

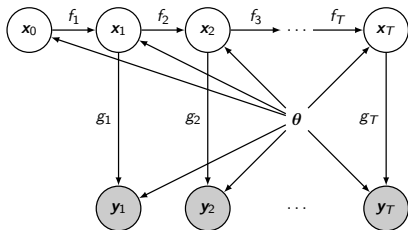
# Bayesian Parameter Estimation of State-Space Models with Intractable Likelihood

Bc. Tomáš Kala

Supervisor: Ing. Kamil Dedecius, PhD.

June 20, 2019

# State-Space Model (SSM)



$$\begin{aligned} \mathbf{x}_0 \mid \boldsymbol{\theta} &\sim p(\mathbf{x}_0 \mid \boldsymbol{\theta}), \\ \mathbf{x}_t \mid \mathbf{x}_{t-1}, \boldsymbol{\theta} &\sim f_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \boldsymbol{\theta}), \\ \mathbf{y}_t \mid \mathbf{x}_t, \boldsymbol{\theta} &\sim g_t(\mathbf{y}_t \mid \mathbf{x}_t, \boldsymbol{\theta}), \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}). \end{aligned}$$

- The posterior of  $\boldsymbol{\theta}$  takes the form of

$$p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

where

$$p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) = \int p(\mathbf{x}_{0:T}, \mathbf{y}_{1:T} \mid \boldsymbol{\theta}) \, d\mathbf{x}_{0:T}.$$

# Particle filter

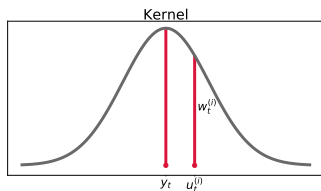
- Use weighted particles  $\left\{ \left( \mathbf{x}_t^{(i)}, w_t^{(i)} \right) : i = 1, \dots, N \right\}$  to approximate the filtering distribution  $p(\mathbf{x}_t \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$ .
- Simulate the particles  $\mathbf{x}_t^{(i)} \sim f_t(\mathbf{x}_t \mid \mathbf{x}_{t-1}^{(i)})$  and weight them according to  $w_t^{(i)} = g_t(\mathbf{x}_t^{(i)} \mid \mathbf{y}_t, \boldsymbol{\theta})$ .
- An unbiased likelihood estimator is

$$\hat{p}(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) = \prod_{t=1}^T \frac{1}{N} \sum_{i=1}^N w_t^{(i)}.$$

- Problem: the observation model  $g_t(\mathbf{x}_t \mid \mathbf{y}_t, \boldsymbol{\theta})$  must be known and probabilistic.

# Approximate Bayesian Computation (ABC)

- Use  $g_t(\mathbf{y}_t \mid \mathbf{x}_t, \boldsymbol{\theta})$  to simulate pseudo-observations  $\mathbf{u}_t$ .
- Determine  $w_t^{(i)}$  based on the closeness of  $\mathbf{u}_t$  to the true  $\mathbf{y}_t$  measured by a kernel function.
- A flexible and robust approach no longer requiring a probabilistic observation model.



## Lotka-Volterra model

- A simplified system of interacting prey ( $\mathcal{X}_1$ ) and predator ( $\mathcal{X}_2$ ) species described by

$$\mathcal{R}_1 : \quad \mathcal{X}_1 \rightarrow 2\mathcal{X}_1,$$

$$\mathcal{R}_2 : \quad \mathcal{X}_1 + \mathcal{X}_2 \rightarrow 2\mathcal{X}_2,$$

$$\mathcal{R}_3 : \quad \mathcal{X}_2 \rightarrow \emptyset.$$

- Assume  $\mathbf{y}_t = \mathbf{x}_t = (x_{1,t}, x_{2,t})^\top$ , Gaussian observation model and the unknown parameters to be  $\boldsymbol{\theta} = (c_1, c_2, c_3)^\top$ .
- Denoting the number of species  $j$  present at the beginning of  $\mathcal{R}_i$  by  $p_{ij}$ ,

$$c_i \prod_{j=1}^2 \binom{x_{j,t}}{p_{ij}}$$

is the mean time to the next occurrence of  $\mathcal{R}_i$  at time  $t$ .

# Lotka-Volterra model

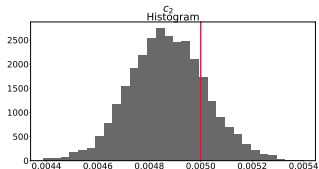


Figure: Well-specified PF.

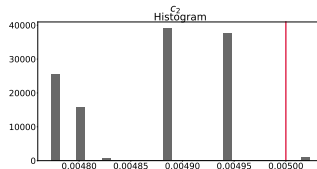


Figure: Misspecified PF.

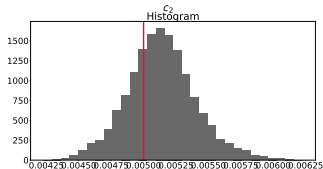


Figure: Well-specified ABC.

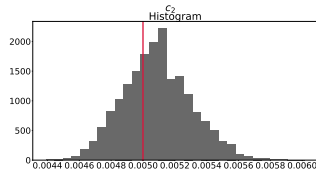
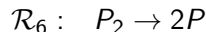
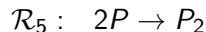
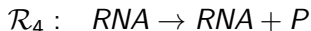
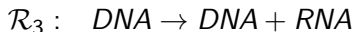
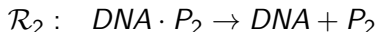
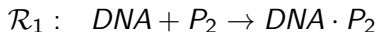


Figure: Misspecified ABC.

# Prokaryotic autoregulation model



- Unknown parameters  $\theta = (c_1, c_2, c_3, c_4, c_7, c_8)^T$ , Gaussian observation model.
- $y_t = P_t + 2(P_2)_t$ .

# Prokaryotic autoregulation model

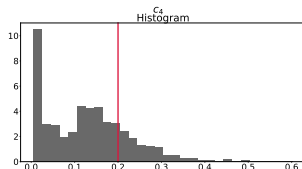


Figure: Well-specified PF.

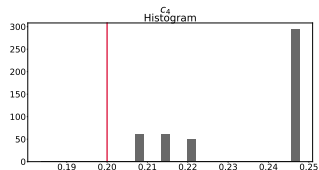


Figure: Misspecified PF.

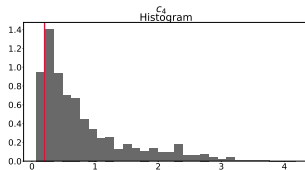


Figure: Well-specified ABC.

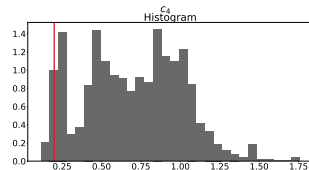


Figure: Misspecified ABC.



# Prokaryotic autoregulation model

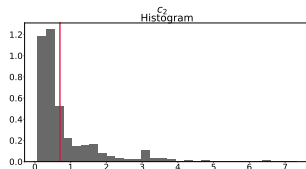


Figure: Well-specified PF.

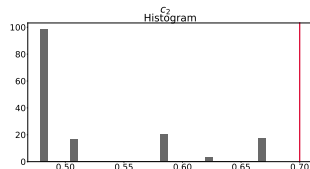


Figure: Misspecified PF.

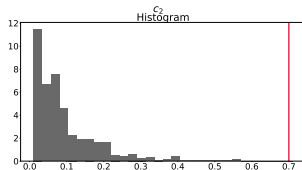


Figure: Well-specified ABC.

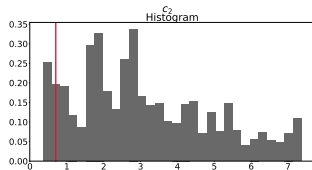


Figure: Misspecified ABC.

# Summary

- Static parameters complicate inference.
- Intractable likelihood can be approximated through the particle filter, assuming a known observation density  $g_t(\mathbf{y}_t \mid \mathbf{x}_t, \boldsymbol{\theta})$ .
- When only a deterministic observation process is available, ABC methods can be used instead.
- Robust to outliers and model misspecification.
- Flexibility in kernel choice.

# Gillespie algorithm

All operations for  $i = 1, \dots, \nu$ , the number of reactions.

1. Set  $t = 0$ , initialize  $\mathbf{x}_t$ .

2. While  $t \leq T$ :

2.1 Calculate  $h_i(\mathbf{x}_t, c_i) = c_i \prod_{j=1}^u \binom{x_{j,t}}{p_{ij}}$ .

2.2 Sample  $dt \sim \text{Exp}(\sum_{i=1}^{\nu} h_i(\mathbf{x}_t, c_i))$ .

2.3 Sample  $i$  with probability  $\propto h_i(\mathbf{x}_t, c_i)$ .

2.4 Set  $\mathbf{x}_{t+dt}$  by updating  $\mathbf{x}_t$  according to  $\mathcal{R}_i$ .

2.5  $t = t + dt$ .

3. Output  $\mathbf{x}_t$ ,  $t$ .

# ABC filter

All operations for  $i = 1, \dots, N$ .

1. Sample  $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0 \mid \boldsymbol{\theta})$ , set  $w_0^{(i)} = \frac{1}{N}$ .
2. For  $t = 1, \dots, T$ :
  - 2.1 Sample  $\mathbf{x}_t^{(i)} \sim f_t(\mathbf{x}_t \mid \mathbf{x}_t^{(i)}, \boldsymbol{\theta})$ .
  - 2.2 Simulate  $\mathbf{u}_t^{(i)}$  from  $g_t(\cdot \mid \mathbf{x}_t^{(i)}, \boldsymbol{\theta})$ .
  - 2.3 Identify  $\mathbf{u}_t^{[\alpha]}$ , the  $\alpha$ th closest pseudo-observation to  $\mathbf{y}_t$ .
  - 2.4 Set the kernel scale  $\epsilon_t = \frac{|\mathbf{u}_t^{[\alpha]} - \mathbf{y}_t|}{F^{-1}(\frac{1+p}{2})}$ .
  - 2.5 Set the weights  $w_t^{(i)} \propto \kappa(\frac{\mathbf{u}_t^{(i)} - \mathbf{y}_t}{\epsilon_t}) w_{t-1}^{(i)}$ .
  - 2.6 Resample  $\mathbf{x}_t^{(i)}$  and reset  $w_t^{(i)}$ .
3. Output  $\{w_1^{(1)}, \dots, w_1^{(N)}, \dots, w_T^{(1)}, \dots, w_T^{(N)}\}$ .

# ABC Metropolis-Hastings

1. Initialize  $\theta^{(0)}$ , estimate  $\hat{p}(\mathbf{y}_{1:T} \mid \theta^{(0)})$ .
2. For  $m = 1, \dots, M$ :
  - 2.1 Propose  $\theta' \sim q(\cdot \mid \theta^{(m-1)})$ .
  - 2.2 Estimate  $\hat{p}(\mathbf{y}_{1:T} \mid \theta')$ .
  - 2.3 Calculate the acceptance ratio

$$\alpha = \min \left\{ 1, \frac{\hat{p}(\mathbf{y}_{1:T} \mid \theta') \pi(\theta')}{\hat{p}(\mathbf{y}_{1:T} \mid \theta^{(m-1)}) \pi(\theta^{(m-1)})} \frac{q(\theta^{(m-1)} \mid \theta')}{q(\theta' \mid \theta^{(m-1)})} \right\}$$

- 2.4 With probability  $\alpha$ , set  $\theta^{(m)} = \theta'$ . Otherwise, set  $\theta^{(m)} = \theta^{(m-1)}$ .
3. Output  $\{\theta^{(1)}, \dots, \theta^{(M)}\}$ .