An introduction to Sequential Monte Carlo Methods (Particle Filters) and Kalman Filters

Deniz Gencaga

University of Texas at Dallas
Center for Space Sciences

Application Areas

- Navigation and Guidance
- Localization and Mapping
- Computer Vision
- Geophysical data modeling
- Meteorology and Oceanography

Particle Filtering (Sequential Monte Carlo method)

- Sequential Bayesian solution to infer the dynamical parameters (states) of a system
- System dynamics is expressed by state-space representation

General State-Space Equations

- Dynamic systems are expressed in state-space equations:

We do not
Know the model in our
Cases:

Propose an
Artificial Model

$$\left\{\begin{aligned} \mathbf{x}_t &= f_t\left(\mathbf{x}_{t-1}, \mathbf{v}_t\right) \\ \mathbf{y}_t &= h_t\left(\mathbf{x}_t, \mathbf{n}_t\right) \end{aligned}\right\}$$

Generally
Nonlinear and/or
Non-Gaussian

- *Objective: Sequentially estimate a posteriori distribution of the states:*

$$p\left(\mathbf{x}_{0:t} | \mathbf{y}_{1:t}\right)$$

$$\mathbf{x}_{0:t} = \left[\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t\right] \qquad \mathbf{y}_{1:t} = \left[\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t\right]$$

Recursive Bayesian Solution

$$p\left(\mathbf{x}_{0:t} \middle| \mathbf{y}_{1:t}\right) = p\left(\mathbf{x}_{0:t} \middle| \mathbf{y}_t, \mathbf{y}_{1:t-1}\right)$$

$$p\left(\mathbf{x} \middle| \mathbf{y}\right) = \frac{p\left(\mathbf{y} \middle| \mathbf{x}\right) p(\mathbf{x})}{p\left(\mathbf{y}\right)} \propto p\left(\mathbf{y} \middle| \mathbf{x}\right) p(\mathbf{x})$$

$$= \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_{0:t}, \mathbf{y}_{1:t-1}\right) p\left(\mathbf{x}_{0:t} \middle| \mathbf{y}_{1:t-1}\right)}{p\left(\mathbf{y}_t \middle| \mathbf{y}_{1:t-1}\right)}$$

$$= \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_{0:t}, \mathbf{y}_{1:t-1}\right) p\left(\mathbf{x}_t \middle| \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t-1}\right) p\left(\mathbf{x}_{0:t-1} \middle| \mathbf{y}_{1:t-1}\right)}{p\left(\mathbf{y}_t \middle| \mathbf{y}_{1:t-1}\right)}$$

$$= \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_t\right) p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}\right)}{p\left(\mathbf{y}_t \middle| \mathbf{y}_{1:t-1}\right)} p\left(\mathbf{x}_{0:t-1} \middle| \mathbf{y}_{1:t-1}\right)$$

$$\propto p\left(\mathbf{y}_t \middle| \mathbf{x}_t\right) p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}\right) p\left(\mathbf{x}_{0:t-1} \middle| \mathbf{y}_{1:t-1}\right)$$

A sub-case: Kalman Filter

- General solution is available through tracking of mean and covariance matrices in case of state-space systems expressible by linear, Gaussian equations:

$$\mathbf{x}_t = \mathbf{F}_t \mathbf{x}_{t-1} + \mathbf{v}_t \qquad \mathbf{v}_t \sim \mathrm{N}\left(\mathbf{0}, \mathbf{Q}_t\right)$$

$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{n}_t \qquad \mathbf{n}_t \sim \mathrm{N}\left(\mathbf{0}, \mathbf{R}_t\right)$$

GAUSSIAN

Linear

$$\text{GAUSSIAN} \longrightarrow \boxed{\text{Linear Time Invariant System}} \longrightarrow \text{GAUSSIAN}$$

Kalman Filter (cont.'d)

- Prediction and update can be performed by propagating mean and covariance only!

MEAN COVARIANCE

$$p(\mathbf{x}_{t-1} \mid \mathbf{y}_{1:t-1}) = \mathrm{N}\,(\mathbf{x}_{t-1}; \mathbf{m}_{t-1|t-1}, \mathbf{P}_{t-1|t-1})$$

$$p(\mathbf{x}_{t} \mid \mathbf{y}_{1:t-1}) = \mathrm{N}\,(\mathbf{x}_{t}; \mathbf{m}_{t|t-1}, \mathbf{P}_{t|t-1})$$

Posterior $\longleftarrow$ $$p(\mathbf{x}_{t} \mid \mathbf{y}_{1:t}) = \mathrm{N}\,(\mathbf{x}_{t}; \mathbf{m}_{t|t}, \mathbf{P}_{t|t})$$

$$\mathbf{m}_{t|t-1} = \mathbf{F}_t \mathbf{m}_{t-1|t-1}$$

$$\mathbf{P}_{t|t-1} = \mathbf{Q}_{t-1} + \mathbf{F}_t \mathbf{P}_{t-1|t-1} \mathbf{F}_t^T$$

$$\mathbf{m}_{t|t} = \mathbf{m}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{H}_t \mathbf{m}_{t|t-1})$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{K}_t \mathbf{H}_t \mathbf{P}_{t|t-1}$$

$$\mathbf{S}_t = \mathbf{H}_t \mathbf{P}_{t|t-1} \mathbf{H}_t^T + \mathbf{R}_t$$

$$\mathbf{K}_t = \mathbf{P}_{t|t-1} \mathbf{H}_t^T \mathbf{S}_t^{-1}$$

General Bayesian solution: Sequential Importance Sampling

- We cannot perform analytical calculations in the most general case: Nonlinear and/or non-Gaussian system
- Solution: Express posterior by samples(particles)

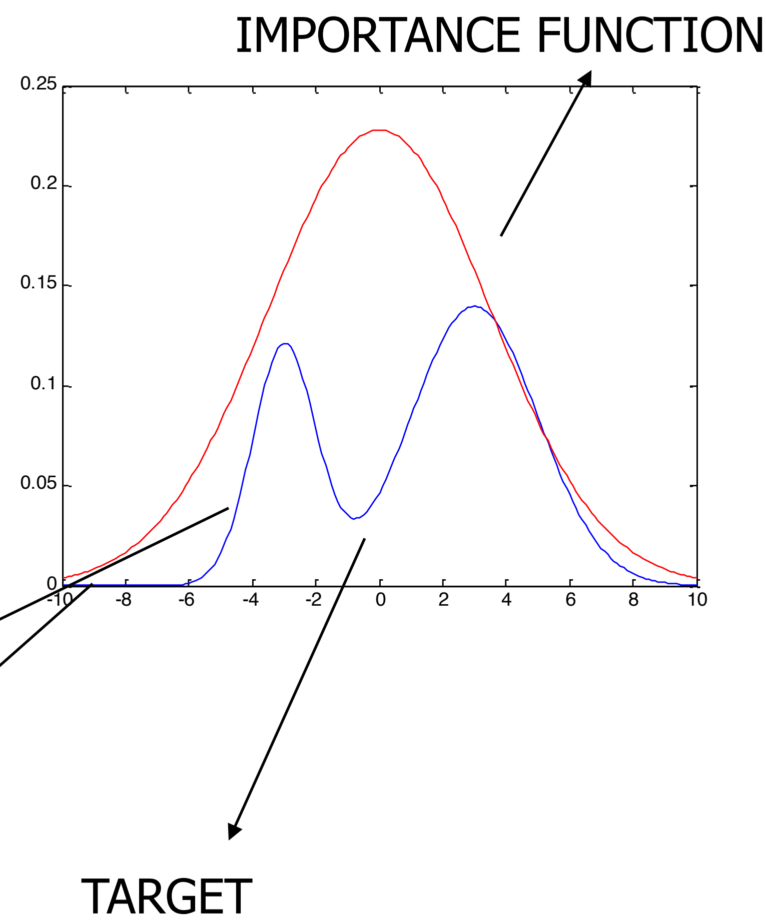
$$p\left(\mathbf{x}_{0:t} \mid \mathbf{y}_{1:t}\right) = \underbrace{\sum_{i=1}^{N} \tilde{w}_t^{(i)}}_{\text{weights}} \delta\left(\mathbf{x}_{0:t} - \underbrace{\mathbf{x}_{0:t}^{(i)}}_{\text{particles}}\right)$$

$$\left\{\mathbf{x}_{0:t}^{(i)}, i = 1, \ldots, N\right\}$$

How to sample? Importance sampling

- Sample from a probability density function (from which sampling is easy)
- Calculate importance weights:

$$\tilde{w}\left(\mathbf{x}_{0:t}^{(i)}\right) \propto \frac{p\left(\mathbf{x}_{0:t}^{(i)} \middle| \mathbf{y}_{1:t}\right)}{q\left(\mathbf{x}_{0:t}^{(i)} \middle| \mathbf{y}_{1:t}\right)}$$



How to perform sequential Importance Sampling?

- Pick an importance function which can be factorized as follows:

$$q\left(\mathbf{x}_{0:t}\middle|\mathbf{y}_{1:t}\right) \triangleq q\left(\mathbf{x}_t\middle|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}\right)q\left(\mathbf{x}_{0:t-1}\middle|\mathbf{y}_{1:t-1}\right)$$

Draw samples from this:

$$\mathbf{x}_t^{(i)} \sim q\left(\mathbf{x}_t\middle|\mathbf{x}_{0:t-1},\mathbf{y}_{1:t}\right) \longrightarrow \mathbf{x}_{0:t}^{(i)} = \left\{\mathbf{x}_{0:t-1}^{(i)},\mathbf{x}_t^{(i)}\right\}$$

- Then we can sequentially update according to the following weights:

$$\tilde{w}\left(\mathbf{x}_{0:t}^{(i)}\right) \propto \frac{p\left(\mathbf{x}_{0:t}^{(i)} \middle| \mathbf{y}_{1:t}\right)}{q\left(\mathbf{x}_{0:t}^{(i)} \middle| \mathbf{y}_{1:t}\right)} \longrightarrow \tilde{w}_t^{(i)} \propto \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) p\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}\right)}{q\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t}\right)} \tilde{w}_{t-1}^{(i)}$$

State transition

Likelihood

$$\mathbf{x}_t = f_t\left(\mathbf{x}_{t-1}, \mathbf{v}_t\right)$$

$$\mathbf{y}_t = h_t\left(\mathbf{x}_t, \mathbf{n}_t\right)$$

Optimal Choice of Importance function:

$$q\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t}\right) = p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t}\right)$$

↓

Dependence on measurement

↓

Very difficult to sample from such a pdf!!!

Use APPROXIMATIONS: BOOTSTRAP PARTICLE FILTER

$$p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}^{(i)}, \mathbf{y}_{1:t}\right) \approx p\left(\mathbf{x}_t \middle| \mathbf{x}_{t-1}^{(i)}\right)$$ → STATE TRANSITION pdf

Importance weight calculation →

$$\tilde{w}_t^{(i)} \propto \frac{p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) \cancel{p\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}\right)}}{\cancel{p\left(\mathbf{x}_t^{(i)} \middle| \mathbf{x}_{t-1}^{(i)}\right)}} \tilde{w}_{t-1}^{(i)}$$

$$\tilde{w}_t^{(i)} \propto p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) \tilde{w}_{t-1}^{(i)}$$

Problem: Degeneracy

- Only one particle can survive in time!!!
- Example:

$$x_t = \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos(1.2t) + v_t$$

$$y_t = \frac{x_t^2}{20} + n_t$$

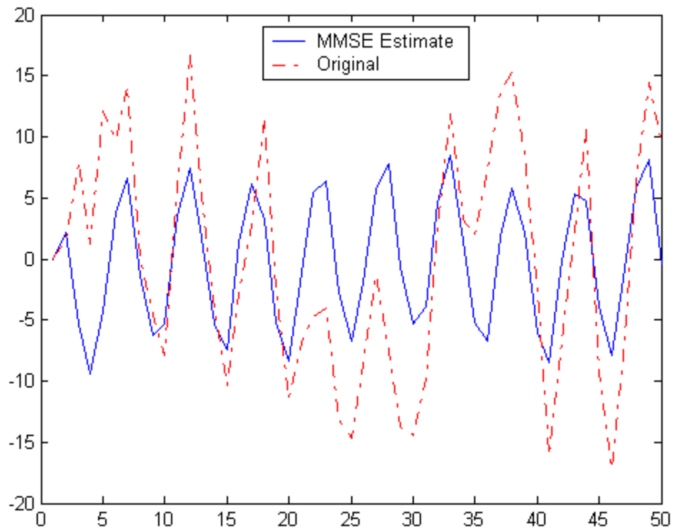
$\sigma_v^2 = 10$ and $\sigma_n^2 = 1$

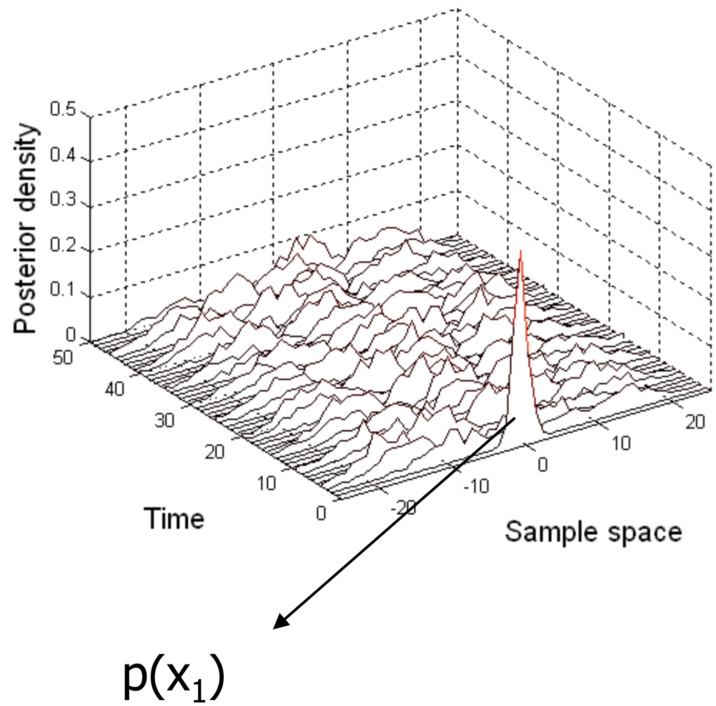
Necessary pdf forms:

$$p(x_t | x_{t-1}) = \mathrm{N}\left(x_t; \frac{x_{t-1}}{2} + \frac{25x_{t-1}}{1+x_{t-1}^2} + 8\cos(1.2t), 10\right)$$

$$p(y_t | x_t) = \mathrm{N}\left(y_t; \frac{x_t^2}{20}, 1\right)$$

Estimation result:

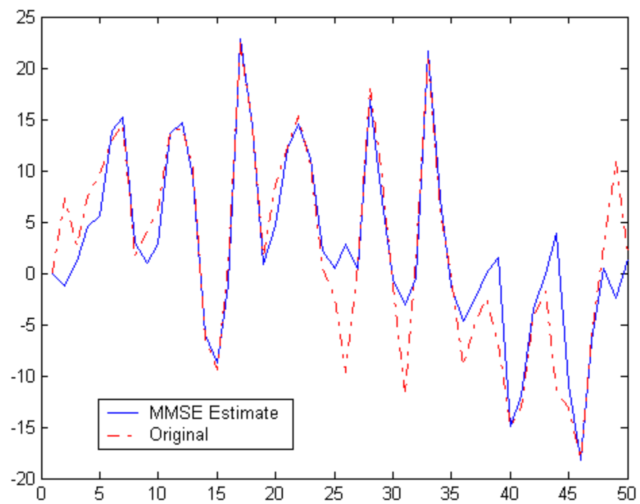


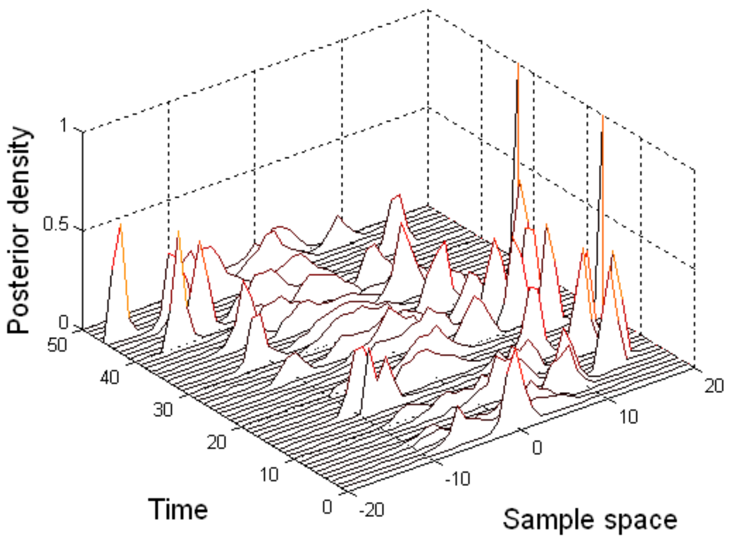


POOR ESTIMATION !

Solution: RESAMPLING

- Kill or Duplicate particles according to their importance weights:





GOOD ESTIMATION !

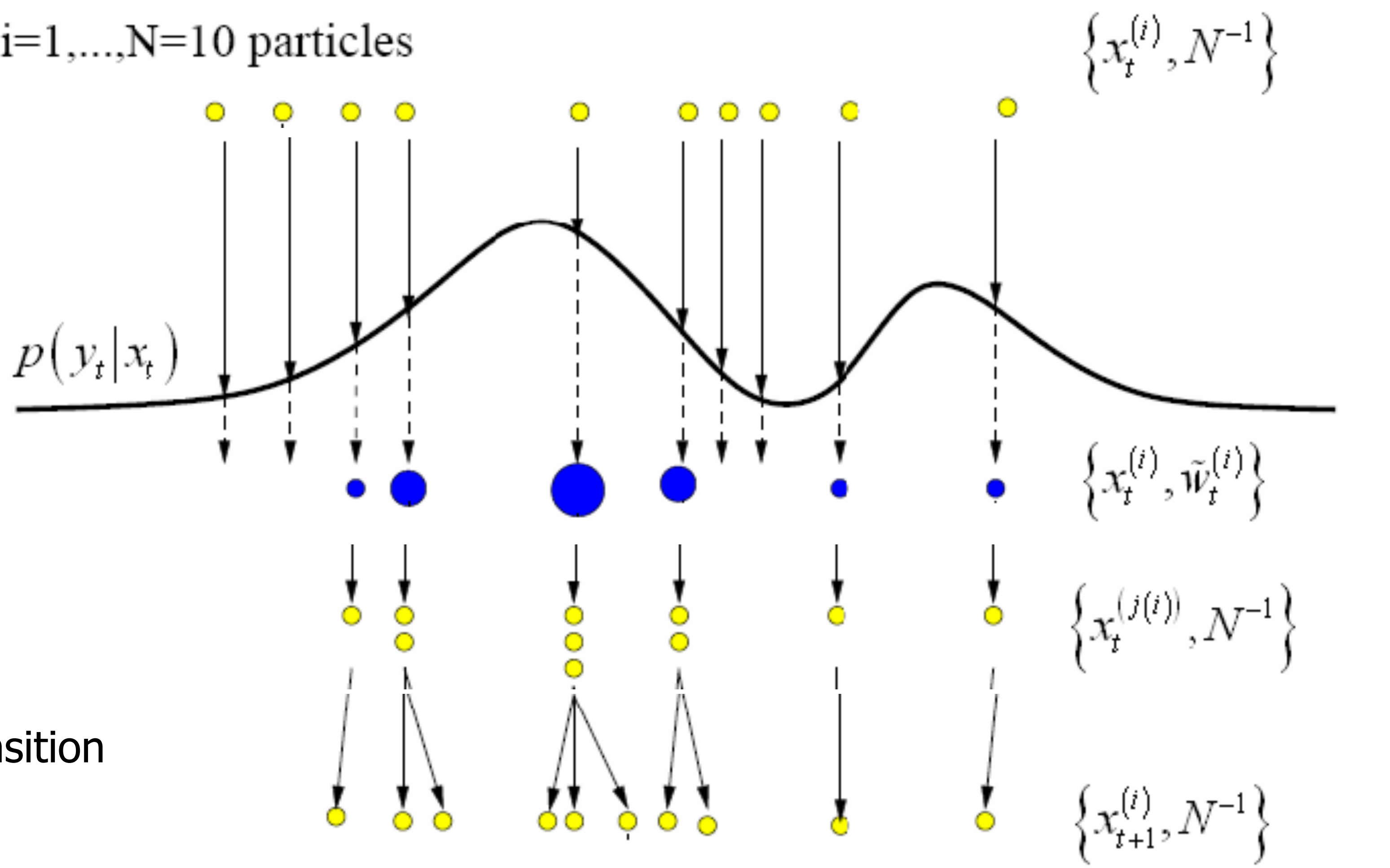
SUMMARY

Sample

Weights:

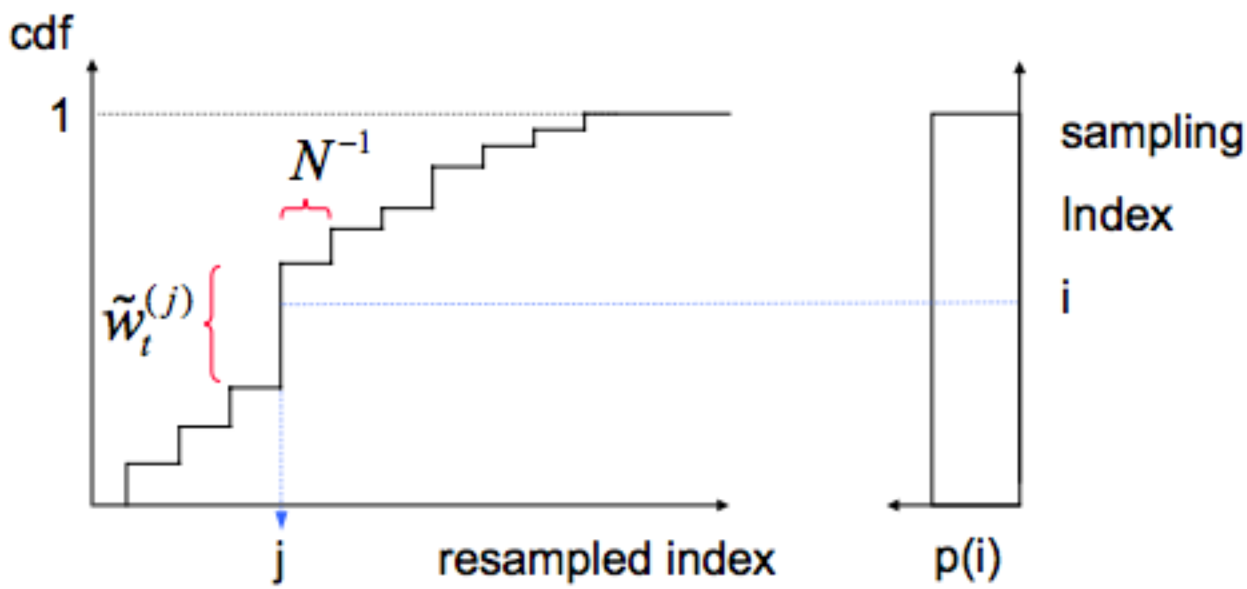
Resample

State Transition

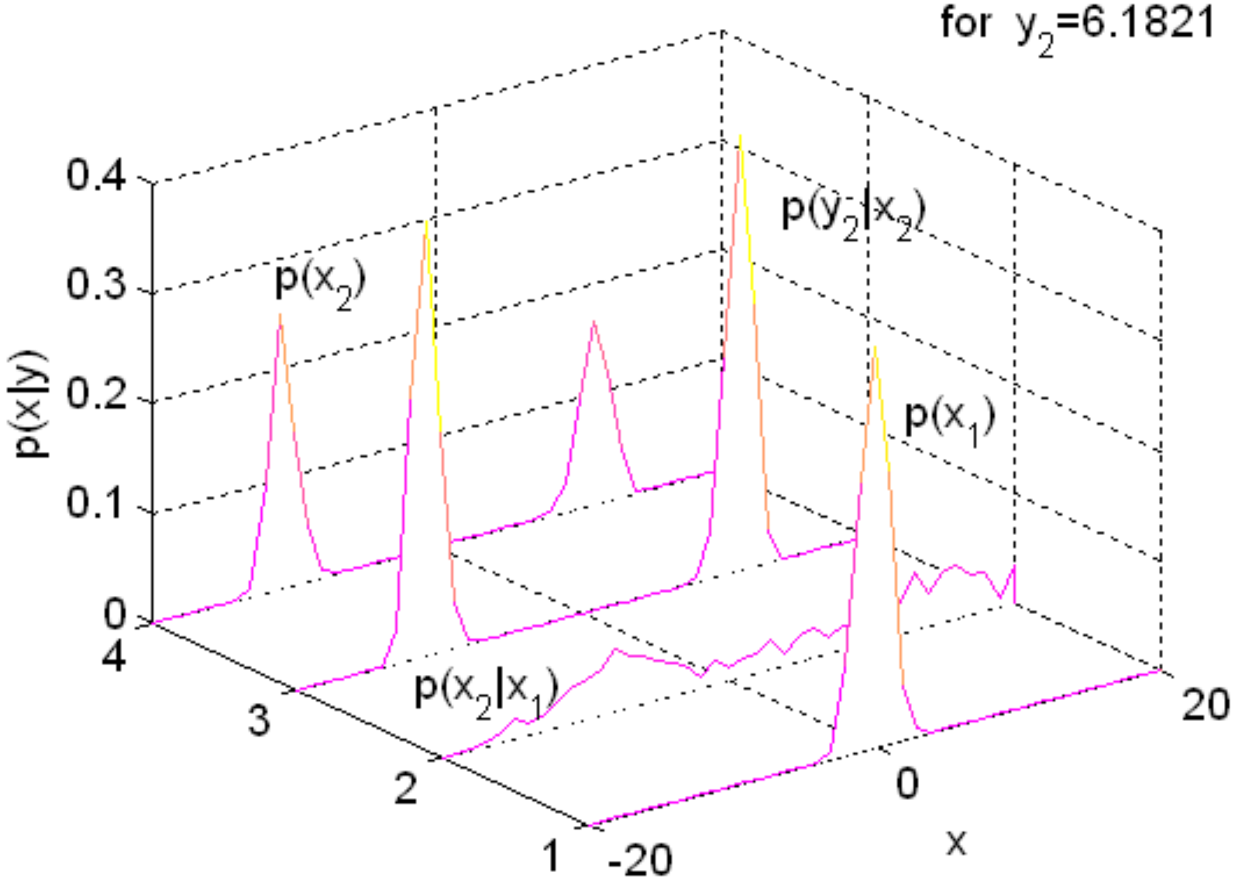


(Van der Merwe, 2000)

Resampling (systematic)



One step propagation



2) Modeling Cross-Correlated Non-Stationary Non-Gaussian Processes

Vector Autoregressive (VAR) Models

- Relationship between cross-correlated AR processes is modeled
- Example: Bivariate, first order:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix}$$

$$\mathbf{\Phi}_1 = \left(\boldsymbol{\varphi}_1^T, \boldsymbol{\varphi}_2^T\right)^T = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix} \quad \text{and} \quad E\left[\mathbf{n}_{t_1} \mathbf{n}_{t_2}^T\right] = \begin{cases} \mathbf{\Sigma}_{\mathbf{n}}, & t_1 = t_2 \\ \mathbf{0}, & otherwise \end{cases}$$

Vector Autoregressive Models

- General Model (K-th order VAR)

$$\boxed{\mathbf{y}_t = \mathbf{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{\Phi}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_K \mathbf{y}_{t-K} + \mathbf{n}_t}$$

$$\mathbf{y}_t = \left[y_{1,t}, y_{2,t}, \ldots, y_{d_1,t} \right]^T \quad \mathbf{n}_t = \left[n_{1,t}, n_{2,t}, \ldots, n_{d_1,t} \right]^T$$

$$E\left[\mathbf{n}_t\right] = \mathbf{0} \text{ and } E\left[\mathbf{n}_{t_1} \mathbf{n}_{t_2}^T\right] = \begin{cases} \mathbf{\Sigma}_{\mathbf{n}}, & t_1 = t_2 \\ \mathbf{0}, & otherwise \end{cases}$$

$\mathbf{\Phi}_j$:j-th order AR coefficientt matrix

Vector Autoregressive Models

- Wide-Sense Stationary VAR processes
  - A VAR(K) is w.s.s. if

$$\left| \mathbf{I}_{d_1} e^{K} - \mathbf{\Phi}_1 e^{K-1} - \mathbf{\Phi}_2 e^{K-2} - \mathrm{L} \; - \mathbf{\Phi}_K \right| = 0$$

is satisfied for all $|e| < 1$

Literature Survey on VAR models

- Stationary cases:
  - Least Squares (Hamilton, 1994; Hsu, 1997)
  - Maximum Likelihood (Lütkepohl, 1993)
- Non-stationary cases:
  - Recursive Least Squares for EEG data (Möller et al., 2001)
  - Wavelet expansion of AR coefficients, fMRI data (Sato et al., 2006)
  - Modified Yule-Walker approach, mobile comm. (Jachan and Matz, 2005)

Literature Survey on VAR models

- Driving processes are modeled by Gaussian distributions in all these cases

- ***OUR SECOND CONTRIBUTION*: Non-Gaussian driving processes can also be modeled in time-varying VAR models: A more general method**

- ***A building block to model non-stationary mixtures of cross-correlated processes (DCA) is developed with successful results***

Modeling Cross-Correlated Non-Stationary Non-Gaussian Processes

- Bivariate time-varying VAR model:

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix}$$

$$\mathbf{\Phi}_{1,t} = \left(\boldsymbol{\varphi}_{1,t}^T, \boldsymbol{\varphi}_{2,t}^T\right)^T = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{bmatrix} \quad \text{and} \quad \mathbf{n}_t = \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix}, \quad \mathbf{n}_t \sim \mathrm{N}\left(\mathbf{0}, \mathbf{\Sigma}_{n_t}\right)$$

Time-varying

Modeling Cross-Correlated Non-Stationary Non-Gaussian Processes

- K-th order time-varying VAR model:

$$\mathbf{y}_t = \mathbf{\Phi}_{1,t}\mathbf{y}_{t-1} + \mathbf{\Phi}_{2,t}\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_{K,t}\mathbf{y}_{t-K} + \mathbf{n}_t$$

$$\mathbf{n}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}_{n_t}\right)$$

In Literature

$$\mathbf{n}_t \sim \sum_{l=1}^{N_n} p_l \mathcal{N}\left(\mathbf{m}_l, \mathbf{\Sigma}_l\right)$$

In our method

Mixture of Gaussians

$$\mathbf{y}_t = \left[y_{1,t}, y_{2,t}, \ldots, y_{d,t}\right]^T$$

$$\mathbf{n}_t = \left[n_{1,t}, n_{2,t}, \ldots, n_{d,t}\right]^T$$

Modeling Cross-Correlated Non-Stationary Non-Gaussian Processes

- Bootstrap PF is used with the following states:

$$\mathbf{x}_t = \left[x_{1,t}, x_{2,t}, \ldots, x_{K',t} \right]^T = \left[\tilde{\Phi}_{1,t}^T, \tilde{\Phi}_{2,t}^T, \ldots, \tilde{\Phi}_{K,t}^T \right]^T$$

$$\tilde{\Phi}_{j,t} = vec\left(\Phi_{j,t} \right)$$

- Their time-variations are unknown!!!

Modeling Cross-Correlated Non-Stationary Non-Gaussian Processes

- Artificial state-transition equation model:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{V}_t$$

$$\mathbf{x}_t = \begin{bmatrix} \tilde{\Phi}_{1,t} \\ \tilde{\Phi}_{2,t} \\ \vdots \\ \tilde{\Phi}_{K,t} \end{bmatrix} = \begin{bmatrix} \tilde{\Phi}_{1,t-1} \\ \tilde{\Phi}_{2,t-1} \\ \vdots \\ \tilde{\Phi}_{K,t-1} \end{bmatrix} + \begin{bmatrix} \nu_{1,t} \\ \nu_{2,t} \\ \vdots \\ \nu_{K',t} \end{bmatrix}$$

$$\mathbf{\Sigma}_{\mathbf{V}_t} = \mathbf{\Sigma}_{\mathbf{x}_{t-1}} \left(\frac{1}{\xi} - 1 \right)$$

$$\mathbf{\Sigma}_{\mathbf{V}_t} = diag\left(\sigma_{1,t}^2, \ldots, \sigma_{K',t}^2 \right)$$

$$\sigma_{k,t}^2 = \left(\frac{1}{\xi} - 1 \right) \mathrm{var}\left(x_{t-1}(k) \right), \qquad k = 1, 2, \ldots, K'$$

Modeling Cross-Correlated Non-Stationary Non-Gaussian Processes

- Bootstrap PF $\longrightarrow$ Use artificial state-transition pdf to draw samples:

$$q\left(\mathbf{x}_t \,\middle|\, \mathbf{x}_{0:t-1}, \mathbf{y}_{1:t}\right) = p\left(\mathbf{x}_t \,\middle|\, \mathbf{x}_{t-1}\right)$$

$$\mathbf{x}_t^{(i)} \sim \mathrm{N}\left(\mathbf{x}_t^{(i)}, \mathbf{\Sigma}_{V_t}\right)\mathbf{I}_{\overline{\mathbf{x}}}$$

Indicator function

$$\mathbf{I}_{\mathbf{x}} = \begin{cases} 1, & \mathbf{x}_t^{(i)} \in \overline{\mathbf{x}} \\ 0, & \mathbf{x}_t^{(i)} \notin \overline{\mathbf{x}} \end{cases}$$

Stability region

Modeling Cross-Correlated Non-Stationary Non-Gaussian Processes

- Calculate the importance weights by evaluating drawn particles at the likelihood function (Bootstrap PF):

$$w_t^{(i)} = p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right)$$

$$p\left(\mathbf{y}_t \middle| \mathbf{x}_t^{(i)}\right) = \mathrm{N}\left(\mathbf{y}_t ; \mathbf{m}_t^{(i)}, \mathbf{\Sigma}_{\mathbf{n}_t}\right)$$

$$\mathbf{m}_t^{(i)} = \mathbf{\Phi}_{1,t}^{(i)}\mathbf{y}_{t-1} + \mathbf{\Phi}_{2,t}^{(i)}\mathbf{y}_{t-2} + \cdots + \mathbf{\Phi}_{K,t}^{(i)}\mathbf{y}_{t-K}$$

Gaussian Driving Process

- Non-stationary cross-correlated first order VAR(1):
- 100 realizations for ensemble averaging

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix}$$

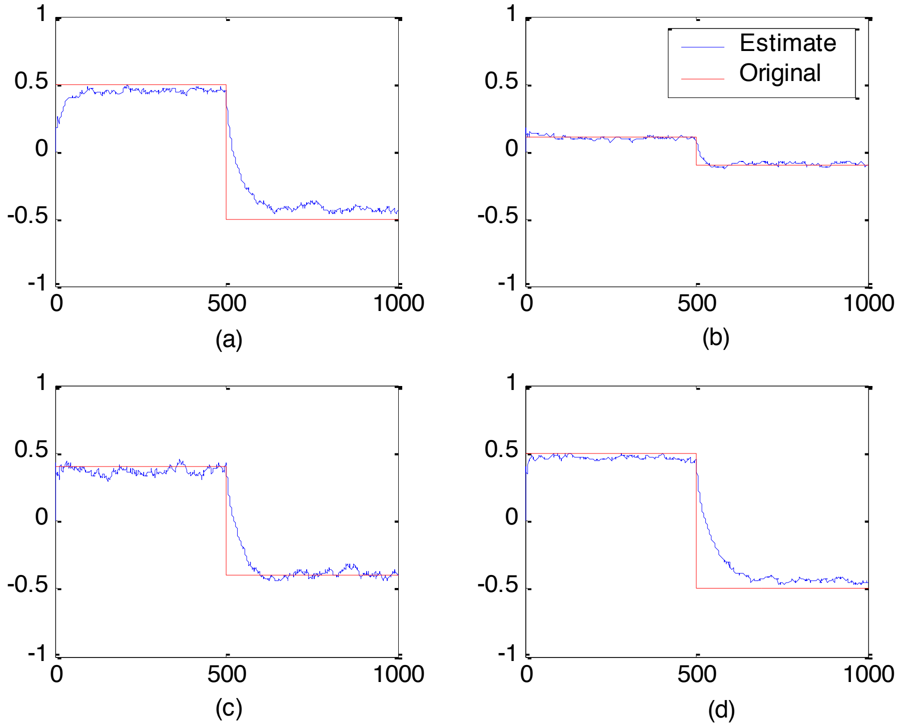
Cross-correlations through AR coefficient matrix

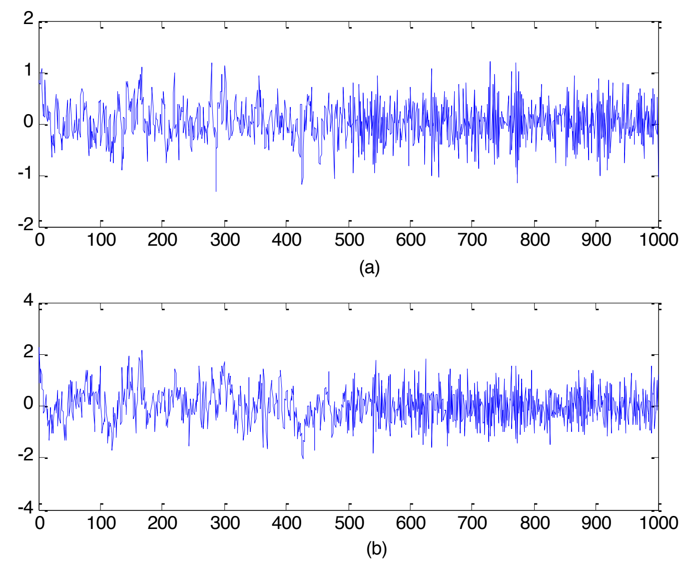
$$\phi_{11,t} = \begin{cases} 0.5 & t < 500 \\ -0.5 & t \geq 500 \end{cases}, \quad \phi_{12,t} = \begin{cases} 0.1 & t < 500 \\ -0.1 & t \geq 500 \end{cases}$$

$$\phi_{21,t} = \begin{cases} 0.4 & t < 500 \\ -0.4 & t \geq 500 \end{cases}, \quad \phi_{22,t} = \begin{cases} 0.5 & t < 500 \\ -0.5 & t \geq 500 \end{cases}$$

$$\mathbf{n}_t \sim \mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 0.1 & 0 \\ 0 & 0.3 \end{bmatrix}\right)$$

Estimates of TVAR coefficients (Gaussian case)





A realization from the ensemble:

Non-Gaussian Driving Process

- Non-stationary cross-correlated first order VAR(1):
- 100 realizations for ensemble averaging

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} \phi_{11}(t) & \phi_{12}(t) \\ \phi_{21}(t) & \phi_{22}(t) \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} n_{1,t} \\ n_{2,t} \end{bmatrix}$$

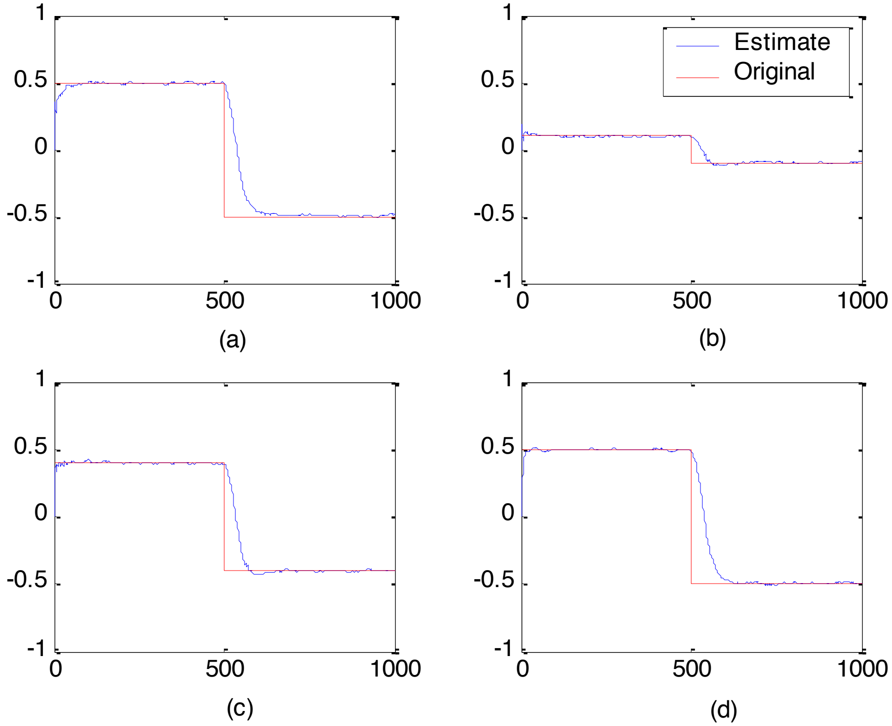
Cross-correlations through AR coefficient matrix

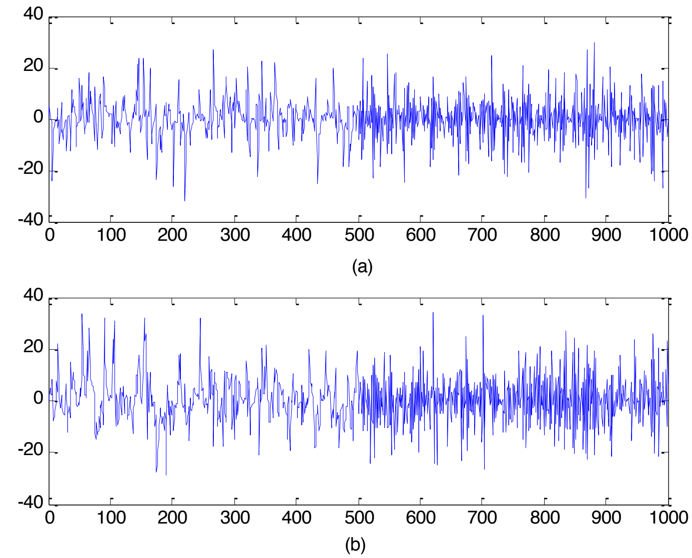
$$\phi_{11,t} = \begin{cases} 0.5 & t < 500 \\ -0.5 & t \geq 500 \end{cases}, \quad \phi_{12,t} = \begin{cases} 0.1 & t < 500 \\ -0.1 & t \geq 500 \end{cases}$$

$$\phi_{21,t} = \begin{cases} 0.4 & t < 500 \\ -0.4 & t \geq 500 \end{cases}, \quad \phi_{22,t} = \begin{cases} 0.5 & t < 500 \\ -0.5 & t \geq 500 \end{cases}$$

$$\mathbf{n}_t \sim 0.5\mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) + 0.5\mathrm{N}\left(\mathbf{0}, \begin{bmatrix} 100 & 0 \\ 0 & 100 \end{bmatrix}\right)$$

Estimates of TVAR coefficients (Non-Gaussian case)





A realization from the ensemble: