticism
  inspect multiple errors jointly, no problem for Monte Carlo methods
- "weighted" prior predictive error according to error magnitude
  - focus on those θ actually inferred
  - criticize posterior model
  - alleviate sensitivity of $L_\rho(\varepsilon_{1:K}|M)$ on $\pi_\theta(\theta|M)$
- computationally feasible

ABCμ: model criticism

Posterior mean shift

[unpublished material taken out]

ABCμ: model criticism

[unpublished material taken out]

kth mean shift re-weighted according to mutual constraints

[unpublished material taken out]

ABCμ: model criticism

Recall: summaries typically co-dependent and not clear if sufficient for θ under *M*

Exploit co-dependency of summary errors in ABCμ

- If ε_k independent, then mutual constraints collapse [unpublished material taken out]
- Mutual constraints require co-dependent summaries
- Mutual constraints do not require summaries to be sufficient for θ under *M*

ABCμ reveals model inconsistency with conflicting, co-dependent summaries

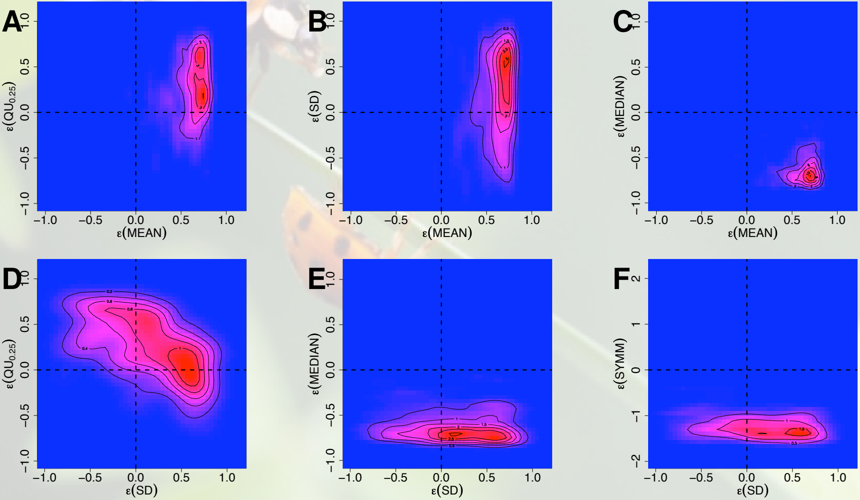
ABCμ under wrong model: example 2

Simulation study

- observe dataset x_0 of 100 sample points, $\overline{x}_0 = 5$, 0.25 Quantile $Q_0 = 1.48$
- we believe $x_i \sim \mathcal{N}(\mu, \sigma^2)$ whereas in reality $x_i \sim \text{Exp}(0.2)$ [unpublished material taken out]

ABCμ under wrong model: example 2

- Std-ABCμ with 5 summary errors



Concluding remarks 1

Utility of posterior summary errors $\varepsilon_{1:K}$

- Exploit the data already generated in ABC also for model criticism
- Formally, dual use of simulated data is described by the joint posterior density

$$f_{\rho,\tau}(\theta, \varepsilon_{1:K} | x_0, M)$$

  that rests on augmenting the likelihood in a particular way
- Marginally in θ, we perform approximate inference exactly as in ABC
- Marginally in $\varepsilon_{1:K}$, we criticize the fitted model with all summaries simultaneously

Concluding remarks 2

Choice of summary errors $\varepsilon_{1:K}$

- we always work with K errors that correspond to K summaries
- we always include as many summaries as possible to preserve co-dependencies;
  compare to [Joyce P, Stat App Gen Mol Biol 2008]
- we do not combine summaries into scalar error
- we always choose τ so that co-dependencies are preserved

Concluding remarks 3

Diagnostics and hypothesis testing

- we use high probability density intervals of marginals $f_{\rho,\tau}(\varepsilon_k|x_0, M)$ to obtain precise information as to how improve a model
- these are not confidence intervals, and marginal posterior errors are typically not independent
- to test $H_0 : \varepsilon_{1:K} = 0$ we compute a Bayes factor [unpublished material taken out]

Concluding remarks 4

Std-ABCμ does not work well in other than very simple settings

- Prob1: $f_{\rho,\tau}(\theta, \varepsilon_{1:K}|x_0, M)$ typically very different from $\pi_\theta(\theta|M)\prod_{k=1}^{K}\pi_{\varepsilon_k}(\varepsilon_k|M)$
- Prob2: By simulating $x \sim f(\cdot|\theta, M)$, we extend the state space $(\theta, \varepsilon_{1:K})$ with x. The volatility of the stochastic process $f(\cdot|\theta, M)$ induces an extra price to pay.
- Prob3: Compared to scalar error, acceptance probability of ABCμ based on several summary errors $\varepsilon_{1:K}$ is further reduced.

But straightforward to modify existing methods of ABC. We also implement a novel ABCμ algorithm . . .

Concluding remarks 5

- We use **repeated sampling** $x_b \sim f(\cdot|\theta, M)$, $b = 1, \ldots, B$ to control the volatility of the data generating process. [Andrieu C, Ann Stat 2009]
- We also replaced the joint likelihood $\xi_{\theta,x_0}(\varepsilon_{1:K})$ with a conservative **combination of its marginals** $\min_k \xi_{k,\theta,x_0}(\varepsilon_k)$ [Ratmann O, PNAS 2009]
- For our applications, we found **MCMC samplers (with annealing & refined proposals)** to work well [Ratmann O, PLOS CompBiol 2007]

Thank you

Sylvia Richardson, Christophe Andrieu, Carsten Wiuf

and

Wellcome Trust