Statistical modeling of summary values leads to accurate Approximate Bayesian Computations

Oliver Ratmann (Imperial College London, UK)

Anton Camacho (London School of Hygiene & Tropical Medicine, UK)
Adam Meijer (National Institute of the Environment & Public Health, NL)
Gé Donker (Netherlands Institute for Health Services Research, NL)
Standard ABC

(Beaumont 2002)

standard ABC rejection sampler (ABC-r)

1: for $i = 1 \ldots N$ do
2: Propose $\theta' \sim \pi$ and simulate $y' \sim \ell(\cdot | \theta')$.
3: Compute summary statistics $S_k(y')$, $k = 1, \ldots, K_S$.
4: Compute summary errors $z'_k = d_k(S_k(y'), S_k(x))$ for all $k$.
5: Accept $(\theta', z')$ if for all $k$ $c^-_k \leq z'_k \leq c^+_k$. Go to line 2.
6: end for

ABC approximation to likelihood

$$\ell(x|\theta) \approx \ell_{abc}(x|\theta) = \int \prod_{k=1}^{K_S} 1 \left\{ c^-_k \leq d_k(S_k(y), S_k(x)) \leq c^+_k \right\} \ell(dy|\theta)$$

is exact if
1) summary statistics are sufficient
2) upper and lower tolerances coincide
Standard ABC
(Beaumont 2002)

standard ABC rejection sampler (ABC-r)
1: for $i = 1 \ldots N$ do
2: Propose $\theta' \sim \pi$ and simulate $y' \sim \ell(\cdot | \theta')$.
3: Compute summary statistics $S_k(y')$, $k = 1, \ldots, K_S$.
4: Compute summary errors $z'_k = d_k(S_k(y'), S_k(x))$ for all $k$.
5: Accept $(\theta', z')$ if for all $k$, $c_k^- \leq z'_k \leq c_k^+$. Go to line 2.
6: end for

ABC approximation to likelihood

$$\ell(x|\theta) \approx \ell_{\text{abc}}(x|\theta) = \int \prod_{k=1}^{K_S} 1 \left\{ c_k^- \leq d_k(S_k(y), S_k(x)) \leq c_k^+ \right\} \ell(dy|\theta)$$

is exact if
1) summary statistics are sufficient
2) upper and lower tolerances coincide

in practice not feasible, ‘asymptotic’ argument
Standard ABC is noisy
even with sufficient summary statistics (Fernhead & Prangle 2012)

Example

estimate $\sigma^2$ for $x \sim \mathcal{N}(0, \sigma^2)$

1. $\theta' \sim \mathcal{U}(0, 30)$, $y'_{1:1000} \sim \mathcal{N}(0, \theta')$
2. compute $S_y^2 = \frac{1}{n-1} \sum (y_i - \bar{y})^2$
3. compute $\rho = S_y^2 - S_x^2$
4. accept $(\theta', y')$ if $c^- \leq \rho \leq c^+$

Tuesday, 7 January 14
Can we construct ABC such that inference is accurate

- wrt point estimate, eg MAP
- wrt overall similarity in distribution, eg KL divergence
- and maintain computational feasibility

If yes, under which conditions?
How general are these?

(Ratmann, Camacho, Meijer, Donker; arXiv 2013)
ABC* step 1

To avoid asymptotics, interpret ABC accept/reject step as the outcome of a decision test

$$ R = \{ c^- \leq T(s^{1:n}(x), s^{1:m}(y)) \leq c^+ \} $$

T-test

objective: declare $\mu(\theta), \mu_x$ unequal
H$_0$: $\mu(\theta), \mu_x$ equal
H$_1$: $\mu(\theta), \mu_x$ unequal
rejection region: $(-\infty, c^-] \cup [c^+, \infty)$

ABC

objective: declare $\mu(\theta), \mu_x$ equal
H$_0$: $\mu(\theta), \mu_x$ unequal
H$_1$: $\mu(\theta), \mu_x$ equal
rejection region: $[c^-, c^+]$
ABC* step 1

To avoid asymptotics, interpret ABC accept/reject step as the outcome of a decision test

\[ R = \{ c^- \leq T(s^{1:n}(x), s^{1:m}(y)) \leq c^+ \} \]

T-test

objective: declare \( \mu(\theta), \mu_x \) unequal
H_0: \( \mu(\theta), \mu_x \) equal
H_1: \( \mu(\theta), \mu_x \) unequal
rejection region: \( (\infty, c^-] \cup [c^+, \infty) \)

ABC

objective: declare \( \mu(\theta), \mu_x \) equal
H_0: \( \mu(\theta), \mu_x \) unequal
H_1: \( \mu(\theta), \mu_x \) equal
rejection region: \( [c^-, c^+] \)

\( c^-, c^+ \) are fully determined
\( \text{sth } P(R|H_0) \leq \alpha \)
ABC* step 1

To avoid asymptotics, interpret ABC accept/reject step as the outcome of a decision test

\[ R = \{ c^- \leq T(s^{1:n}(x), s^{1:m}(y)) \leq c^+ \} \]

**T-test**

objective: declare \( \mu(\theta), \mu_x \) unequal

\[ H_0: \mu(\theta), \mu_x \text{ equal} \]

\[ H_1: \mu(\theta), \mu_x \text{ unequal} \]

rejection region: \( (-\infty, c^-] \cup [c^+, \infty) \)

**ABC**

objective: declare \( \mu(\theta), \mu_x \) equal

\[ H_0: \mu(\theta), \mu_x \text{ unequal} \]

\[ H_1: \mu(\theta), \mu_x \text{ equal} \]

rejection region: \([c^-, c^+]\)

\( c^-, c^+ \) are fully determined

s.t. \( P(R | H_0) \leq \alpha \)

Let \( \rho = \mu - \mu_x \) then ABC approximation to likelihood is the power function

\( \rho \to P(R | \rho) \) of the test
ABC* step 1

To avoid asymptotics, interpret ABC accept/reject step as the outcome of a decision test.

\[ R = \left\{ c^- \leq T(s^{1:n}(x), s^{1:m}(y)) \leq c^+ \right\} \]

**T-test**

- \( \mu(\theta), \mu_x \) unequal
- \( H_0: \mu(\theta), \mu_x \) equal
- \( H_1: \mu(\theta), \mu_x \) unequal
- Rejection region: \(( -\infty, c^- ] \cup [ c^+, \infty )\)

**ABC**

- \( \mu(\theta), \mu_x \) equal
- \( H_0: \mu(\theta), \mu_x \) unequal
- \( H_1: \mu(\theta), \mu_x \) equal
- Rejection region: \([ c^-, c^+ ]\)

\( c^-, c^+ \) are fully determined

\[ \text{s.t.} \quad P( R \mid H_0 ) \leq \alpha \]

holds for specific test:

- Two sided, one sample equivalence hypothesis test

Let \( \rho = \mu - \mu_x \)

then ABC approximation to likelihood is the power function

\[ \rho \rightarrow P( R \mid \rho ) \]
Example: test variance

suppose
\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

then
Example: test variance

suppose
\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

then
\[ \rho = \sigma^2 / \hat{\sigma}_x^2 \]
\[ \rho^* = 1 \]
\[ H_0: \rho \notin [\tau^-, \tau^+] \]
\[ H_1: \rho \in [\tau^-, \tau^+] \]

\[ T = S^2(y^{1:m}) / S^2(x^{1:n}) = \rho \frac{1}{n-1} \sum_{i=1}^m \frac{(y_i - \bar{y})^2}{\sigma^2} \]
\[ \sim \frac{\rho}{n-1} \chi^2_{m-1} \]
Example: test variance

Suppose
\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

Then
\[ \rho = \sigma^2 / \hat{\sigma}_x^2 \]
\[ \rho^* = 1 \]
\[ H_0: \rho \notin [\tau^-, \tau^+] \]
\[ H_1: \rho \in [\tau^-, \tau^+] \]

\[ T = S^2(y^{1:m}) / S^2(x^{1:n}) = \rho \frac{1}{n-1} \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \]
\[ \sim \frac{\rho}{n-1} \chi^2_{m-1} \]

For simplicity, summary values equal data point of equality.
Example: test variance

Suppose

\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

Then

\[ \rho = \frac{\sigma^2}{\hat{\sigma}_x^2} \]
\[ \rho^* = 1 \]
\[ H_0: \rho \notin [\tau^-, \tau^+] \]
\[ H_1: \rho \in [\tau^-, \tau^+] \]

\[ T = \frac{S^2(y^{1:m})}{S^2(x^{1:n})} = \rho \frac{1}{n-1} \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \]

\[ \sim \frac{\rho}{n-1} \chi^2_{m-1} \]
Example: test variance

suppose
\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

then
\[ \rho = \sigma^2 / \hat{\sigma}_x^2 \]
\[ \rho^* = 1 \]
\[ H_0: \rho \notin [\tau^-, \tau^+] \]
\[ H_1: \rho \in [\tau^-, \tau^+] \]

\[ T = S^2(y^{1:m}) / S^2(x^{1:n}) = \rho \frac{1}{n-1} \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \]
\[ \sim \frac{\rho}{n-1} \chi^2_{m-1} \]

know distribution of T, can work out \( c^-, c^+ \)
Example: test variance

suppose
\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

then
\[
\rho = \sigma^2 / \hat{\sigma}_x^2 \\
\rho^* = 1 \\
H_0: \rho \notin [\tau^-, \tau^+] \\
H_1: \rho \in [\tau^-, \tau^+] \\
\]

\[ T = S^2(y^{1:m}) / S^2(x^{1:n}) = \rho \frac{1}{n-1} \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \]

\[ \sim \frac{\rho}{n-1} \chi^2_{m-1} \]

know distribution of T, can work out \( c^-, c^+ \)

\[
\text{density of } T_k^- \text{ given } \rho_k = \tau_k^- \\
\text{density of } T_k^+ \text{ given } \rho_k = \tau_k^+ \\
\]

\([c_k^-, c_k^+]\) critical region of test statistic

\([\tau_k^-, \tau_k^+]\) tolerance region of \( \rho_k \)
Example: test variance

suppose
\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

then
\[
\rho = \frac{\sigma^2}{\hat{\sigma}_x^2} \\
\rho^* = 1 \\
H_0: \rho \notin [\tau^-, \tau^+] \\
H_1: \rho \in [\tau^-, \tau^+] \\
\]

\[
T = \frac{S^2(y^{1:m})}{S^2(x^{1:n})} = \rho \frac{1}{n-1} \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \\
\sim \frac{\rho}{n-1} \chi^2_{m-1} \\
\]

know distribution of \( T \), can work out \( c^- \), \( c^+ \) and power function

\[
[\tau^-, \tau^+] = [1 - 0.65, 1 + 0.65] 
\]
Example: test variance

suppose
\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

then
\[ \rho = \sigma^2 / \hat{\sigma}_x^2 \]
\[ \rho^* = 1 \]
\[ H_0: \rho \notin [\tau^-, \tau^+] \]
\[ H_1: \rho \in [\tau^-, \tau^+] \]

\[ T = S^2(y^{1:m}) / S^2(x^{1:n}) = \rho \frac{1}{n-1} \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \]
\[ \sim \frac{\rho}{n-1} \chi_{m-1}^2 \]

know distribution of T, can work out \( C^-, C^+ \) and power function and calibrate

\[ [\tau^-, \tau^+] = [1 - 0.65, 1 + 0.65] \]
Example: test variance

\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

then
\[ \rho = \frac{\sigma^2}{\hat{\sigma}_x^2} \]
\[ \rho^* = 1 \]
\[ H_0: \rho \notin [\tau^-, \tau^+] \]
\[ H_1: \rho \in [\tau^-, \tau^+] \]

\[ T = \frac{S^2(y^{1:m})}{S^2(x^{1:n})} = \rho \frac{1}{n-1} \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \]
\[ \sim \frac{\rho}{n-1} \chi^2_{m-1} \]

know distribution of T, can work out \( c^-, c^+ \) and power function and calibrate

\[ [\tau^-, \tau^+] = [1 - 0.65, 1 + 0.65] \]
Example: test variance

Suppose

\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

then

\[ \rho = \frac{\sigma^2}{\hat{\sigma}_x^2} \]
\[ \rho^* = 1 \]

\[ H_0: \rho \notin [\tau^-, \tau^+] \]
\[ H_1: \rho \in [\tau^-, \tau^+] \]

\[ T = S^2(y^{1:m})/S^2(x^{1:n}) = \rho \frac{1}{n-1} \sum_{i=1}^m \frac{(y_i - \bar{y})^2}{\sigma^2} \]
\[ \sim \frac{\rho}{n-1} \chi^2_{m-1} \]

**Calibrated tolerances**

\[
\begin{align*}
\tau^- &= 0.477 \\
\tau^+ &= 2.2 \\
\pi(\sigma^2|x) &= \text{argmax}_{\sigma^2} \pi(\sigma^2|x) \\
\end{align*}
\]

\[
\begin{align*}
\text{naive tolerances} \\
\tau^- &= 0.35 \\
\tau^+ &= 1.65 \\
\end{align*}
\]
Example: test variance

suppose
\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

then
\[ \rho = \sigma^2 / \hat{\sigma}_x^2 \]
\[ \rho^* = 1 \]
\[ H_0: \rho \notin [\tau^-, \tau^+] \]
\[ H_1: \rho \in [\tau^-, \tau^+] \]

\[ T = S^2(y^{1:m}) / S^2(x^{1:n}) = \rho \frac{1}{n-1} \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \]
\[ \sim \frac{\rho}{n-1} \chi^2_{m-1} \]

exact posterior

n-ABC estimate of \( \pi(\sigma^2|x) \)

calibrated tolerances
\( \tau^- = 0.572 \quad \tau^+ = 1.808 \)
\( m = 97 \)

calibrated tolerances
\( \tau^- = 0.726 \quad \tau^+ = 1.392 \)
\( m = 300 \)

tighten

\[ \arg\max_{\sigma^2} \pi(\sigma^2|x) \]
Example: test variance

Suppose

\[ x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2) \]

Then

\[ \rho = \frac{\sigma^2}{\hat{\sigma}_x^2} \]

\[ \rho^* = 1 \]

\[ H_0: \rho \notin [\tau^-, \tau^+] \]

\[ H_1: \rho \in [\tau^-, \tau^+] \]

\[ T = \frac{S^2(y^{1:m})/S^2(x^{1:n})}{\rho} = \frac{1}{n-1} \sum_{i=1}^{m} \frac{(y_i - \bar{y})^2}{\sigma^2} \]

\[ \sim \frac{\rho}{n-1} \chi^2_{m-1} \]

Calibrated tolerances

Calibrated m
Example: test variance

\[
x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2)
\]

suppose

\[
x^{1:n} \sim \mathcal{N}(0, \sigma_x^2) \quad y^{1:m} \sim \mathcal{N}(0, \sigma^2)
\]

then

\[
\rho = \sigma^2 / \hat{\sigma}_x^2 \\
\rho^* = 1 \\
H_0: \rho \notin [\tau^-, \tau^+] \\
H_1: \rho \in [\tau^-, \tau^+]
\]

\[
T = S^2(y^{1:m}) / \hat{\sigma}_x^2
\]

\[
\sim \frac{\rho}{n-1} \chi^2_{m-1}
\]

Conclusions-1

using statistical decision theory, the ABC accept/reject step can be set up such that

- the ABC* MAP equals the MAP of the exact posterior
- the KL divergence of the ABC* posterior to the exact posterior is minimal

calibrated tolerances calibrated \( m \)
1. repeat data points on summary level “summary values”
   ➢ can model their distribution, eg
   \[ s^{1:n}(x) \sim \mathcal{N}(\mu_x, \sigma_x^2) \quad \text{and} \quad s^{1:m}(y) \sim \mathcal{N}(\mu(\theta), \sigma^2(\theta)) \]

2. testing on auxiliary space
   ➢ given \( s^{1:n}(x), s^{1:m}(y) \) is the underlying \( \rho = \mu(\theta) - \mu_x \) small ?

3. indirect inference
   ➢ link auxiliary space back to original space
ABC* step 2
Statistical decision testing on summary level

1. repeat data points on summary level "summary values"
   ➣ can model their distribution, e.g.
     \[ s^{1:n}(x) \sim \mathcal{N}(\mu_x, \sigma_x^2) \]

2. testing on auxiliary space
   ➣ given \( s^{1:n}(x), s^{1:m}(y) \) is small?
     \[ s^{1:n}(x) \sim \mathcal{N}(\mu_x, \sigma_x^2) \]

3. indirect inference
   ➣ link auxiliary space back to original space

Assumptions
summary values can be found sth
A1 they are sufficient for \( \theta \)
A2 their distribution can be modeled in an elementary way so that test statistics are available and can be calibrated

Further conditions to transport the accurate ABC* density to the original space
**ABC* step 2**

Statistical decision testing on summary level

1. repeat data points on summary level, can model their distribution:
   \[ s_{1:n}(x) \sim \mathcal{N}(\mu_x, \sigma_x^2) \]

**Assumptions**

- summary values can be found sth
- **A1** they are sufficient for \( \theta \)
- **A2** their distribution can be modeled in an elementary way so that test statistics are available and can be calibrated

- further conditions to transport the accurate ABC* density to the original space

**Auxiliary space**

- \( s_{1:n}(x) \), \( s_{1:m}(y) \) is reference

- can model the auxiliary space back to original space:

\[ s_{1:n}(x), s_{1:m}(y) \mapsto \mu(\theta) \]

\[ \sigma(\theta) \]

\[ \mathcal{N}(\mu(\theta), \sigma^2(\theta)) \]
Summary values

suitable data points on a summary level can be found

H3N2 regression estimate, Netherlands
H3N2 ILI under compartment model

Tuesday, 7 January 14
Summary values

suitable data points on a summary level can be found

data time series is biennial

A

B

C

Tuesday, 7 January 14
Summary values

suitable data points on a summary level can be found

data time series is biennial

odd and even time series values can be modeled as iid Gaussian
Modeling summary values

constructs an auxiliary probability space

\[
\begin{align*}
\text{obs} & \quad s_{1:n}(x) \sim \mathcal{N}(\mu_x, \sigma_x^2) \\
\text{sim} & \quad s_{1:n}(y) \sim \mathcal{N}(\mu(\theta), \sigma^2(\theta)) \\
\text{population error} & \quad \rho = \mu(\theta) - \mu_x
\end{align*}
\]

D orig parameters \quad \theta = (\theta_1, \ldots, \theta_D)
K error parameters \quad \rho = (\rho_1, \ldots, \rho_K)
Link function
\[
\mathcal{L} : \Theta \subset \mathbb{R}^D \rightarrow \Delta \subset \mathbb{R}^K
\]
\[
\begin{align*}
\theta & \rightarrow (\rho_1, \ldots, \rho_K) \\
\rho_k & = \delta_k(\nu_{x_k}, \nu_k(\theta))
\end{align*}
\]

Discussion wrt indirect inference (Gouriéroux 1993)

- difficulty in indirect inference: which aux space chosen
  here constructed empirically from distr of summary values
ABC* indirect inference

using assumptions A1, A2:

\[ \pi_{\text{true posterior}}(\theta | x) \propto \ell(x | \theta) \pi(\theta) \]

\[ \propto \ell(s_{k}^{1:n_k}(x), k = 1, \ldots, K | \theta) \pi(\theta) \]
ABC* indirect inference

using assumptions A1, A2:

$$\pi_{\text{true posterior}}(\theta|x) \propto \ell(x|\theta) \pi(\theta)$$

$$\propto \ell(s_{1:n_k}^k(x), k = 1, \ldots, K|\theta) \pi(\theta)$$

$$= \ell(s_{1:n_k}^k(x), k = 1, \ldots, K|\rho) \pi(\rho) |\partial \mathcal{L}(\theta)|$$
\textbf{ABC* indirect inference}

using assumptions A1, A2:

\[ p_{\text{true posterior}}(\theta|x) \propto \ell(x|\theta) \pi(\theta) \]
\[ \propto \ell(s_{k}^{1:n_k}(x), k = 1, \ldots, K|\theta) \pi(\theta) \]
\[ = \ell(s_{k}^{1:n_k}(x), k = 1, \ldots, K|\rho) \pi(\rho) |\partial L(\theta)| \]

\textbf{ABC* approximation on }\rho\text{ -space is}

\[ p_{\text{abc}}(\theta|x) \propto P_x(\text{ABC accept}|\rho) \pi(\rho) |\partial L(\theta)| \]
ABC* indirect inference

using assumptions A1, A2:

\[
\pi_{\text{true posterior}}(\theta|x) \propto \ell(x|\theta) \pi(\theta) \\
\propto \ell(s_{k}^{1:n_k}(x), k = 1, \ldots, K|\theta) \pi(\theta) \\
= \ell(s_{k}^{1:n_k}(x), k = 1, \ldots, K|\rho) \pi(\rho) |\partial \mathcal{L}(\theta)|
\]

ABC* approximation on \(\rho\)-space is

\[
\pi_{\text{abc}}(\theta|x) \propto P_x(\text{ABC accept}|\rho) \pi(\rho) |\partial \mathcal{L}(\theta)|
\]

match through calibration of ABC tolerances and m
ABC* indirect inference

using assumptions A1, A2:

$$\pi_{\text{true posterior}}(\theta|x) \propto \ell(x|\theta) \pi(\theta)$$

$$\propto \ell(s_{k}^{1:n_k}(x), k = 1, \ldots, K|\theta) \pi(\theta)$$

$$= \ell(s_{k}^{1:n_k}(x), k = 1, \ldots, K|\rho) \pi(\rho) |\partial \mathcal{L}(\theta)|$$

ABC* approximation on $\rho$-space is

$$\pi_{abc}(\theta|x) \propto P_x(\text{ABC accept}|\rho) \pi(\rho) |\partial \mathcal{L}(\theta)|$$

Regularity conditions on the link function

A3 the link function is bijective and continuously differentiable

match through calibration of ABC tolerances and $m$
Example: moving average

no sufficient statistics other than data, simple enough so that link function is analytically known

\[ x_t = u_t + au_{t-1}, \quad u_t \sim \mathcal{N}(0, \sigma^2) \]

\[ \theta = (a, \sigma^2) \]
Example: moving average

no sufficient statistics other than data, simple enough so that link function is analytically known

\[ x_t = u_t + au_{t-1}, \quad u_t \sim \mathcal{N}(0, \sigma^2) \]

\[ \theta = (a, \sigma^2) \]

\[ \nu_1 = (1 + a^2)\sigma^2 \]

\[ \nu_2 = a/(1 + a^2) \]
Example: moving average

no sufficient statistics other than data, simple enough so that link function is analytically known

\[ x_t = u_t + au_{t-1}, \quad u_t \sim \mathcal{N}(0, \sigma^2) \]
\[ \theta = (a, \sigma^2) \]

\[ \nu_1 = (1 + a^2)\sigma^2 \]
\[ \nu_2 = a/(1 + a^2) \]

\[ \rho_1 = (1 + a^2)\sigma^2 / \hat{\nu}_{x1} \]
\[ \rho_2 = \text{atanh}(a/(1 + a^2)) - \text{atanh}(\hat{\nu}_{x2}) \]
Example: moving average

no sufficient statistics other than data, simple enough so that link function is analytically known

\[ x_t = u_t + au_{t-1}, \quad u_t \sim \mathcal{N}(0, \sigma^2) \]

\[ \theta = (a, \sigma^2) \]

\[ \nu_1 = (1 + a^2)\sigma^2 \]

\[ \nu_2 = a/(1 + a^2) \]

\[ \rho_1 = (1 + a^2)\sigma^2 / \hat{v}_{x1} \]

\[ \rho_2 = \text{atanh}(a/(1 + a^2)) - \text{atanh}(\hat{v}_{x2}), \]

exact posterior

Testing only variance: link not bijective

Testing variance and autocorrelation with even values: summary values not sufficient
Example: moving average

no sufficient statistics other than data, simple enough so that link function is analytically known

\[ x_t = u_t + au_{t-1}, \quad u_t \sim \mathcal{N}(0, \sigma^2) \]

\[ \theta = (a, \sigma^2) \]

\[ \nu_1 = (1 + a^2)\sigma^2 \]

\[ \nu_2 = a/(1 + a^2) \]

\[ \rho_1 = (1 + a^2)\sigma^2/\hat{\nu}_{x1} \]

\[ \rho_2 = \text{atanh}(a/(1 + a^2)) - \text{atanh}(\hat{\nu}_{x2}), \]

exact posterior

Testing only variance: link not bijective

Testing variance and autocorrelation with even values: summary values not sufficient

Tuesday, 7 January 14
Example: moving average

no sufficient statistics other than data, simple enough so that link function is analytically known

\[ x_t = u_t + au_{t-1}, \quad u_t \sim \mathcal{N}(0, \sigma^2) \]

\[ \theta = (a, \sigma^2) \]

\[ \nu_1 = (1 + a^2)\sigma^2 \]

\[ \nu_2 = a/(1 + a^2) \]

\[ \rho_1 = (1 + a^2)\sigma^2 / \hat{\nu}_{x1} \]

\[ \rho_2 = \text{atanh}(a/(1 + a^2)) - \text{atanh}(\hat{\nu}_{x2}) \]

5 tests: link bijective and summary values sufficient
Example: flu time series data

stochastic transmission model, derived from ODEs

\[
\begin{align*}
\frac{dS^i}{dt} &= \mu^i(N^i - S^i) - \beta^i \frac{S^i}{N^i} (I_1^i + I_2^i + M^i) + \gamma R^i \\
\frac{dE^i}{dt} &= \beta^i \frac{S^i}{N^i} (I_1^i + I_2^i + M^i) - (\mu^i + \phi) E^i \\
\frac{dI_1^i}{dt} &= \phi E^i - (\mu^i + 2\nu) I_1^i \\
\frac{dI_2^i}{dt} &= 2\nu I_1^i - (\mu^i + 2\nu) I_2^i \\
\frac{dR^i}{dt} &= 2\nu I_2^i - (\mu^i + \gamma) R^i
\end{align*}
\]
Example: flu time series data

stochastic transmission model, derived from ODEs

\[
\begin{align*}
\frac{dS_i}{dt} &= \mu_i(N_i - S_i) - \beta_i \frac{S_i}{N_i}(I_1^i + I_2^i + M^i) + \gamma R^i \\
\frac{dE_i}{dt} &= \beta_i \frac{S_i}{N_i}(I_1^i + I_2^i + M^i) - (\mu^i + \phi)E_i \\
\frac{dI_1^i}{dt} &= \phi E_i - (\mu^i + 2\nu)I_1^i \\
\frac{dI_2^i}{dt} &= 2\nu I_1^i - (\mu^i + 2\nu)I_2^i \\
\frac{dR^i}{dt} &= 2\nu I_2^i - (\mu^i + \gamma)R^i
\end{align*}
\]

three parameters of interest:
reproductive number $R_0$, duration of immunity, reporting rate

6 sets of iid summary values, from 3 time series, subsetting odd and even values
Example: flu time series data

Test if link bijective from ABC* output
Example: flu time series data

Test if link bijective from ABC* output

previous standard MCMC ABC
Example: flu time series data

Test if link bijective from ABC* output

Previous standard MCMC ABC

MCMC ABC* with calibrated tolerances
Conclusions

using statistical decision theory in ABC,

• we can entirely avoid previous asymptotic arguments
• and construct accurate ABC algorithms by calibrating the decision tests appropriately

necessary to understand the distribution of the data on a summary level

identifying replicate structures and modeling them is key in ABC as in any other approaches for which the likelihood is tractable
Thank you

coworkers on this project
Anton Camacho (London School of Hygiene & Tropical Medicine, UK)
Adam Meijer (National Institute of the Environment & Public Health, NL)
Gé Donker (Netherlands Institute for Health Services Research, NL)

acknowledgements
Ioanna Manolopoulou (University College London)
Christian Robert (Paris Dauphine)