# K2-ABC: Approximate Bayesian Computation with Kernel Embeddings

Mijung Park[1]
Wittawat Jitkrittum[1]
Dino Sejdinovic
(University of Oxford)

Gatsby Tea Talk

10 Feb 2015

[1]MP and WJ contributed equally to this work.

# Approximate Bayesian Computation (ABC)

## Setting

- Evaluation of likelihood $p(y|\theta)$ is intractable.
- Given $\theta$, possible to sample $y \sim p(y|\theta)$

**Example**: a complicated dynamical system for blowfly population

$$N_{t+1} = PN_{t-\tau} \exp\left(-\frac{N_{t-\tau}}{N_0}\right) e_t + N_t \exp(-\delta \epsilon_t)$$

where $e_t \sim$ Gamma $\left(\frac{1}{\sigma_P^2}, \sigma_P^2\right)$ and $\epsilon_t \sim$ Gamma $\left(\frac{1}{\sigma_d^2}, \sigma_d^2\right)$.

- $\theta := \{P, N_0, \sigma_d, \sigma_p, \tau, \delta\}$
- Can simulate (sample). No explicit form of a likelihood.

# Approximate Bayesian Computation (ABC)

**Setting**

- Evaluation of likelihood $p(y|\theta)$ is intractable.
- Given $\theta$, possible to sample $y \sim p(y|\theta)$

**Example**: a complicated dynamical system for blowfly population

$$N_{t+1} = PN_{t-\tau} \exp\left(-\frac{N_{t-\tau}}{N_0}\right) e_t + N_t \exp(-\delta \epsilon_t)$$

where $e_t \sim$ Gamma $\left(\frac{1}{\sigma_P^2}, \sigma_P^2\right)$ and $\epsilon_t \sim$ Gamma $\left(\frac{1}{\sigma_d^2}, \sigma_d^2\right)$.

- $\theta := \{P, N_0, \sigma_d, \sigma_p, \tau, \delta\}$
- Can simulate (sample). No explicit form of a likelihood.

# Rejection ABC

- Observe a dataset $y^*$

$$\begin{aligned} p(\theta|y^*) &\propto \pi(\theta)p(y^*|\theta) \\ &= \pi(\theta)\int p(y|\theta)\, d\delta_{y^*}(y) \\ &\approx \pi(\theta)\int p(y|\theta)\kappa_\epsilon(y,y^*)\, dy \end{aligned}$$

  where $\kappa_\epsilon(y, y^*)$ defines an $\epsilon$-ball around $y^*$.
- Sample $\theta \sim \pi(\theta)$.
- Keep $\theta$ such that $y \sim p(y|\theta)$ and $y^*$ are close.

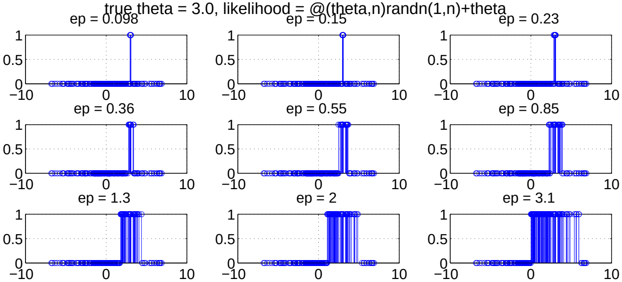
# Rejection ABC Algorithm

- **Input:** observed dataset $y^*$, distance $\rho$, threshold $\epsilon$
- **Output:** posterior sample $\{\theta_i\}_{i=1}^M$ from approximate posterior $\approx p(\theta|y^*) \propto \pi(\theta)p(y^*|\theta)$

```
repeat
    Sample θ' ~ π(·)
    Sample a dataset y' ~ p(·|θ')
    if ρ(y', y*) < ε then
        Keep θ'
    end if
until we have M samples
```

# Rejection ABC Example

$$
\begin{aligned}
p(y|\theta) &= \mathcal{N}(y;\theta,1) \\
\pi(\theta) &= \mathcal{N}(\theta,0,8) \\
\theta^* &= 3.0 \\
\rho(y,y^*) &= \|\hat{\mathbb{E}}[y] - \hat{\mathbb{E}}[y^*]\|_2
\end{aligned}
$$




- High $\epsilon \Rightarrow$ get $\theta$ sample from prior
- Low $\epsilon \Rightarrow$ sample closely follows true posterior. High rejection rate.

# Problem of ABC: Summary Statistic

- Often distance function depends on summary statistics

$$\rho(y, y^*) = \|s(y) - s(y^*)\|_2.$$

- Difficult to choose summary statistics $s$.
  - More statistics give high sufficiency.
  - But higher rejection rate.
- Contribution: Use a kernel-based distance function defined directly on $y$ and $y^*$.
- No need to design $s$.

# Maximum Mean Discrepancy (MMD)

- **Mean embedding**: $\mu_{F_x} := \mathbb{E}_{F_x(x)}\phi(x)$ where $\phi$ is the feature map associated with a kernel $k$.

$$
\begin{aligned}
& \mathrm{MMD}^2(F_x, F_y) \\
= \; & \|\mu_{F_x} - \mu_{F_y}\|^2_{\mathcal{H}} \\
= \; & \mathbb{E}_X \mathbb{E}_{X'} k(X, X') - 2\mathbb{E}_X \mathbb{E}_Y k(X, Y) + \mathbb{E}_Y \mathbb{E}_{Y'} k(Y, Y').
\end{aligned}
$$

  where we used $k(X, Y) = \langle \phi(X), \phi(Y) \rangle_{\mathcal{H}}$.
- $\mathrm{MMD}(F_y, F_{y^*}) = \rho(y, y^*)$ previously
- Simple empirical estimator.
- If $k$ is characteristic (e.g., Gaussian kernel), MMD defines a distance on space of distributions.
- $\phi$ plays the same role as $s$.
- $\phi$ never needs to be formed explicitly.

# K2-ABC (proposed method)

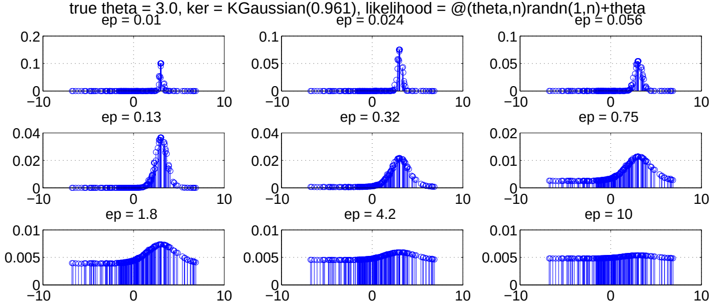
- **Input:** observed data $y^*$, threshold $\epsilon$
- **Output:** Empirical posterior $\sum_{i=1}^{M} w_i \delta_{\theta_i}$

1: **for** $i = 1, \ldots, M$ **do**
2: Sample $\theta_i \sim \pi$
3: Sample pseudo dataset $y_i \sim p(\cdot|\theta_i)$
4: $\widetilde{w}_i = \kappa_\epsilon(y_i, y^*) = \exp\left(-\frac{\widehat{\text{MMD}}^2(F_{y_i}, F_{y^*})}{\epsilon}\right)$
5: **end for**
6: $w_i = \widetilde{w}_i / \sum_{j=1}^{M} \widetilde{w}_j$ for $i = 1, \ldots, M$

- No rejection.
- $0 \leq w_i \leq 1$ instead of $w_i \in \{0, 1\}$ as in rejection ABC.
- "K2" because we use two kernels. $k$ (in MMD) and $\kappa_\epsilon$.
- On Arxiv: `http://arxiv.org/abs/1502.02558`

# Same Example with K2-ABC

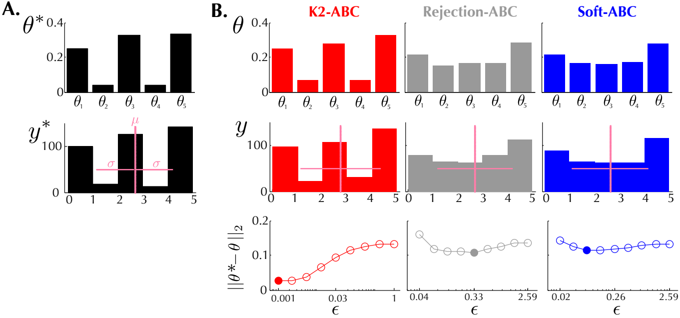
$$p(y|\theta) = \mathcal{N}(y;\theta,1)$$
$$\pi(\theta) = \mathcal{N}(\theta,0,8)$$
$$\theta^* = 3.0$$




- High $\epsilon \Rightarrow$ get $\theta$ sample from prior
- Low $\epsilon \Rightarrow$ sample closely follows true posterior.

# Toy Problem: Failure of Insufficient Statistics

$$
\begin{aligned}
p(y|\theta) &= \sum_{i=1}^{5} \theta_i \mathsf{Uniform}(y; [i-1, i]) \\
\pi(\theta) &= \mathsf{Dirichlet}(\theta; \mathbf{1}) \\
\theta^* &= \text{(see figure A)}
\end{aligned}
$$



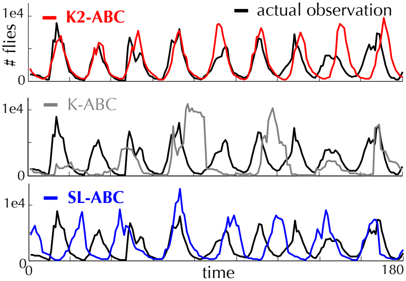

- $s(y) = (\hat{\mathbb{E}}[y], \hat{\mathbb{V}}[y])^\top$ for Rejection and Soft ABC. Insufficient to represent $p(y|\theta)$.

# Blowfly Population Modelling

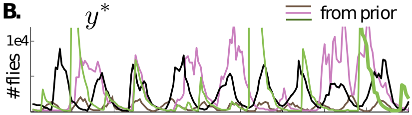
A model for blowfly population over time

$$N_{t+1} = P N_{t-\tau} \exp\left(-\frac{N_{t-\tau}}{N_0}\right) e_t + N_t \exp(-\delta \epsilon_t).$$

- $e_t \sim$ Gam $\left(\frac{1}{\sigma_P^2}, \sigma_P^2\right)$ and $\epsilon_t \sim$ Gam $\left(\frac{1}{\sigma_d^2}, \sigma_d^2\right)$
- $\theta := \{P, N_0, \sigma_d, \sigma_p, \tau, \delta\}$




- $\leftarrow$ Simulated trajectories with inferred posterior mean of $\theta$
  - Observed sample of size 180.
  - Other methods use handcrafted 10-dimensional summary statistics.
- $\downarrow$ 3 random draws from prior.



# Conclusion

- **Problem:** intractable $p(y|\theta)$ but can be sampled.
- **Proposed:** summary statistic free kernel-based ABC algorithm


- "Approximate Bayesian Computations Done Exactly: Towards a Thousand Human Genomes. Dr. Raazesh Sainudiin , Dr. Amandine Véber, 2011."
- **Our future plan**: ABCDEFG ?

# References I

- K2-ABC on Arxiv: http://arxiv.org/abs/1502.02558