

# Simulation Based Inference: Gibbs Sampling

Chris Men

chrismen4@gmail.com

September 29, 2017

# Outline

Motivation

Gibbs Algorithm

Examples

Summary

## Why Gibbs sampling

- ▶ In statistics, direct maximum likelihood estimation (MLE) algorithms sometimes are difficult to derive.

## Why Gibbs sampling

- ▶ In statistics, direct maximum likelihood estimation (MLE) algorithms sometimes are difficult to derive.
- ▶ Gibbs sampling is commonly used in statistical inference especially in Bayesian inference.

## Why Gibbs sampling

- ▶ In statistics, direct maximum likelihood estimation (MLE) algorithms sometimes are difficult to derive.
- ▶ Gibbs sampling is commonly used in statistical inference especially in Bayesian inference.
- ▶ Gibbs sampling is an alternative of the expectation-maximization (EM), and most of times is much easier to implement.

## Why Gibbs sampling

- ▶ In statistics, direct maximum likelihood estimation (MLE) algorithms sometimes are difficult to derive.
- ▶ Gibbs sampling is commonly used in statistical inference especially in Bayesian inference.
- ▶ Gibbs sampling is an alternative of the expectation-maximization (EM), and most of times is much easier to implement.
- ▶ Applied widely in finance and financial econometrics.

## Why Gibbs sampling

- ▶ In statistics, direct maximum likelihood estimation (MLE) algorithms sometimes are difficult to derive.
- ▶ Gibbs sampling is commonly used in statistical inference especially in Bayesian inference.
- ▶ Gibbs sampling is an alternative of the expectation-maximization (EM), and most of times is much easier to implement.
- ▶ Applied widely in finance and financial econometrics.
- ▶ Sometimes are time consuming. C/C++, MATLAB call C, or R call C have to be used.

## Why Gibbs sampling (Cont'd)

- ▶ Gibbs sampling is a Markov Chain Monte Carlo (MCMC) algorithm.



## Why Gibbs sampling (Cont'd)

- ▶ Gibbs sampling is a Markov Chain Monte Carlo (MCMC) algorithm.
- ▶ Gibbs sampling generates correlated time series.

## Why Gibbs sampling (Cont'd)

- ▶ Gibbs sampling is a Markov Chain Monte Carlo (MCMC) algorithm.
- ▶ Gibbs sampling generates correlated time series.
- ▶ A burn in period is needed before performing inference.

## Why Gibbs sampling (Cont'd)

- ▶ Gibbs sampling is a Markov Chain Monte Carlo (MCMC) algorithm.
- ▶ Gibbs sampling generates correlated time series.
- ▶ A burn in period is needed before performing inference.
- ▶ The joint distribution is not known explicitly or is difficult to sample from directly.

## Gibbs Algorithm

- ▶ Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a distribution or a model indexed by a vector of parameters  $\theta^T = (\theta_1, \dots, \theta_m)$ .

## Gibbs Algorithm

- ▶ Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a distribution or a model indexed by a vector of parameters  $\theta^T = (\theta_1, \dots, \theta_m)$ .
- ▶ Suppose also that each parameter can be sampled conditional on other parameters in the model.

## Gibbs Algorithm

- ▶ Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a distribution or a model indexed by a vector of parameters  $\theta^T = (\theta_1, \dots, \theta_m)$ .
- ▶ Suppose also that each parameter can be sampled conditional on other parameters in the model.
- ▶ We have generated a sample  $\{\theta_i^{(1)}, \dots, \theta_i^{(L)}\}$  from the conditional distribution of  $\theta_i$ .

## Gibbs Algorithm

- ▶ Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a distribution or a model indexed by a vector of parameters  $\theta^T = (\theta_1, \dots, \theta_m)$ .
- ▶ Suppose also that each parameter can be sampled conditional on other parameters in the model.
- ▶ We have generated a sample  $\{\theta_i^{(1)}, \dots, \theta_i^{(L)}\}$  from the conditional distribution of  $\theta_i$ .
- ▶ The estimate of  $\theta_i$  is

$$\hat{\theta}_i = \frac{1}{L-l} \sum_{k=l+1}^L \theta_i^{(k)},$$

where the first  $l$  generated numbers are discarded as burn in.

## Gibbs Algorithm (cont'd)

Table: Gibbs algorithm.

---

**Step 0.** Initialize  $\{\theta_1^{(0)}, \dots, \theta_m^{(0)}\}$ , set  $k = 0$ .

**Step 1.** Sample  $\theta_1^{(k+1)} | \theta_2^{(k)}, \dots, \theta_m^{(k)}$ .

Sample  $\theta_2^{(k+1)} | \theta_1^{(k+1)}, \theta_3^{(k)}, \dots, \theta_m^{(k)}$ .

.

.

.

Sample  $\theta_m^{(k+1)} | \theta_1^{(k+1)}, \theta_2^{(k+1)}, \dots, \theta_{m-1}^{(k+1)}$ .

**Step 2.** Go to Step 1.

- 
- Gibbs sampler generates posterior samples by sweeping through each variable to sample from its conditional distribution with the remaining variables fixed to their current values.



## Example 1 – fit a univariate normal distribution

- ▶ Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a univariate normal distribution  $y \sim N(\mu, \sigma^2)$  with pdf
$$p(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right).$$

## Example 1 – fit a univariate normal distribution

- ▶ Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a univariate normal distribution  $y \sim N(\mu, \sigma^2)$  with pdf  $p(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$ .
- ▶ The likelihood of the data based on the model is

$$f(\mathbf{y}|\mu, \sigma^2) = \prod_{i=1}^n f(y_i|\mu, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}\right)$$

## Example 1 – fit a univariate normal distribution

- Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a univariate normal distribution  $y \sim N(\mu, \sigma^2)$  with pdf  $p(y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$ .

- The likelihood of the data based on the model is

$$f(\mathbf{y}|\mu, \sigma^2) = \prod_{i=1}^n f(y_i|\mu, \sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}\right)$$

- The posterior distribution of the two parameters  $\mu$  and  $\sigma^2$  is

$$\begin{aligned} f(\mu, \sigma^2|\mathbf{y}) &\propto f(\mu|\mu_*, \sigma_*^2) \text{IG}(\sigma^2|\alpha_\sigma, \beta_\sigma) \prod_{i=1}^n f(y_i|\mu, \sigma^2) \\ &\propto \left(\frac{1}{\sigma_*^2}\right)^{\frac{1}{2}} \exp\left(-\frac{(\mu - \mu_*)^2}{2\sigma_*^2}\right) \frac{(\beta_\sigma)^{\alpha_\sigma} e^{-\beta_\sigma/\sigma^2}}{\Gamma(\alpha_\sigma)(\sigma^2)^{\alpha_\sigma+1}} \\ &\quad \times \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}\right), \end{aligned}$$

where  $f(\mu|\mu_*, \sigma_*^2)$  and  $\text{IG}(\sigma^2|\alpha_\sigma, \beta_\sigma)$  are two prior distributions.

## Example 1 – fit a univariate normal distribution (Cont'd) – simulate $\mu$

- The conditional distribution of  $\mu$  is

$$\begin{aligned} f(\mu|\sigma^2, \mathbf{y}) &\propto \exp\left(-\frac{(\mu - \mu_*)^2}{2\sigma_*^2}\right) \exp\left(-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}\right) \\ &\propto \exp\left(-\frac{1}{2}\left[\mu^2\left(\frac{1}{\sigma_*^2} + \frac{n}{\sigma^2}\right) - 2\mu\left[\frac{\mu_*}{\sigma_*^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2}\right]\right]\right) \end{aligned}$$

## Example 1 – fit a univariate normal distribution (Cont'd) – simulate $\mu$

- ▶ The conditional distribution of  $\mu$  is

$$\begin{aligned} f(\mu|\sigma^2, \mathbf{y}) &\propto \exp\left(-\frac{(\mu - \mu_*)^2}{2\sigma_*^2}\right) \exp\left(-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}\right) \\ &\propto \exp\left(-\frac{1}{2}\left[\mu^2\left(\frac{1}{\sigma_*^2} + \frac{n}{\sigma^2}\right) - 2\mu\left[\frac{\mu_*}{\sigma_*^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2}\right]\right]\right) \end{aligned}$$

- ▶ The conditional distribution of  $\mu$  is then

$$f(\mu|\sigma^2, \mathbf{y}) \sim N\left(\frac{B}{A}, \frac{1}{A}\right)$$

where

$$B = \frac{\mu_*}{\sigma_*^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2}, A = \frac{1}{\sigma_*^2} + \frac{n}{\sigma^2}$$

## Example 1 – fit a univariate normal distribution (Cont'd) – simulate $\sigma^2$

- The conditional distribution of  $\sigma^2$  is

$$\begin{aligned} f(\sigma^2 | \mu, \mathbf{y}) &\propto \frac{(\beta_\sigma)^{\alpha_\sigma} e^{-\beta_\sigma / \sigma^2}}{\Gamma(\alpha_\sigma)(\sigma^2)^{\alpha_\sigma+1}} \left( \frac{1}{\sigma^2} \right)^{\frac{n}{2}} \exp \left( - \frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2} \right) \\ &\propto \frac{e^{-\beta_\sigma / \sigma^2}}{(\sigma^2)^{\alpha_\sigma+1}} \left( \frac{1}{\sigma^2} \right)^{\frac{n}{2}} \exp \left( - \frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2} \right) \\ &\propto \left( \frac{1}{\sigma^2} \right)^{(\alpha_\sigma + \frac{n}{2})+1} \exp \left( - \frac{\beta_\sigma + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2}{\sigma^2} \right) \end{aligned}$$

## Example 1 – fit a univariate normal distribution (Cont'd) – simulate $\sigma^2$

- ▶ The conditional distribution of  $\sigma^2$  is

$$\begin{aligned} f(\sigma^2 | \mu, \mathbf{y}) &\propto \frac{(\beta_\sigma)^{\alpha_\sigma} e^{-\beta_\sigma / \sigma^2}}{\Gamma(\alpha_\sigma)(\sigma^2)^{\alpha_\sigma+1}} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}\right) \\ &\propto \frac{e^{-\beta_\sigma / \sigma^2}}{(\sigma^2)^{\alpha_\sigma+1}} \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{\sum_{i=1}^n (y_i - \mu)^2}{2\sigma^2}\right) \\ &\propto \left(\frac{1}{\sigma^2}\right)^{(\alpha_\sigma + \frac{n}{2})+1} \exp\left(-\frac{\beta_\sigma + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2}{\sigma^2}\right) \end{aligned}$$

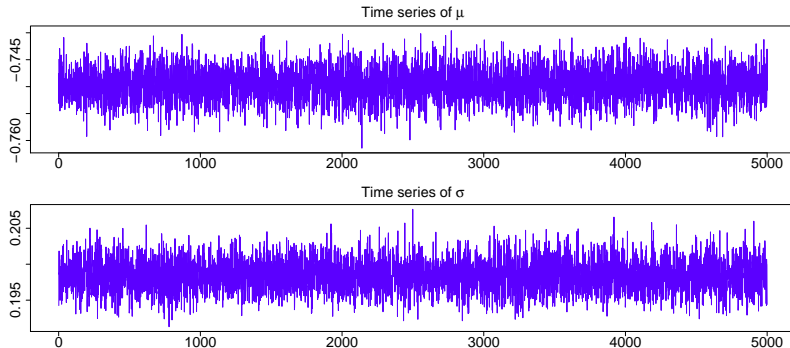
- ▶ The conditional distribution of  $\sigma^2$  is

$$f(\sigma^2 | \mu, \mathbf{y}) \sim \text{IG}(\hat{\alpha}_\sigma, \hat{\beta}_\sigma)$$

where

$$\hat{\alpha}_\sigma = \alpha_\sigma + \frac{n}{2}, \hat{\beta}_\sigma = \beta_\sigma + \frac{1}{2} \sum_{i=1}^n (y_i - \mu)^2$$

## Example 1 – fit a univariate normal distribution (Cont'd) – sampled time series



The algorithm was iterated 5000 times to fit a data set artificially generated from  $N(-0.75, 0.2^2)$ .



## Example 1 – fit a univariate normal distribution (Cont'd)

**Table:** Estimate  $\mu$  and  $\sigma$  of a normal distribution  $N(\mu, \sigma^2)$ .

|           |              | Case 1  |         | Case 2     |         |         |
|-----------|--------------|---------|---------|------------|---------|---------|
| Parameter | True         | Average | SD      | True       | Average | SD      |
| $\mu$     | <b>-0.75</b> | -0.7495 | 0.00307 | <b>0.5</b> | 0.48144 | 0.03935 |
| $\sigma$  | <b>0.2</b>   | 0.1984  | 0.00223 | <b>2.5</b> | 2.51564 | 0.02814 |

The Gibbs algorithm was iterated 5000 times. After the first 500 iterations were discarded, parameters were estimated by sample means.

## Example 2 – fit a mixture of two normal distributions

- Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a mixture of two normal distributions.

$$f(y) = \rho \times f_1(y|\mu_1, \sigma_1^2) + (1 - \rho) \times f_2(y|\mu_2, \sigma_2^2),$$

where  $0 < \rho < 1$ . The component functions  $f_i(y|\mu_i, \sigma_i^2)$ ,  $i = 1, 2$ , are normal pdfs.

## Example 2 – fit a mixture of two normal distributions

- ▶ Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a mixture of two normal distributions.

$$f(y) = \rho \times f_1(y|\mu_1, \sigma_1^2) + (1 - \rho) \times f_2(y|\mu_2, \sigma_2^2),$$

where  $0 < \rho < 1$ . The component functions  $f_i(y|\mu_i, \sigma_i^2), i = 1, 2$ , are normal pdfs.

- ▶ For the distribution to be identifiable, we assume that  $\mu_1 < \mu_2$  and  $\sigma_1^2 < \sigma_2^2$ .

## Example 2 – fit a mixture of two normal distributions

- ▶ Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a mixture of two normal distributions.

$$f(y) = \rho \times f_1(y|\mu_1, \sigma_1^2) + (1 - \rho) \times f_2(y|\mu_2, \sigma_2^2),$$

where  $0 < \rho < 1$ . The component functions  $f_i(y|\mu_i, \sigma_i^2)$ ,  $i = 1, 2$ , are normal pdfs.

- ▶ For the distribution to be identifiable, we assume that  $\mu_1 < \mu_2$  and  $\sigma_1^2 < \sigma_2^2$ .
- ▶ The likelihood of the proposed model takes a complicated form.

## Example 2 – fit a mixture of two normal distributions

- ▶ Suppose that we have a data set  $\mathbf{y} = \{y_i, i = 1, \dots, n\}$ , which is generated from a mixture of two normal distributions.

$$f(y) = \rho \times f_1(y|\mu_1, \sigma_1^2) + (1 - \rho) \times f_2(y|\mu_2, \sigma_2^2),$$

where  $0 < \rho < 1$ . The component functions  $f_i(y|\mu_i, \sigma_i^2)$ ,  $i = 1, 2$ , are normal pdfs.

- ▶ For the distribution to be identifiable, we assume that  $\mu_1 < \mu_2$  and  $\sigma_1^2 < \sigma_2^2$ .
- ▶ The likelihood of the proposed model takes a complicated form.
- ▶ To simplify this, we introduce latent variables  $\mathbf{z} = (z_1, \dots, z_n)$ , which are defined below,

$$z_i = \begin{cases} 0, & \text{with probability } \rho, \\ 1, & \text{with probability } 1 - \rho, \end{cases}$$

for  $i = 1, \dots, n$ .

## Example 2 – fit a mixture of two normal distributions (Cont'd)

- Conditional on these latent variables, we have

$$y_i | \theta \sim \begin{cases} \mathcal{N}(\mu_1, \sigma_1^2), & \text{if } z_i = 0, \\ \mathcal{N}(\mu_2, \sigma_2^2), & \text{if } z_i = 1. \end{cases}$$

## Example 2 – fit a mixture of two normal distributions (Cont'd)

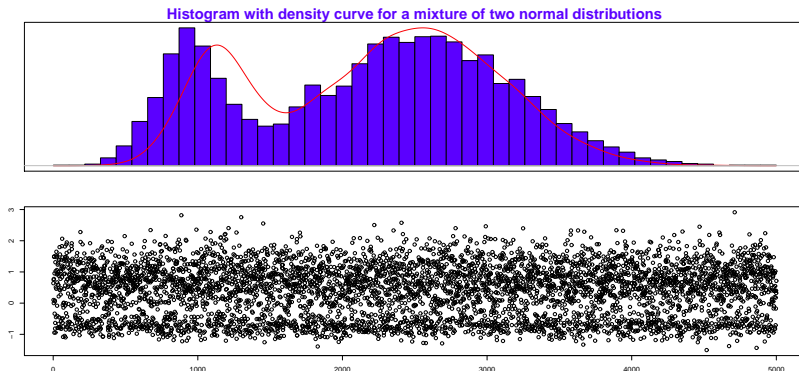
- Conditional on these latent variables, we have

$$y_i|\theta \sim \begin{cases} \mathcal{N}(\mu_1, \sigma_1^2), & \text{if } z_i = 0, \\ \mathcal{N}(\mu_2, \sigma_2^2), & \text{if } z_i = 1. \end{cases}$$

- The likelihood of the data  $\mathbf{y}$  given the model separates into two parts, each of them corresponding to each of the two mixture components,

$$l(\theta|\mathbf{y}, \mathbf{z}) \propto \prod_{i=1, z_i=0}^n \frac{\rho}{\sigma_1} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_1)^2}{\sigma_1^2} \right\} \\ \prod_{i=1, z_i=1}^n \frac{(1 - \rho)}{\sigma_2} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_2)^2}{\sigma_2^2} \right\}$$

## Example 2 – fit a mixture of two normal distributions (Cont'd)



**Figure:** The data was generated from a mixture of two normal distributions with pdf:  $f(y) = 0.25 \times f_1(y | -0.75, 0.2^2) + 0.75 \times f_2(y | 0.75, 0.6^2)$



# Gibbs algorithm

Table: The Gibbs estimation procedure.

---

**Step 0.** Initialize  $\mu_1, \mu_2, \sigma_1, \sigma_2$  and  $\rho$ .

**Step 1.** Sample  $z_i, i = 1, \dots, n$ .

**Step 2.** Sample  $\rho$ .

**Step 3.** Sample  $\mu_1, \sigma_1^2$ .

**Step 4.** Sample  $\mu_2, \sigma_2^2$ .

**Step 5.** Go to **Step 1**.

---

## Example 2 – fit a mixture of two normal distributions (Cont'd)

**Step 1.** The conditional probability that the observed  $y_i$  has been generated by the first mixture component is

$$f(z_i = 0 | \mathbf{y}, \theta) = \frac{\frac{\rho}{\sigma_1} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_1)^2}{\sigma_1^2} \right\}}{\frac{\rho}{\sigma_1} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_1)^2}{\sigma_1^2} \right\} + \frac{(1-\rho)}{\sigma_2} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_2)^2}{\sigma_2^2} \right\}}.$$

## Example 2 – fit a mixture of two normal distributions (Cont'd)

**Step 1.** The conditional probability that the observed  $y_i$  has been generated by the first mixture component is

$$f(z_i = 0 | \mathbf{y}, \theta) = \frac{\frac{\rho}{\sigma_1} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_1)^2}{\sigma_1^2} \right\}}{\frac{\rho}{\sigma_1} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_1)^2}{\sigma_1^2} \right\} + \frac{(1-\rho)}{\sigma_2} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_2)^2}{\sigma_2^2} \right\}}.$$

- The simulation of the latent variables  $z_i, i = 1, \dots, n$ , can be carried out by simulating Bernoulli distributions.

## Example 2 – fit a mixture of two normal distributions (Cont'd)

**Step 2.** The conditional posterior density of  $f(\rho|\theta_{-\rho}, \mathbf{y}, \mathbf{z})$ , where  $\theta_{-\rho}$  denotes the remaining parameters except  $\rho$ , has the following kernel,

$$f(\rho|\mathbf{y}, \theta_{-\rho}, \mathbf{z}) \propto \rho^{T_0} (1 - \rho)^{T_1},$$

where  $T_k = \#\{z_i = k\}$ ,  $k = 0, 1$ , the number of observations assigned to the  $k$ -th component.

## Example 2 – fit a mixture of two normal distributions (Cont'd)

**Step 2.** The conditional posterior density of  $f(\rho|\theta_{-\rho}, \mathbf{y}, \mathbf{z})$ , where  $\theta_{-\rho}$  denotes the remaining parameters except  $\rho$ , has the following kernel,

$$f(\rho|\mathbf{y}, \theta_{-\rho}, \mathbf{z}) \propto \rho^{T_0} (1 - \rho)^{T_1},$$

where  $T_k = \#\{z_i = k\}$ ,  $k = 0, 1$ , the number of observations assigned to the  $k$ -th component.

- It is obvious that  $\rho \sim \text{Beta}(T_0 + 1, T_1 + 1)$ , and the simulation of  $\rho$  is easier to carry out.

## Example 2 – fit a mixture of two normal distributions (Cont'd)

**Step 3.** The conditional distribution of  $\mu_1$  is

$$\begin{aligned} f(\mu_1 | \mathbf{y}, \theta_{-\mu_1}, \mathbf{z}) \\ &\propto \prod_{i=1, z_i=0}^n \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_1)^2}{\sigma_1^2} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \frac{T_0}{\sigma_1^2} \mu_1^2 - 2 \frac{\mu_1}{\sigma_1^2} \sum_{i=1, z_i=0}^n y_i \right] \right\}. \end{aligned}$$

## Example 2 – fit a mixture of two normal distributions (Cont'd)

**Step 3.** The conditional distribution of  $\mu_1$  is

$$\begin{aligned} f(\mu_1 | \mathbf{y}, \theta_{-\mu_1}, \mathbf{z}) \\ &\propto \prod_{i=1, z_i=0}^n \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_1)^2}{\sigma_1^2} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[ \frac{T_0}{\sigma_1^2} \mu_1^2 - 2 \frac{\mu_1}{\sigma_1^2} \sum_{i=1, z_i=0}^n y_i \right] \right\}. \end{aligned}$$

- It is easier to see that  $\mu_1 \sim \mathcal{N}(\frac{b_1}{a_1}, \frac{1}{a_1})$ , where

$$a_1 = \frac{T_0}{\sigma_1^2}, \quad b_1 = \frac{1}{\sigma_1^2} \sum_{i=1, z_i=0}^n y_i.$$

## Example 2 – fit a mixture of two normal distributions (Cont'd)

**Step 4.** Given an Inverse Gamma prior distribution  $\sigma_1^2 \sim \mathcal{IG}(\alpha, \delta)$ , the conditional distribution of  $\sigma_1^2$  is

$$\begin{aligned} f(\sigma_1^2 | \mathbf{y}, \mathbf{z}, \theta_{\sigma_1^2}) &\propto \frac{1}{(\sigma_1^2)^{\alpha+1}} \exp \left\{ -\frac{\delta}{\sigma_1^2} \right\} \prod_{i=1, z_i=0}^n \frac{1}{\sigma_1^2} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_1)^2}{\sigma_1^2} \right\} \\ &\propto \frac{1}{(\sigma_1^2)^{\alpha+1+T_0/2}} \exp \left\{ -\frac{\delta + \frac{1}{2} \sum_{i=1, z_i=0}^n (y_i - \mu_1)^2}{\sigma_1^2} \right\}. \end{aligned}$$



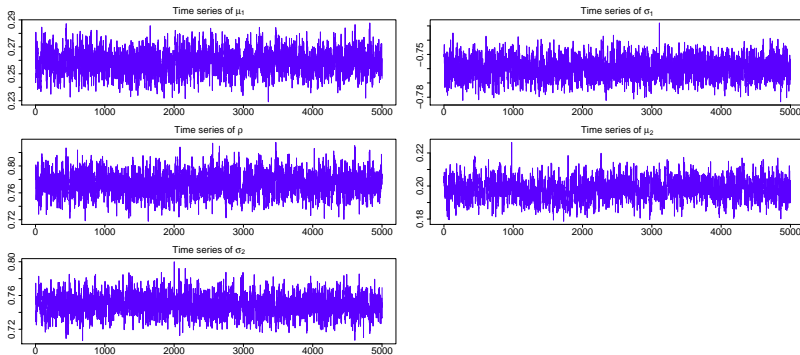
## Example 2 – fit a mixture of two normal distributions (Cont'd)

**Step 4.** Given an Inverse Gamma prior distribution  $\sigma_1^2 \sim \mathcal{IG}(\alpha, \delta)$ , the conditional distribution of  $\sigma_1^2$  is

$$\begin{aligned} f(\sigma_1^2 | \mathbf{y}, \mathbf{z}, \theta_{\sigma_1^2}) &\propto \frac{1}{(\sigma_1^2)^{\alpha+1}} \exp \left\{ -\frac{\delta}{\sigma_1^2} \right\} \prod_{i=1, z_i=0}^n \frac{1}{\sigma_1^2} \exp \left\{ -\frac{1}{2} \frac{(y_i - \mu_1)^2}{\sigma_1^2} \right\} \\ &\propto \frac{1}{(\sigma_1^2)^{\alpha+1+T_0/2}} \exp \left\{ -\frac{\delta + \frac{1}{2} \sum_{i=1, z_i=0}^n (y_i - \mu_1)^2}{\sigma_1^2} \right\}. \end{aligned}$$

- It is seen that  $\sigma_1^2 \sim \mathcal{IG}(c, d)$ , where  $c = \alpha + T_0/2$  and  $d = \delta + \frac{1}{2} \sum_{i=1, z_i=0}^n (y_i - \mu_1)^2$ , respectively.

## Example 2 – fit a mixture of two normal distributions – sampled time series



The algorithm was iterated 5000 times to fit a data set artificially generated from  $0.25 \times N(-0.75, 0.2^2) + (1 - 0.25) \times N(0.75, 0.6^2)$ .

## Example 2 – fit a mixture of two normal distributions (Cont'd)

**Table:** Fit a mixture distribution:  $\rho \times N(\mu_1, \sigma_1^2) + (1 - \rho) \times N(\mu_2, \sigma_2^2)$ .

| Parameter  | Case 1       |         |        | Case 2      |         |        |
|------------|--------------|---------|--------|-------------|---------|--------|
|            | True         | Average | SD     | True        | Average | SD     |
| $\mu_1$    | <b>-0.75</b> | -0.7624 | 0.0075 | <b>0.25</b> | 0.2451  | 0.0048 |
| $\sigma_1$ | <b>0.2</b>   | 0.2023  | 0.0059 | <b>0.2</b>  | 0.1943  | 0.0045 |
| $\rho$     | <b>0.25</b>  | 0.2439  | 0.0077 | <b>0.5</b>  | 0.4913  | 0.0100 |
| $\mu_2$    | <b>0.75</b>  | 0.7241  | 0.0128 | <b>1.25</b> | 1.2096  | 0.0271 |
| $\sigma_2$ | <b>0.6</b>   | 0.6187  | 0.0098 | <b>1.2</b>  | 1.2061  | 0.0169 |

The Gibbs algorithm was iterated 5000 times. After the first 500 iterations were discarded, parameters were estimated by sample means.

## Summary

- ▶ Joint distribution of parameters are hard to simulate directly.

## Summary

- ▶ Joint distribution of parameters are hard to simulate directly.
- ▶ The conditional or posterior distribution of each individual parameter is easier to simulate.

## Summary

- ▶ Joint distribution of parameters are hard to simulate directly.
- ▶ The conditional or posterior distribution of each individual parameter is easier to simulate.
- ▶ Gibbs algorithms are usually easier to implement.

## Summary

- ▶ Joint distribution of parameters are hard to simulate directly.
- ▶ The conditional or posterior distribution of each individual parameter is easier to simulate.
- ▶ Gibbs algorithms are usually easier to implement.
- ▶ Gibbs methods are widely used in solving time series models.

## Summary

- ▶ Joint distribution of parameters are hard to simulate directly.
- ▶ The conditional or posterior distribution of each individual parameter is easier to simulate.
- ▶ Gibbs algorithms are usually easier to implement.
- ▶ Gibbs methods are widely used in solving time series models.
- ▶ Mixture distributions can be imposed to time series and non-time series models.



# Thank you!

## References

- [1] Robert, C. and G., Casella. 2013. Monte Carlo Statistical Methods: Edition 2. London: Chapman and Hall.
- [2] Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. 1995. Markov Chain Monte Carlo in Practice. London: Chapman and Hall.
- [3] Paul Glasserman. 2013. Monte Carlo Methods in Financial Engineering (Stochastic Modelling and Applied Probability) . ([Springer](#)).