Stat 535 C - Statistical Computing & Monte Carlo Methods

Lecture 15 - 9th March 2006

Arnaud Doucet

Email: arnaud@cs.ubc.ca

1.1– Outline

- More on the probit model
- Conditional prior proposals for time series.
- Advanced proposals for time series.

2.1– General hybrid algorithm

- Generally speaking, to sample from $\pi(\theta)$ where $\theta = (\theta_1, ..., \theta_p)$, we can use the following algorithm at iteration i.

- Iteration i; $i \geq 1$:

For $k = 1 : p$

- Sample $\theta_k^{(i)}$ using an MH step of proposal distribution

$$q_k\left(\left(\theta_{-k}^{(i)}, \theta_k^{(i-1)}\right), \theta_k'\right) \text{ and target } \pi\left(\theta_k \middle| \theta_{-k}^{(i)}\right).$$

where $\theta_{-k}^{(i)} = \left(\theta_1^{(i)}, ..., \theta_{k-1}^{(i)}, \theta_{k+1}^{(i-1)}, ..., \theta_p^{(i-1)}\right)$.

3.1– Probit model

- Banknotes data modelled using a probit regression model

$$\Pr\left(Y=1|\, x\right)=\Phi\left(x^{1}\beta_{1}+\ldots+x^{4}\beta_{4}\right)$$

where

$$\Phi\left(u\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{u}\exp\left(-\frac{v^{2}}{2}\right)dv$$

- For n data, the likelihood is then given by

$$f\left(y_{1:n}|\,\beta,x_{1:n}\right)=\prod_{i=1}^{n}\Phi\left(x_{i}^{T}\beta\right)^{y_{i}}\left(1-\Phi\left(x_{i}^{T}\beta\right)\right)^{1-y_{i}}.$$

3.1– Probit model

• One can use the MH algorithm where $q(\beta, \beta') = \mathcal{N}\left(\beta'; \beta, \tau^2 \widehat{\Sigma}\right)$ or use the Gibbs sampler by introducing additional latent variables.

• "Extended" model

$$Z_i \sim \mathcal{N}\left(x_i^{\mathrm{T}} \beta, 1\right), \; Y_i = \begin{cases} 1 & \text{if } Z_i > 0 \\ 0 & \text{otherwise.} \end{cases}$$

• We are now going to sample from $\pi\left(\beta, z_{1:n} \middle| x_{1:n}, y_{1:n}\right)$ instead of $\pi\left(\beta \middle| x_{1:n}, y_{1:n}\right)$

3.2– Gibbs Sampler for Probit model

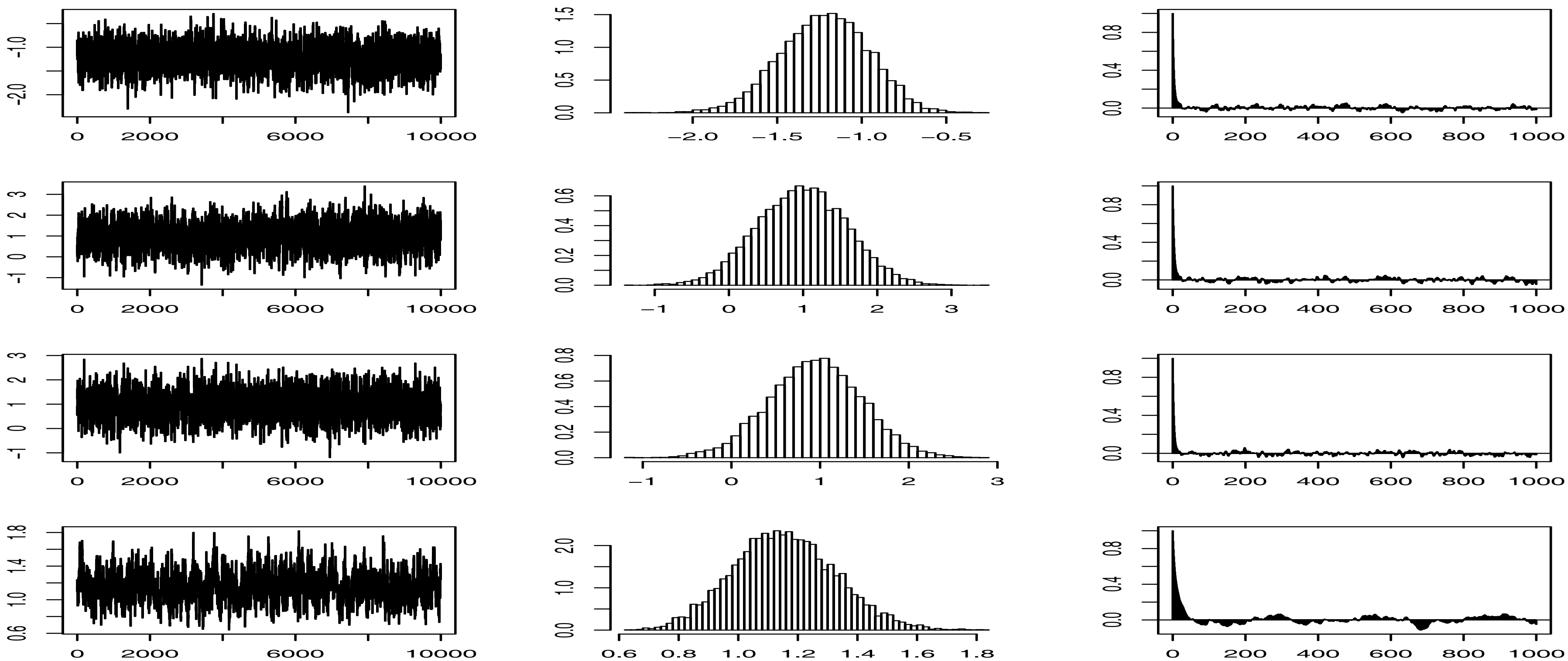
- The full conditional distributions are simple

$$\begin{aligned}
\pi\left(\beta \middle| y_{1:n}, x_{1:n}, z_{1:n}\right) &= \pi\left(\beta \middle| x_{1:n}, z_{1:n}\right) \text{ (standard Gaussian!)}, \\
\pi\left(z_{1:n} \middle| y_{1:n}, x_{1:n}, \beta\right) &= \prod_{i=1}^{n} \pi\left(z_k \middle| y_k, x_k, \beta\right)
\end{aligned}$$

where

$$z_k \mid y_k, x_k, \beta \sim \begin{cases} \mathcal{N}_+\left(x_k^{\mathrm{T}}\beta, 1\right) & \text{if } y_k = 1 \\ \mathcal{N}_-\left(x_k^{\mathrm{T}}\beta, 1\right) & \text{if } y_k = 0. \end{cases}$$

3.2– Gibbs Sampler for Probit model



Traces (left), Histograms (middle) and Autocorrelations (right) for $\left(\beta_1^{(i)}, .., \beta_4^{(i)}\right)$.

3.2– Gibbs Sampler for Probit model

- The results obtained through Gibbs are very similar to MH.

- We can also adopt an Zellner's type prior and obtain very similar results.

- Very similar were also obtained using a logistic fonction using the MH (Gibbs is feasible but more difficult).

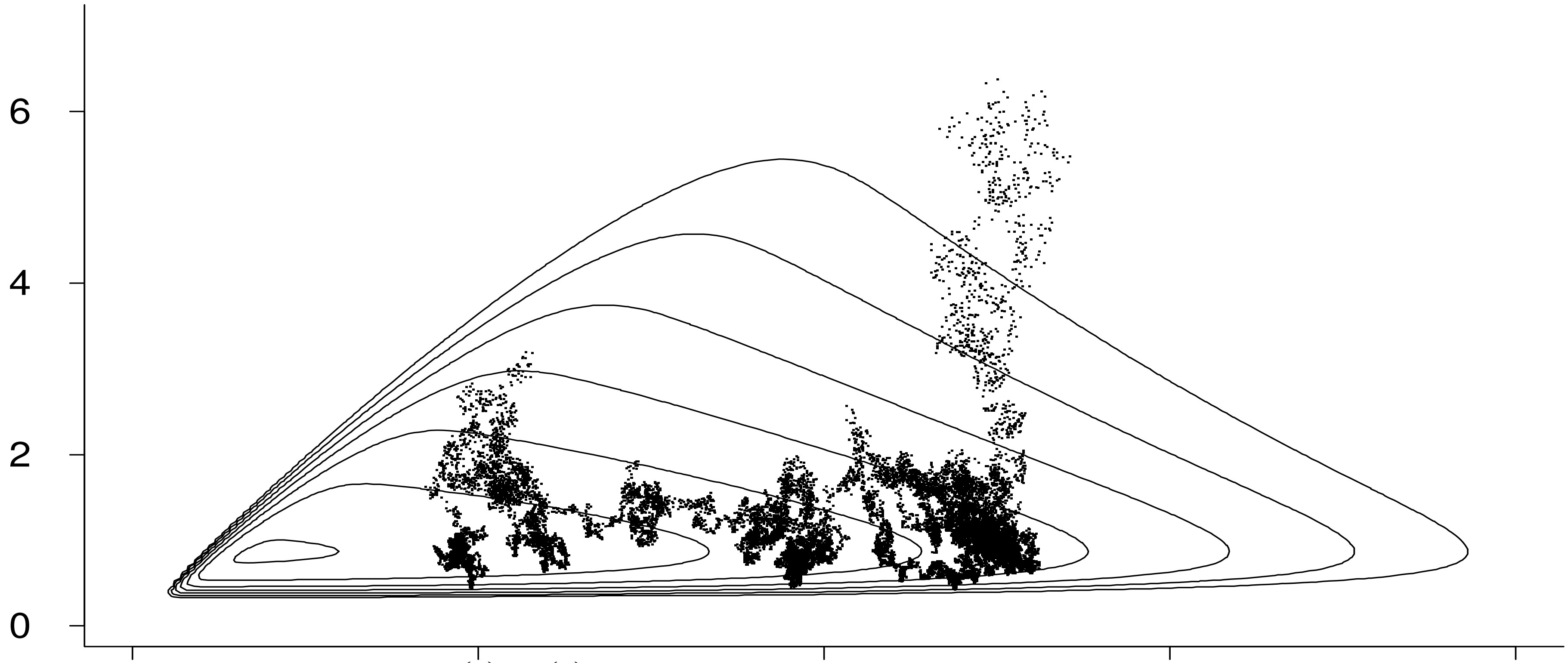
3.3– Gibbs sampling and Hybrid algorithm for Probit Regression

- Although the introduction of latent variables can be attractive, it can be also very inefficient.
- It is not because you can use the Gibbs sampler that everything works well!
- Consider the following simple generalization of the previous model

$$Z_i \sim \mathcal{N}\left(x_i\beta, \sigma^2\right),\ Y_i = \begin{cases} 1 & \text{if } Z_i > 0 \\ 0 & \text{otherwise.} \end{cases}$$

- We complete the model by $\sigma^2 \sim \mathcal{IG}\,(1.5, 1.5)$ and $\beta|\,\sigma^2 \sim \mathcal{N}\,(0, 100)$.

3.3– Gibbs sampling and Hybrid algorithm for Probit Regression



Samples of $\left(\beta^{(i)}, \sigma^{(i)}\right)$ obtained by the Gibbs sampler plotted with some contours of the posterior.

- Not only the data Z_i and $\left(\beta, \sigma^2\right)$ are very correlated but we have

$$\Pr\left(Y_i = 1 | \, x_i, \beta, \sigma^2\right) = \Phi\left(\frac{x_i \beta}{\sigma}\right)$$

- The likelihood only depends on β/σ and the parameters β and σ are not identifiable.

3.3– Gibbs sampling and Hybrid algorithm for Probit Regression

- One way to improve the mixing consists of using an additional MH step

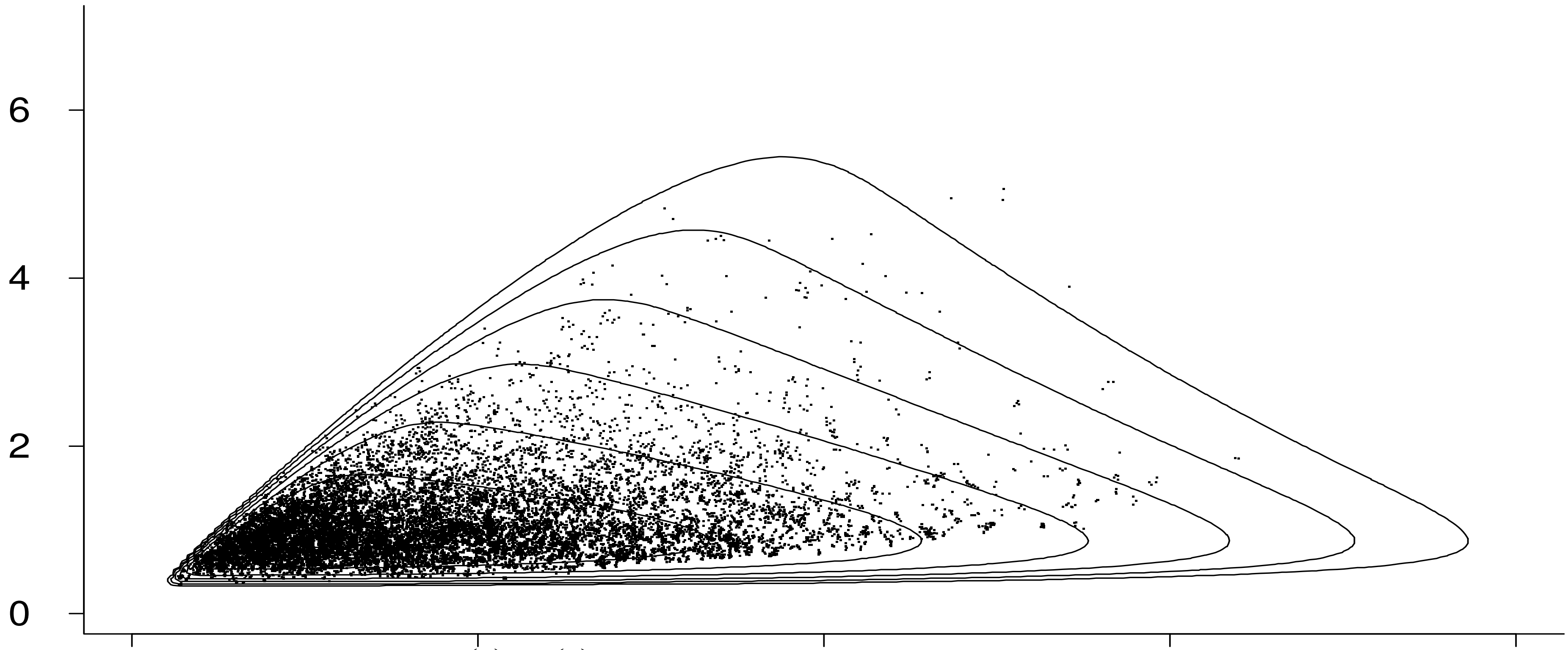
that proposes to randomly rescale the current value.

- We use a proposal distribution such that

$$(\beta', \sigma') = \lambda\,(\beta, \sigma) \;\text{ with } \lambda \sim \mathcal{E}\text{xp}\,(1)$$

that proposes to randomly rescale the current value.

3.3– Gibbs sampling and Hybrid algorithm for Probit Regression



Samples of $\left(\beta^{(i)}, \sigma^{(i)}\right)$ obtained by the Gibbs sampler+MH step plotted with some contours of the posterior.

4.1– Back to Hidden Markov Models

- Consider the following hidden Markov model

$$X_k | (X_{k-1} = x_{k-1}) \sim f_\theta(\cdot | x_{k-1}), \; X_1 \sim \mu$$

$$Y_n | (X_k = x_k) \sim g_\theta(\cdot | x_k),$$

and we set a prior $\pi(\theta)$ on the unknown hyperparameters θ.

- Given n data, we are interested in the joint posterior

$$\pi(\theta, x_{1:n} | y_{1:n}).$$

- There is no closed-form expression for this joint distribution even in the model is linear Gaussian or for finite state-space model.

4.1– Back to Hidden Markov Models

• In previous lectures, we propose sampling from $\pi(\theta, x_{1:n} | y_{1:n})$ using the Gibbs sampler where the variables are updated according to

$$X_k \sim \pi(x_k | y_{1:n}, x_{-k}, \theta)$$

with for $2 < k < n$

$$\begin{aligned}
\pi(x_k | y_{1:n}, x_{-k}, \theta) &\propto \pi(x_{1:n}, y_{1:n}, \theta) \\
&\propto \underbrace{\pi(\theta)\,\mu(x_1)\prod_{i=2}^{n} f_\theta(x_i | x_{i-1})}_{\text{prior}}\underbrace{\prod_{i=1}^{n} g_\theta(y_i | x_i)}_{\text{likelihood}} \\
&\propto f_\theta(x_k | x_{k-1})\, f_\theta(x_{k+1} | x_k)\, g_\theta(y_k | x_k)
\end{aligned}$$

and $\theta \sim \pi(\theta | y_{1:n}, x_{1:n})$ (or by subblocks).

4.1– Back to Hidden Markov Models

• It is often possible to implement the Gibbs sampler even if this can be expensive; e.g. if you use Accept/Reject to sample from $\pi\left(x_k \mid y_{1:n}, x_{-k}, \theta\right)$ using the proposal $\pi\left(x_k \mid x_{-k}, \theta\right) \propto f_\theta\left(x_k \mid x_{k-1}\right) f_\theta\left(x_{k+1} \mid x_k\right)$.

• Even if it is possible to implement the Gibbs sampler, one can expect a very slow convergence of the algorithm is successive variables are highly correlated.

• Indeed, as you update x_k with x_{k-1} and x_{k+1} being fixed, then you cannot move much into the space.

4.2– Illustrative Example

- Consider the very simple case where $\theta = \left(\sigma_v^2, \sigma_w^2\right)$

$$
\begin{aligned}
X_k &= X_{k-1} + V_k \text{ where } V_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_v^2\right), \\
Y_k &= X_k + W_k \text{ where } W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma_w^2\right)
\end{aligned}
$$

then we have

$$
\begin{aligned}
\pi\left(x_k \mid x_{-k}, \theta\right) &\propto f_\theta\left(x_k \mid x_{k-1}\right) f_\theta\left(x_{k+1} \mid x_k\right) \\
&= \mathcal{N}\left(x_k; \frac{x_{k-1} + x_{k+1}}{2}, \frac{\sigma_v^2}{2}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
\pi\left(x_k \mid y_{1:n}, x_{-k}, \theta\right) &\propto \pi\left(x_k \mid x_{-k}, \theta\right) g_\theta\left(y_k \mid x_k\right) \\
&= \mathcal{N}\left(x_k; \frac{\sigma_v^2 \sigma_w^2}{\sigma_v^2 + 2\sigma_w^2}\left(\frac{x_{k-1} + x_{k+1}}{\sigma_v^2} + \frac{y_k}{\sigma_w^2}\right), \frac{\sigma_v^2 \sigma_w^2}{\sigma_v^2 + 2\sigma_w^2}\right)
\end{aligned}
$$

4.2– Illustrative Example

- Assume for the time being that instead of sampling from $\pi\left(x_k | \, y_{1:n}, x_{-k}, \theta\right)$ directly, we use rejection sampling with $\pi\left(x_k | \, x_{-k}, \theta\right)$ as a proposal distribution.

- In this case we have to bound

$$g_\theta\left(y_k | \, x_k\right) = \frac{1}{\sqrt{2\pi}\sigma_w} \exp\left(-\frac{\left(y_k - x_k\right)^2}{2\sigma_w^2} \right) \leq \frac{1}{\sqrt{2\pi}\sigma_w}.$$

- We accept each proposal $X^* \sim \pi\left(x_k | \, x_{-k}, \theta\right)$ with probability $\exp\left(-\frac{(y_k - X^*)^2}{2\sigma_w^2} \right)$, so the (unconditional) acceptance probability is given by

$$\int \pi\left(x_k | \, x_{-k}, \theta\right) \exp\left(-\frac{\left(y_k - x_k\right)^2}{2\sigma_w^2} \right) dx_k = \frac{\sigma_w \exp\left(-\frac{1}{2}\left(y_k^2/\sigma_w^2 - \left(x_{k-1} + x_{k+1}\right)^2/\sigma_v^2 \right.\right.}{\sqrt{\sigma_v^2 + 2\sigma_w^2}}$$

4.3– Block sampling strategies

• To improve the algorithm, we would like to be able to sample a whole block of variables simultaneously; i.e. being able to sample for $1 < k < k + L < n$ from

$$\begin{aligned}\pi\left(x_{k:k+L}\middle|\, y_{1:n}, x_{-(k:k+L)}, \theta\right) &= \pi\left(x_{k:k+L}\middle|\, y_{k:k+L}, x_{k-1}, x_{k+L+1}, \theta\right) \\ &\propto \prod_{i=k}^{k+L+1} f_\theta\left(x_i\middle|\, x_{i-1}\right) \prod_{i=k}^{k+L} g_\theta\left(y_i\middle|\, x_i\right).\end{aligned}$$

• In this case, it is typically impossible to sample from $\pi\left(x_{k:k+L}\middle|\, y_{1:n}, x_{-(k:k+L)}, \theta\right)$ exactly as L is large, say 5 or 10.

• We are propose to use a MH step of invariant distribution $\pi\left(x_{k:k+L}\middle|\, y_{1:n}, x_{-(k:k+L)}, \theta\right)$ instead, hence we need to build a proposal distribution $q\left(\left(x_{1:n}, \theta\right), x'_{k:k+L}\right)$.

4.4– Conditional prior proposals

• We first propose to use the conditional prior

$$q\left((x_{1:n},\theta),x'_{k:k+L}\right) = \pi\left(x_{k:k+L}\middle|\, x_{-(k:k+L)},\theta\right) = \pi\left(x_{k:k+L}\middle|\, x_{k-1},x_{k+L+1},\theta\right)$$

$$\propto \prod_{i=k}^{k+L+1} f_\theta\left(x_i|\, x_{i-1}\right).$$

• In this case, the candidate $X'_{k:k+L} \sim \pi\left(x_{k:k+L}|\, x_{k-1},x_{k+L+1},\theta\right)$ is accepted with probability

$$\min\left(1,\frac{\pi\left(x'_{k:k+L}\middle|\, y_{k:k+L},x_{k-1},x_{k+L+1},\theta\right)\pi\left(x_{k:k+L}|\, x_{k-1},x_{k+L+1},\theta\right)}{\pi\left(x_{k:k+L}|\, y_{k:k+L},x_{k-1},x_{k+L+1},\theta\right)\pi\left(x'_{k:k+L}\middle|\, x_{k-1},x_{k+L+1},\theta\right)}\right)$$

$$= \min\left(1,\frac{\prod_{i=k}^{k+L} g_\theta\left(y_i|\, x'_i\right)}{\prod_{i=k}^{k+L} g_\theta\left(y_i|\, x_i\right)}\right).$$

• Simple but one cannot expect it to be too efficient when the observations are very informative compared to the prior.

4.5– Illustrative example

- Consider the case where

$$X_k = AX_{k-1} + BV_k, \text{ where } V_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, I).$$

- Particular cases include

$$X_k = X_{k-1} + \sigma V_k, \text{ where } V_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1),$$

$$X_k = \begin{pmatrix} \alpha_k \\ \alpha_{k-1} \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} X_{k-1} + \begin{pmatrix} \sigma \\ 0 \end{pmatrix} V_k, \text{ where } V_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1).$$

• In this case, it is simple to see that $\pi\left(x_{k:k+L}\middle|\, x_{k-1}, x_{k+1}, \theta\right)$ is a Gaussian distribution.

• In (Knorr-Held, 1999), one samples from this distribution by computing directly the parameters of this joint distribution: complexity $O\left(L^2\right)$.

• We can derive a simpler method of complexity $O\left(L\right)$ based on the following decomposition (omitting θ in the notation)

$$\begin{aligned}\pi\left(x_{k:k+L}\middle|\, x_{k-1}, x_{k+L+1}\right) &= \prod_{i=k}^{k+L-1} \pi\left(x_i\middle|\, x_{k-1}, x_{k+L+1}, x_{i+1}\right). \\ &= \prod_{i=k}^{k+L-1} \pi\left(x_i\middle|\, x_{k-1}, x_{i+1}\right)\end{aligned}$$

- Moreover it is easy to establish the expression for $\pi\left(x_i \mid x_{k-1}, x_{i+1}\right)$

$$\pi\left(x_i \mid x_{k-1}, x_{i+1}\right) \propto \pi\left(x_i \mid x_{k-1}\right) f\left(x_{i+1} \mid x_i\right)$$

as

$$\pi\left(x_i \mid x_{k-1}\right) = \int \pi\left(x_{k:i} \mid x_{k-1}\right) dx_{k:i-1} = \mathcal{N}\left(x_k; \mu_i\left(x_{k-1}\right), \Sigma_i\right)$$

with, for $X_n = AX_{n-1} + BV_n$, $\mu_{k-1}\left(x_{k-1}\right) = x_{k-1}$, $\Sigma_{k-1} = 0$ and for $i \geq k$

$$\begin{aligned} \mu_i\left(x_{k-1}\right) &= A\mu_{i-1}\left(x_{k-1}\right), \\ \Sigma_i &= A\Sigma_{i-1}A^{\mathrm{T}} + \Sigma \text{ with } \Sigma = BB^{\mathrm{T}}. \end{aligned}$$

- To obtain $\pi\left(x_i \mid x_{k-1}, x_{i+1}\right)$, we combine the prior $\pi\left(x_i \mid x_{k-1}\right)$ with the "likelihood" $f\left(x_{i+1} \mid x_i\right)$.

• We have $\pi\left(x_i \mid x_{k-1}\right) = \mathcal{N}\left(x_k; \mu_i\left(x_{k-1}\right), \Sigma_i\right)$ and $f\left(x_{i+1} \mid x_i\right) = \mathcal{N}\left(x_{i+1}; Ax_i, \Sigma\right)$ then

$$\pi\left(x_i \mid x_{k-1}, x_{i+1}\right) = \mathcal{N}\left(x_i; \mu_i\left(x_{k-1}, x_{i+1}\right), \widetilde{\Sigma}_i\right)$$

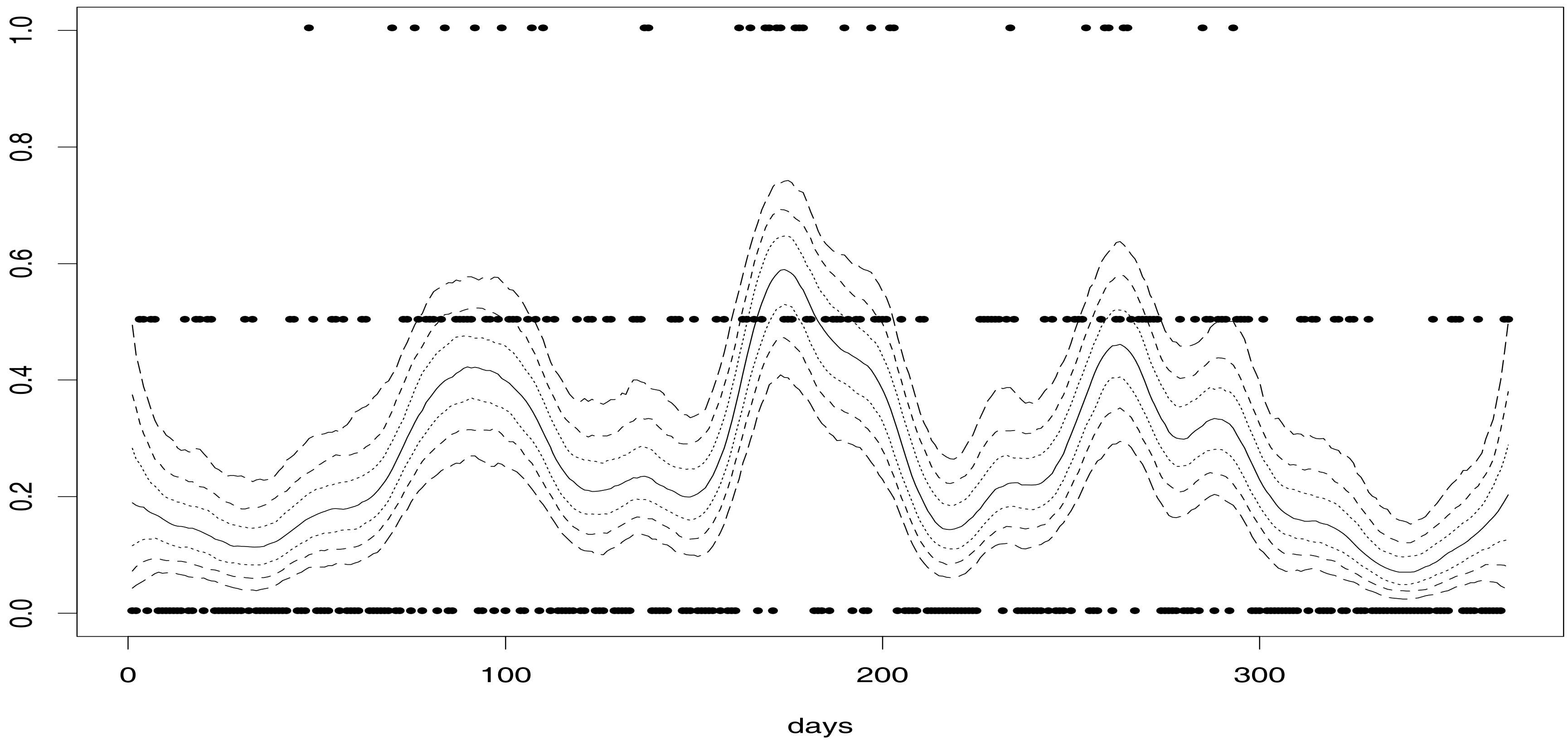
where

$$\begin{aligned} \widetilde{\Sigma}_i &= \left(\Sigma_i^{-1} + A^{\mathrm{T}}\Sigma^{-1}A\right)^{-1}, \\ \mu_i\left(x_{k-1}, x_{k+L+1}\right) &= \widetilde{\Sigma}_i\left(A^{\mathrm{T}}\Sigma^{-1}x_{i+1} + \Sigma_i^{-1}\mu_i\left(x_{k-1}\right)\right). \end{aligned}$$

• To sample a realization of $\pi\left(x_{k:k+L} \mid x_{k-1}, x_{k+L+1}\right)$, first compute $\mu_i\left(x_{k-1}\right), \Sigma_i$ for $i = k, ..., k+L$ using a forward recursion. Then sample backward

$$X_{k+L} \sim \pi\left(\cdot \mid x_{k-1}, x_{k+L+1}\right),\; X_{k+L-1} \sim \pi\left(\cdot \mid x_{k-1}, X_{k+L}\right), ...,\; X_k \sim \pi\left(\cdot \mid x_{k-1}, X_{k+1}\right)$$

Number of occurences of rainfall in Tokyo for each day during 1983-1984 reproduced as relative frequencies between 0, 0.5 and 1 ($n = 366$)



- Consider the following model

$$X_k = \begin{pmatrix} \alpha_k \\ \\ \alpha_{k-1} \end{pmatrix} = \begin{pmatrix} 2 & \text{-1} \\ \\ 1 & 0 \end{pmatrix} X_{k-1} + \begin{pmatrix} \sigma \\ \\ 0 \end{pmatrix} V_k, \text{ where } V_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$$

and

$$Y_k | X_k \sim \begin{cases} B(2, \pi_k) & k \neq 60, \\ \\ B(1, \pi_k) & k = 60 \text{ (February 29)} \end{cases}, \text{ where } \pi_k = \frac{\exp(\alpha_k)}{1 + \exp(\alpha_k)}.$$

- We also use for $\sigma^2 \sim \mathcal{IG}\left(\frac{\nu_0}{2}, \frac{\gamma_0}{2}\right)$.

• We use the block sampling strategies discussed before where candidates are sampled according to $\pi\left(x_{k+1:k+L}\middle|\, x_{k-1}, x_{k+L+1}\right)$ and accepted with proba

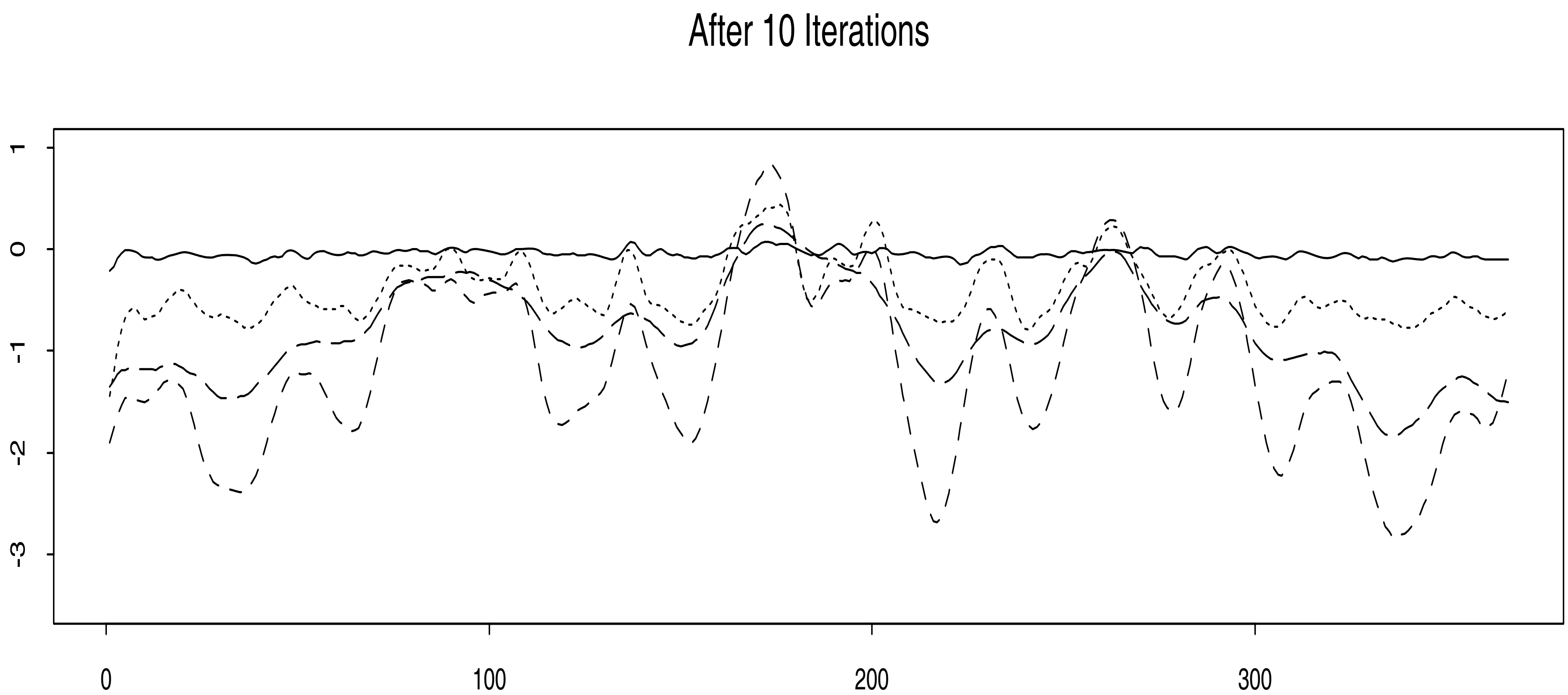
$$\min\left(1, \frac{\prod_{i=k}^{k+L} g\left(y_i\middle|\, x_i'\right)}{\prod_{i=k}^{k+L} g\left(y_i\middle|\, x_i\right)}\right).$$

• The parameter σ^2 is updated through a simple Gibbs step

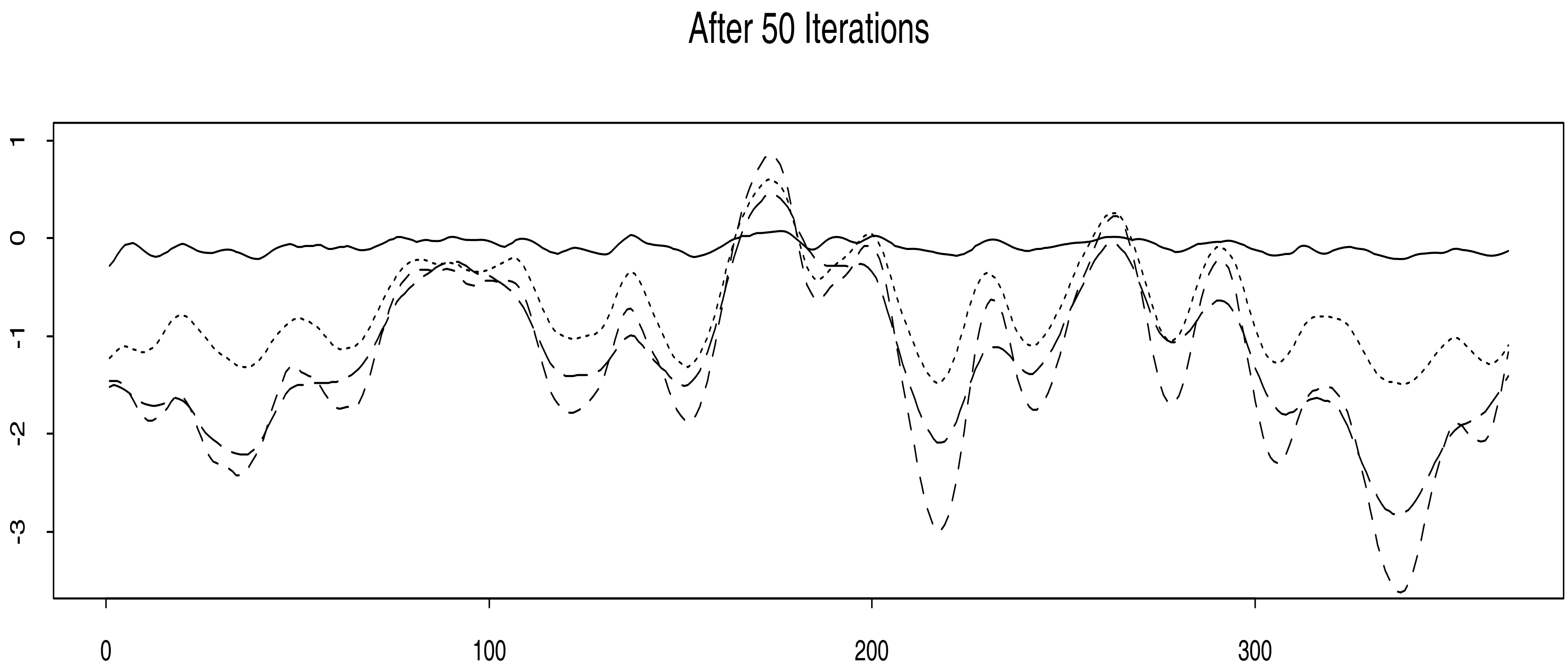
$$\begin{aligned} \sigma^2 &\sim \pi\left(\sigma^2\middle|\, x_{1:n}, y_{1:n}\right) = \pi\left(\sigma^2\middle|\, x_{1:n}\right) \\ &= \mathcal{IG}\left(\frac{\nu_0 + n - 1}{2}, \frac{\gamma_0 + \sum_{k=2}^{n}\left(\alpha_k - 2\alpha_{k-1} + \alpha_{k-2}\right)^2}{2}\right) \end{aligned}$$

• For block size $L = 1, 5, 20$ and 40, we compute the average trajectories of 100 parallel chains after 10, 50, 100 and 500 iterations with initialization $x_k = 0$ for all $k, \sigma^2 = 0.1$.

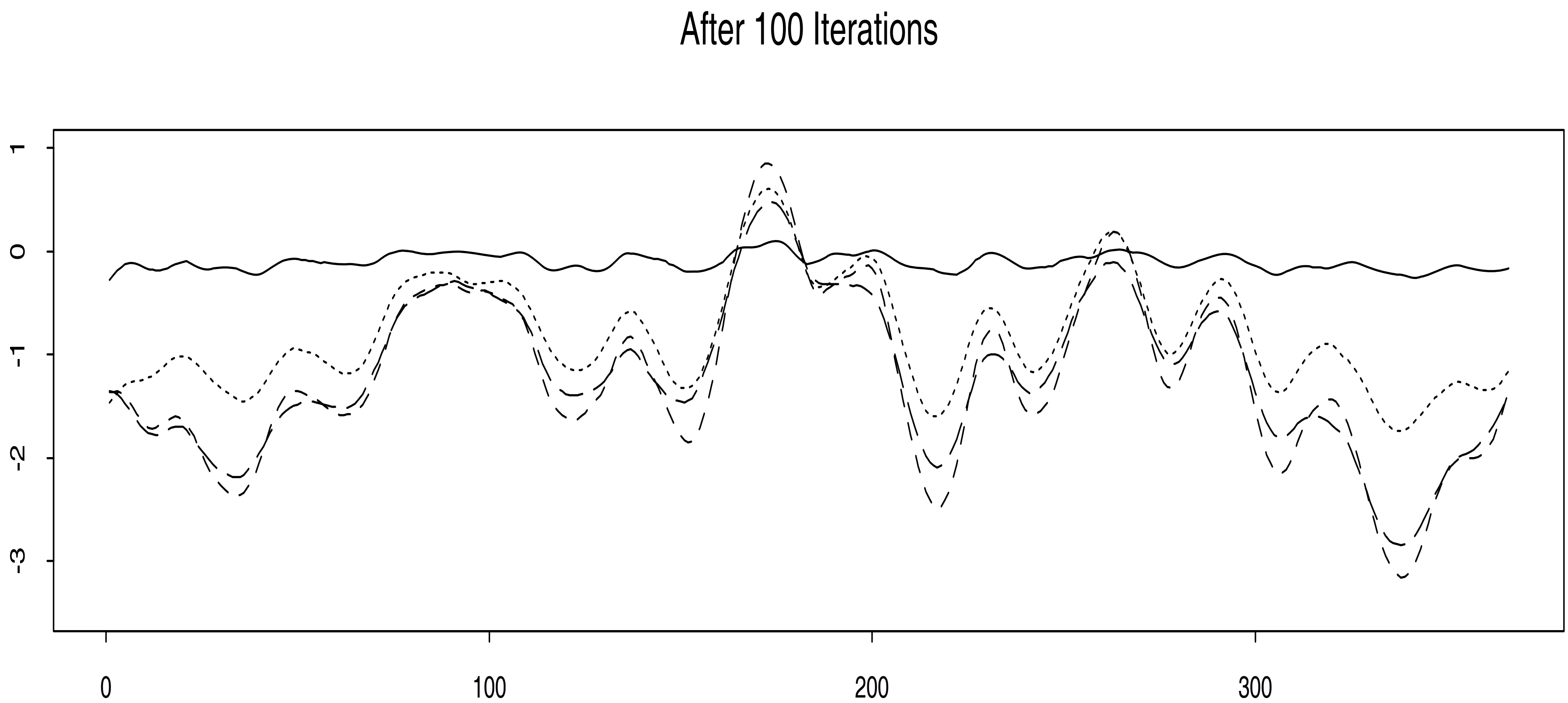
Average trajectories over 100 chains for $L = 1, 5, 20$ and 40 from top to bottom.



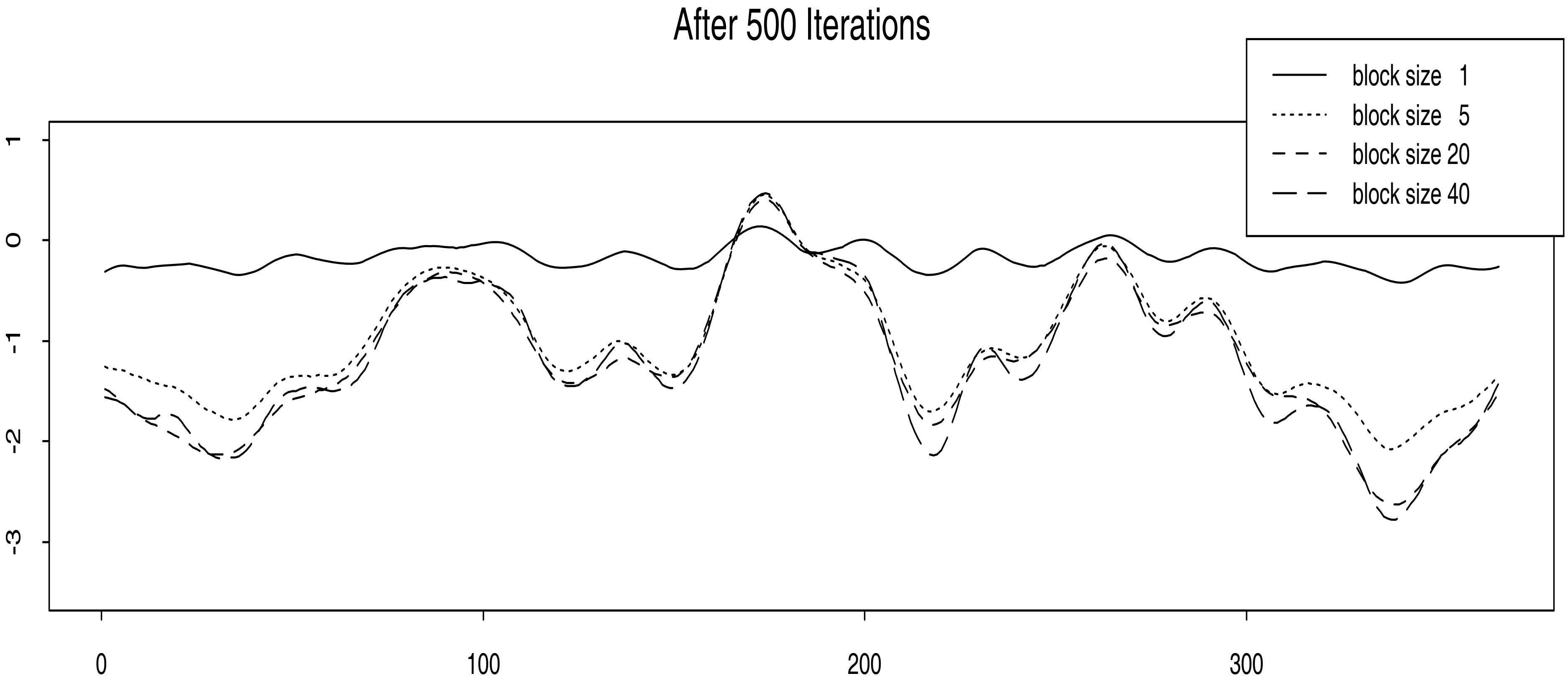
Average trajectories over 100 chains for $L = 1, 5, 20$ and 40 from top to bottom.



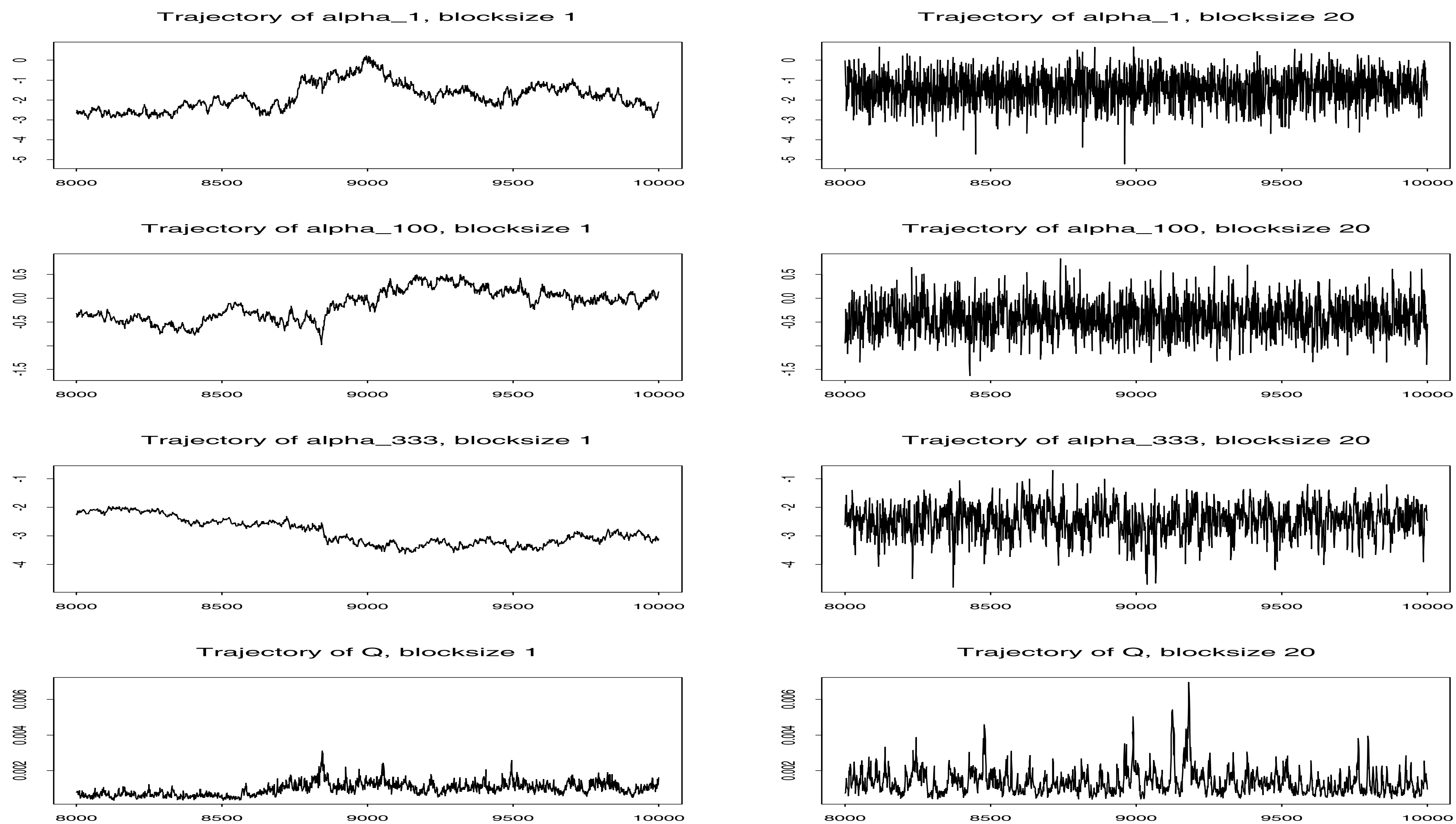
Average trajectories over 100 chains for $L = 1, 5, 20$ and 40 from top to bottom.



Average trajectories over 100 chains for $L = 1, 5, 20$ and 40 from top to bottom.



Traces of $\alpha_1, \alpha_{100}, \alpha_{333}$ and σ^2 for $L = 1$ (left) and $L = 20$ (right).



4.9– Simulation Results

- This (naive!) block sampling strategy performs well here because the likelihood of the observations is fairly flat.

- For a linear Gaussian observation equation, Knorr-Held compares this strategy to a direct Gibbs sampling implementation. As expected, the conditional proposal strategy is competitive when the observations are not very informative compared to the prior.

- For more complex problems, such strategies are inefficient and we will need to use the observations to build the proposal.

• (Pitt & Shephard, 1999) propose a more efficient strategy... also more computationally intensive.

• Consider the log full conditional distribution

$$\log \pi \left(x_{k:k+L} \middle| y_{k:k+L}, x_{k-1}, x_{k+1} \right) = \sum_{i=k}^{k+L} \log g \left(y_i \middle| x_i \right) + \sum_{i=k}^{k+L+1} \log f \left(x_{i+1} \middle| x_i \right)$$

$$\equiv \sum_{i=k}^{k+L} \log g \left(y_i \middle| x_i \right) - \frac{1}{2} \sum_{i=k}^{k+L+1} \left(x_{i+1} - A x_i \right)^{\mathrm{T}} \Sigma^{-1} \left(x_{i+1} - A x_i \right)$$

which is not quadratic in x_i hence $\pi \left(x_{k:k+L} \middle| y_{k:k+L}, x_{k-1}, x_{k+1} \right)$ is not Gaussian.

• The idea is to expand the log-likelihood part around some point estimates

$$\begin{aligned} \log g \left(y_i \middle| x_i \right) \quad \simeq \quad & \log g \left(y_i \middle| \widehat{x}_i \right) + \nabla \log g \left(y_i \middle| \widehat{x}_i \right) . \left(x_i - \widehat{x}_i \right) \\ & + \frac{1}{2} \left(x_i - \widehat{x}_i \right)^{\mathrm{T}} \nabla^2 \log g \left(y_i \middle| \widehat{x}_i \right) \left(x_i - \widehat{x}_i \right) \end{aligned}$$

• By doing this, we have a Gaussian approximation of the log-likelihood and then we obtain a Gaussian proposal $q\left(x_{1:n}, x'_{k:k+L}\right) = q\left(x_{-(k:k+L)}, x'_{k:k+L}\right)$

$$\log q\left(x_{-(k:k+L)}, x'_{k:k+L}\right) \equiv \sum_{i=k}^{k+L} \nabla \log g\left(y_i \middle| \widehat{x}_i\right) . \left(x_i - \widehat{x}_i\right)$$

$$+\frac{1}{2}\left(x_i - \widehat{x}_i\right)^{\mathrm{T}} \nabla^2 \log g\left(y_i \middle| \widehat{x}_i\right)\left(x_i - \widehat{x}_i\right) - \frac{1}{2}\sum_{i=k}^{k+L+1}\left(x_{i+1} - Ax_i\right)^{\mathrm{T}} \Sigma^{-1}\left(x_{i+1} - Ax_i\right)$$

• (Pitt & Shepard, 1999) propose to select

$$\widehat{x}_{k:k+1} = \arg\max \pi\left(x_{k:k+L} \middle| y_{k:k+L}, x_{k-1}, x_{k+1}\right)$$

and a scheme to sample from $q\left(x_{-(k:k+L)}, x'_{k:k+L}\right)$ which is of complexity $O\left(L\right)$.

4.9– Simulation Results

- This algorithm is applied to the SV model where

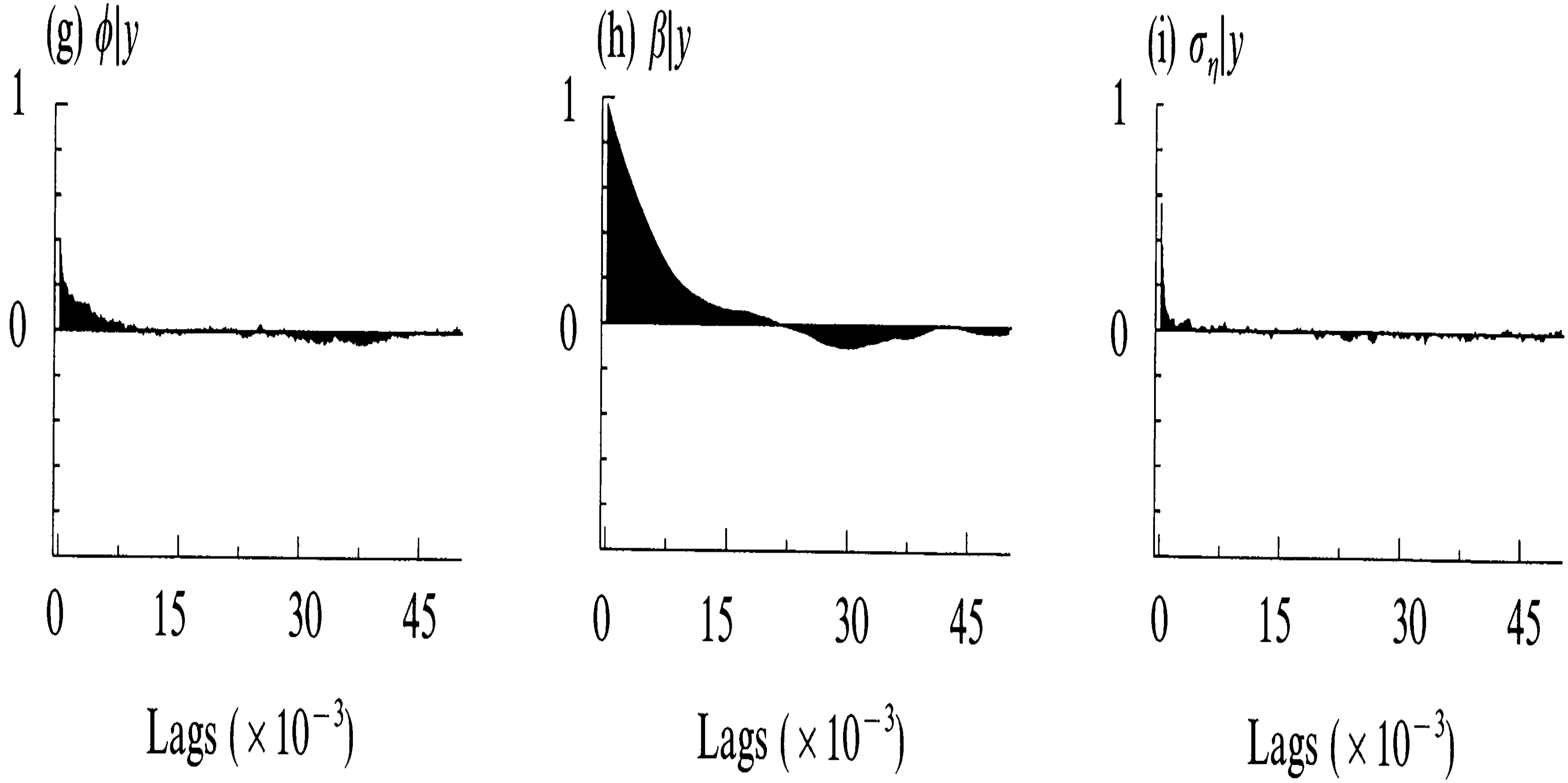
$$\begin{aligned} X_k &= \phi X_{k-1} + \sigma V_k,\ V_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \\ Y_k &= \beta \exp(X_k/2)\, W_k,\ W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1). \end{aligned}$$

- Prior are set to $\phi \sim \mathcal{U}[-1,1]$, $\sigma^2 \sim \mathcal{IG}\left(\frac{\nu_\sigma}{2}, \frac{\gamma_\sigma}{2}\right)$ and $\beta \sim \mathcal{IG}\left(\frac{\nu_\beta}{2}, \frac{\gamma_\beta}{2}\right)$.

- Full conditional distributions of the parameters given $x_{1:n}, y_{1:n}$ are standard.

- Compared to standard single move strategies, the authors report significant improvement.

Autocorrelation plots for (ϕ, σ^2, β) with $L = 1$



Autocorrelation plots for (ϕ, σ^2, β) with $L = 50$ on average