# CPSC 535
# Gibbs Sampling

AD

February 2007

# Finite State-Space Hidden Markov Models

- Mixture models cannot model dependent data; one straightforward extension consists of picking for $\{X_n\}$ a finite state-space Markov chain.

# Finite State-Space Hidden Markov Models

- Mixture models cannot model dependent data; one straightforward extension consists of picking for $\{X_n\}$ a finite state-space Markov chain.
- We have say $X_n \in \{1, ..., K\}$ with

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

but $\Pr(X_1 = i) = \mu_i$ and

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

# Finite State-Space Hidden Markov Models

- Mixture models cannot model dependent data; one straightforward extension consists of picking for $\{X_n\}$ a finite state-space Markov chain.
- We have say $X_n \in \{1, ..., K\}$ with

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

but $\Pr(X_1 = i) = \mu_i$ 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.

# Finite State-Space Hidden Markov Models

- Mixture models cannot model dependent data; one straightforward extension consists of picking for $\{X_n\}$ a finite state-space Markov chain.
- We have say $X_n \in \{1, ..., K\}$ with

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

but $\Pr(X_1 = i) = \mu_i$ 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.

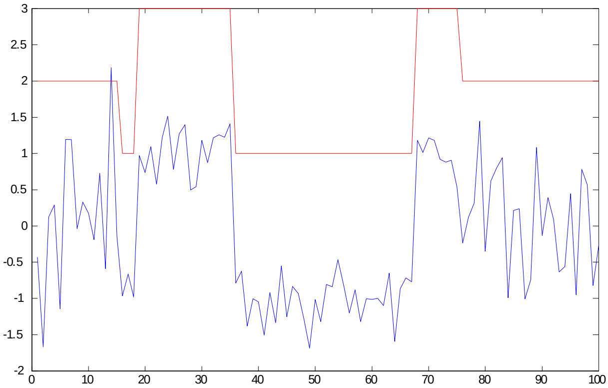


Figure: Realization of 100 observations for $K = 3$, $\mu_1 = -1, \sigma_1^2 = 0.1$, $\mu_2 = 0, \sigma_2^2 = 1$ $\mu_3 = 1$, $\sigma_2^2 = 0.1$ with $p_{i,i} = 0.90$, $p_{i,j} = 0.05$ for $i \neq j$. $\{X_n\}$ is displayed in red, $\{Y_n\}$ in blue.

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

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

where $\theta$ includes all the unknown parameters.

- 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 $\theta$ 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

$$\begin{aligned} p\left(y_{1:T} \mid \theta, x_{1:T}\right) &= \textstyle\prod_{n=1}^{T} p\left(y_n \mid \theta, x_n\right), \\ p\left(x_{1:T} \mid \theta\right) &= p\left(x_1 \mid \theta\right) \textstyle\prod_{n=2}^{T} p\left(x_n \mid \theta, x_{n-1}\right). \end{aligned}$$

- Typically, one uses the EM algorithm to estimate the maximum likelihood estimate of the unknown parameter $\theta$.

- Typically, one uses the EM algorithm to estimate the maximum likelihood estimate of the unknown parameter $\theta$.
- Alternatively, given a prior distribution $p(\theta)$ on $\theta$, then we can perform Bayesian inference and estimate

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

- Typically, one uses the EM algorithm to estimate the maximum likelihood estimate of the unknown parameter $\theta$.
- Alternatively, given a prior distribution $p(\theta)$ on $\theta$, 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. The Gibbs sampler can be implemented for this class of models by sampling iteratively from $p(\theta | y_{1:T}, x_{1:T})$ and $p(x_{1:T} | y_{1:T}, \theta)$.

# Extension to General State-Space HMM

- It is important to realize that this class of models can be significantly extended by taking a latent process $\{X_n\}$ which is not discrete-valued.

# Extension to General State-Space HMM

- It is important to realize that this class of models can be significantly extended by taking a latent process $\{X_n\}$ which is not discrete-valued.
- A simple example correspond to the case where

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

# Extension to General State-Space HMM

- It is important to realize that this class of models can be significantly extended by taking a latent process $\{X_n\}$ which is not discrete-valued.
- A simple example correspond to the case where

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

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

$$X_n | X_{n-1} \sim f_\theta(x_n | x_{n-1})$$

and $Y_n | X_n \sim g_\theta(y_n | x_n)$ where

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

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

- Suppose you have

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

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 $var(B(t)) = 1$.

- 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 $(g(t_1) \;\; dg(t_1)/dt) \sim \mathcal{N}(0, kI_2)$

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

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

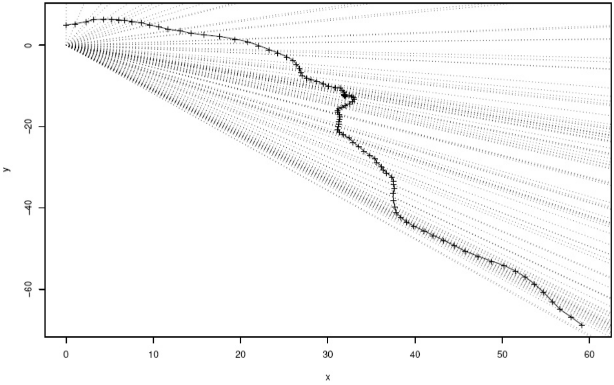


Figure: Bearings-only-tracking data

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

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

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

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

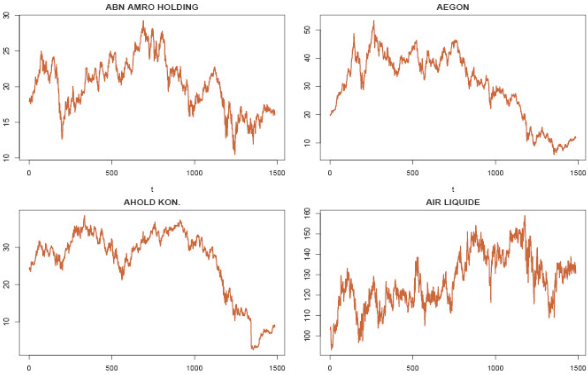


Figure: Four stock prices

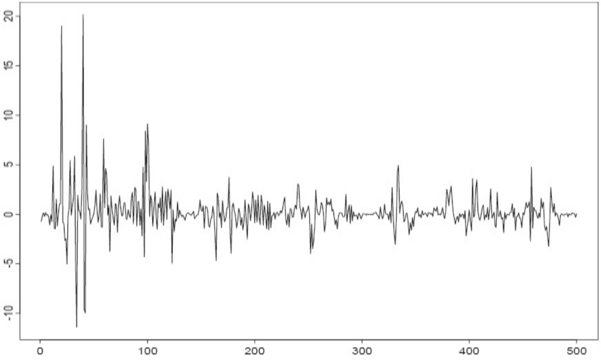


Figure: Log-return of a stock price

- 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.

- 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

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

- Many real-world problems can be rewritten as

$$\begin{aligned} X_n | X_{n-1} &\sim f_\theta(x_n | x_{n-1}), \; X_1 \sim \mu(x_1), \\ Y_n | X_n &\sim g_\theta(y_n | x_n) \end{aligned}$$

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

- Many real-world problems can be rewritten as

$$\begin{aligned} X_n | X_{n-1} &\sim f_\theta ( x_n | x_{n-1} ), \ X_1 \sim \mu (x_1), \\ Y_n | X_n &\sim g_\theta ( y_n | x_n ) \end{aligned}$$

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

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

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

where

$$\begin{aligned} p ( y_{1:T} | \theta, x_{1:T} ) &= \prod_{n=1}^{T} g_\theta ( y_n | x_n ), \\ p ( x_{1:T} | \theta ) &= \mu (x_1) \prod_{n=2}^{T} f_\theta ( x_n | x_{n-1} ). \end{aligned}$$

- Assume you have

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

where $X_1 \sim \mathcal{N}(0,1)$, $\alpha \sim \mathcal{N}\left(0, \sigma_0^2\right)$, $\sigma_v^2 \sim \mathcal{IG}\left(\frac{\upsilon_0}{2}, \frac{\gamma_0}{2}\right)$ and $\sigma_w^2 \sim \mathcal{IG}\left(\frac{\upsilon_0}{2}, \frac{\gamma_0}{2}\right)$.

- Assume you have

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

where $X_1 \sim \mathcal{N}(0,1)$, $\alpha \sim \mathcal{N}\left(0, \sigma_0^2\right)$, $\sigma_v^2 \sim \mathcal{IG}\left(\frac{\upsilon_0}{2}, \frac{\gamma_0}{2}\right)$ and $\sigma_w^2 \sim \mathcal{IG}\left(\frac{\upsilon_0}{2}, \frac{\gamma_0}{2}\right)$.

- Gibbs sampler based on

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

- We have for $1 < k < T$

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

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 for $1 < k < T$

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

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

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

- We have

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

- We have

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

- We have

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

- Finally we have

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

where

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

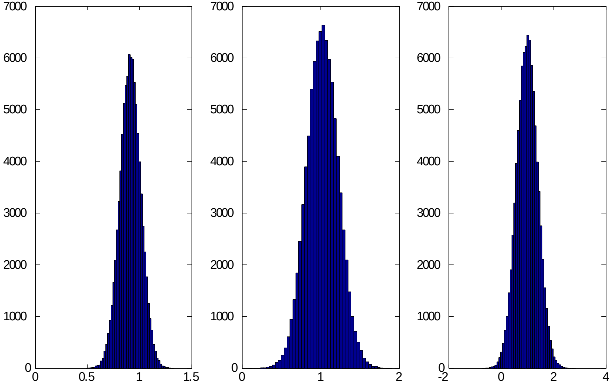


Figure: 100,000 samples after 10,000 burn in with $\alpha = 0.9$, $\sigma_w = 1$ and $\sigma_v = 1$ for $T = 100$. Approximations of $p\left(\alpha \mid y_{1:T}\right)$, $p\left(\sigma_w^2 \mid y_{1:T}\right)$ and $p\left(\sigma_v^2 \mid y_{1:T}\right)$

- We have

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

- 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

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

- We sample from the full conditional distributions

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

- 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

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

- 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

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

- To sample from

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

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

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

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

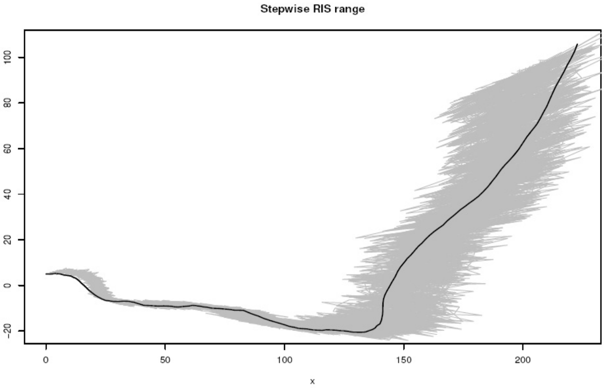


Figure: MCMC for state estimation using bearings-only-tracking data. Mean and credible intervals for $p\left(x_n \middle| Y_{1:n}\right)$.

- We have

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

$$Y_n = \beta \exp(X_n/2) W_n \text{ where } W_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

- 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 $\beta \sim \mathcal{IG}\left(\frac{\upsilon_0}{2}, \frac{\gamma_0}{2}\right)$.

- We want to sample from

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

where

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

- We want to sample from

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

where

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

- We have

$$\begin{aligned} &\log g\left(y_k \mid x_k, \beta\right) \equiv -\tfrac{x_k}{2} - \tfrac{y_k^2}{2\beta^2}\exp\left(-x_k\right) \\ &\leq -\tfrac{x_k}{2} - \tfrac{y_k^2}{2\beta^2}\left(\exp\left(-m_k\right)\left(1+m_k\right) - x_k \exp\left(-m_k\right)\right) \quad \left[\text{as } \exp\left(u\right) \geq 1+u\right. \\ &= \log g^*\left(y_k \mid x_k, \beta\right) \end{aligned}$$

- 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

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

- 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

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

- Then we accept the proposal with probability

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

- 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

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

- 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.

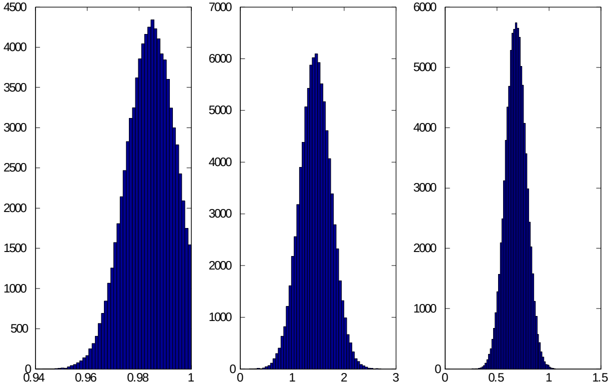


Figure: UK Sterling/US dollar exhange rates from 1/10/81 to 28/6/85: 200,000 samples after 20,000 burn-in. Approximations of $p(\alpha|\, y_{1:T})$, $p(\sigma^2 \big|\, y_{1:T})$ and $p(\beta|\, y_{1:T})$.

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

- These Gibbs sampling algorithms are simple but once more they are not very efficient 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 $p\left(\theta \middle| y_{1:T}, x_{1:T}\right)$.

- These Gibbs sampling algorithms are simple but once more they are not very efficient 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 and linear Gaussian models.

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

- 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.

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