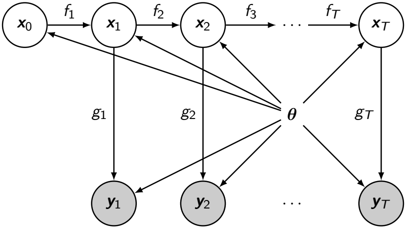
Bayesian Parameter Estimation of State-Space Models with Intractable Likelihood

Bc. Tomáš Kala
Supervisor: Ing. Kamil Dedecius, PhD.

June 20, 2019

State-Space Model (SSM)



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

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

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

where

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

Particle filter

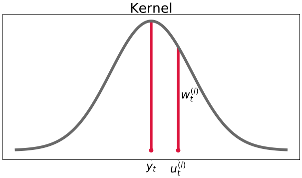
- Use weighted particles $\left\{\left(\boldsymbol{x}_t^{(i)}, w_t^{(i)}\right) : i = 1, \ldots, N\right\}$ to approximate the filtering distribution $p(\boldsymbol{x}_t \mid \boldsymbol{y}_{1:t}, \boldsymbol{\theta})$.
- Simulate the particles $\boldsymbol{x}_t^{(i)} \sim f_t(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}^{(i)})$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

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

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

Approximate Bayesian Computation (ABC)

- Still approximating the likelihood, different weight calculation.
- Simulate pseudo-observations $\boldsymbol{u}_t$ from $g_t(\boldsymbol{y}_t \mid \boldsymbol{x}_t, \boldsymbol{\theta})$.
- Determine $w_t^{(i)}$ based on the distance of $\boldsymbol{u}_t$ to the true $\boldsymbol{y}_t$ 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

$$\begin{aligned} \mathcal{R}_1 :& \quad \mathcal{X}_1 \to 2\mathcal{X}_1, \\ \mathcal{R}_2 :& \quad \mathcal{X}_1 + \mathcal{X}_2 \to 2\mathcal{X}_2, \\ \mathcal{R}_3 :& \quad \mathcal{X}_2 \to \emptyset. \end{aligned}$$

- Assume $\boldsymbol{y}_t = \boldsymbol{x}_t = (x_{1,t}, x_{2,t})^\intercal$, Gaussian observation model and the unknown parameters to be $\boldsymbol{\theta} = (c_1, c_2, c_3)^\intercal$.
- 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

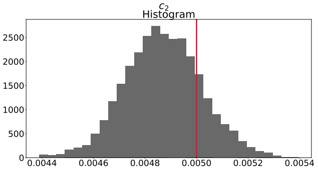


Figure: Well-specified PF.

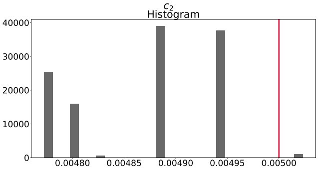


Figure: Misspecified PF.

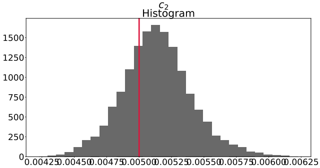


Figure: Well-specified ABC.

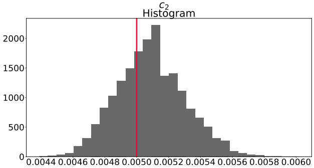


Figure: Misspecified ABC.

Prokaryotic autoregulation model

$$\begin{aligned}
\mathcal{R}_1 &: \quad DNA + P_2 \to DNA \cdot P_2 & \mathcal{R}_5 &: \quad 2P \to P_2 \\
\mathcal{R}_2 &: \quad DNA \cdot P_2 \to DNA + P_2 & \mathcal{R}_6 &: \quad P_2 \to 2P \\
\mathcal{R}_3 &: \quad DNA \to DNA + RNA & \mathcal{R}_7 &: \quad RNA \to \emptyset \\
\mathcal{R}_4 &: \quad RNA \to RNA + P & \mathcal{R}_8 &: \quad P \to \emptyset
\end{aligned}$$

- Unknown parameters $\boldsymbol{\theta} = (c_1, c_2, c_3, c_4, c_7, c_8)^\intercal$, Gaussian observation model.
- $\boldsymbol{x}_t = (RNA_t, P_t, (P_2)_t, DNA_t)^\intercal$.
- $\boldsymbol{y}_t = P_t + 2(P_2)_t$.

Prokaryotic autoregulation model

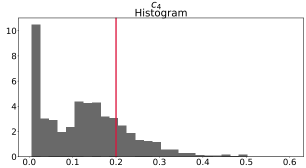


Figure: Well-specified PF.

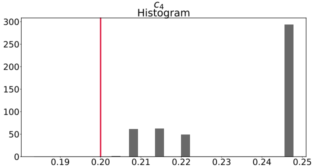


Figure: Misspecified PF.

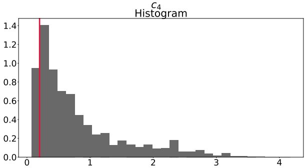


Figure: Well-specified ABC.

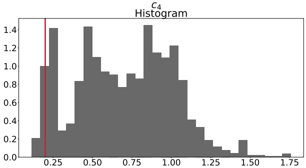


Figure: Misspecified ABC.

Prokaryotic autoregulation model

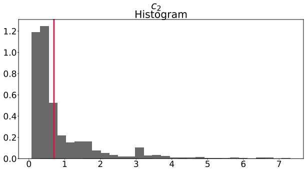


Figure: Well-specified PF.

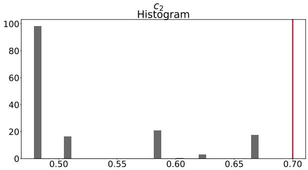


Figure: Misspecified PF.

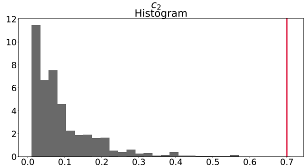


Figure: Well-specified ABC.

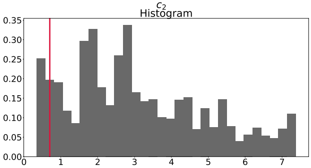


Figure: Misspecified ABC.

Summary

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

Prokaryotic autoregulation model

- The model assumes that the dimer of a protein P, denoted P_2, represses the transcription of its coding gene by binding to a regulatory region in the gene.

$\mathcal{R}_1$:	$DNA + P_2 \rightarrow DNA \cdot P_2$	Repression
$\mathcal{R}_2$:	$DNA \cdot P_2 \rightarrow DNA + P_2$	(Reversible)
$\mathcal{R}_3$:	$DNA \rightarrow DNA + RNA$	Transcription
$\mathcal{R}_4$:	$RNA \rightarrow RNA + P$	Translation
$\mathcal{R}_5$:	$2P \rightarrow P_2$	Dimerization
$\mathcal{R}_6$:	$P_2 \rightarrow 2P$	(Reversible)
$\mathcal{R}_7$:	$RNA \rightarrow \emptyset$	RNA degradation
$\mathcal{R}_8$:	$P \rightarrow \emptyset$	Protein degradation

Gillespie algorithm

All operations for $i = 1, \ldots, v$, the number of reactions.

1. Set $t = 0$, initialize $\boldsymbol{x}_t$.
2. While $t \leq T$:
   2.1 Calculate $h_i(\boldsymbol{x}_t, c_i) = c_i \prod_{j=1}^{u} \binom{x_{j,t}}{p_{ij}}$.
   2.2 Sample $\mathrm{d}t \sim \mathcal{E}xp(\sum_{i=1}^{v} h_i(\boldsymbol{x}_t, c_i))$.
   2.3 Sample i with probability $\propto h_i(\boldsymbol{x}_t, c_i)$.
   2.4 Set $\boldsymbol{x}_{t+\mathrm{d}t}$ by updating $\boldsymbol{x}_t$ according to $\mathcal{R}_i$.
   2.5 $t = t + \mathrm{d}t$.
3. Output $\boldsymbol{x}_t$, t.

ABC filter

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

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

ABC Metropolis-Hastings

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

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

   2.4 With probability α, set $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}'$. Otherwise, set $\boldsymbol{\theta}^{(m)} = \boldsymbol{\theta}^{(m-1)}$.
3. Output $\left\{\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M)}\right\}$.