Particle Filtering for Nonlinear State Space Models

Hans-Peter Höllwirth
Supervisor: Christian Brownlees

Barcelona Graduate School of Economics

June 29, 2017

Outline

- State Space Models
- Filtering
  - Kalman Filter
  - Sequential Importance Resampling (SIR)
  - Continuous Sequential Importance Resampling (CSIR)
  - Importance Sampling Particle Filter
- Evaluation
- Illustration
  - Trivariate Local Level Model
  - Hierarchical Dynamic Poisson Model

State Space Models

Local Level Model

Formulation

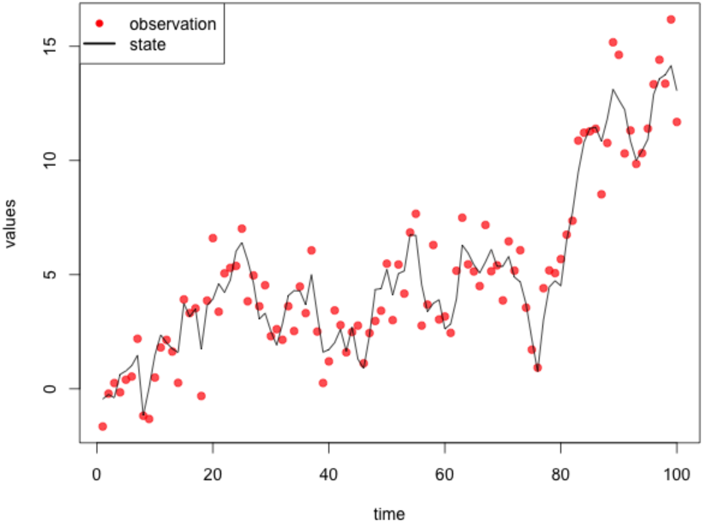
$$\begin{aligned} \text{observation:} \quad & y_t = x_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \\ \text{state:} \quad & x_{t+1} = x_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2) \end{aligned}$$

$$\boldsymbol{\theta} = [\sigma_\eta^2, \sigma_\epsilon^2]^T$$

$$\begin{aligned} \text{transition density:} \quad & x_{t+1} | x_t, \boldsymbol{\theta} \quad \sim N(x_t, \sigma_\epsilon^2) \\ \text{measurement density:} \quad & y_t | x_t, \boldsymbol{\theta} \quad \sim N(x_t, \sigma_\eta^2) \end{aligned}$$

Local Level Realization

$\sigma^2_\eta = 1.4$, $\sigma^2_\epsilon = 1.0$



Filtering

Filtering

Let $\mathcal{I}_t = \{y_1, y_2, \dots, y_t\}$. The objective of filtering is to update our knowledge of the system $p(x_{0:t}|\mathcal{I}_t, \boldsymbol{\theta})$ each time a new observation y_t is brought in. $p(x_{0:t}|\mathcal{I}_t, \boldsymbol{\theta})$ can be decomposed in recursive form:

$$p(x_{0:t}|\mathcal{I}_t, \boldsymbol{\theta}) = \left[\frac{p(y_t|x_t, \boldsymbol{\theta})p(x_t|x_{t-1}, \boldsymbol{\theta})}{p(y_t|\mathcal{I}_{t-1}, \boldsymbol{\theta})}\right] p(x_{0:(t-1)}|\mathcal{I}_{t-1}, \boldsymbol{\theta})$$

where

- $p(y_t|x_t, \boldsymbol{\theta})$ is the measurement density
- $p(x_t|x_{t-1}, \boldsymbol{\theta})$ is the transition density

Particle filtering: recursively simulate the transition density and evaluate the measurement density

Kalman filter

by Kalman (1960)

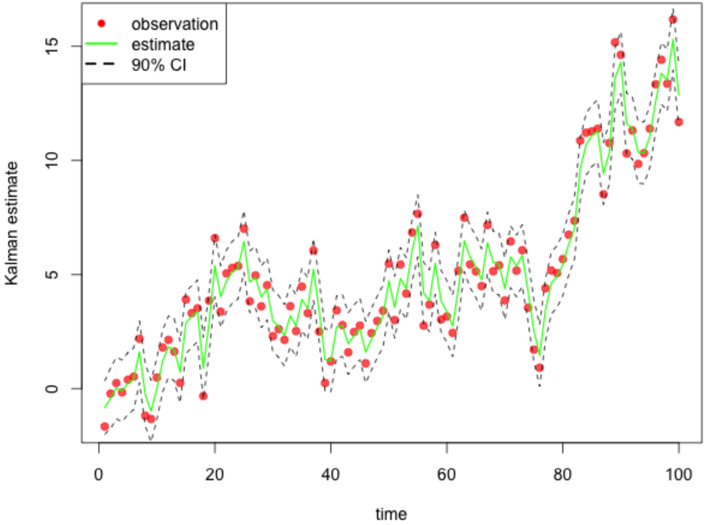
For linear Gaussian state space models, $p(x_{0:t}|\mathcal{I}_t, \boldsymbol{\theta})$ is analytically tractable. The **Kalman filter** infers latent states analytically by recursively updating:

1. the prediction density $x_t|\mathcal{I}_{t-1}, \boldsymbol{\theta} \sim N(\mu_{t|t-1}, \Sigma_{t|t-1})$
2. the filtering density $x_t|\mathcal{I}_t, \boldsymbol{\theta} \sim N(\mu_{t|t}, \Sigma_{t|t})$

Latent state estimates statistically minimize the error and hence are *optimal*.

Kalman filter

Latent State Inference



Sequential Importance Resampling (SIR)

by Gordon et al. (1993)

Recursively computes P prediction and filtering particles:

1. **Prediction step**: draw prediction particles from transition density:

$$x^i_{t|t-1} \sim p(x_t | x^i_{t-1|t-1}, \boldsymbol{\theta}) \quad \text{for } i = 1, \dots, P$$

2. **Filtering step**: draw filtering particles via multinomial sampling:

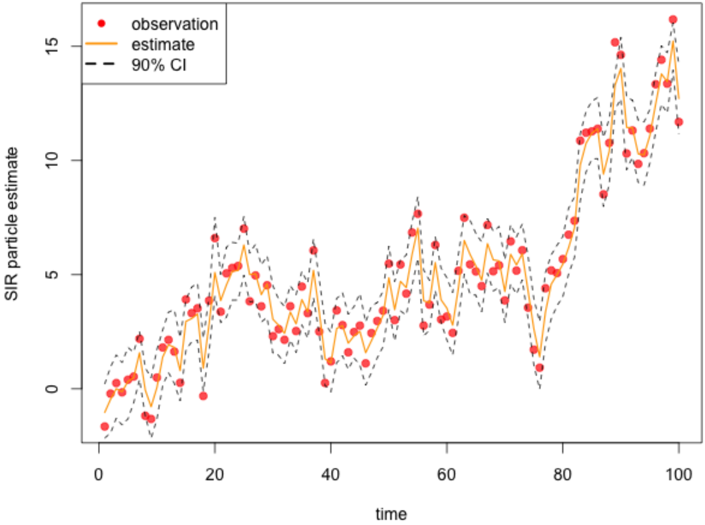
$$x^j_{t|t} \sim MN(w^1_t, \dots, w^P_t) \quad \text{for } j = 1, \dots, P$$

where importance weights are (normalized) evaluations of the measurement density

$$w^i_t = \frac{p(y_t | x^i_{t|t-1}, \boldsymbol{\theta})}{\sum_{j=1}^P p(y_t | x^j_{t|t-1}, \boldsymbol{\theta})} \quad \text{for } i = 1, \dots, P$$

Sequential Importance Resampling (SIR)

Latent State Inference



Sequential Importance Resampling (SIR)

Parameter Inference

Maximum likelihood estimation on the approximated log-likelihood of the observations, given all previous observations:

$$
\begin{aligned}
\log \hat{\mathcal{L}}(\mathcal{I}_T, \boldsymbol{\theta}) &= \log \prod_{t=1}^{T} \hat{p}(y_t|\mathcal{I}_{t-1}, \boldsymbol{\theta}) \\
&= \sum_{t=1}^{T} \log \hat{p}(y_t|\mathcal{I}_{t-1}, \boldsymbol{\theta}) \\
&= \sum_{t=1}^{T} \log \frac{1}{P} \sum_{i=1}^{P} p(y_t|x^i_{t|t-1}, \boldsymbol{\theta})
\end{aligned}
$$

Continuous Sequential Importance Resampling (CSIR)

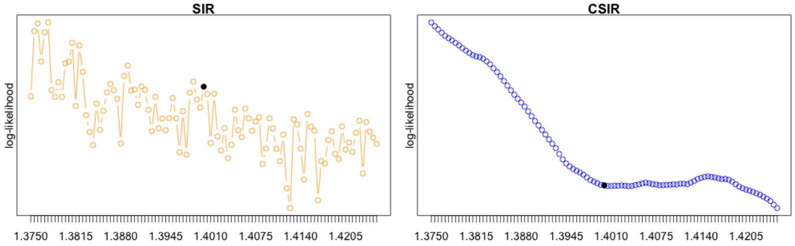
by Malik & Pitt (2011)

Problem: Log-likelihood estimator of SIR method w.r.t. $\boldsymbol{\theta}$ is not smooth. Poor parameter inference results.

Solution: Smooth it! Sort prediction particles in ascending order, form (discrete) CDF using associated weights, interpolate CDF, and finally draw filtering particles from this inverted, smoothed CDF.

Continuous Sequential Importance Resampling (CSIR)

Comparison of log-likelihoods w.r.t. $\boldsymbol{\theta}$



Caveat: CSIR method only works for univariate state space models.

Importance Sampling Particle Filter

by Brownlees & Kristensen (2017)

Key idea: Use an auxiliary, misspecified particle filter with parameter vector $\tilde{\boldsymbol{\theta}}$ to compute the log-likelihood with respect to $\boldsymbol{\theta}$ via recursive importance sampling.

$$\begin{aligned}
p(y_t|\mathcal{I}_{t-1},\boldsymbol{\theta}) &= \int p(y_t|\tilde{x}_t,\boldsymbol{\theta})p(\tilde{x}_t|\mathcal{I}_{t-1},\boldsymbol{\theta})d\tilde{x}_t \\
&= \int p(y_t|\tilde{x}_t,\boldsymbol{\theta})\left[\frac{p(\tilde{x}_t|\mathcal{I}_{t-1},\boldsymbol{\theta})}{p(\tilde{x}_t|\mathcal{I}_{t-1},\tilde{\boldsymbol{\theta}})}\right]p(\tilde{x}_t|\mathcal{I}_{t-1},\tilde{\boldsymbol{\theta}})d\tilde{x}_t \\
\hat{p}(y_t|\mathcal{I}_{t-1},\boldsymbol{\theta}) &\approx \frac{1}{P}\sum_{i=1}^{P} p(y_t|\tilde{x}^i_{t|t-1},\boldsymbol{\theta})\left[\frac{p(\tilde{x}^i_{t|t-1}|\mathcal{I}_{t-1},\boldsymbol{\theta})}{p(\tilde{x}^i_{t|t-1}|\mathcal{I}_{t-1},\tilde{\boldsymbol{\theta}})}\right] \\
&= \frac{1}{P}\sum_{i=1}^{P} p(y_t|\tilde{x}^i_{t|t-1},\boldsymbol{\theta})is^i_{t|t-1}
\end{aligned}$$

Importance Sampling Particle Filter

Importance Weights

Recursively computes P prediction and filtering importance weights:

1. **Prediction importance weight**: compares transition densities

$$is^i_{t|t-1} = \left[\frac{p(\tilde{x}^i_{t|t-1} | \tilde{x}^i_{t-1|t-1}, \boldsymbol{\theta})}{p(\tilde{x}^i_{t|t-1} | \tilde{x}^i_{t-1|t-1}, \tilde{\boldsymbol{\theta}})} \right] is^i_{t-1|t-1}$$

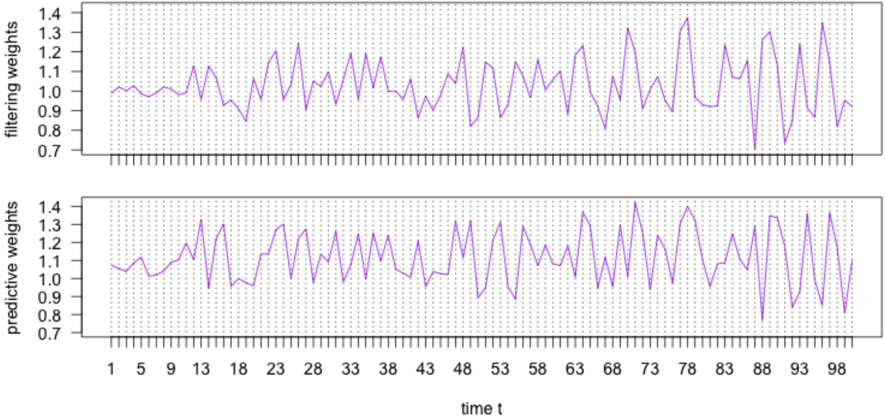
2. **Filtering importance weight**: compares measurement densities and likelihood

$$is^i_{t|t} = \left[\frac{p(y_t | \tilde{x}^i_{t|t}, \boldsymbol{\theta})}{p(y_t | \tilde{x}^i_{t|t}, \tilde{\boldsymbol{\theta}})} \right] \left[\frac{p(y_t | \mathcal{I}_{t-1}, \tilde{\boldsymbol{\theta}})}{p(y_t | \mathcal{I}_{t-1}, \boldsymbol{\theta})} \right] is^j_{t|t-1}$$

where j is such that $\tilde{x}^i_{t|t} = \tilde{x}^j_{t|t-1}$

Importance Sampling Particle Filter

Importance Weights



Importance Sampling Particle Filter

Parameter Inference

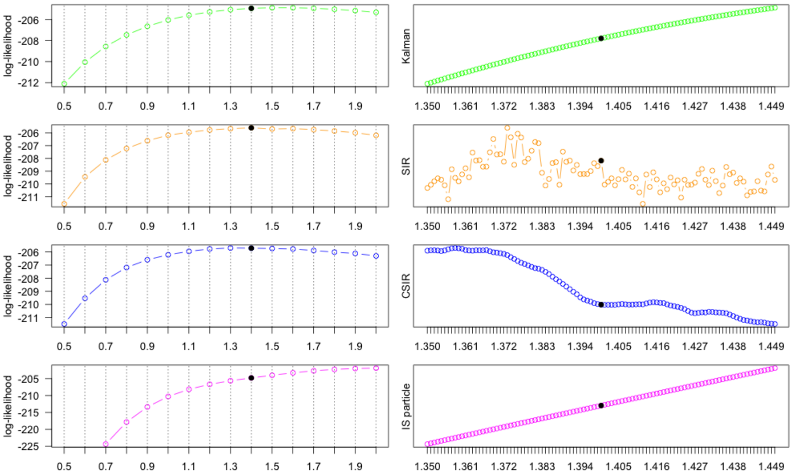
The estimated log-likelihood of $\boldsymbol{\theta}$ weighs each auxiliary particle $\tilde{x}^i_{t|t-1}$ with predictive importance weight $is^i_{t|t-1}$:

$$\begin{aligned}
\log \hat{\mathcal{L}}(\mathcal{I}_T, \boldsymbol{\theta}) &= \log \prod_{t=1}^{T} \hat{p}(y_t|\mathcal{I}_{t-1}, \boldsymbol{\theta}) \\
&= \sum_{t=1}^{T} \log \hat{p}(y_t|\mathcal{I}_{t-1}, \boldsymbol{\theta}) \\
&= \sum_{t=1}^{T} \log \frac{1}{P} \sum_{i=1}^{P} p(y_t|\tilde{x}^i_{t|t-1}, \boldsymbol{\theta}) is^i_{t|t-1}
\end{aligned}$$

Evaluation

Method Comparison

Log-likelihood plots

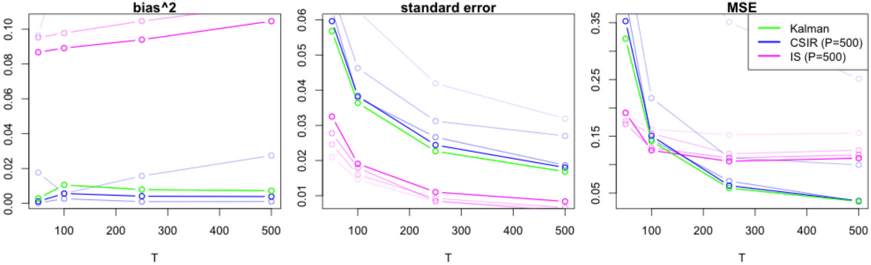


Method Comparison
Filter Choice

Filter	Latent state	Parameter	Comment
Kalman	x	x	linear Gaussian models only
SIR	x		
CSIR	x	x	univariate models only
IS		x	

Monte Carlo Simulations

...on random realizations of the local level model with $\sigma_\eta^2 = 1.4$ and $\sigma_\epsilon^2 = 1.0$ of several different lengths T
(and different values of P: 20, 50, 200, 500)



Illustration

Trivariate Local Level Model

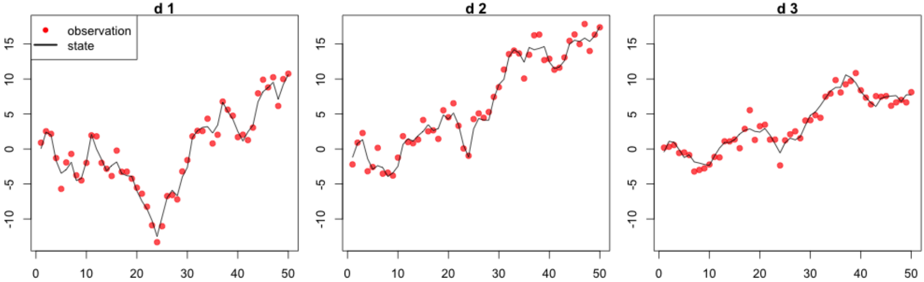
Formulation

$$\begin{aligned} \text{observation:} \quad & \boldsymbol{y}_t = \boldsymbol{x}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \sigma_\epsilon^2 I_3) \\ \text{state:} \quad & \boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \Sigma_\eta) \end{aligned}$$

$$\Sigma_\eta = \begin{bmatrix} \sigma_{\eta 1}^2 & \rho\sigma_{\eta 1}\sigma_{\eta 2} & \rho\sigma_{\eta 1}\sigma_{\eta 3} \\ \rho\sigma_{\eta 1}\sigma_{\eta 2} & \sigma_{\eta 2}^2 & \rho\sigma_{\eta 2}\sigma_{\eta 3} \\ \rho\sigma_{\eta 1}\sigma_{\eta 3} & \rho\sigma_{\eta 2}\sigma_{\eta 3} & \sigma_{\eta 3}^2 \end{bmatrix}$$

$$\boldsymbol{\theta} = [\rho, \sigma_{\eta 1}^2, \sigma_{\eta 2}^2, \sigma_{\eta 3}^2, \sigma_\epsilon^2]^T$$

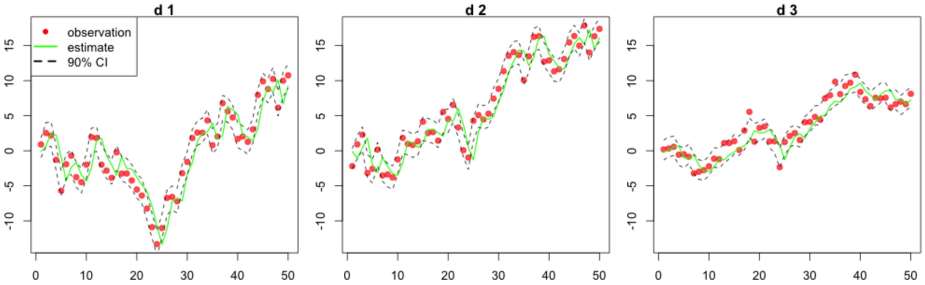
Trivariate Local Level Realization



$$\boldsymbol{\theta} = [\rho = 0.7, \sigma^2_{\eta 1} = 4.2, \sigma^2_{\eta 2} = 2.8, \sigma^2_{\eta 3} = 0.9, \sigma^2_{\epsilon} = 1.0]^T$$

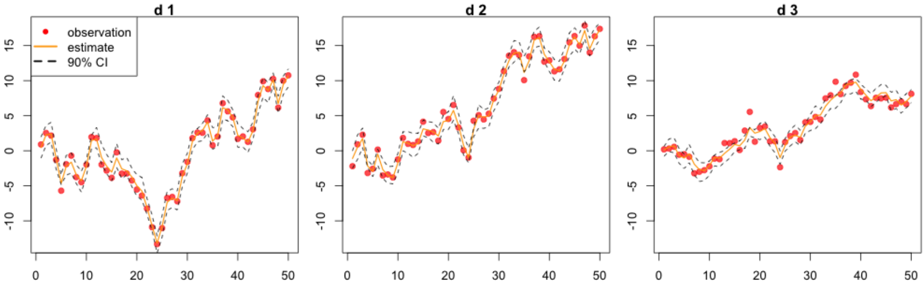
Latent State Inference

Kalman filter



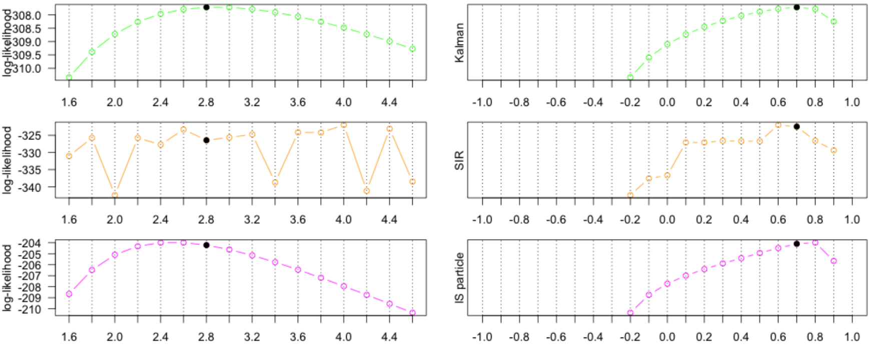
Latent State Inference

SIR particle filter



Parameter Inference

Log-likelihood plots for $\sigma^2_{\eta 2}$ and ρ



Parameter Inference

Results

	$\sigma^2_{\eta 1}$	$\sigma^2_{\eta 2}$	$\sigma^2_{\eta 3}$	ρ
True	4.20	2.80	0.90	0.70
Kalman	4.96	3.10	1.01	0.73
SIR	2.27	1.53	1.29	0.52
IS	2.69	2.09	1.06	0.42

Hierarchical Dynamic Poisson Model

Formulation

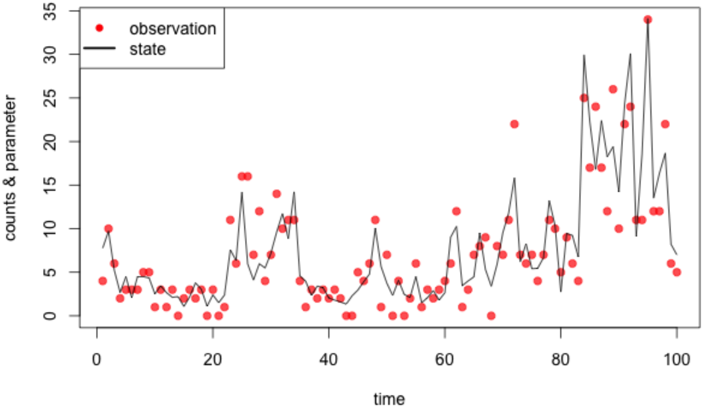
$$\begin{aligned} \text{observation:} \quad y_{m,n} &\sim \text{Poisson}(\lambda_{m,n}) \\ \text{state:} \quad \log \lambda_{m,n} &= \log \lambda_m^{(D)} + \log \lambda_{m,n}^{(I)} + \log \lambda_n^{(P)} \end{aligned}$$

$$\begin{aligned} \text{daily:} \quad \log \lambda_{m+1}^{(D)} &= \phi_0^{(D)} + \phi_1^{(D)} \log \lambda_m^{(D)} + \eta_m^{(D)} & \eta_t \sim N(0, \sigma_{(D)}^2) \\ \text{intra-daily:} \quad \log \lambda_{m,n+1}^{(I)} &= \phi_1^{(I)} \log \lambda_{m,n}^{(I)} + \eta_{m,n}^{(I)} & \eta_{m,n} \sim N(0, \sigma_{(I)}^2) \\ \text{periodic:} \quad \log \lambda_n^{(P)} &= \phi_1^{(P)} \sin(\pi(n-1)/M) \end{aligned}$$

$$\boldsymbol{\theta} = [\phi_0^{(D)}, \phi_1^{(D)}, \sigma_{(D)}^2, \phi_1^{(I)}, \sigma_{(I)}^2, \phi_1^{(P)}]^T$$

Hierarchical Dynamic Poisson Realization

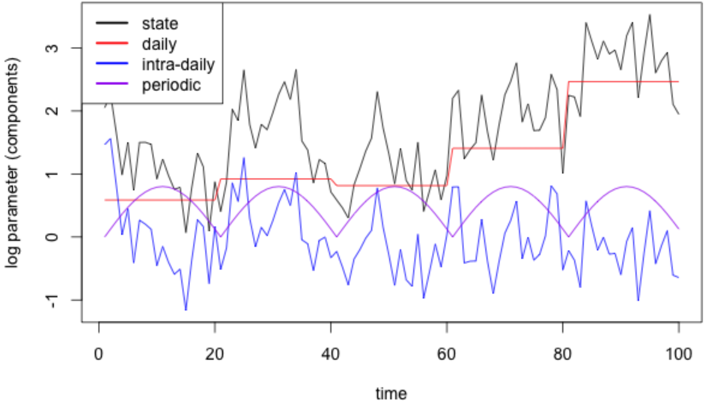
$N = 5,\ M = 20$



$$\boldsymbol{\theta} = [\phi_0^{(D)} = 0.7, \phi_1^{(D)} = 0.6, \sigma^2_{(D)} = 0.6, \phi_1^{(I)} = 0.3, \sigma^2_{(I)} = 0.2, \phi_1^{(P)} = 0.8]^T$$

Hierarchical Dynamic Poisson Realization

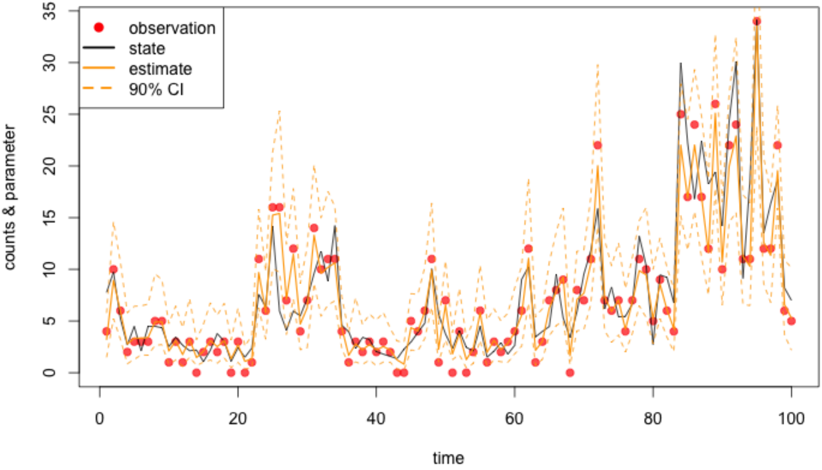
Components



$$\boldsymbol{\theta} = [\phi_0^{(D)} = 0.7, \phi_1^{(D)} = 0.6, \sigma^2_{(D)} = 0.6, \phi_1^{(I)} = 0.3, \sigma^2_{(I)} = 0.2, \phi_1^{(P)} = 0.8]^T$$

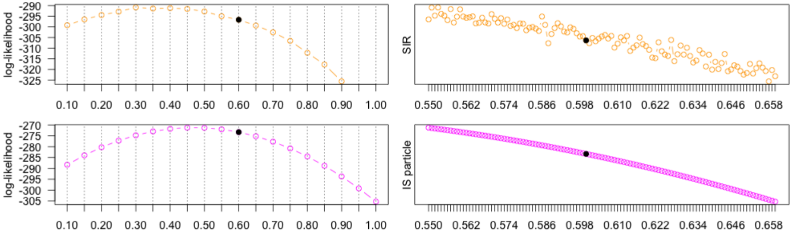
Latent State Inference

SIR particle filter



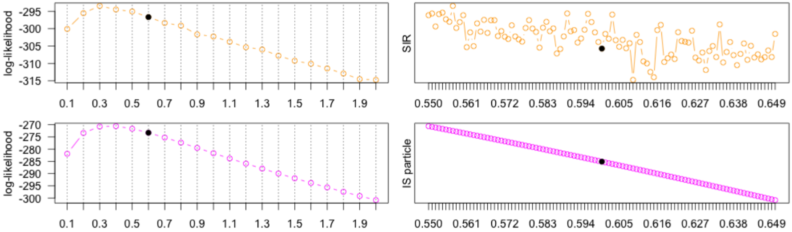
Parameter Inference

Log-likelihood plots for $\phi_1^{(D)}$



Parameter Inference

Log-likelihood plots for $\sigma^2_{(D)}$



Parameter Inference

Results

	$\phi_0^{(D)}$	$\phi_1^{(D)}$	$\sigma^2_{(D)}$	$\phi_1^{(I)}$	$\sigma^2_{(I)}$	$\phi_1^{(P)}$
True	0.70	0.60	0.30	0.80	0.60	0.20
SIR	0.76	0.56	0.85	0.48	0.83	1.13
IS	0.65	0.59	0.40	0.63	0.35	0.31

Q & A