Stat 535 C - Statistical Computing & Monte Carlo Methods

Arnaud Doucet

Email: arnaud@cs.ubc.ca

1.1– Outline

- Hidden Markov Models & Nonlinear Non-Gaussian State-Space Models.
- Gibbs Sampling in Nonlinear Non-Gaussian models.

2.1– Finite Mixture of Distributions

- Let $X_n \in \{1, ..., K\}$ be some missing data such that

$$Y_n | X_n \sim g_{X_n}(y)$$

and

$$\Pr(X_n = k) = p_k,$$

then

$$Y_n \sim \sum_{k=1}^{K} p_k g_k(y).$$

- If $g_k(y) = \mathcal{N}\left(y; \mu_k, \sigma_k^2\right)$ then, by setting priors on $\left\{p_i, \mu_i, \sigma_i^2\right\}$, we have a Bayesian model.

2.1– Finite Mixture of Distributions

- Mixture models are ubiquitous but can only model i.i.d. data.

- In numerous applications, we have statistically dependent data.

- A potential extension consists of assuming that $\{X_n\}$ is a finite state-space Markov chain.

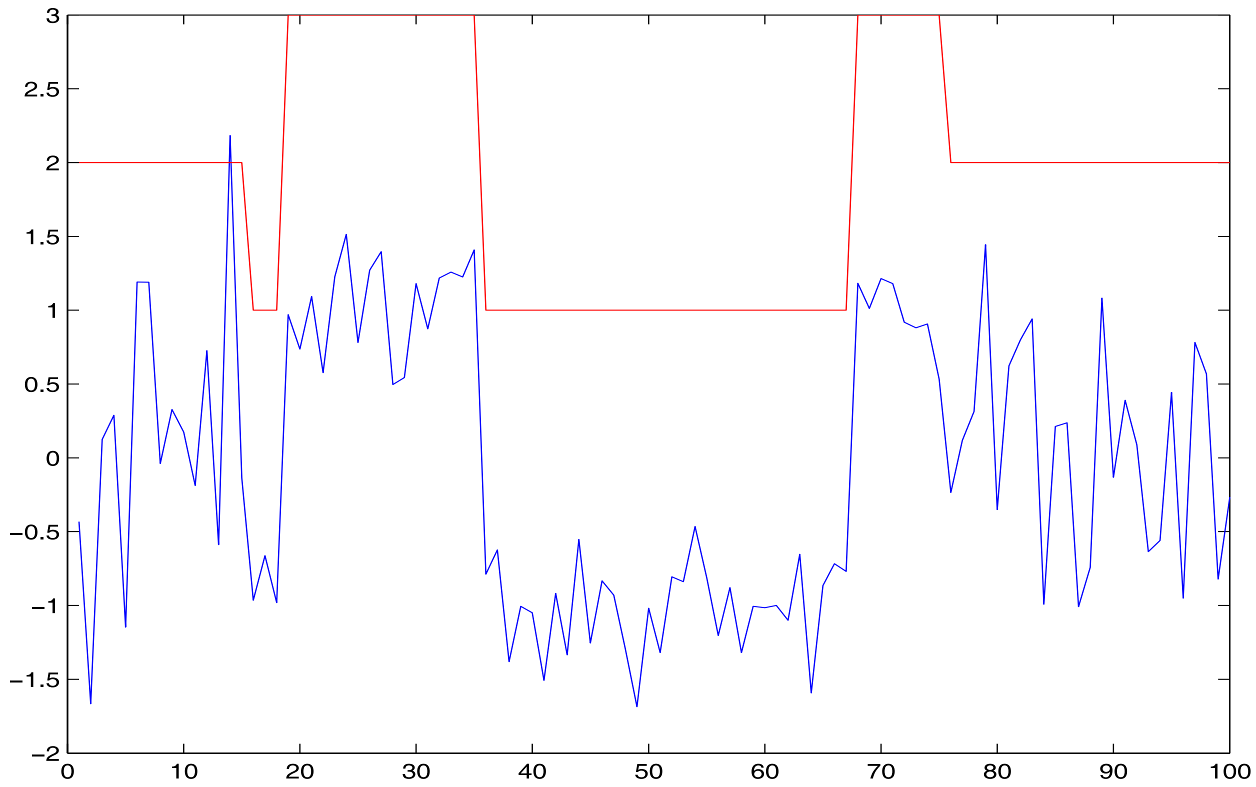
- We have say $\Pr(X_1 = i) = \mu_i$ and

$$\Pr(X_{n+1} = j | X_n = i) = p_{i,j}.$$

- In this case, the probability to stay in a given state is geometric.

- Simple model (over)used in speech processing, DNA sequence analysis, communications etc.

- Realization of 100 observations for $K = 3$, $\mu_1 = -1, \sigma_1^2 = 0.1$, $\mu_2 = 0, \sigma_2^2 = 1$, $\mu_3 = 1$, $\sigma_2^2 = 0.1$ with $p_{i,i} = 0.90$, $p_{i,j} = 0.05$ for $i \neq j$.



$\{X_n\}$ red line and $\{Y_n\}$ observations

• Given T observations $y_1, ..., y_T$ then the likelihood of the observations is given by

$$p\left(y_1, ..., y_T \mid \theta\right)$$

where θ includes all the unknown parameters.

• The likelihood can be computed exactly using a simple recursion. However, we limit ourselves first to the complete likelihood

$$p\left(y_{1:T}, x_{1:T} \mid \theta\right) = p\left(y_{1:T} \mid \theta, x_{1:T}\right) p\left(x_{1:T} \mid \theta\right)$$

where

$$p\left(y_{1:T} \mid \theta, x_{1:T}\right) = \prod_{n=1}^{T} p\left(y_n \mid \theta, x_n\right),$$

$$p\left(x_{1:T} \mid \theta\right) = p\left(x_1 \mid \theta\right) \prod_{n=2}^{T} p\left(x_n \mid \theta, x_{n-1}\right).$$

• Typically, one uses the EM algorithm to estimate the maximum likelihood estimate of the unknown parameter θ.

• Alternatively, given a prior distribution $p(\theta)$ on θ, then we can perform Bayesian inference and estimate

$$p(\theta, x_{1:T} | y_{1:T}) = \frac{p(y_{1:T} | \theta, x_{1:T})\, p(x_{1:T} | \theta)\, p(\theta)}{p(y_{1:T})}$$

• For mixture, there is no closed-form. Hence there is none for HMM!!!

2.5– Hidden Markov Models

- Next assignement consists of developing efficient MCMC algorithms for HMM.

- The introduction of $x_{1:T}$ allows to establish an efficient MCMC sampler.

- Similarly to mixtures, the posterior is exchangeable if the prior is exchangeable.

- Today, we look at models admitting a similar structure.

2.6– Linear Gaussian State-Space Models

• It is important to realize that this class of models can be significantly extended.

• In particular, the latent process $\{X_n\}$ does not have to be discrete-valued.

• A simple example correspond to the case where

$$\begin{aligned} X_n &= \alpha X_{n-1} + \sigma_v V_n,\ V_n \overset{\text{i.i.d.}}{\sim} N(0,1) \\ Y_n &= X_n + \sigma_w W_n,\ W_n \overset{\text{i.i.d.}}{\sim} N(0,1) \end{aligned}$$

2.6– Linear Gaussian State-Space Models

- Clearly, we are in the case where $\{X_n\}$ is a Markov process

$$\begin{aligned} X_n | X_{n-1} &\sim f_\theta (x_n | x_{n-1}), \\ Y_n | X_n &\sim g_\theta (y_n | x_n). \end{aligned}$$

where

$$\begin{aligned} f_\theta (x_n | x_{n-1}) &= \mathcal{N}\left(x_n; \alpha x_{n-1}, \sigma_v^2\right), \\ g_\theta (y_n | x_n) &= \mathcal{N}\left(x_n; \alpha x_{n-1}, \sigma_w^2\right). \end{aligned}$$

and $\theta = \left(\alpha, \sigma_v^2, \sigma_w^2\right)$.

2.7– Spline Smoothing using Linear Gaussian State-Space Model

- Suppose you have

$$Y_n = g(t_n) + W_n \text{ where } W_n \sim \mathcal{N}\left(0, \sigma^2\right)$$

with
$$\frac{d^2 g(t)}{dt^2} = \tau \frac{dB(t)}{dt} \text{ where } B(t) \text{ Wiener process}$$
with $B(0) = 0$ and $var(B(t)) = 1$.

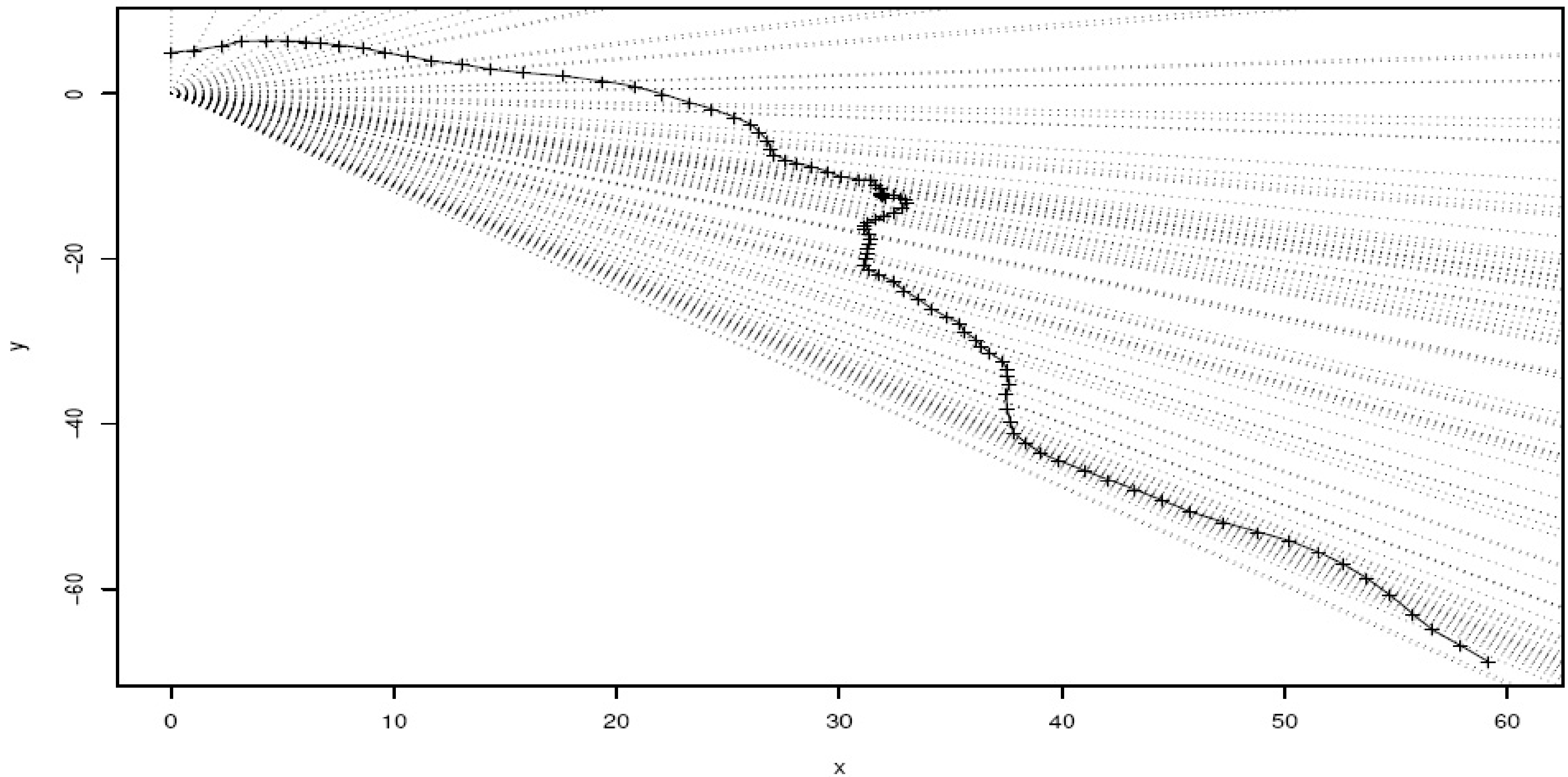
- With initial conditions such that $(g(t_1) \;\; dg(t_1)\, dt) \sim N(0, kI_2)$

$$Y_n = (1\ 0)\, X(t_n) + W_n,$$

$$X(t_n) = \begin{pmatrix} 1 & \delta_n \\ 0 & 1 \end{pmatrix} X(t_{n-1}) + V_n,\ V_n \sim \mathcal{N}\left(0, \begin{pmatrix} \delta_n^3/3 & \delta_n^2/2 \\ \delta_n^2/2 & \delta_n \end{pmatrix}\right)$$

where $\delta_n = t_n - t_{n-1}$.

2.8– Target tracking



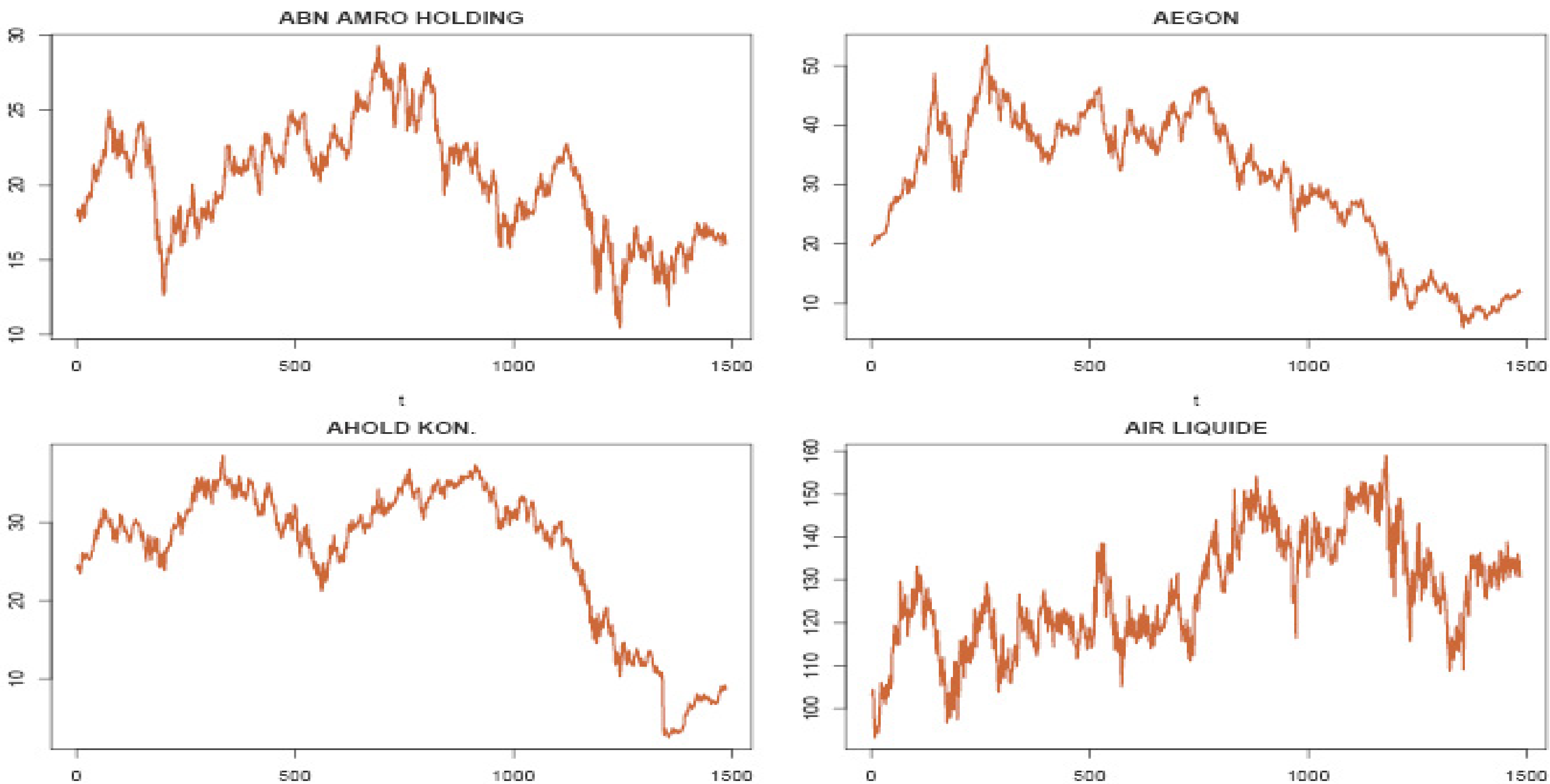
Bearings-only-tracking data

- Consider the coordinates of a target observed through a radar.

$$\begin{pmatrix} X_n^1 \\ \dot{X}_n^1 \\ X_n^2 \\ \dot{X}_n^2 \end{pmatrix} = \Delta \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_{n-1}^1 \\ \dot{X}_{n-1}^1 \\ X_{n-1}^2 \\ \dot{X}_{n-1}^2 \end{pmatrix} + V_n,\ V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \Sigma_v\right),$$

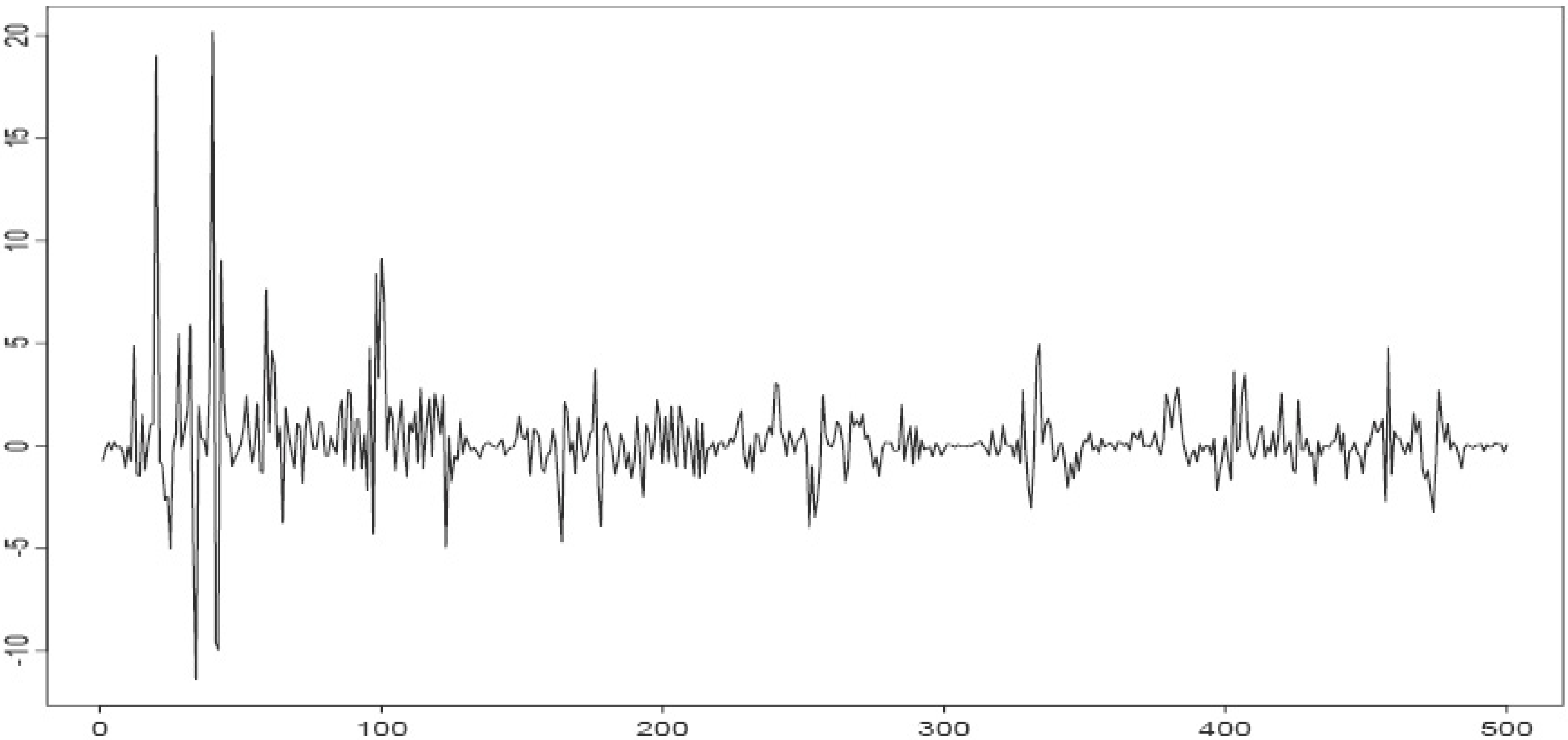
$$Y_n = \tan^{-1}\left(\frac{X_n^1}{X_n^2}\right) + W_n,\ W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right).$$

where the process $\{Y_n\}$ is observed but $\{X_n\}$ is unknown.



Four stock data

2.9– Financial Time Series



Log return of a stock

• Consider the log-return sequence of a stock then a popular model in financial econometrics is the stochastic volatility model

$$\begin{aligned} X_n &= \alpha X_{n-1} + \sigma V_n \text{ where } V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \\ Y_n &= \beta \exp(X_n/2)\, W_n \text{ where } W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \end{aligned}$$

where the process $\{Y_n\}$ is observed but $\{X_n\}$ and $\theta = (\alpha, \sigma, \beta)$ are unknown.

• We have

$$\begin{aligned} f_\theta(x_n | x_{n-1}) &= \mathcal{N}\left(x_n; \alpha x_{n-1}, \sigma_v^2\right), \\ g_\theta(y_n | x_n) &= \mathcal{N}\left(y_n; 0, \beta^2 \exp(x_n)\right). \end{aligned}$$

2.10– Summar: Nonlinear non-Gaussian State-Space Models

- Many real-world problems can be rewritten as

$$X_n|\,X_{n-1} \sim f_\theta\left(x_n|\,x_{n-1}\right),\ X_1 \sim \mu\left(x_1\right),$$

$$Y_n|\,X_n \sim g_\theta\left(y_n|\,x_n\right)$$

where $\theta \sim p\left(\theta\right)$.

- In a Bayesian framework, given $y_{1:T}$, we are interested in estimating the posterior

$$p\left(x_{1:T},\theta|\,y_{1:T}\right) \propto p\left(y_{1:T}|\,\theta,x_{1:T}\right) p\left(x_{1:T}|\,\theta\right) p\left(\theta\right)$$

where

$$p\left(y_{1:T}|\,\theta,x_{1:T}\right) = \prod_{n=1}^{T} g_\theta\left(y_n|\,x_n\right),$$

$$p\left(x_{1:T}|\,\theta\right) = \mu\left(x_1\right) \prod_{n=2}^{T} f_\theta\left(x_n|\,x_{n-1}\right).$$

2.11– Simple Gibbs Sampler for Linear Gaussian Model

- Assume you have

$$X_n = \alpha X_{n-1} + \sigma_v V_n, \; V_n \overset{\text{i.i.d.}}{\sim} N(0,1)$$

$$Y_n = X_n + \sigma_w W_n, \; W_n \overset{\text{i.i.d.}}{\sim} N(0,1)$$

where $\alpha \sim \mathcal{N}\left(0, \sigma_0^2\right)$, $\sigma_v^2 \sim \mathcal{IG}\left(\frac{\upsilon_0}{2}, \frac{\gamma_0}{2}\right)$ and $\sigma_w^2 \sim \mathcal{IG}\left(\frac{\upsilon_0}{2}, \frac{\gamma_0}{2}\right)$.

- Gibbs samplers based on

$$p\left(x_k | y_{1:T}, x_{-k}, \alpha, \sigma_v^2, \sigma_w^2\right), \; p\left(\sigma_v^2, \sigma_w^2 \middle| y_{1:T}, x_{1:T}, \alpha\right),$$

$$p\left(\alpha | y_{1:T}, x_{1:T}, \sigma_v^2, \sigma_w^2\right).$$

2.11– Simple Gibbs Sampler for Linear Gaussian Model

- We have for $1 < k < T$

$$
\begin{aligned}
p\left(x_k \middle| y_{1:T}, x_{-k}, \alpha, \sigma_v^2, \sigma_w^2\right) &\propto g\left(y_k \middle| x_k, \sigma_w^2\right) f\left(x_k \middle| x_{k-1}, \alpha, \sigma_v^2\right) f\left(x_{k+1} \middle| x_k, \alpha, \sigma_v^2\right) \\
&= \mathcal{N}\left(x_k; m_k, \sigma_k^2\right)
\end{aligned}
$$

where

$$
\begin{aligned}
m_k &= \sigma_k^2 \left(\frac{y_k^2}{\sigma_k^2} + \alpha \frac{x_{k+1} + x_{k-1}}{\sigma_v^2} \right), \\
\frac{1}{\sigma_k^2} &= \frac{1}{\sigma_w^2} + \frac{\alpha^2 + 1}{\sigma_v^2}.
\end{aligned}
$$

- We have

$$
p\left(\sigma_v^2, \sigma_w^2 \middle| y_{1:T}, x_{1:T}, \alpha\right) = p\left(\sigma_v^2 \middle| x_{1:T}, \alpha\right) p\left(\sigma_w^2 \middle| y_{1:T}, x_{1:T}\right)
$$

2.11– Simple Gibbs Sampler for Linear Gaussian Model

- We have

$$\begin{aligned}
p\left(\sigma_v^2\middle|\, x_{1:T}, \alpha\right) &\propto p\left(x_{1:T}\middle|\, \alpha, \sigma_v^2\right) p\left(\sigma_v^2\right) \\
&\propto \frac{1}{\sigma_v^{T-1}} \exp\left(-\frac{\sum_{n=2}^{T}\left(x_k - \alpha x_{k-1}\right)^2}{2\sigma_v^2}\right) \frac{1}{\sigma_v^{\upsilon_0}} \exp\left(-\frac{\gamma_0}{2\sigma_v^2}\right) \\
&= \mathcal{IG}\left(\sigma_v^2; \frac{\upsilon_0 + T - 1}{2}, \frac{\gamma_0 + \sum_{n=2}^{T}\left(x_k - \alpha x_{k-1}\right)^2}{2}\right).
\end{aligned}$$

2.11– Simple Gibbs Sampler for Linear Gaussian Model

- We have

$$
\begin{aligned}
p\left(\sigma_w^2 \middle| y_{1:T}, x_{1:T}\right) &\propto p\left(y_{1:T} \middle| x_{1:T}, \sigma_w^2\right) p\left(\sigma_w^2\right) \\
&\propto \frac{1}{\sigma_w^T} \exp\left(-\frac{\sum_{n=2}^{T}\left(y_k - x_k\right)^2}{2\sigma_w^2}\right) \frac{1}{\sigma_w^{\upsilon_0}} \exp\left(-\frac{\gamma_0}{2\sigma_w^2}\right) \\
&= \mathcal{IG}\left(\sigma_w^2; \frac{\upsilon_0 + T}{2}, \frac{\gamma_0 + \sum_{n=1}^{T}\left(y_k - x_k\right)^2}{2}\right).
\end{aligned}
$$

2.11– Simple Gibbs Sampler for Linear Gaussian Model

- Finally we have

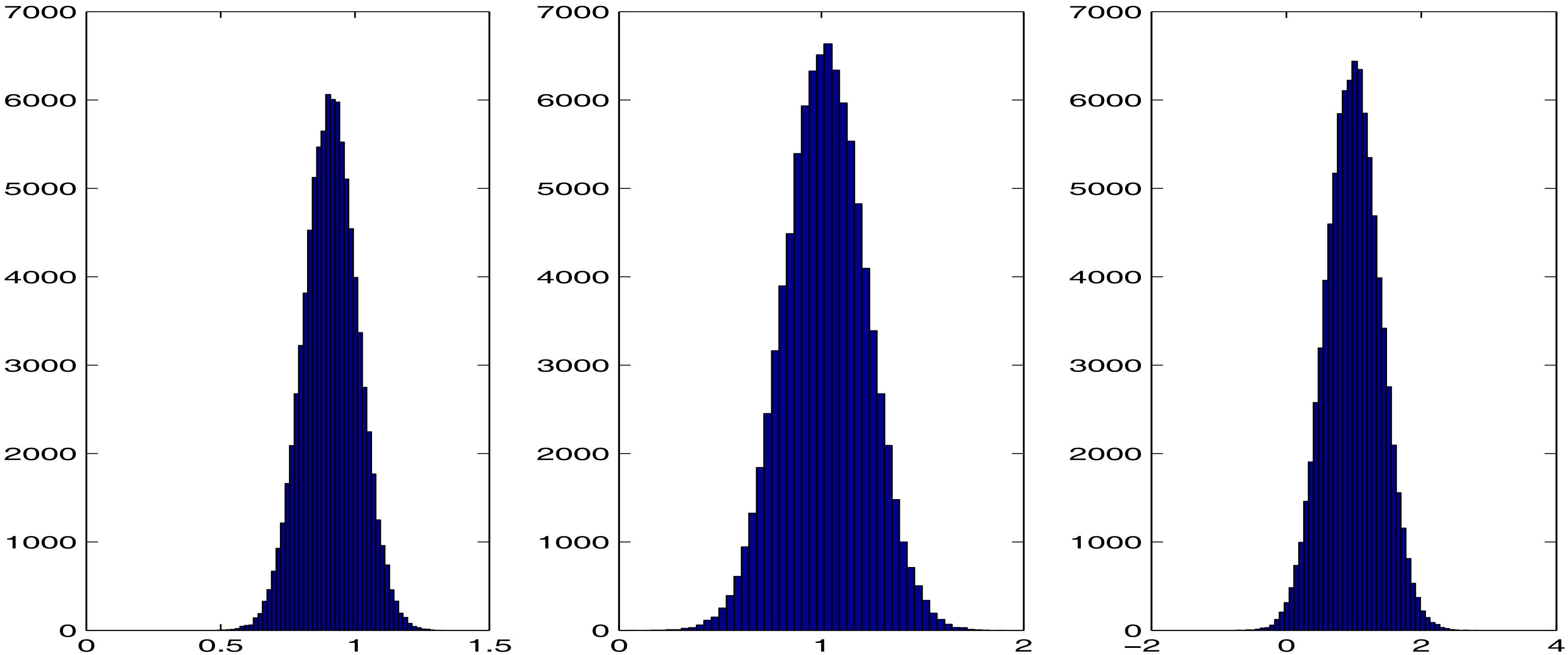
$$
\begin{aligned}
p\left(\alpha \mid y_{1:T}, x_{1:T}, \sigma_v^2, \sigma_w^2\right) &= p\left(\alpha \mid x_{1:T}, \sigma_v^2\right) \propto p\left(x_{1:T} \mid \alpha, \sigma_v^2\right) p(\alpha) \\
&\propto \frac{1}{\sigma_v^{T-1}} \exp\left(-\frac{\sum_{n=2}^{T}\left(x_k - \alpha x_{k-1}\right)^2}{2\sigma_v^2}\right) \exp\left(-\frac{\alpha^2}{2\sigma_0^2}\right) \\
&= \mathcal{N}\left(\alpha; m_\alpha, \sigma_\alpha^2\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\frac{1}{\sigma_\alpha^2} &= \frac{1}{\sigma_0^2} + \frac{\sum_{n=1}^{T-1} x_k^2}{\sigma_v^2}, \\
m_\alpha &= \sigma_\alpha^2 \left(\sum_{n=2}^{T} x_k x_{k-1}\right).
\end{aligned}
$$

2.11– Simple Gibbs Sampler for Linear Gaussian Model

• 100,000 samples after 10,000 burn in with $\alpha = 0.9$, $\sigma_w = 1$ and $\sigma_v = 1$ for $T = 100$.



Approximation of $p\left(\alpha \middle| y_{1:T}\right)$, $p\left(\sigma_w^2 \middle| y_{1:T}\right)$ and $p\left(\sigma_v^2 \middle| y_{1:T}\right)$.

2.12– Simple Gibbs Sampler for Bearings-only-Tracking

- We have

$$
\begin{aligned}
X_n &= AX_{n-1} + V_n,\ V_n \sim \mathcal{N}(0, \Sigma),\\
Y_n &= \tan^{-1}\left(\frac{X_n^1}{X_n^2}\right) + W_n,\ W_n \sim \mathcal{N}\left(0, \sigma^2\right)
\end{aligned}
$$

- Assume for sake of simplicity that only $x_{1:T}$ are unknown, we want to estimate

$$p\left(x_{1:T}\middle| y_{1:T}\right).$$

2.12– Simple Gibbs Sampler for Bearings-only-Tracking

- We sample from the full conditional distributions

$$\begin{aligned} p\left(x_k \mid y_{1:T}, x_{-k}\right) & \propto p\left(x_k \mid x_{-k}\right) g\left(y_k \mid x_k\right) \\ & \propto f\left(x_{k+1} \mid x_k\right) f\left(x_k \mid x_{k-1}\right) g\left(y_k \mid x_k\right). \end{aligned}$$

- We have

$$p\left(x_k \mid x_{-k}\right) \propto f\left(x_{k+1} \mid x_k\right) f\left(x_k \mid x_{k-1}\right) = \mathcal{N}\left(x_k; m_k, \Sigma_k\right)$$

where

$$\begin{aligned} \Sigma_k^{-1} & = \Sigma^{-1} + A^{\mathsf{T}} \Sigma^{-1} A, \\ m_k & = \Sigma_k \left(\Sigma^{-1} A x_{k-1} + A^{\mathsf{T}} \Sigma^{-1} x_{k+1}\right) \end{aligned}$$

2.12– Simple Gibbs Sampler for Bearings-only-Tracking

- To sample from

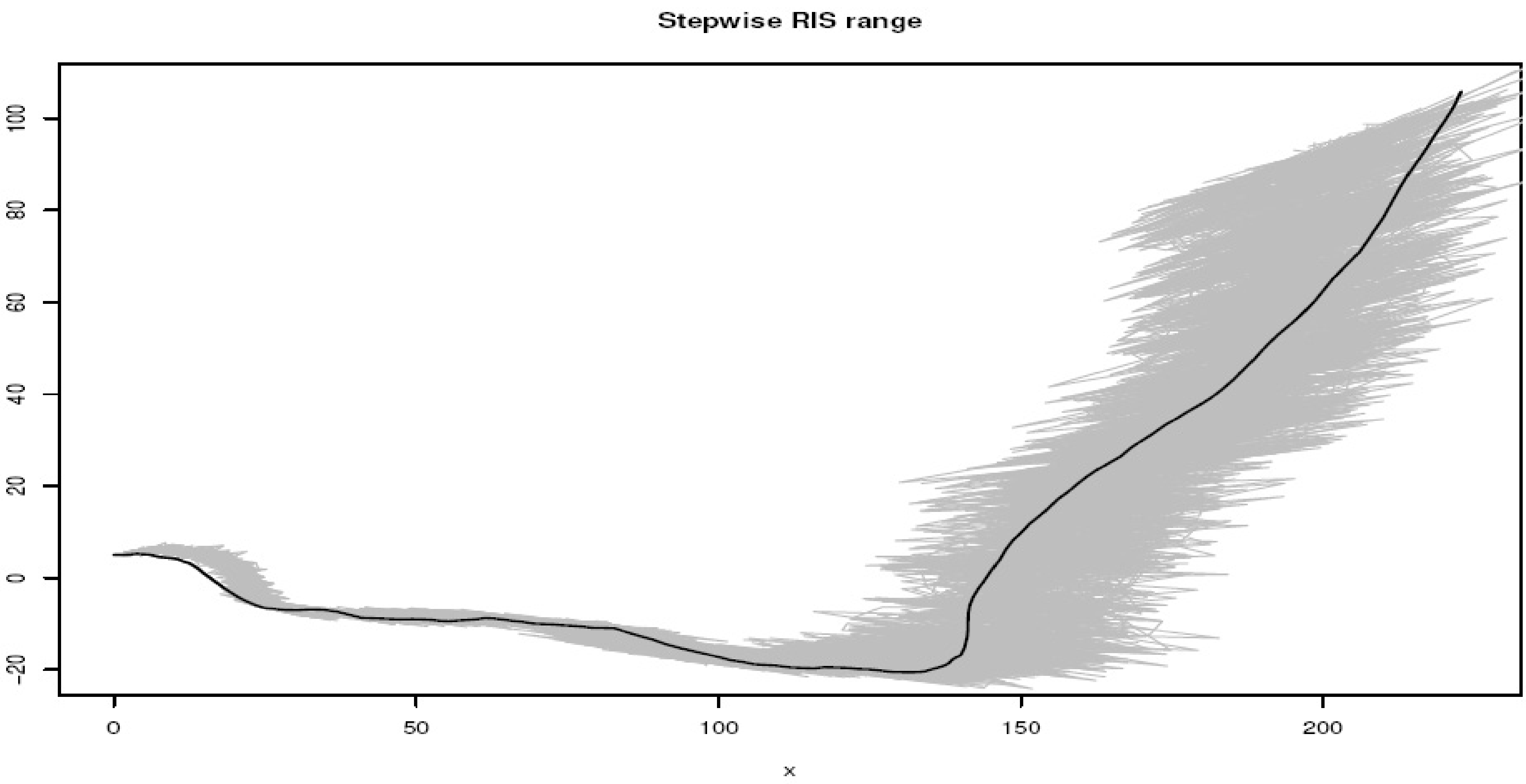
$$p\left(x_k \mid y_{1:T}, x_{-k}\right) \propto p\left(x_k \mid x_{-k}\right) g\left(y_k \mid x_k\right)$$

we can use rejection sampling as you can sample from $p\left(x_k \mid x_{-k}\right)$ and

$$g\left(y_k \mid x_k\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\left(y_k - \tan^{-1}\left(\frac{x_k^1}{x_k^2}\right)\right)^2 / \left(2\sigma^2\right)\right) \leq \frac{1}{\sqrt{2\pi}\sigma}.$$

- Gibbs sampling can be implemented even for non-linear models

2.12– Simple Gibbs Sampler for Bearings-only-Tracking



MCMC for state estimation using bearings-only-tracking data

2.13– Simple Gibbs Sampler for Stochastic Volatility Models

- We have

$$
\begin{aligned}
X_n &= \alpha X_{n-1} + \sigma V_n \text{ where } V_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \\
Y_n &= \beta \exp(X_n/2) W_n \text{ where } W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)
\end{aligned}
$$

- Prior model: $\alpha \sim \mathcal{U}(-1,1)$, $\sigma^2 \sim \mathcal{IG}\left(\frac{\upsilon_0}{2}, \frac{\gamma_0}{2}\right)$ and $\beta \sim \mathcal{IG}\left(\frac{\upsilon_0}{2}, \frac{\gamma_0}{2}\right)$.

2.13– Simple Gibbs Sampler for Stochastic Volatility Models

- We want to sample from

$$p\left(x_k | x_{-k}, y_{1:T}, \alpha, \sigma^2, \beta\right) \propto f\left(x_k | x_{k-1}, \alpha, \sigma^2\right) f\left(x_{k+1} | x_k, \alpha, \sigma^2\right) g\left(y_k | x_k, \beta\right)$$

where

$$\begin{aligned} p\left(x_k | x_{-k}, \alpha, \sigma^2\right) &\propto f\left(x_k | x_{k-1}, \alpha, \sigma^2\right) f\left(x_{k+1} | x_k, \alpha, \sigma^2\right) \\ &= \mathcal{N}\left(x_k; m_k = \frac{\alpha\left(x_{k-1} + x_{k+1}\right)}{1+\alpha^2}, \sigma_k^2 = \frac{\sigma^2}{1+\alpha^2}\right). \end{aligned}$$

- We have

$$\log g\left(y_k | x_k, \beta\right) \equiv -\tfrac{x_k}{2} - \tfrac{y_k^2}{2\beta^2}\exp\left(-x_k\right)$$

$$\leq -\tfrac{x_k}{2} - \tfrac{y_k^2}{2\beta^2}\left(\exp\left(-m_k\right)\left(1+m_k\right) - x_k \exp\left(-m_k\right)\right) \quad [\text{as } \exp(u) \geq 1+u]$$

$$= \log g^*\left(y_k | x_k, \beta\right)$$

2.13– Simple Gibbs Sampler for Stochastic Volatility Models

- We propose to sample from $p\left(x_k \middle| x_{k-1}, x_{k+1}, y_k, \alpha, \sigma^2, \beta \right)$ using rejection by sampling from where

$$\begin{aligned} q(x_k) &\propto p\left(x_k \middle| x_{-k}, \alpha, \sigma^2 \right) g^* \left(y_k \middle| x_k, \beta \right) \\ &= \mathcal{N}\left(x_k; m_k + \frac{\sigma_k^2}{2} \left[\frac{y_k^2}{\beta_2} \exp\left(-m_k^2\right) - 1 \right], \sigma_k^2 \right). \end{aligned}$$

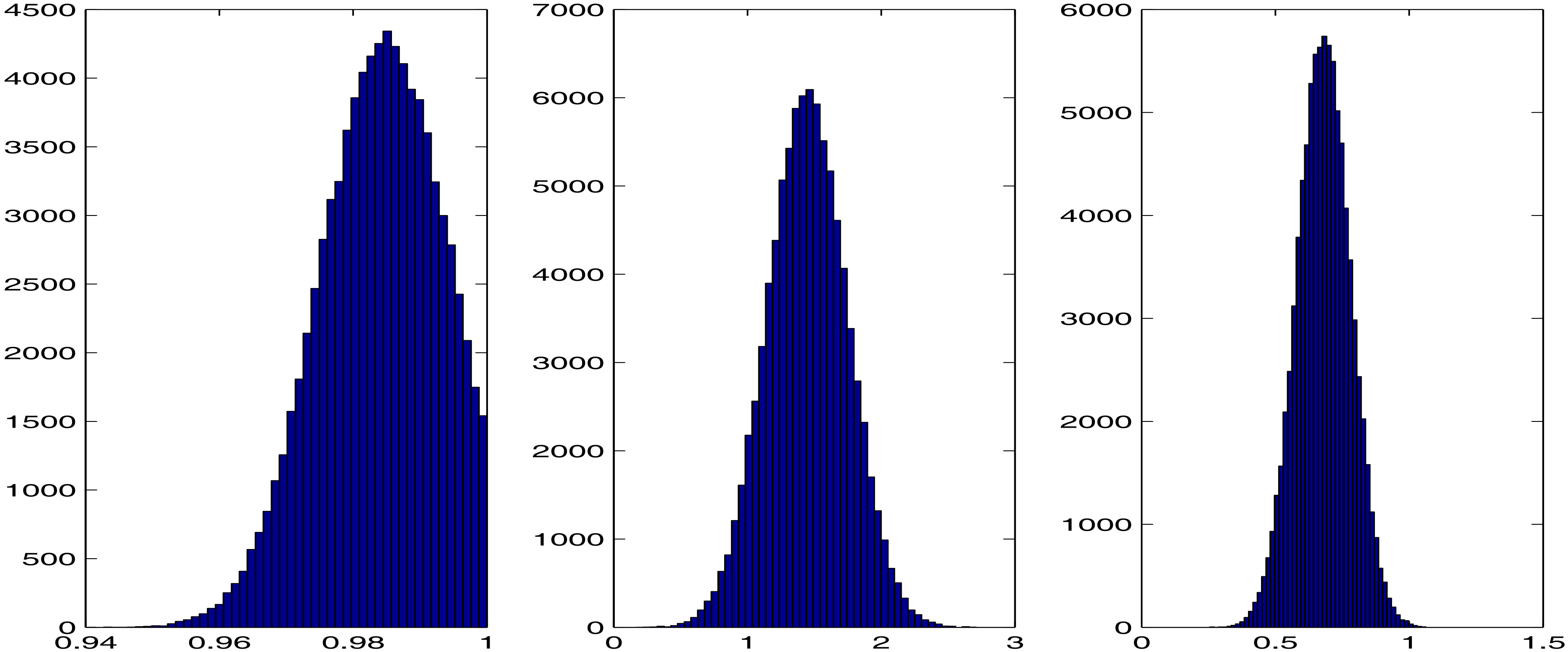
- Then we accept the proposal with probability

$$\frac{g\left(y_k \middle| x_k, \beta \right)}{g^*\left(y_k \middle| x_k, \beta \right)}.$$

- Update of the hypeparameters are straightforward.

2.13– Simple Gibbs Sampler for Stochastic Volatility Models

• UK Sterling/US dollar exhange rates from 1/10/81 to 28/6/85: 200,000 samples after 20,000 burn.



Approximations of $p\left(\alpha | y_{1:T}\right)$, $p\left(\sigma^2 \middle| y_{1:T}\right)$ and $p\left(\beta | y_{1:T}\right)$.

2.13– Simple Gibbs Sampler for Stochastic Volatility Models

- These Gibbs sampling algorithms are simple but once more they are not very effiicient as we sample typically $p\left(x_k \middle| y_{1:T}, x_{-k}, \theta \right)$ then $p\left(\theta \middle| y_{1:T}, x_{1:T} \right)$.

- We would like to be able to sample all the states variables jointly; i.e. sampling iteratively from $p\left(x_{1:T} \middle| y_{1:T}, \theta \right)$ then $p\left(\theta \middle| y_{1:T}, x_{1:T} \right)$.

- Generally sampling exactly from $p\left(x_{1:T} \middle| y_{1:T}, \theta \right)$ is impossible except for HMM (assignement) and linear Gaussian models.

2.13– Simple Gibbs Sampler for Stochastic Volatility Models

• All the models we have seen rely on the ability to sample from some full conditional distribution $\pi\left(\theta_k \middle| \theta_{-k}\right)$.

• Although it is possible in numerous models, there are also numerous models where one CANNOT do it.

$\Rightarrow$ In such cases, alternative methods relying on the Metropolis-Hastings algorithm have to be developed.