On the use of ABC-MCMC with inflated tolerance and post-correction

Matti Vihola
Department of Mathematics and Statistics, University of Jyväskylä

One world ABC seminar, 10 December 2020

Joint work with Jordan Franks (Newcastle)

Approximate Bayesian computation

Model:

- $\mathrm{pr}(\theta)$ — the prior
- $g(y \mid \theta)$ — the observation model
- y^* — the data (observed values of y)
- posterior $\pi(\theta) \propto \mathrm{pr}(\theta) g(y^* \mid \theta)$

Approximate Bayesian computation

Model:

- $\mathrm{pr}(\theta)$ — the prior
- $g(y \mid \theta)$ — the observation model
- y^* — the data (observed values of y)
- posterior ~~$\pi(\theta) \propto \mathrm{pr}(\theta) g(y^* \mid \theta)$~~

ABC is inference methods for models that

× have intractable (or extremely expensive) likelihood $g(y^* \mid \theta)$, but

Approximate Bayesian computation

Model:

- $\mathrm{pr}(\theta)$ — the prior
- $g(y \mid \theta)$ — the observation model
- y^* — the data (observed values of y)
- posterior $\pi(\theta) \propto \mathrm{pr}(\theta) g(y^* \mid \theta)$

ABC is inference methods for models that

- × have intractable (or extremely expensive) likelihood $g(y^* \mid \theta)$, but
- ✓ (cheap) simulations of pseudo-data $Y \sim g(\,\cdot \mid \theta)$.

Algorithmic description

ABC rejection sampler:

(R1) Draw $\Theta \sim \mathrm{pr}(\,\cdot\,)$ and $Y \sim g(\,\cdot \mid \Theta)$.

(R2) If $\|s(Y) - s(y^*)\| \leq \epsilon$, output (Θ, Y); otherwise go to (R1).

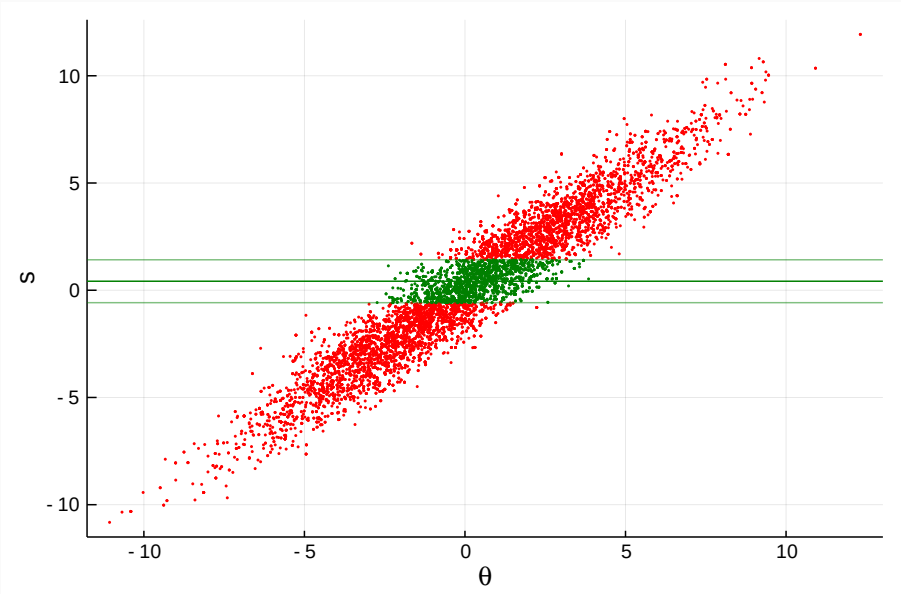
where

- $\|\,\cdot\,\|$ is Euclidean distance on $\mathbb{R}^d$
- s extracts d summary statistics from the (pseudo-)observations.

It is not hard to see that the distribution of (Θ, Y) is

$$\tilde{\pi}_\epsilon(\theta, y) \propto \mathrm{pr}(\theta) g(y \mid \theta) \mathbb{I}\{\|s(y) - s(y^*)\| \leq \epsilon\}.$$

Example: $\mathrm{pr}(\theta) = N(\theta; 0, 3^2)$, $g(y \mid \theta) = N(y \mid \theta, 1)$, $s(y) = y$, $s(y^*) = 0.42$, $\epsilon = 1.0$



Pseudo-posterior and pseudo-likelihood

If we only consider Θ, its marginal is the 'pseudo-posterior'

$$\pi_\epsilon(\theta) = \frac{1}{Z_\epsilon}\mathrm{pr}(\theta)\int g(y \mid \theta)\mathbb{I}\{\|s(y) - s(y^*)\| \leq \epsilon\}\mathrm{d}y = \mathrm{pr}(\theta)\frac{\ell_\epsilon(y^* \mid \theta)}{Z_\epsilon},$$

where $\ell_\epsilon(y^* \mid \theta)$ is a 'pseudo-likelihood'

$$\ell_\epsilon(y^* \mid \theta) = \mathbb{P}\big(\|s(Y) - s(y^*)\| \leq \epsilon\big) \qquad \text{where} \qquad Y \sim g(\,\cdot \mid \theta)$$

How tolerance parameter ϵ affects the pseudo-posterior?

- If $\epsilon \to \infty$, then $\ell_\epsilon(y^* \mid \theta) \nearrow 1 \implies \pi_\epsilon(\theta) \to \mathrm{pr}(\theta)$.

Pseudo-posterior and pseudo-likelihood

If we only consider Θ, its marginal is the 'pseudo-posterior'

$$\pi_\epsilon(\theta) = \frac{1}{Z_\epsilon}\mathrm{pr}(\theta)\int g(y \mid \theta)\mathbb{I}\{\|s(y) - s(y^*)\| \le \epsilon\}\mathrm{d}y = \mathrm{pr}(\theta)\frac{\ell_\epsilon(y^* \mid \theta)}{Z_\epsilon},$$

where $\ell_\epsilon(y^* \mid \theta)$ is a 'pseudo-likelihood'

$$\ell_\epsilon(y^* \mid \theta) = \underbrace{\mathbb{P}\big(\|s(Y) - s(y^*)\| \le \epsilon\big)}_{\approx 2\epsilon g(y^*|\theta) \text{ in the example}} \qquad \text{where} \qquad Y \sim g(\,\cdot \mid \theta)$$

How tolerance parameter ϵ affects the pseudo-posterior?

- If $\epsilon \to \infty$, then $\ell_\epsilon(y^* \mid \theta) \nearrow 1 \implies \pi_\epsilon(\theta) \to \mathrm{pr}(\theta)$.

If $s(\,\cdot\,)$ are sufficient (& some regularity conditions):

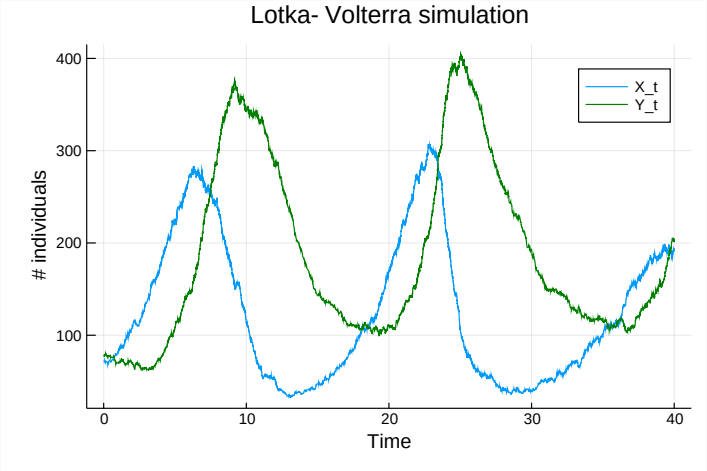
- if $\epsilon \to 0$, then $Z_\epsilon^{-1}\ell_\epsilon(y^* \mid \theta) \to Z^{-1}\ell(y^* \mid \theta) \implies \pi_\epsilon(\theta) \to \pi(\theta)$.

Example: Lotka-Volterra model i

- Markov process $(X_t, Y_t)_{t\geq 0}$ with jump rates:

$$\theta_1 \ : \ X_t \to X_t + 1$$
$$\theta_2 \ : \ X_t \to X_t - 1 \ \& \ Y_t \to Y_t + 1$$
$$\theta_3 \ : \ Y_t \to Y_t - 1$$

- Simulation with:
  - $(X_0, Y_0) = (71, 79)$
  - $\boldsymbol{\theta} = (0.5, 0.0025, 0.3)$.



Example: Lotka-Volterra model ii

Consider ABC with prior

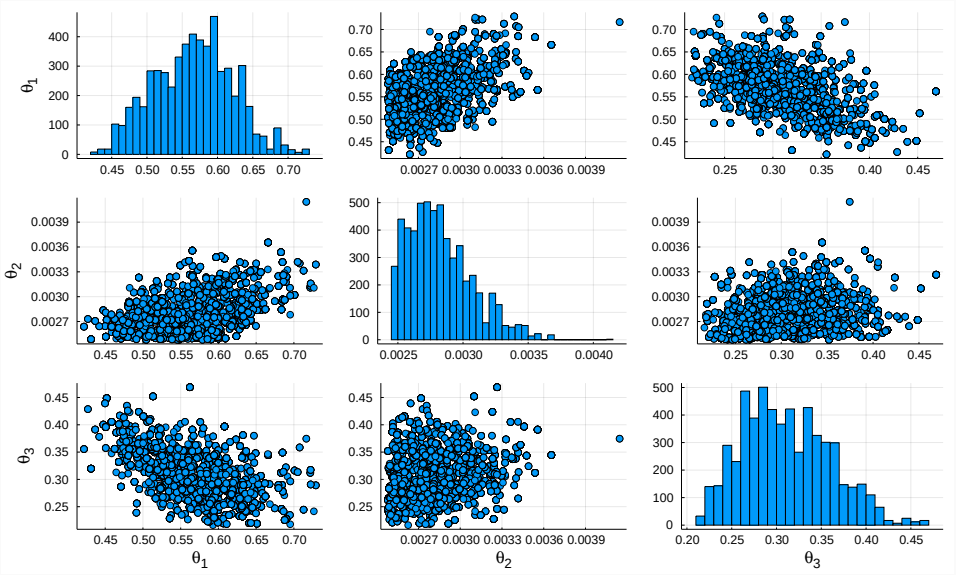
$$(\log \theta_1, \log \theta_2, \log \theta_3) \sim U([-6, 0]^3)$$

and (rather arbitrary?) summaries based on discrete-time observations $\tilde{X}_k = X_{5k}$ and $\tilde{Y}_k = Y_{5k}$:

- sample autocorrelation of $\tilde{X}_k$ at lag 2
- mean and standard deviations of $\tilde{X}_k$ and $\tilde{Y}_k$

Observed values $(-46.32, 129.0, 88.64, 202.6, 140.8)$.

Lotka-Volterra: Inference with $\epsilon = 40$ (data from $\theta = (0.5, 0.0025, 0.3)$)



ABC Markov chain Monte Carlo (Marjoram et al., 2003)

Let $q(y \mid x)$ be a proposal, start with some Θ_0 and $L_0 = 1$, and for $k = 1, \ldots, n$:

- Draw $\Theta'_k \sim q(\,\cdot \mid \Theta_{k-1})$ and $Y'_k \sim g(\,\cdot \mid \Theta'_k)$.
- Calculate $L'_k := \mathbb{I}\{\|s(Y'_k) - s(y^*)\| \leq \epsilon\}$.
- With probability

$$\min\left\{1, \frac{\mathrm{pr}(\Theta'_k) L'_k q(\Theta_{k-1} \mid \Theta'_k)}{\mathrm{pr}(\Theta_{k-1}) L_{k-1} q(\Theta'_k \mid \Theta_{k-1})}\right\},$$

  accept and set $(\Theta_k, L_k) \leftarrow (\Theta'_k, L'_k)$;
  otherwise reject and set $(\Theta_k, L_k) \leftarrow (\Theta_{k-1}, L_{k-1})$.

ABC can be viewed as a pseudo-marginal MCMC (Andrieu & Roberts 2009)...

This is the (purely theoretical!) 'marginal' variant:

- Draw $\Theta'_k \sim q(\,\cdot\mid \Theta_{k-1})$
- Calculate $\ell_\epsilon(y^* \mid \Theta'_k) = \mathbb{P}\big(\|s(Y'_k) - s(y^*)\| \le \epsilon\big)$.
- With probability

$$\min\left\{1, \frac{\mathrm{pr}(\Theta'_k)\ell_\epsilon(y^* \mid \Theta'_k)q(\Theta_{k-1} \mid \Theta'_k)}{\mathrm{pr}(\Theta_{k-1})\ell_\epsilon(y^* \mid \Theta_{k-1})q(\Theta'_k \mid \Theta_{k-1})}\right\},$$

accept and set $\Theta_k \leftarrow \Theta'_k$; otherwise reject and set $\Theta_k \leftarrow \Theta_{k-1}$.

Why ABC-MCMC instead of Rejection sampling ABC, SMC-ABC, BOLFI??

ABC-MCMC is good because:

- Can use diffuse (uninformative, even improper) prior unlike rejection sampling.
- Admits convergence guarantees (to π_ϵ).
- Simple to use (hopefully!)

Why ABC-MCMC instead of Rejection sampling ABC, SMC-ABC, BOLFI??

ABC-MCMC is good because:

- Can use diffuse (uninformative, even improper) prior unlike rejection sampling.
- Admits convergence guarantees (to π_ϵ).
- Simple to use (hopefully!)

Limitations of ABC-MCMC:

- Cheap-to-simulate models: MCMC requires many simulations.
- Moderate dimensional parameter (and summary).

Other likelihood-free methods may be more useful in more challenging settings!

Adaptive ABC-MCMC (V & Franks, Biometrika, to appear) i

Need to specify:

- Initial value θ_0
- The (log-)prior $\mathrm{pr}(\theta)$
- The observed summary statistic $s(y^*)$
- Simulator of summaries $s(Y_\theta)$
- Number of MCMC iterations

Adaptive ABC-MCMC (V & Franks, Biometrika, to appear) i

Need to specify:

- Initial value θ_0 `-ones(3)`
- The (log-)prior $\mathrm{pr}(\theta)$ function `prior(th)`
- The observed summary statistic $s(y^*)$ `-s_obs`
- Simulator of summaries $s(Y_\theta)$ function `sim!(s, th, rng)`
- Number of MCMC iterations 30000

```
julia> using AdaptiveToleranceABC_MCMC
julia> out = abc_mcmc(-ones(3), prior, s_obs, sim!, 30000)
```

`https://github.com/mvihola/AdaptiveToleranceABC_MCMC.jl`

Adaptive ABC-MCMC (V & Franks, Biometrika, to appear) ii

Adaptive state $\xi_k = (\mu_k, \Gamma_k, \log \delta_k)$:

- Draw $\Theta'_k \sim q_{\Gamma_{k-1}}(\cdot \mid \Theta_{k-1})$ and $Y'_k \sim g(\cdot \mid \Theta'_k)$.
- Calculate $L'_k := \mathbb{I}\{\|s(Y'_k) - s(y^*)\| \leq \delta_{k-1}\}$.
- With probability

$$\alpha_k := \min\left\{1, \frac{\mathrm{pr}(\Theta'_k) L'_k q(\Theta_{k-1} \mid \Theta'_k)}{\mathrm{pr}(\Theta_{k-1}) L_{k-1} q(\Theta'_k \mid \Theta_{k-1})}\right\},$$

  accept and set $(\Theta_k, L_k) \leftarrow (\Theta'_k, L'_k)$;
  otherwise reject and set $(\Theta_k, L_k) \leftarrow (\Theta_{k-1}, L_{k-1})$.

- Adapt $\xi_k = \xi_{k-1} + \eta_k H(\xi_{k-1}; \alpha_k, \Theta_k)$.

Adaptive ABC-MCMC (V & Franks, Biometrika, to appear) iii

- $q_\Gamma(y \mid x) = N\left(y; x, \frac{2.38^2}{n_{\text{par}}}\Gamma\right)$ — random-walk proposal with covariance $\propto \Gamma$.
- Adaptation:

$$\begin{aligned}
\mu_k &= (1-\eta_k)\mu_{k-1} + \eta_k \Theta_k \\
\Gamma_k &= (1-\eta_k)\Gamma_{k-1} + \eta_k(\Theta_k - \mu_{k-1})(\Theta_k - \mu_{k-1})^T \\
\log \delta_k &= \log \delta_{k-1} + \eta_k(\alpha^* - \alpha_k)
\end{aligned}$$

where $\eta_k = k^{-2/3}$ are step sizes.

Adaptive ABC-MCMC (V & Franks, Biometrika, to appear) iii

- $q_\Gamma(y \mid x) = N\left(y; x, \frac{2.38^2}{n_{\text{par}}}\Gamma\right)$ — random-walk proposal with covariance $\propto \Gamma$.
- Adaptation:

$$\begin{aligned}
\mu_k &= (1-\eta_k)\mu_{k-1} + \eta_k \Theta_k \\
\Gamma_k &= (1-\eta_k)\Gamma_{k-1} + \eta_k(\Theta_k - \mu_{k-1})(\Theta_k - \mu_{k-1})^T \\
\log \delta_k &= \log \delta_{k-1} + \eta_k(\alpha^* - \alpha_k)
\end{aligned}$$

where $\eta_k = k^{-2/3}$ are step sizes. That is,

- Adaptive Metropolis (Haario, Saksman, Tamminen 2001; Andrieu & Moulines 2006) mean-covariance update.

Adaptive ABC-MCMC (V & Franks, Biometrika, to appear) iii

- $q_\Gamma(y \mid x) = N\left(y; x, \frac{2.38^2}{n_{\text{par}}}\Gamma\right)$ — random-walk proposal with covariance $\propto \Gamma$.
- Adaptation:

$$\begin{aligned}
\mu_k &= (1-\eta_k)\mu_{k-1} + \eta_k \Theta_k \\
\Gamma_k &= (1-\eta_k)\Gamma_{k-1} + \eta_k(\Theta_k - \mu_{k-1})(\Theta_k - \mu_{k-1})^T \\
\log \delta_k &= \log \delta_{k-1} + \eta_k(\alpha^* - \alpha_k)
\end{aligned}$$

where $\eta_k = k^{-2/3}$ are step sizes. That is,

- Adaptive Metropolis (Haario, Saksman, Tamminen 2001; Andrieu & Moulines 2006) mean-covariance update.
- Acceptance-rate adaptation in the spirit of Andrieu & Robert (2001).

Acceptance rate adaptation: rationale

Must have a reasonable acceptance rate, otherwise ABC-MCMC will not mix.

- Optimise $\delta_k \to \delta_*$ such that the ABC-MCMC has a desired acceptance rate α^*.
- Observation: $\delta \mapsto \alpha_\delta := \mathbb{E}_{\pi_\delta}[\alpha_k]$ (mean accept rate w/ δ) is increasing.
  - If too small accept rate, increase δ_k (and vice versa).

$\rightsquigarrow$ Convergent adaptation (for fixed Γ). But how to choose α^*?

Acceptance rate adaptation: rationale

Must have a reasonable acceptance rate, otherwise ABC-MCMC will not mix.

- Optimise $\delta_k \to \delta_*$ such that the ABC-MCMC has a desired acceptance rate α^*.
- Observation: $\delta \mapsto \alpha_\delta := \mathbb{E}_{\pi_\delta}[\alpha_k]$ (mean accept rate w/ δ) is increasing.
  - If too small accept rate, increase δ_k (and vice versa).

$\rightsquigarrow$ Convergent adaptation (for fixed Γ). But how to choose α^*?

If $\{\pi_\delta\}_{\delta>0}$ were Gaussian, then AM adaptation for fixed δ would lead to mean accept rate $\leq \alpha^{\mathsf{RWM}}_{\mathsf{opt}}$ where $\alpha^{\mathsf{RWM}}_{\mathsf{opt}} \in [0.234, 0.44]$ (depending on $n_{\mathrm{par}} \dots$)

- $\delta \to \infty \implies$ random-walk Metropolis with accept rate $\approx \alpha^{\mathsf{RWM}}_{\mathsf{opt}}$.
- For any δ the 'marginal' variant of ABC-MCMC has accept rate $\approx \alpha^{\mathsf{RWM}}_{\mathsf{opt}}$.
  The accept rate of pseudo-marginal (ABC-MCMC) is less.
- When $\delta \to 0$, accept rate $\to 0$.

This suggests $\alpha^* \in (0, 0.234) \subset (0, \alpha^{\mathsf{RWM}}_{\mathsf{opt}})$.

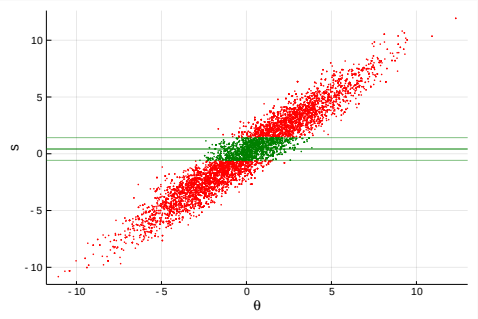
Acceptance rate adaptation: design choices

- We use $\alpha^* = 0.10$ as 'balanced' acceptance rate.
- Initialise $\delta_0 = \|s(Y_{\theta_0}) - s(y^*)\|$
  $\implies \delta_0 \gg \delta_*$ — this ensures acceptances during the initial phase;
  $\sim$ the explicit annealing scheme of Ratmann et al. (2007)?
- Changing tolerance changes target distribution
  $\rightsquigarrow$ stop tolerance adaptation after burn-in.
- Continue covariance adaptation also after burn-in.

This leads to adaptive ABC-MCMC with automatically chosen tolerance $\delta \approx \delta_* \ldots$

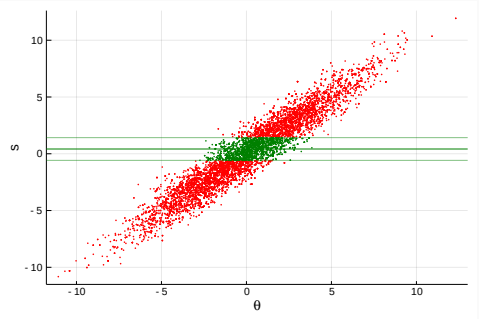
Post-refining tolerances $\delta \to \epsilon$

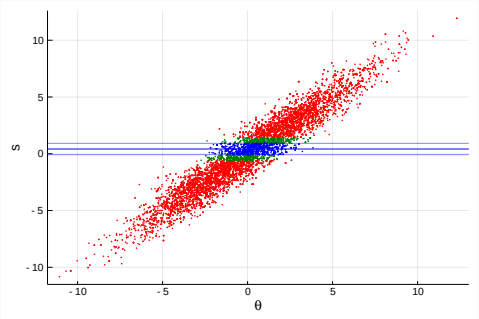
- Rejection sampling: Tolerance $\delta = 1 \implies$ samples



Post-refining tolerances $\delta \to \epsilon$

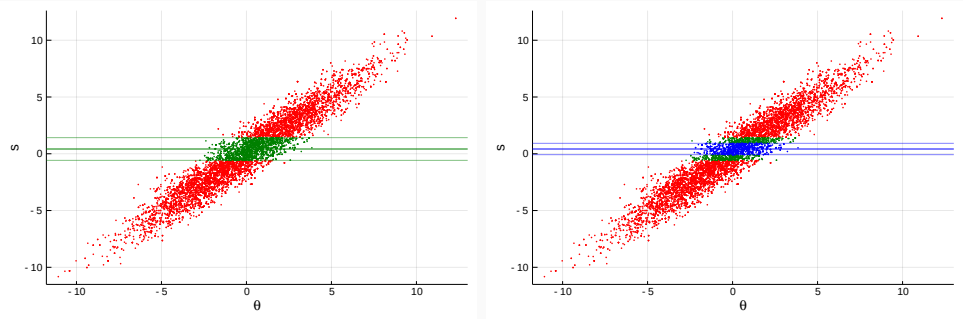
- Rejection sampling: Tolerance $\delta = 1 \implies$ samples $\implies$ refined for $\epsilon = 0.5 < \delta$





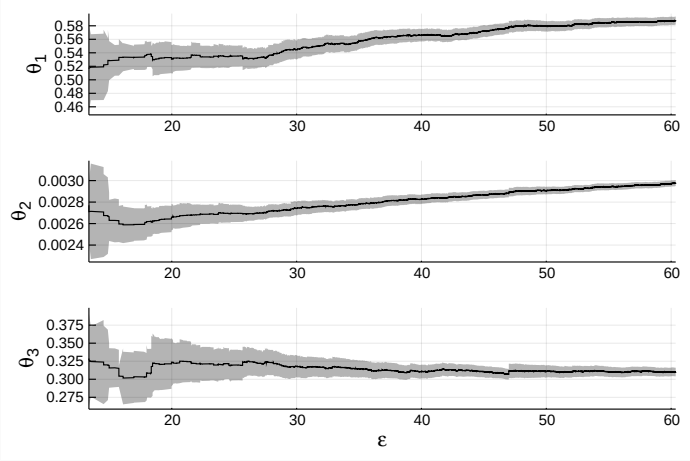
Post-refining tolerances $\delta \to \epsilon$

- Rejection sampling: Tolerance $\delta = 1 \implies$ samples $\implies$ refined for $\epsilon = 0.5 < \delta$



- The same can be done with ABC-MCMC (Wegmann et al. *Genetics*, 2009)... but how does it affect posterior mean uncertainty?

Lotka-Volterra example: Posterior mean estimates & CIs for a range of $\varepsilon \leq \delta$




```
julia> est = abc_postprocess(out, x -> exp.(x))
julia> plot_abc(est; labels=["\\theta_$(i)" for i in 1:3])
```

Posterior mean estimates & CIs for a range of $\varepsilon \leq \delta$

MCMC output $(\Theta_k, D_k)_{k=1,\ldots,n}$ with $D_k = \|s(Y_k) - s(y^*)\|$;
test function $f : \mathbb{R}^{n_{\text{par}}} \to \mathbb{R}$.

- Calculate self-normalised IS estimator:

$$U_k^{(\delta,\epsilon)} := \frac{\mathbb{I}\{D_k \leq \epsilon\}}{\mathbb{I}\{D_k \leq \delta\}}, \qquad W_k^{(\delta,\epsilon)} := \frac{U_k^{(\delta,\epsilon)}}{\sum_{j=1}^n U_j^{(\delta,\epsilon)}},$$

$$E_{\delta,\epsilon}(f) := \sum_{k=1}^n W_k^{(\delta,\epsilon)} f(\Theta_k), \qquad S_{\delta,\epsilon}(f) := \sum_{k=1}^n (W_k^{(\delta,\epsilon)})^2 \big\{f(\Theta_k) - E_{\delta,\epsilon}(f)\big\}^2.$$

- Calculate $\hat{\tau}_\delta(f) = \mathrm{IACT}\big(f(X_1), \ldots, f(X_n)\big)$.
- Report (say) 95% approximate CI

$$\Big[E_{\delta,\epsilon}(f) \;\pm\; 1.96 S_{\delta,\epsilon}(f)\hat{\tau}_\delta(f)\Big] \qquad \text{for any} \qquad \epsilon \in [\min_k(D_k), \delta]$$

(In the case of simple cut-off, the CI may be calculated for all ϵ in $O(n \log n)$ time.)

Posterior mean estimates & CIs: loose theoretical justification

- Under general conditions, for $\epsilon \leq \delta$

$$\overbrace{nS_{\delta,\epsilon}(f)}^{\to v^2_{\delta,\epsilon,\mathsf{IS}}} \underbrace{\hat{\tau}_\delta(f)}_{\to \tau_\delta(f)} \xrightarrow{n\to\infty} \breve{\sigma}^2(f) \gtrapprox \sigma^2_{\delta,\epsilon}(f) = v^2_{\delta,\epsilon,\mathsf{IS}} \tau_{\delta,\epsilon}(f),$$

  where $\tau_{\delta,\epsilon}(f)$ is integrated autocorrelation of $W_k^{(\delta,\epsilon)} f(\Theta_k)$.
- The approximate upper bound is due to

$$\tau_\delta(f) \gtrapprox \tau_{\delta,\epsilon}(f),$$

  by continuity for $\delta \approx \epsilon$, and loosely analytically justified for $\epsilon \ll \delta$.
- Direct estimation of the asymptotic variance of $W_k^{(\delta,\epsilon)} f(\Theta_k)$ is also possible, but
  - likely unstable for smaller ϵ (lots of zeros...), and
  - computationally more demanding...

About some generalisations

Above, we had "$\|s(Y) - s(y^*)\|$" and "$\mathbb{I}\{ \cdot \leq \epsilon\}$", but in general we might have:

- Any 'dissimilarity' function "$D(Y, y^*)$"
- Any 'kernel' "$k_\epsilon(\cdot)$" with values in $[0, 1]$
- Or we might merge both and define directly $K_\epsilon(Y, y^*) \in [0, 1] \ldots$

The adaptive tolerance ABC-MCMC and the post-processing generalise directly.[1]

It is also possible to accomodate to ABC-MCMC with averaged pseudo-data:

$$L'_k = \frac{1}{m} \sum_{i=1}^{m} \mathbb{I}\{\|s(Y_k^{(i)}) - s(y^*)\| \leq \epsilon\}, \qquad Y_k^{(1)}, \ldots, Y_k^{(m)} \overset{\text{i.i.d.}}{\sim} g(\cdot \mid \Theta'_k)$$

Theoretical findings of Bornn et al. (2017) suggest that $m = 1$ might be enough...

[1] The loose theoretical justification of CI is lost, but empirical results remain promising.

Regression adjustment

- *Regression correction/adjustment* is often useful with ABC.
- Instead of looking at $(\Theta_k) \sim \pi_\delta$ only, consider (Θ_k, S_k) where $S_k = s(Y_k)$ are summaries of accepted pseudo-data.
- Fit a regression model such as:

$$\theta_k = \beta_0 + \beta^T (S_k - s(y^*)) + \eta_k, \qquad \text{where } \eta_k \text{ are residuals}$$

  $\implies$ least squares estimates $\hat{\beta}_0$, $\hat{\beta}$

  $\implies$ empirical residuals $\hat{\eta}_k = \theta_k - \hat{\beta}_0 - \hat{\beta}^T (S_k - s(y^*))$

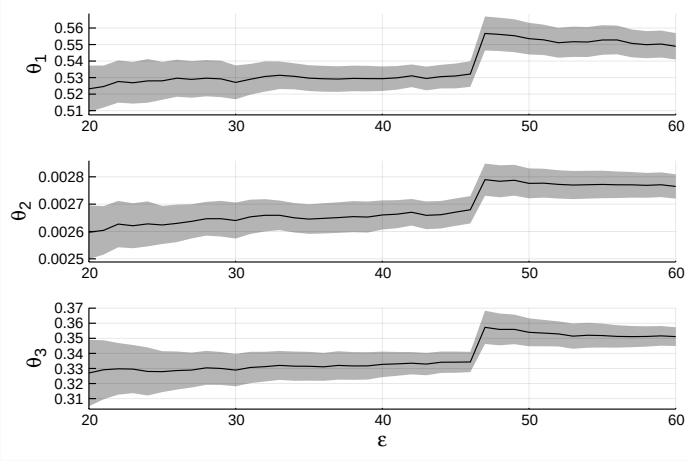
- The regression corrected samples are

$$\hat{\theta}_k := \hat{\beta}_0 + \hat{\eta}_k = \theta_k - \hat{\beta}^T (S_k - s(y^*));$$

  (if model and estimates correct, $\hat{\theta}_k$ follow $\theta \mid S = s(y^*) \ldots$)

- Post-correction $\delta \to \epsilon \implies$ weighted regression $\implies$ correction.
  (Experimental CI analogous to plain ABC...)

Regression adjustment: Lotka-Volterra example




```
julia> est_r = abc_postprocess(out, x->exp.(x), 20:out.cutoff.tol; regress=true)
julia> plot_abc(est_r; labels=["\\theta_$(i)" for i in 1:3])
```

Conclusions

- Easy-to-use ABC-MCMC
  - The only (mandatory) user-defined parameter is n — the number of MCMC steps.
  - Julia package — give it a try!
  - (Julia because it is good for implementing a *fast model simulator*...)
- Implicit assumptions:
  - Large n.
  - Low-dimensional parameter & summary.
  - Roughly unimodal (pseudo-)posterior — because of random-walk MCMC.
- No pre-defined 'target' tolerance ϵ.
- Instead, automatically chosen 'simulation tolerance' δ and possibility to inspect a range of $\epsilon \leq \delta$.
  - Can reveal bias due to still-too-large δ...

References

- **Marjoram, P., Molitor, J., Plagnol, V., & Tavaré, S. (2003)**
  Markov chain Monte Carlo without likelihoods.
  Proceedings of the National Academy of Sciences, 100(26), 15324–15328.
- **Andrieu, C., & Roberts, G. O. (2009)**
  The pseudo-marginal approach for efficient Monte Carlo computations.
  The Annals of Statistics, 37(2), 697–725.
- **Wegmann, D., Leuenberger, C., & Excoffier, L. (2009)**
  Efficient approximate Bayesian computation coupled with Markov chain Monte Carlo without likelihood.
  Genetics, 182(4), 1207–1218.
- **Ratmann, O., Jørgensen, O., Hinkley, T., Stumpf, M., Richardson, S., & Wiuf, C. (2007).**
  Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *H. pylori* and *P. falciparum*.
  PLoS Comput Biol, 3(11), e230.
- **Bornn, L., Pillai, N. S., Smith, A., & Woodard, D. (2017)**
  The use of a single pseudo-sample in approximate Bayesian computation.
  Statistics and Computing, 27(3), 583–590.
- **Haario, H., Saksman, E. & Tamminen, J. (2001)**
  An adaptive Metropolis algorithm.
  Bernoulli, 7(2), 223-242.
- **Andrieu, C., & Moulines, É. (2006).**
  On the ergodicity properties of some adaptive MCMC algorithms.
  The Annals of Applied Probability, 16(3), 1462-1505.
- **Andrieu, C., & Robert, C. P. (2001)**
  Controlled MCMC for optimal sampling.
  Tech. Rep. Ceremade 0125, Université Paris Dauphine.
- **Vihola, M., & Franks, J. (2020)**
  On the use of approximate Bayesian computation Markov chain Monte Carlo with inflated tolerance and post-correction.
  Biometrika, 107(2), 381-395.

Thank you!

Julia package: `https://github.com/mvihola/AdaptiveToleranceABC_MCMC.jl`