Politecnico di Milano

Mathematical Engineering - Statistical Learning

A.Y. 2020/2021

Bayesian Registration of Functional Data

Bayesian Statistics Project of:

Paolo Bighignoli
Federica Principe
Federica Mattina

Supervisors:

Prof. Raffaele Argiento
Prof. Lucia Paci
Prof. Alessia Pini

Contents

1 Introduction

Functional data often exhibit a common shape, but with variations in amplitude and phase across curves. Time variation is an important aspect of functional data because it explains a large portion of their variability. So the analysis often proceeds by synchronization of the data through curve registration.
In this project, we analyze data representing the angle generated by a knee during one hoop over time. We have data coming from three groups of people: people who just had surgery, people that are in physical therapy (or that had done it in the past), and healthy people.
Even through everyone perform a similar sequence of moves during one hoop, such steps do not occur at the same time in all people. Therefore, the goal of our analysis lies in finding, for each curve, a warping function that synchronizes all curves, modeling the amplitude of the features and estimating the velocity function. In this way, we can understand if and how having knee problems is related to having difficulty in jumping.

2 Dataset

The data we have been given come from the University Hospital of Umeä, Sweden. The data represent the angle generated by a knee during a one leg hoop over time. They are splitted into three groups: people who just had surgery, people that are in physical therapy (or that had done it in the past), and healthy people.
Observations have been divided into categories and they are shown in Figure 1, Figure 2 and Figure 3.

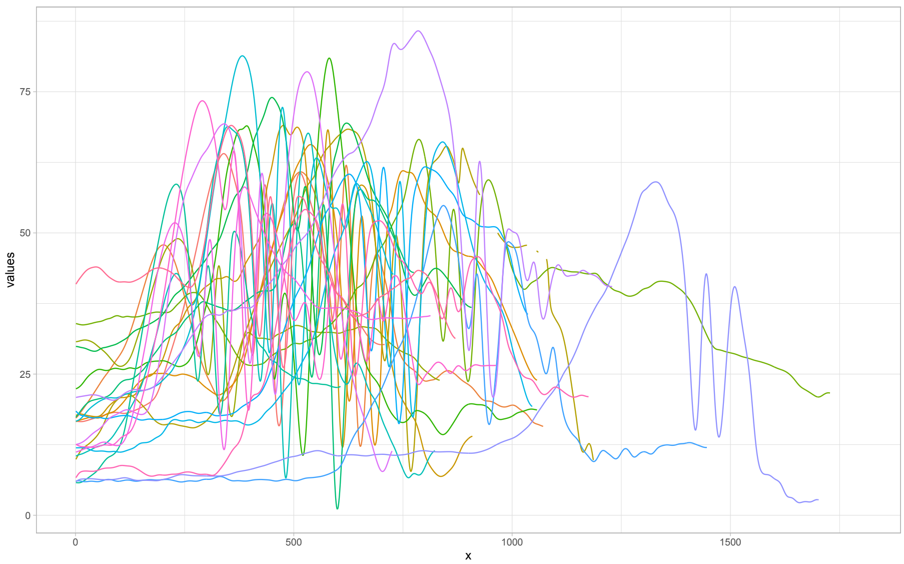


Figure 1: People who underwent surgery.

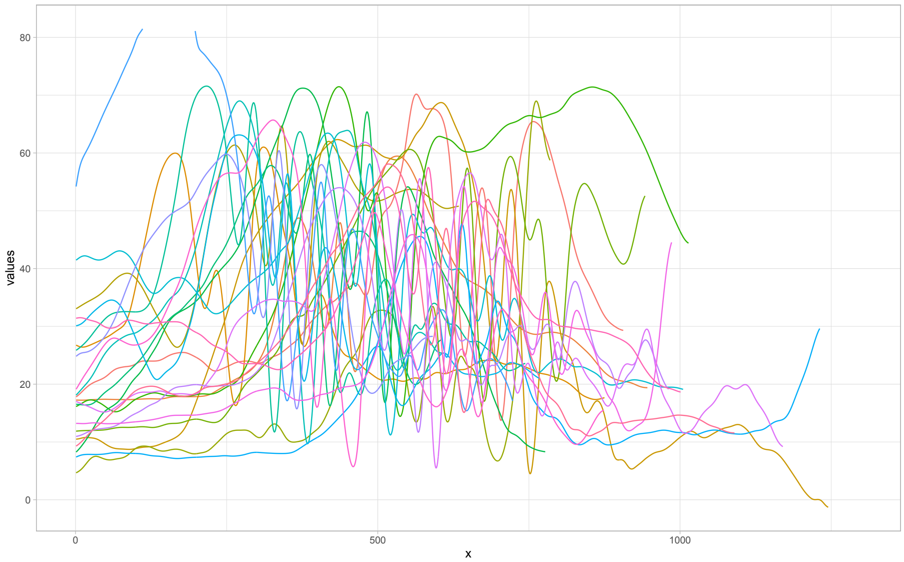


Figure 2: People who underwent physiotherapy.

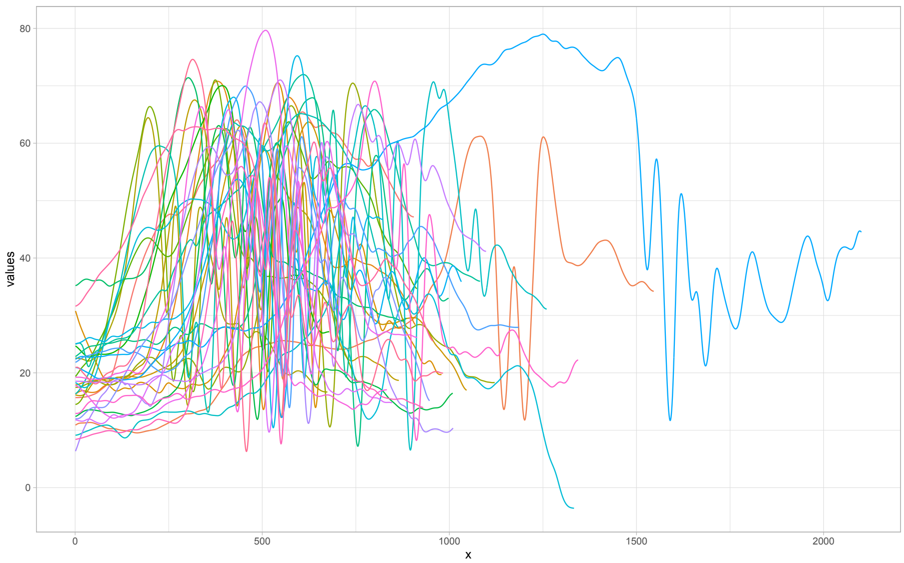


Figure 3: Healthy people.

3 Model formulation

We refer to self-modeling regression (SEMOR) methods. They are semiparametric models in which the subject-specific regression function is a parametric transformation of a common smooth regression function.

3.1 Hierarchical Model

Let $y_i(t)$ denote the observation of the i-th curve at time t, with $i = 1, 2, ..., N$ and $t \in T = [t_1, t_N]$.
We introduce a three-stage hierarchical model.

- *Stage one*

Consider:

$$m_i(t;\boldsymbol{\theta}_i) = c_i + a_i m(t;\boldsymbol{\beta}) = c_i + a_i \boldsymbol{\mathcal{B}}'_m(t)\boldsymbol{\beta}$$
$$\mu_i(t) = \mu_i(t; f_i, g_i) = f_i + g_i \cdot t$$

And the composite function:

$$m_i(\mu_i(t);\theta_i) = c_i + a_i \boldsymbol{\mathcal{B}}'_m(\mu_i(t))\boldsymbol{\beta}$$

The observed value of each curve i at time t is modeled as:

$$\begin{aligned} y_i(t) &= m_i(\mu_i(t);\theta_i) + \epsilon_i \\ &= c_i + a_i \boldsymbol{\mathcal{B}}'_m(\mu_i(t))\boldsymbol{\beta} + \epsilon_i \end{aligned}$$

where $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\epsilon), i = 1, ..., N$.
That is, we assume that the error terms are independent and normally distributed with null mean and *common* variance σ^2_ϵ. We assumed $\sigma^2_\epsilon = 0.3$.
The common shape function m_i is modeled as a linear combination of a set of basis function $\boldsymbol{\mathcal{B}}$ and a set of coefficients $\boldsymbol{\beta}$.
The curve-specific time transformation function $\mu_i(t)$ is modeled as a linear transformation of the original times t through the coefficients f_i and g_i.
We identify the i-th aligned curve at time t as:

$$y^*_i(t) = y_i(t) \circ \mu_i^{-1}(t)$$

- *Stage two*

  Assume:

  $$c_i \sim \mathcal{N}(c_0, \sigma^2_c) \qquad a_i \sim \mathcal{N}(a_0, \sigma^2_a)$$

  In this way, we define a curve-specific random linear transformation of the common shape function m_i. Moreover, assume:

  $$f_i \sim \mathcal{N}(f_0, \sigma^2_f) \qquad g_i \sim \mathcal{N}(g_0, \sigma^2_g)\mathbb{1}_{\{g_i>0\}}$$

  where the indicator function on g ensures that we do not have time reversion. In this way, we have a one-to-one correspondence between physical times and transformed times.

- *Stage three*

  Assume:

  $$a_0 \sim \mathcal{N}(m_{a_0}, \sigma^2_{a_0}) \qquad c_0 \sim \mathcal{N}(m_{c_0}, \sigma^2_{c_0})$$

  Moreover, for the precision parameters:

  $$1/\sigma^2_a \sim Gamma(a_a, b_a) \qquad 1/\sigma^2_c \sim Gamma(a_c, b_c)$$
  $$1/\sigma^2_\epsilon \sim Gamma(a_\epsilon, b_\epsilon)$$

  where $X \sim Gamma(a, b)$ is parameterized so that $E(X) = \frac{a}{b}$.

3.2 Penalized regression splines implementation

It is well known that, when using B-splines, the main issue is the choice of the positions and number of interior knots. We refer to an alternative approach that relies on penalized regression splines (Eilers and Marx, 1996).
Under this formulation, a large number of equidistant nodes is selected and a penalty, dependent on a smooth parameter λ, is placed on coefficients of adjacent B-spines. Specifically, we place a first order random walk shrinkage prior on the shape coefficients β, so that:

$$\beta_k = \beta_{k-1} + \epsilon_k \qquad \epsilon_k \sim \mathcal{N}(0, \lambda)$$

Then:

$$\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$$

where:

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{\mathbf{-1}} = \frac{\boldsymbol{\Omega}}{\lambda}$$

$$\boldsymbol{\Omega} = \begin{pmatrix} 2 & 1 & 0 & & & 0 \\ -1 & 2 & -1 & \ddots & & \\ 0 & -1 & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & -1 & 0 \\ & & \ddots & -1 & 2 & -1 \\ 0 & & & 0 & -1 & 1 \end{pmatrix}$$

Moreover, we assume as prior for the parameter λ:

$$\lambda \sim IG(a_\lambda, b_\lambda).$$

4 Full conditional distributions

The parameters vector of the presented model is:

$$\boldsymbol{\theta} = (\boldsymbol{c}', \boldsymbol{a}', \boldsymbol{\beta}', \boldsymbol{f}', \boldsymbol{g}', c_0, a_0, \sigma_\epsilon^2, \sigma_c^2, \sigma_a^2, \lambda)$$

The observation vector is $\boldsymbol{Y} = (\boldsymbol{Y_1}', ..., \boldsymbol{Y_N}')'$, i.e. a vector where each component $\boldsymbol{Y_i}'$ is itself a vector containing the observations of the i-th curve, $i = 1, ..., N$.
Thus, the posterior distribution of $\boldsymbol{\theta}$ is:

$$\begin{aligned}
\pi(\boldsymbol{\theta}|\boldsymbol{Y}) &= \pi(\boldsymbol{Y}|\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{f}, \boldsymbol{g}, c_0, a_0, \sigma_\epsilon^2, \sigma_c^2, \sigma_a^2, \lambda) \\
&\propto \pi(\boldsymbol{Y}|\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{f}, \boldsymbol{g}, \sigma_\epsilon^2) \cdot \pi(\boldsymbol{c}, \boldsymbol{a}|c_0, a_0, \sigma_c^2, \sigma_a^2) \\
&\cdot \pi(\boldsymbol{\beta}|\lambda) \cdot \pi(\boldsymbol{f}, \boldsymbol{g}|g_0, f_0, \sigma_g^2, \sigma_f^2) \cdot \pi(\sigma_c^2|a_c, b_c) \\
&\cdot \pi(a_0|\sigma_{a_0}^2) \cdot \pi(c_0|\sigma_{c_0}^2) \cdot \pi(\sigma_\epsilon^2|a_\epsilon, b_\epsilon) \\
&\cdot \pi(\sigma_a^2|a_a, b_a) \cdot \pi(\lambda|a_\lambda, b_\lambda)
\end{aligned}$$

It's clear that the joint posterior density is analytically intractable. Therefore, our posterior inferences are based on MCMC simulation and, specifically, on a Gibbs-within-Metropolis algorithm.
In order to implement it, we first derive all the full conditional distributions.

4.1 Full conditional distribution of β

In order to derive the full conditional distribution of β, observe that:

$$\pi(\boldsymbol{\beta}|\boldsymbol{Y}, \theta_{-\boldsymbol{\beta}}) \propto \pi(\boldsymbol{Y}|\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{g}, \boldsymbol{f}, \sigma_\epsilon^2) \cdot \pi(\boldsymbol{\beta}|\lambda)$$

And remember the definition of $y_i(t)$:

$$\begin{aligned}
& y_i(t) = c_i + a_i \boldsymbol{\mathcal{B}}'_m(\mu_i(t))\boldsymbol{\beta} + \epsilon_i \\
\Rightarrow \quad & y_i(t)|c_i, a_i, \boldsymbol{\beta}, f_i, g_i, \sigma_\epsilon \sim \mathcal{N}(c_i + a_i \boldsymbol{\mathcal{B}}'_m(\mu_i(t))\boldsymbol{\beta}, \sigma_\epsilon^2)
\end{aligned}$$

It is straightforward the extension to multiple dimensions:

$$\begin{aligned}
\boldsymbol{Y}_i|c_i, a_i, \boldsymbol{\beta}, f_i, g_i, \sigma_\epsilon &\sim \mathcal{N}\left(\begin{pmatrix} c_i + a_i \boldsymbol{\mathcal{B}}'_m(\mu(t_1))\boldsymbol{\beta} \\ c_i + a_i \boldsymbol{\mathcal{B}}'_m(\mu(t_2))\boldsymbol{\beta} \\ \vdots \end{pmatrix}, \sigma_\epsilon^2 \boldsymbol{I_N}\right) \\
&\sim \mathcal{N}(c_i \mathbb{1}_n + a_i \boldsymbol{\mathcal{B}}'_m(\mu(\boldsymbol{t}))\boldsymbol{\beta}, \sigma_\epsilon^2 \boldsymbol{I_N})
\end{aligned}$$

In order to derive the distribution of the whole vector $\boldsymbol{Y}$, we define the vectors $\boldsymbol{C}$ and $\boldsymbol{M}$ as:

$$\begin{aligned}
C &= \left[(c_1 \mathbb{1}_n)', \dots, (c_N \mathbb{1}_n)'\right]' \\
M &= \left[(a_1 \boldsymbol{\mathcal{B}}'_m(\mu(\boldsymbol{t}))\boldsymbol{\beta})', \dots, (a_N \boldsymbol{\mathcal{B}}'_m(\mu(\boldsymbol{t}))\boldsymbol{\beta})'\right]' \\
&\Rightarrow \boldsymbol{Y}|\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\beta}, \boldsymbol{f}, \boldsymbol{g}, \sigma_\epsilon^2 \sim \mathcal{N}(\boldsymbol{C} + \boldsymbol{M}, \sigma_\epsilon^2 \boldsymbol{I_{N_n}})
\end{aligned}$$

Therefore, the full conditional of β is proportional to the product of two normal distributions:

$$\begin{aligned}
\pi(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\theta}_{-\boldsymbol{\beta}}) \propto & \frac{1}{\sqrt{(2\pi)^{N_n} \mathrm{Det}(\sigma_\epsilon^2 \boldsymbol{I_{N_n}})}} \cdot \exp\left\{-\frac{1}{2}\left[\boldsymbol{Y} - (\boldsymbol{C} + \boldsymbol{M})\right]' (\sigma_\epsilon^2 \boldsymbol{I_{N_n}})^{-1} \left[\boldsymbol{Y} - (\boldsymbol{C} + \boldsymbol{M})\right]\right\} \\
& \cdot \frac{1}{(2\pi)^k \mathrm{Det}\left[\left(\frac{\boldsymbol{\Omega}}{\lambda}\right)^{-1}\right]} \cdot \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}' \frac{\boldsymbol{\Omega}}{\lambda} \boldsymbol{\beta}\right)\right\} \\
& \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left[\boldsymbol{Y} - (\boldsymbol{C} + \boldsymbol{M})\right]' \left[\boldsymbol{Y} - (\boldsymbol{C} + \boldsymbol{M})\right]\right\} \cdot \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}' (\frac{\boldsymbol{\Omega}}{\lambda}) \boldsymbol{\beta}\right)\right\}
\end{aligned}$$

Observe that M can be rewritten as:

$$\boldsymbol{M} = \begin{pmatrix} a_1\boldsymbol{\mathcal{B}}'_m(\mu(\boldsymbol{t}))\boldsymbol{\beta} \\ \vdots \\ a_N\boldsymbol{\mathcal{B}}'_m(\mu(\boldsymbol{t}))\boldsymbol{\beta} \end{pmatrix} = \begin{pmatrix} a_1\boldsymbol{\mathcal{B}}'_m(\mu_1(t_1))\boldsymbol{\beta} \\ a_1\boldsymbol{\mathcal{B}}'_m(\mu_1(t_2))\boldsymbol{\beta} \\ \vdots \\ a_2\boldsymbol{\mathcal{B}}'_m(\mu_2(t_1))\boldsymbol{\beta} \\ a_2\boldsymbol{\mathcal{B}}'_m(\mu_2(t_2))\boldsymbol{\beta} \\ \vdots \\ a_N\boldsymbol{\mathcal{B}}'_m(\mu_N(t_1))\boldsymbol{\beta} \\ a_N\boldsymbol{\mathcal{B}}'_m(\mu_N(t_2))\boldsymbol{\beta} \\ \vdots \end{pmatrix} = \begin{pmatrix} a_1\boldsymbol{\mathcal{B}}'_m(\mu_1(t_1)) \\ a_1\boldsymbol{\mathcal{B}}'_m(\mu_1(t_2)) \\ \vdots \\ a_2\boldsymbol{\mathcal{B}}'_m(\mu_2(t_1)) \\ a_2\boldsymbol{\mathcal{B}}'_m(\mu_2(t_2)) \\ \vdots \\ a_N\boldsymbol{\mathcal{B}}'_m(\mu_N(t_1)) \\ a_N\boldsymbol{\mathcal{B}}'_m(\mu_N(t_2)) \\ \vdots \end{pmatrix} \cdot \boldsymbol{\beta} = \widetilde{\boldsymbol{M}}\boldsymbol{\beta}$$

$$\begin{aligned}
\implies \pi(\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{\theta}_{-\boldsymbol{\beta}}) &\propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}(\boldsymbol{Y}-\boldsymbol{C}-\tilde{\boldsymbol{M}}\boldsymbol{\beta})'\cdot(\boldsymbol{Y}-\boldsymbol{C}-\tilde{\boldsymbol{M}}\boldsymbol{\beta})\right\}\cdot\exp\left\{-\frac{1}{2}\boldsymbol{\beta}'\frac{\boldsymbol{\Omega}}{\lambda}\boldsymbol{\beta}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left[-\boldsymbol{Y}'\tilde{\boldsymbol{M}}\boldsymbol{\beta}+\boldsymbol{C}'\tilde{\boldsymbol{M}}\boldsymbol{\beta}-\boldsymbol{\beta}'\tilde{\boldsymbol{M}}'\boldsymbol{Y}+\boldsymbol{\beta}'\tilde{\boldsymbol{M}}'\boldsymbol{C}+\boldsymbol{\beta}'\tilde{\boldsymbol{M}}'\tilde{\boldsymbol{M}}\boldsymbol{\beta}\right]-\frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}\right\} \\
&= \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left[-2\boldsymbol{Y}'\tilde{\boldsymbol{M}}'\boldsymbol{\beta}+2\boldsymbol{C}'\tilde{\boldsymbol{M}}\boldsymbol{\beta}+\boldsymbol{\beta}'\tilde{\boldsymbol{M}}'\tilde{\boldsymbol{M}}\boldsymbol{\beta}\right]-\frac{1}{2}(\boldsymbol{\beta}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta})\right\} \\
&= \exp\left\{-\frac{1}{2}\boldsymbol{\beta}'\left(\frac{\tilde{\boldsymbol{M}}'\tilde{\boldsymbol{M}}}{\sigma_\epsilon^2}+\boldsymbol{\Sigma}^{-1}\right)\boldsymbol{\beta}+\frac{1}{\sigma_\epsilon^2}\left(\boldsymbol{Y}-\boldsymbol{C}\right)'\tilde{\boldsymbol{M}}\boldsymbol{\beta}\right\}
\end{aligned}$$

It can be shown, with some standard calculations, that the full conditional distribution of $\boldsymbol{\beta}$ is still normal:

$$\boldsymbol{\beta}|\boldsymbol{Y},\boldsymbol{\theta}_{-\boldsymbol{\beta}} \sim \mathcal{N}(\boldsymbol{m}_{\boldsymbol{\beta}},\boldsymbol{V}_{\boldsymbol{\beta}})$$

where:

$$\boldsymbol{m}_{\boldsymbol{\beta}} = \boldsymbol{V}_{\boldsymbol{\beta}}\left[\frac{1}{\sigma_\epsilon^2}\tilde{M}^\intercal\left(\boldsymbol{Y}-\boldsymbol{C}\right)\right] \qquad \boldsymbol{V}_{\boldsymbol{\beta}}^{-1} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^{-1}+\frac{1}{\sigma_\epsilon^2}\boldsymbol{X}'\boldsymbol{X}$$
$$\boldsymbol{X} = \left(Vec\left[a_1\boldsymbol{\mathcal{B}}(\mu(\boldsymbol{t}))\right]\ldots Vec\left[a_N\boldsymbol{\mathcal{B}}(\mu(\boldsymbol{t}))\right]\right)$$

4.2 Full conditional of a_0

In order to derive the full conditional distribution of a_0, observe that:

$$\pi(a_0|\boldsymbol{Y},\boldsymbol{\theta}_{-a_0}) \propto \pi(\boldsymbol{c},\boldsymbol{a}|c_0,a_0,\sigma_a^2,\sigma_c^2)\cdot\underbrace{\pi(a_0|\sigma_{a_0}^2)}_{\mathcal{N}(m_{a_0},\sigma_{a_0}^2)}$$

Assume parameters $\boldsymbol{c}$ and $\boldsymbol{a}$ to be a priori independent and normally distributed:

$$\boldsymbol{z} = \begin{pmatrix}\boldsymbol{c}\\ \boldsymbol{a}\end{pmatrix} = \begin{pmatrix} c_1 \\ \vdots \\ c_N \\ a_1 \\ \vdots \\ a_N \end{pmatrix} \sim \mathcal{N}\left(\boldsymbol{m},\boldsymbol{V}\right)$$

where:

$$\boldsymbol{m} = \begin{pmatrix} c_0\mathbb{1}_N \\ a_0\mathbb{1}_N \end{pmatrix} \qquad \boldsymbol{V} = \begin{pmatrix} \sigma_c^2\boldsymbol{I}_N & 0 \\ 0 & \sigma_a^2\boldsymbol{I}_N \end{pmatrix}$$

Then:

$$\pi(a_0|\boldsymbol{Y},\boldsymbol{\theta}_{-a_0}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{m})^{\intercal}\boldsymbol{V}^{-1}(\boldsymbol{z}-\boldsymbol{m})\right\}\cdot\exp\left\{-\frac{1}{2}\frac{(a_0-m_{a_0})^2}{\sigma^2_{a_0}}\right\}$$

Since $\boldsymbol{z}$ doesn't depend on a_0, terms that rely just on it can be discarded, so:

$$\begin{aligned}\pi(a_0|\boldsymbol{Y},\boldsymbol{\theta}_{-a_0}) &\propto \exp\left\{(\boldsymbol{z}^{\intercal}\boldsymbol{V}^{-1}\boldsymbol{m})-\frac{1}{2}(\boldsymbol{m}^{\intercal}\boldsymbol{V}^{-1}\boldsymbol{m})\right\}\cdot\exp\left\{-\frac{1}{2}\frac{(a_0-m_{a_0})^2}{\sigma^2_{a_0}}\right\}\\ &\propto \exp\left\{\boldsymbol{a'}(\sigma^2_a\boldsymbol{I_N})^{-1}\boldsymbol{m}_a-\frac{1}{2}\boldsymbol{m}'_a(\sigma^2_a\boldsymbol{I_N})^{-1}\boldsymbol{m}_a\right\}\cdot\exp\left\{-\frac{1}{2}\frac{(a_0-m_{a_0})^2}{\sigma^2_{a_0}}\right\}\end{aligned}$$

where:

$$\boldsymbol{m}_a = a_0\mathbb{1}_N.$$

It can be shown, with some standard calculations, that the full conditional distribution of a_0 is still normal:

$$\begin{aligned}f(a_0|\boldsymbol{Y},\boldsymbol{\theta}_{-a_0}) &\propto \exp\left\{-\frac{1}{2}\underbrace{\left(\frac{1}{\sigma^2_a}N+\frac{1}{\sigma^2_{a_0}}\right)}_{\text{W}}a_0^2-2\underbrace{\left(\frac{1}{\sigma^2_a}\sum_{i=1}^{N}a_i+\frac{m_{a_0}}{\sigma^2_{a_0}}\right)}_{\text{V}}a_0\right\}\\ &= \exp\left\{-\frac{1}{2}\left[W\left(a_0-\frac{V}{2W}\right)^2\right]\right\}\end{aligned}$$

Therefore:

$$\begin{gathered}a_0|\boldsymbol{Y},\boldsymbol{\theta_{-a_0}} \sim \mathcal{N}(a_0^*,\sigma^*_{a_0})\\ a_0^* = \frac{V}{2W} = \frac{1}{2}\left(\frac{1}{\sigma^2_a}N+\frac{1}{\sigma^2_{a_0}}\right)^{-1}\cdot 2\left(\frac{1}{\sigma^2_a}\sum_{i=1}^{N}a_i+\frac{m_{a_0}}{\sigma^2_{a_0}}\right)\\ \sigma^*_{a_0} = W^{-1} = \left(\frac{1}{\sigma^2_a}N+\frac{1}{\sigma^2_{a_0}}\right)^{-1}\end{gathered}$$

The full conditional distribution of c_0 can be derived in a similar way. Namely:

$$\begin{gathered}c_0|\boldsymbol{Y},\boldsymbol{\theta_{-c_0}} \sim \mathcal{N}(c_0^*,\sigma^*_{c_0})\\ c_0^* = \frac{V}{2W} = \frac{1}{2}\left(\frac{1}{\sigma^2_a}N+\frac{1}{\sigma^2_{c_0}}\right)^{-1}\cdot 2\left(\frac{1}{\sigma^2_c}\sum_{i=1}^{N}c_i+\frac{m_{c_0}}{\sigma^2_{c_0}}\right)\\ \sigma^*_{c_0} = W^{-1} = \left(\frac{1}{\sigma^2_c}N+\frac{1}{\sigma^2_{c_0}}\right)^{-1}\end{gathered}$$

4.3 Full conditional of a_i and c_i

Assume c_i and a_i to be a priori independent and normally distributed, i.e.:

$$\begin{pmatrix}c_i\\a_i\end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix}c_0\\a_0\end{pmatrix},\begin{pmatrix}\sigma^2_c & 0\\ 0 & \sigma^2_a\end{pmatrix}\right)$$

Then, the full conditional of a_i and c_i is proportional to the product of two normal distributions:

$$\pi\left(\begin{pmatrix}c_i\\a_i\end{pmatrix}|\boldsymbol{\theta}_{-c_i,a_i},\boldsymbol{Y}\right)\propto\pi\left(\begin{pmatrix}c_i\\a_i\end{pmatrix}|a_0,c_0,\sigma^2_a,\sigma^2_c\right)\cdot\pi\left(\boldsymbol{Y}|\boldsymbol{c},\boldsymbol{a},\boldsymbol{\beta},g,f,\sigma^2_\epsilon\right)$$

Indeed, recall that:

$$\boldsymbol{Y}|\boldsymbol{c},\boldsymbol{a},\boldsymbol{\beta},g,f,\sigma_\epsilon^2 \sim \mathcal{N}(\boldsymbol{C}+\tilde{\boldsymbol{M}}\boldsymbol{\beta},\sigma_\epsilon^2 I_{N_n})$$
$$\boldsymbol{Y}_i \sim \mathcal{N}\left(c_i\mathbb{1}_n + a_i\boldsymbol{\mathcal{B}}'_m(\mu_i(t))\boldsymbol{\beta},\sigma_\epsilon^2\boldsymbol{I}_N\right)$$
$$\boldsymbol{m_i} = c_i\mathbb{1}_n + a_i\boldsymbol{\mathcal{B}}'_m(\mu_i(t))\boldsymbol{\beta}$$

Therefore:

$$\pi\left(\begin{pmatrix}c_i\\a_i\end{pmatrix}|\boldsymbol{\theta}_{-c_i,a_i},\boldsymbol{Y}\right) \propto \exp\left\{-\frac{1}{2}\left[\begin{pmatrix}c_i\\a_i\end{pmatrix}-\begin{pmatrix}c_0\\a_0\end{pmatrix}\right]^\intercal \Sigma_{c,a}^{-1}\left[\begin{pmatrix}c_i\\a_i\end{pmatrix}-\begin{pmatrix}c_0\\a_0\end{pmatrix}\right]\right\}$$
$$\cdot\exp\left\{-\frac{1}{2}\left(\boldsymbol{Y}_i-\boldsymbol{m}_i\right)'\frac{1}{\sigma_\epsilon^2}\boldsymbol{I}_N^{-1}\left(\boldsymbol{Y}_i-\boldsymbol{m}_i\right)\right\}$$

It can be shown with (a lot of) standard calculations that the full conditional distribution of the vector $\begin{pmatrix}c_i\\a_i\end{pmatrix}$ is still normal:

$$\begin{pmatrix}c_i\\a_i\end{pmatrix}|\boldsymbol{\theta}_{-c_i,a_i},\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{m}_l,\boldsymbol{\Sigma}_l)$$

where:

$$\boldsymbol{m}_l = \boldsymbol{\Sigma}_l\cdot\left[\Sigma_{c,a}^{-1}\begin{pmatrix}c_0\\a_0\end{pmatrix}+\frac{1}{\sigma_\epsilon^2}\boldsymbol{W}'\boldsymbol{Y}_i\right]$$
$$\boldsymbol{\Sigma}_l^{-1} = \Sigma_{c,a}^{-1}+\frac{1}{\sigma_\epsilon^2}\boldsymbol{W}'\boldsymbol{W}$$
$$\boldsymbol{W} = \begin{pmatrix}\mathbb{1}_n & \boldsymbol{\mathcal{B}'}_m(\mu(t))\boldsymbol{\beta}\end{pmatrix}$$

4.4 Full conditional of λ

For sake of simplicity, we refer to $\frac{1}{\lambda}$ instead of λ.
Assume as prior for $\frac{1}{\lambda}$ a gamma distribution:

$$\frac{1}{\lambda}\sim Gamma\left(a_\lambda,b_\lambda\right)$$

The full conditional of $\frac{1}{\lambda}$ is proportional to the product of a normal and a gamma distribution. Indeed:

$$\pi\left(\frac{1}{\lambda}|\boldsymbol{\theta}_{-\lambda},\boldsymbol{Y}\right)\propto\underbrace{\pi(\boldsymbol{\beta}|\lambda)}_{\mathcal{N}(\boldsymbol{0},(\frac{\boldsymbol{\Omega}}{\lambda})^{-1})}\cdot\pi\left(\frac{1}{\lambda}|a_\lambda,b_\lambda\right)$$

$$\Longrightarrow\pi\left(\frac{1}{\lambda}|\boldsymbol{\theta}_{-\lambda},\boldsymbol{Y}\right)\propto\frac{1}{\sqrt{\mathrm{Det}\left(\frac{\boldsymbol{\Omega}}{\lambda}\right)^{-1}}}\cdot\exp\left\{-\frac{1}{2}\boldsymbol{\beta}^\intercal\left(\frac{\boldsymbol{\Omega}}{\lambda}\right)\boldsymbol{\beta}\right\}\cdot\left(\frac{1}{\lambda}\right)^{a_\lambda-1}\cdot\exp\left\{-b_\lambda\frac{1}{\lambda}\right\}$$
$$\propto\frac{1}{\lambda^{\frac{k}{2}}}\cdot\left(\frac{1}{\lambda}\right)^{a_\lambda-1}\cdot\exp\left\{-\frac{1}{\lambda}\left(\frac{1}{2}\boldsymbol{\beta}^\intercal\boldsymbol{\Omega}\boldsymbol{\beta}+b_\lambda\right)\right\}$$
$$=\left(\frac{1}{\lambda}\right)^{a_\lambda+\frac{k}{2}-1}\cdot\exp\left\{-\frac{1}{\lambda}\left(\frac{1}{2}\boldsymbol{\beta}^\intercal\boldsymbol{\Omega}\boldsymbol{\beta}+b_\lambda\right)\right\}$$

Then:

$$\frac{1}{\lambda}\sim\mathrm{Gamma}(a_\lambda^*,b_\lambda^*)$$
$$a_\lambda^*=a_\lambda+\frac{k}{2}\qquad b_\lambda^*=\frac{1}{2}\boldsymbol{\beta}^\intercal\boldsymbol{\Omega}\boldsymbol{\beta}+b_\lambda$$

Since we are referring to the shape-rate Gamma parameterization:

$$\lambda\sim\text{Inv-Gamma}(a_\lambda^*,b_\lambda^*)$$

4.5 Full conditional of σ_c^2 and σ_a^2

Similarly to what done for the parameter λ, we initially refer to $\frac{1}{\sigma_c^2}$ ($\frac{1}{\sigma_a^2}$) and then we shift to σ_c^2 (σ_a^2).
Assume as prior for $\frac{1}{\sigma_c^2}$ a gamma distribution:

$$\frac{1}{\sigma_c^2} \sim \text{Gamma}(a_c, b_c)$$

So:

$$\pi\left(\frac{1}{\sigma_c^2}|\boldsymbol{\theta}_{-\sigma_c^2}, \boldsymbol{Y}\right) \propto \pi(\boldsymbol{c}, \boldsymbol{a}|c_0, a_0, \sigma_c^2, \sigma_a^2) \cdot \pi\left(\frac{1}{\sigma_c^2}|a_c, b_c\right)$$

$$\begin{aligned}
\implies \pi\left(\frac{1}{\sigma_c^2}|\boldsymbol{\theta}_{-\sigma_c^2}, \boldsymbol{Y}\right) &\propto \frac{1}{\sqrt{\text{Det}(\sigma_c^2 \boldsymbol{I}_N)^{-1}}} \cdot \exp\left\{-\frac{1}{2}(\boldsymbol{c} - \boldsymbol{c_0})' \frac{1}{\sigma_c^2}(\boldsymbol{c} - \boldsymbol{c_0})\right\} \cdot \left(\frac{1}{\sigma_c^2}\right)^{a_c - 1} \exp\left\{-b_c \frac{1}{\sigma_c^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2\sigma_c^2}\sum_{i=1}^{N}(c_i - c_0)^2\right\} \cdot \left(\frac{1}{\sigma_c^2}\right)^{a_c + \frac{N}{2} - 1} \exp\left\{-b_c \frac{1}{\sigma_c^2}\right\} \\
&= \exp\left\{-\frac{1}{\sigma_c^2}\left[\frac{1}{2}\sum_{i=1}^{N}(c_i - c_0)^2 + b_c\right]\right\} \left(\frac{1}{\sigma_c^2}\right)^{a_c + \frac{N}{2} - 1}
\end{aligned}$$

Then:

$$\frac{1}{\sigma_c^2}|\boldsymbol{\theta}_{-\sigma_c^2}, \boldsymbol{Y} \sim \text{Gamma}(a_c^*, b_c^*)$$

$$a_c^* = a_c + \frac{N}{2} \qquad b_c^* = b_c + \frac{1}{2}\sum_{i=1}^{N}(c_i - c_0)^2$$

Since we are referring to the shape-rate Gamma parameterization:

$$\sigma_c^2|\boldsymbol{\theta}_{-\sigma_c^2}, \boldsymbol{Y} \sim \text{Inv-Gamma}(a_c^*, b_c^*)$$

Similarly, we get for σ_a^2:

$$\sigma_a^2|\boldsymbol{\theta}_{-\sigma_a^2}, \boldsymbol{Y} \sim \text{Inv-Gamma}(a_a^*, b_a^*)$$

$$a_a^* = a_a + \frac{N}{2} \qquad b_a^* = b_a + \frac{1}{2}\sum_{i=1}^{N}(a_i - a_0)^2$$

4.6 Full conditional of σ_ϵ^2

Similarly to what done in the previous sections, assume as prior for $\frac{1}{\sigma_\epsilon^2}$ a gamma distribution:

$$\frac{1}{\sigma_\epsilon^2} \sim Gamma(a_\epsilon, b_\epsilon)$$

Then the full conditional distribution for $\frac{1}{\sigma_\epsilon^2}$ is proportional to the product of a normal and a gamma distribution:

$$\pi\left(\frac{1}{\sigma_\epsilon^2}|\boldsymbol{\theta}_{-\sigma_\epsilon^2}, \boldsymbol{Y}\right) \propto \pi(\boldsymbol{Y}|\dots) \cdot \pi\left(\frac{1}{\sigma_\epsilon^2}|a_\epsilon, b_\epsilon\right)$$

$$\boldsymbol{Y}_i \sim \mathcal{N}\left(\underbrace{c_i \mathbb{1}_n + a_i \boldsymbol{\mathcal{B}}'_m(\mu_i(t))\boldsymbol{\beta}}_{\boldsymbol{m}_i}, \sigma_\epsilon^2 \boldsymbol{I}_N\right)$$

$$
\begin{aligned}
\implies \pi\left(\frac{1}{\sigma_\epsilon^2}|\boldsymbol{\theta}_{-\sigma_\epsilon^2}, \boldsymbol{Y}\right) \propto \cdot & \prod_{i=1}^{n} \frac{1}{\sqrt{\mathrm{Det}(\sigma_\epsilon^2 \boldsymbol{I}_N)^{-1}}} \exp\left\{-\frac{1}{2}(\boldsymbol{Y}_i - \boldsymbol{m}_i)'(\sigma_\epsilon^2 \boldsymbol{I}_N)^{-1}(\boldsymbol{Y}_i - \boldsymbol{m}_i)\right\} \\
& \cdot \left(\frac{1}{\sigma_\epsilon^2}\right)^{a_\epsilon - 1} \exp\left\{-b_\epsilon \frac{1}{\sigma_\epsilon^2}\right\} \\
& \propto \left(\frac{1}{\sigma_\epsilon^2}\right)^{a_\epsilon + \frac{\sum_{i=1}^N n_i}{2} - 1} \exp\left\{-\frac{1}{\sigma_\epsilon^2}\left[b_\epsilon + \frac{1}{2}\sum_{i=1}^{N}\left(\boldsymbol{Y}_i - \boldsymbol{m}_i\right)'\left(\boldsymbol{Y}_i - \boldsymbol{m}_i\right)\right]\right\}
\end{aligned}
$$

Then:

$$
\frac{1}{\sigma_\epsilon^2}|\boldsymbol{\theta}_{-\sigma_\epsilon^2}, \boldsymbol{Y} \sim Gamma(a_\epsilon^*, b_\epsilon^*)
$$

$$
a_\epsilon^* = a_\epsilon + \frac{\sum_{i=1}^N n_i}{2} \qquad b_\epsilon^* = b_\epsilon + \frac{1}{2}\sum_{i=1}^{N}(\boldsymbol{Y}_i - \boldsymbol{m}_i)'(\boldsymbol{Y}_i - \boldsymbol{m}_i)
$$

Since we are referring to the shape-rate gamma parameterization:

$$
\frac{1}{\sigma_\epsilon^2}|\boldsymbol{\theta}_{-\sigma_\epsilon^2}, \boldsymbol{Y} \sim Inv - Gamma(a_\epsilon^*, b_\epsilon^*)
$$

5 Gibbs-within-Metropolis algorithm

In the previous section, we have shown that all but two of the parameters allow for a closed-form full conditional distribution. The exception is for parameters f and g. Thus, we decided to implement a Gibbs algorithm with a step of Metropolis-Hastings in order to sample from the posterior distributions of f and g.

5.1 Metropolis-Hastings step

Assume:

- f_i and g_i to be independent $\forall i$, i.e. $\pi(f_i, g_i) = \pi(f_i)\pi(g_i) \quad \forall i$.
- Normal and truncated normal prior distributions respectively for f_i and g_i:

$$f_i \sim \mathcal{N}(f_0, \sigma_f^2) \qquad g_i \sim \mathcal{N}(g_0, \sigma_g^2)\mathbb{1}_{(0,+\infty)} \qquad \forall i.$$

Observe that the parameter g_i needs to be positive in order to avoid time reversion.
In this way, we have a one-to-one correspondence between the transformed times and the original ones. Then the posterior distributions of f_i and g_i are given by:

$$\pi(f_i|\boldsymbol{Y}) \propto \pi(\boldsymbol{Y}|f_i) \cdot \pi(f_i) \qquad \pi(g_i|\boldsymbol{Y}) \propto \pi(\boldsymbol{Y}|g_i) \cdot \pi(g_i) \qquad \forall i$$

Finally, we chose as proposal:

$$q(\cdot|f_i^{(t-1)}) = \mathcal{N}\left(f_i^{(t-1)}, \tilde{\sigma_f}^2\right) \qquad q(\cdot|g_i^{(t-1)}) = \mathcal{N}\left(g_i^{(t-1)}, \tilde{\sigma_g}^2\right) \quad \forall i$$

5.2 Adaptive Metropolis-Hastings

It is well known that, when dealing with Metropolis-Hastings algorithm, the key point is the choice of the proposal distribution and, in particular, the setting of its variance.
At first, we kept the proposal variances $\tilde{\sigma_f}^2$ and $\tilde{\sigma_g}^2$ as fixed.
However, the exploration of the state space was not satisfactory at all, so we decided to go for an adaptive approach. Specifically, we set the variance of the proposal as constant until a specific time t_0 and then we started to estimate it using all the previous states:

$$q(\cdot|X^{(1)}, \dots, X^{(t-1)}) \sim \mathcal{N}(X^{(t-1)}, \sigma_t^2)$$
$$\sigma_t^2 = \begin{cases} \sigma_0^2 & t \le t_0 \\ \frac{1}{t-1}\left(\sum_{j=0}^{t-1} x_j^2 - t \cdot \bar{x}^2\right) & t > t_0 \end{cases}$$

The choice for the length of the initial segment t_0 is free, but, the bigger it is chosen, the more slowly the effect of the adaptation is felt. In a sense, the size of t_0 reflects our trust in the initial variance σ_0^2. We set t_0 equal to 2000.
The just presented adaptive approach increased a lot the exploration of the state space of parameters f and g. In this way, we have been able to perform a proper posterior inference on these two parameters as well.

6 Fake data simulation

The dataset includes about 2 thousand time observations for each jump. As a result of both the dense sampling and the high number of parameters, the code was very slow and it would have taken days to provide the output. Therefore, fitting the model directly to our data was definitely not feasible.
To overcome this issue, we decided first to fit the model to some simulated data and then to switch to the original ones just when everything was working properly.
Specifically, we simulated a set of smooth curves as transformations of a common Beta density function. Of course, we used a much less thick sampling in order to decrease the computational cost of the algorithm.
Once we built up the new dataset, we started working on it step by step. Our approach was to go simpler and simpler until we got the whole algorithm to work.
We first checked the algorithm by applying it to:

- Data obtained from the same base function modified just via shape parameters a and c:

$$y(\boldsymbol{t}) = c\mathbb{1}_n + a \cdot \boldsymbol{\mathcal{B}}'_m(\boldsymbol{t})\boldsymbol{\beta} + \epsilon$$

- Data obtained from the same base function modified just via time transformation parameters f and g:

$$y(\boldsymbol{t}) = \boldsymbol{\mathcal{B}}'_m(f\mathbb{1}_n + g \cdot \boldsymbol{t})\boldsymbol{\beta} + \epsilon$$

Once we got both the algorithm implemented in these two preliminary steps working, we switched to the original simulated data and we tested the whole algorithm.
As can be seen in Figure 5, the curves are reconstructed and aligned pretty well.

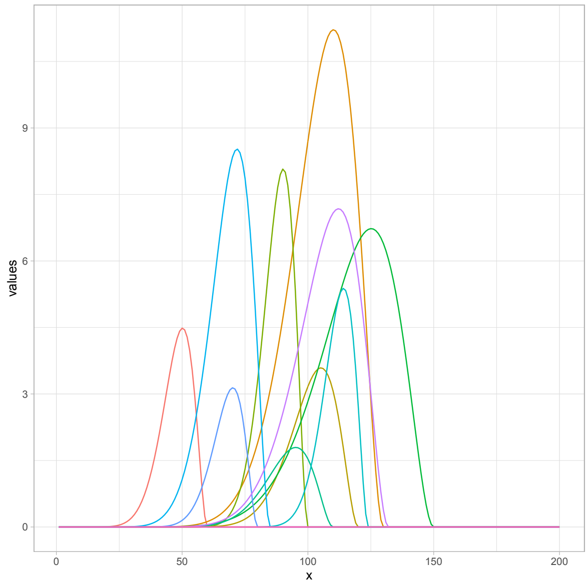


Figure 4: Simulated Curves

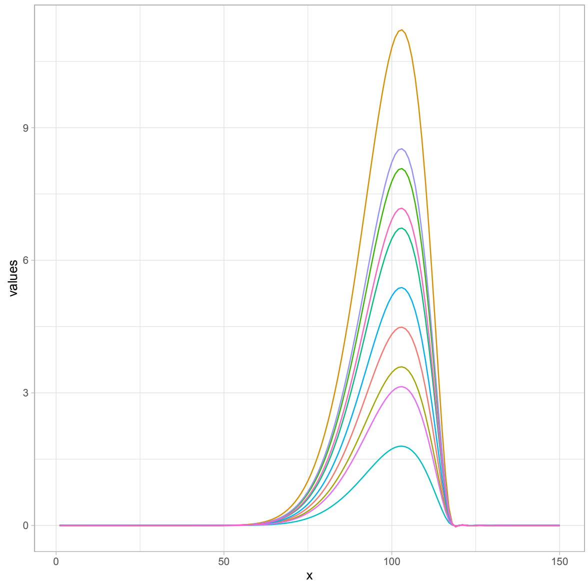


Figure 5: Aligned Curves

7 Identifiability of the chain

At this point, our posterior inference on parameters a and c was not satisfactory. Indeed, getting an iid sample from their posterior distributions would have required a lot of iterations.
The trouble was that the chain was not identifiable. Namely:

- The product of the parameters a and $\boldsymbol{\beta}$ could be synthesized into one single parameter. For this reason, the common shape function as presented in the last chapters turns out to be defined just up to the product to a constant. This provides an additional degree of freedom to the model.

- The algorithm estimates the common shape function in such a way that its value in the origin could be any value. Then, the common shape function is also defined up to an additional constant and this provides an extra degree of freedom to the model.

To solve the first issue, we fix a_1 and specifically we set $a_1 = 1$. In this way, the amplitude of the first curve results to be defined just through the product , i.e. the first curve and the common shape function have the same amplitude.
In the following plot we see an example of curves with different amplitudes, in which we fix the one with a_1=1.
To solve the second issue, we add a constraint that fixes the position of the common shape function. In order to do it, we set $\beta_0 = 0$.
In the following plot we see an example of curves with different intercept, in which we fix the one with β_0=0.
Thanks to these two adjustments, the problem is no more over-parametrized.

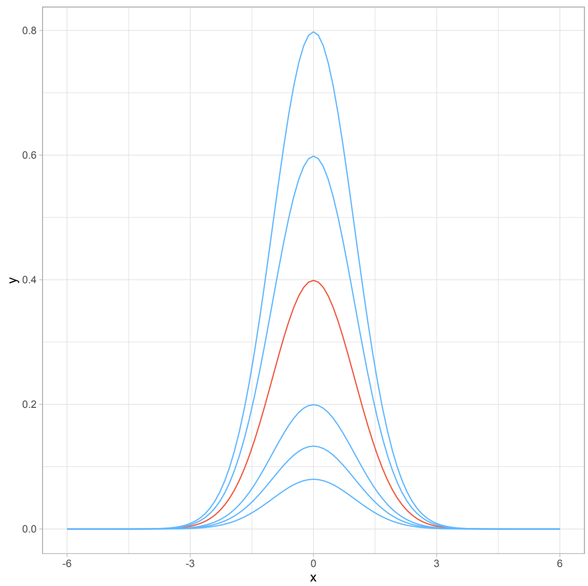


Figure 6: $a_1 = 1$

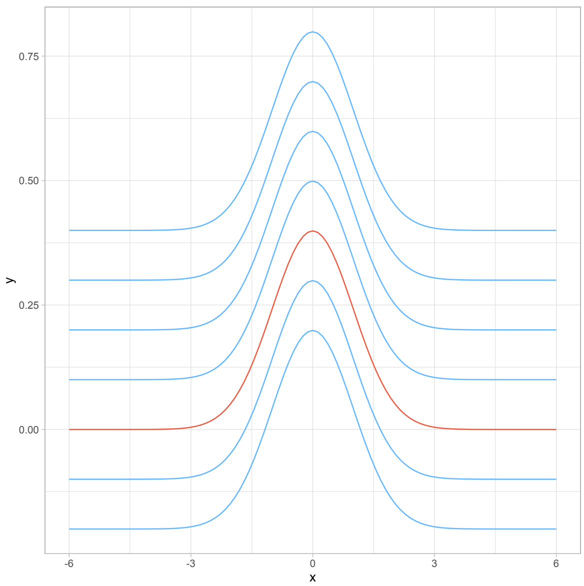


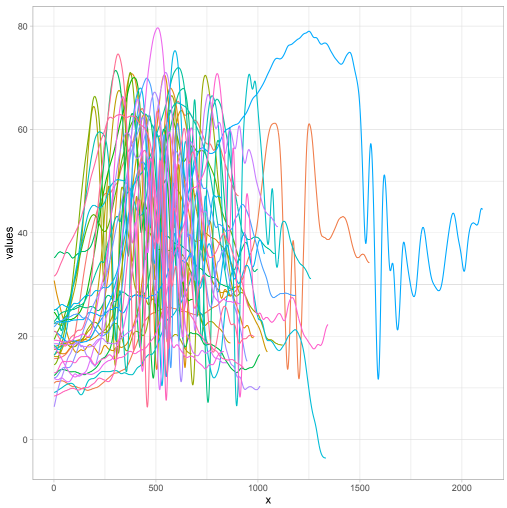
Figure 7: $\beta_0 = 0$

8 Posterior inference

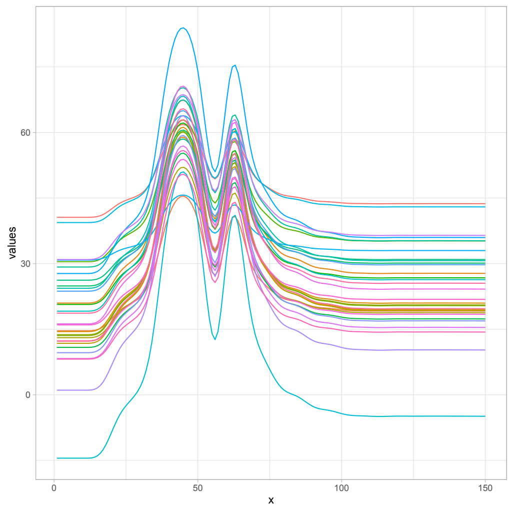
Now we finally come to the transition from the fake data to the original ones. Our inferences are based on 15000 samples from the posterior distribution obtained after discarding the initial 15000 MCMC iterations for burn-in. The posterior inference is presented separately for each group.

8.1 Healthy people

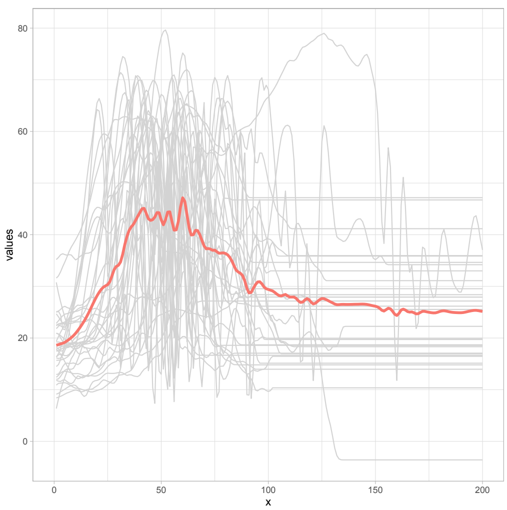
Figure 8 - Panel (c) shows trajectories superimposed with a cross-sectional mean. The cross-sectional mean curve does not resemble any of the individual curves, because phase variation is not taken into account.
Using draws from the posterior distribution of the model parameters we computed the posterior scaled shape function following the procedure described in Section 3.
Figure 8 - Panel (d) shows registered trajectories superimposed with the posterior mean scaled shape function. The figure shows that, accounting for phase variability, we are now able to recover the features of the originating signal.



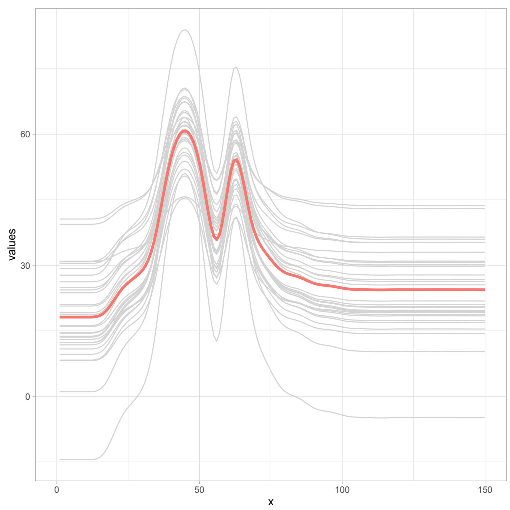
(a) *Original curves*



(b) *Aligned curves*



(c) *Original curves, superimposed with the cross-sectional mean*

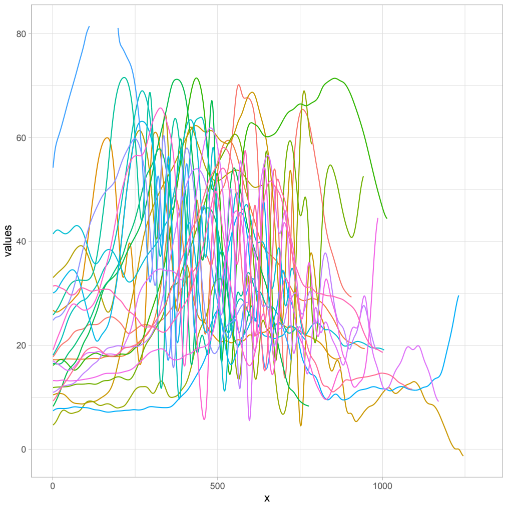


(d) *Aligned curves, superimposed with the posterior mean of the scaled shape function*

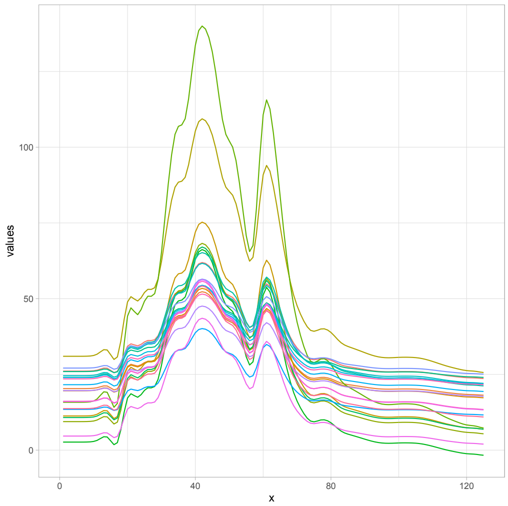
Figure 8

8.2 People who had physiotherapy

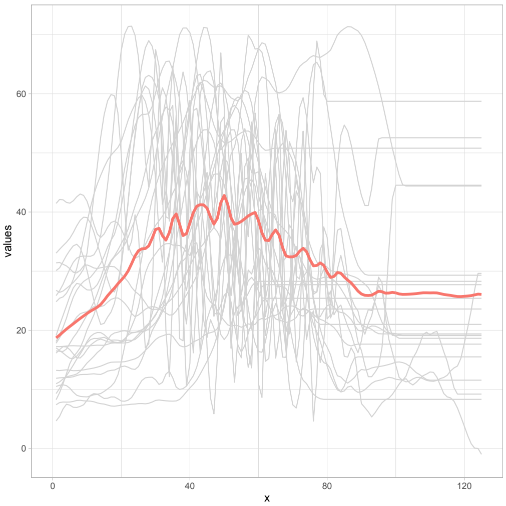
Similar reasoning applies as provided in Section 8.1 for the group of people who had physiotherapy.
Our posterior inference is reported below.



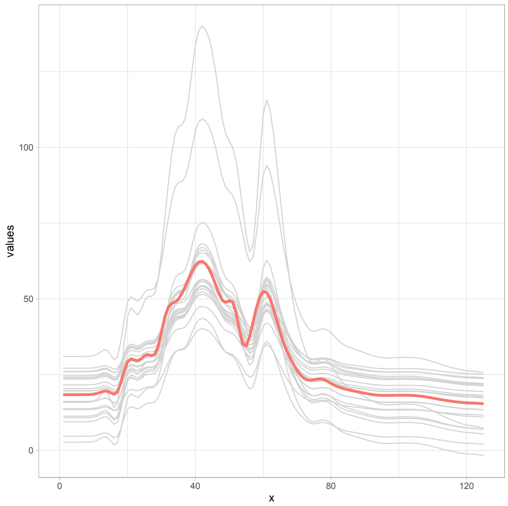
(a) *Original curves*



(b) *Aligned curves*



(c) *Original curves, superimposed with the cross-sectional mean*

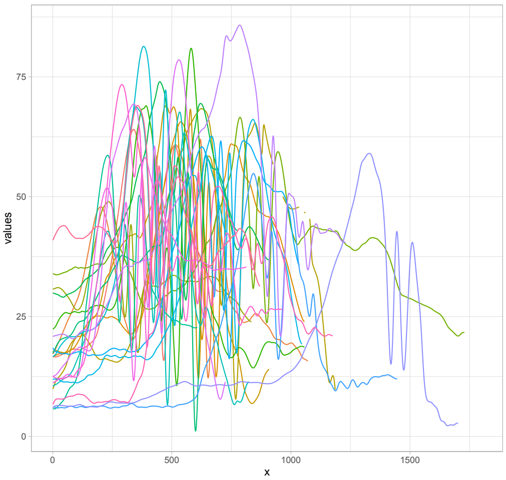


(d) *Aligned curves, superimposed with the posterior mean of the scaled shape function*

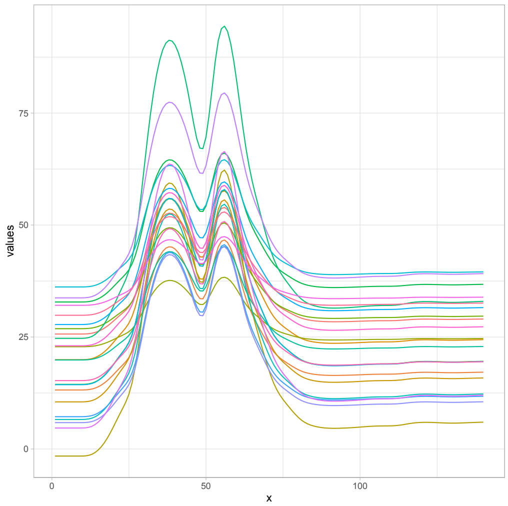
Figure 9

8.3 People who just had surgery

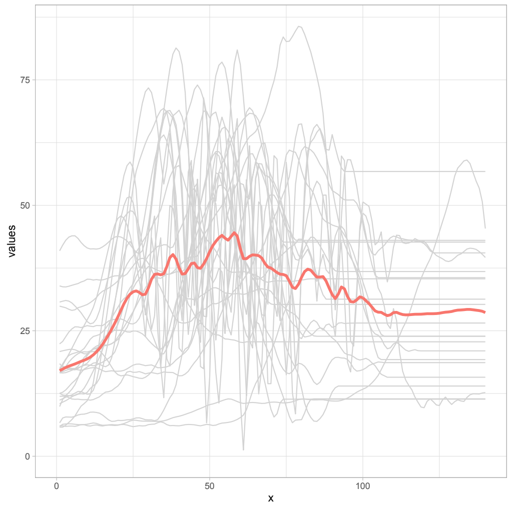
Similar reasoning applies as provided in Section 8.1 for the group of people who just had surgery. Our posterior inference is reported below.



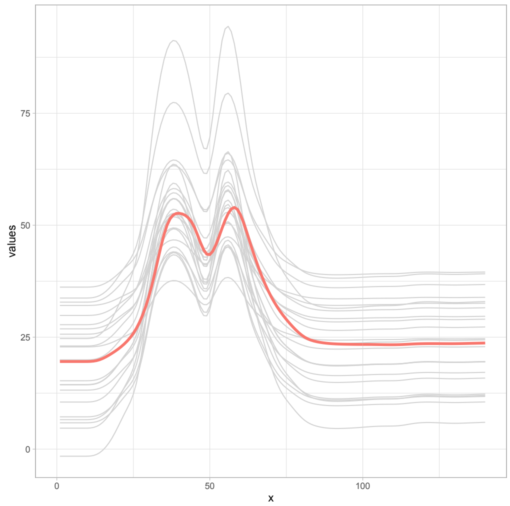
(a) *Original curves*



(b) *Aligned curves*



(c) *Original curves, superimposed with the cross-sectional mean*

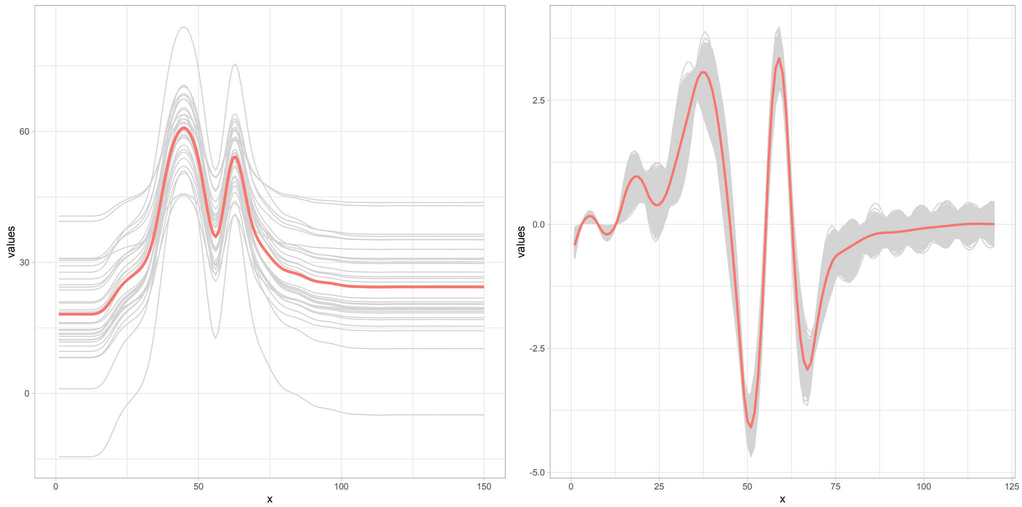


(d) *Aligned curves, superimposed with the posterior mean of the scaled shape function*

Figure 10

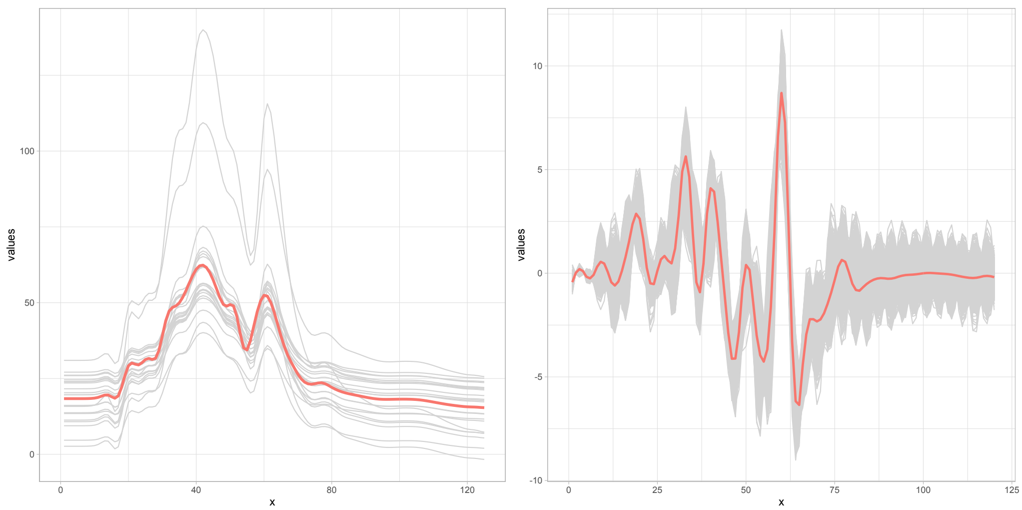
8.4 Comparison among groups

Now we are going to compare what we got from the individual analysis of each group.
In order to do this, we computed the mean posterior derivative in each group and we found out that they were quite different. In particular, the mean posterior derivative of the group of people who had physiotherapy looks different from the ones of the other two groups.
Indeed, it has many more bumps with respect to the others. Oscillations in the derivative could be interpreted as having a weak knee, i.e. a knee that is not able to support one leg hoop in a solid way. It seems like the knee shakes during the jump.
Moreover, notice that the derivative of the group of healthy people shows an higher first downwards peak with respect to the one who had surgery. Therefore, people who had surgery flex the knee less quickly than the healthy people. Our interpretation is that people who had surgery could be afraid of jumping due to their past experiences and this reflects in their slower movements.



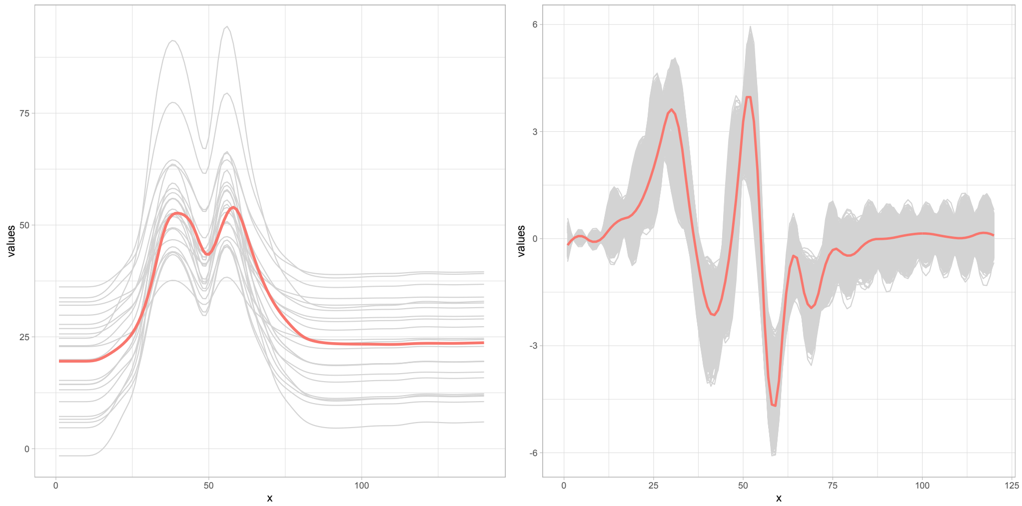
(a) *Aligned curves* (b) *Derivatives of the aligned curves*

Figure 11: Healthy people.



(a) *Aligned curves.* (b) *Derivatives of the aligned curves.*

Figure 12: People who underwent physiotherapy.



(a) *Aligned curves.*

(b) *Derivatives of the aligned curves.*

Figure 13: People who underwent surgery.

9 Conclusions

We achieved our very first goal: the algorithm works pretty well and it alignes the curves properly.
But why should we go for an alignment model in a bayesian framework instead of one in a frequentist framework? The plus of our model is that it does not requires pre-smoothing of the data, that often implies a loss of information.

We were told that doctors believe there are differences among people who had physiotherapy and the other two groups. From our analysis we have evidence to support this medical belief. Indeed, the differences in the posterior mean derivatives in each group is quite evident: the physiotherapy group is characterized by a much more oscillating derivative. As we already said in Section 8.4, oscillations in the derivative could be interpreted as having a weak knee, i.e. a knee that is not able to support one leg hoop in a solid way. It seems like the knee shakes during the jump. Moreover, a weaker difference can be pointed out between the group of healthy people and the one of people who had surgery. Namely, notice that the derivative of the group of healthy people shows an higher first downwards peak with respect to the one who had surgery. Therefore, people who had surgery flex the knee less quickly than the healthy people. Our interpretation is that people who had surgery could be afraid of jumping due to their past experiences and this reflects in their slower movements.

A further possible development of the project is related to the Metropolis-Hastings step. Thanks to the the adaptive approach presented in Section 5.2, the posterior inference on the time transformation parameters f and g is reasonably good, but it could be better. In order to improve the model, the proposal density could be chosen by exploiting some prior knowledge of data.

10 GitHub

Visit our GitHub repository at the following link:
`https://github.com/PrincipeFederica/Bayesian-Principe-Mattina-Bighignoli`

11 Additional plots

11.1 Parameter c

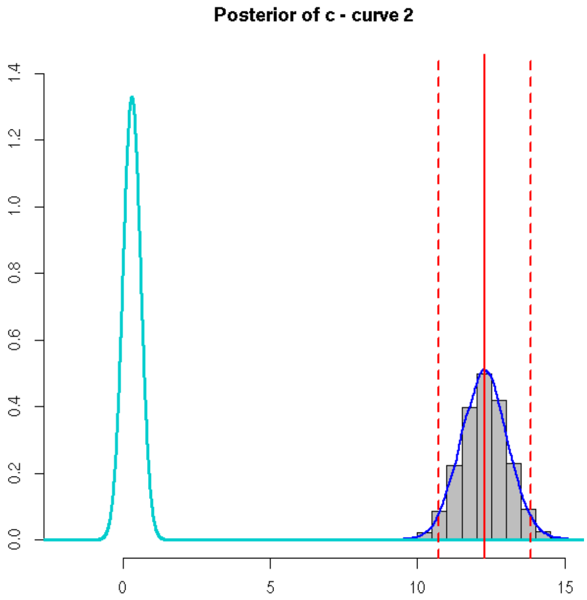


Figure 14: Second curve - Prior (cyan) and posterior (blue) of parameter c.

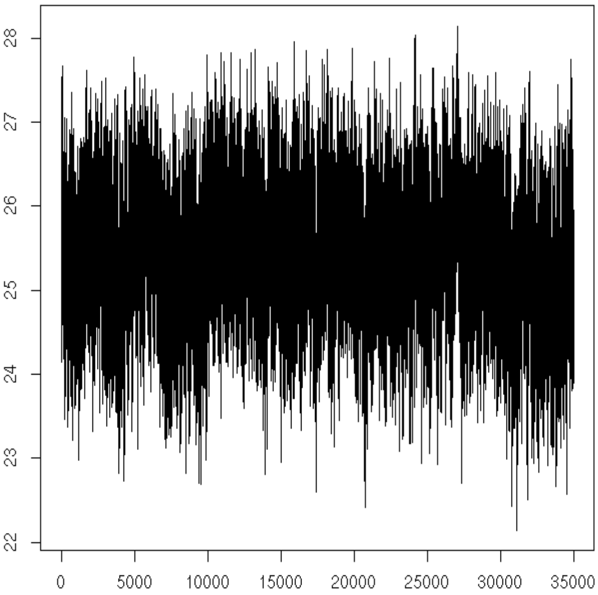


Figure 15: Second curve - Traceplot of parameter c.

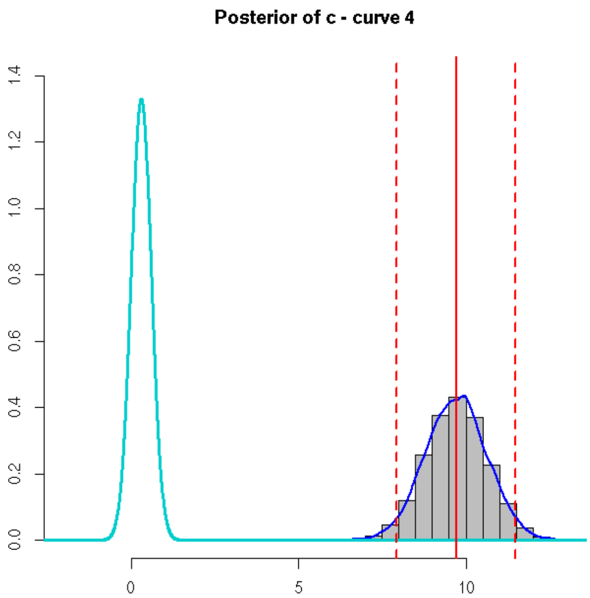


Figure 16: Fourth curve - Prior (cyan) and posterior (blue) of parameter c.

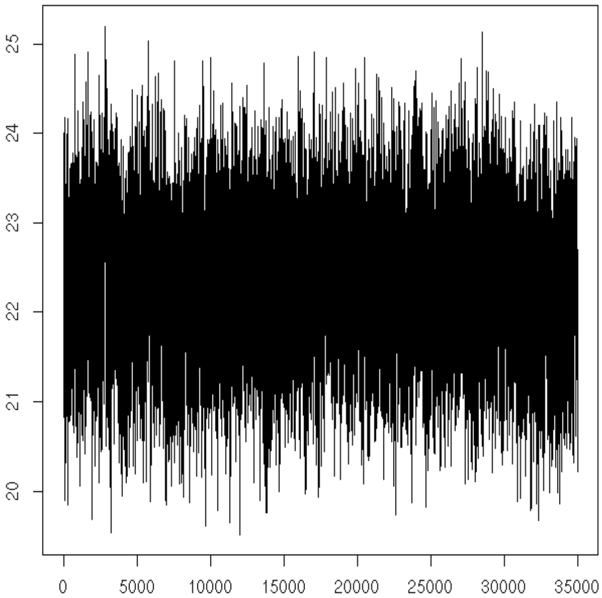


Figure 17: Fourth curve - Traceplot of parameter c.

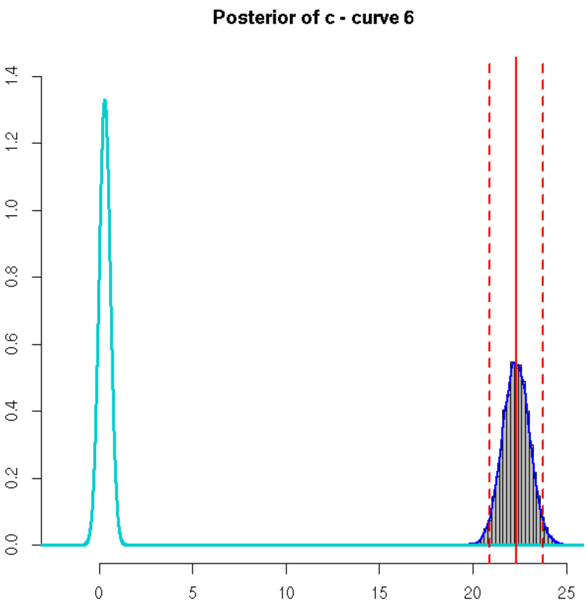


Figure 18: Sixth curve - Prior (cyan) and posterior (blue) of parameter c.

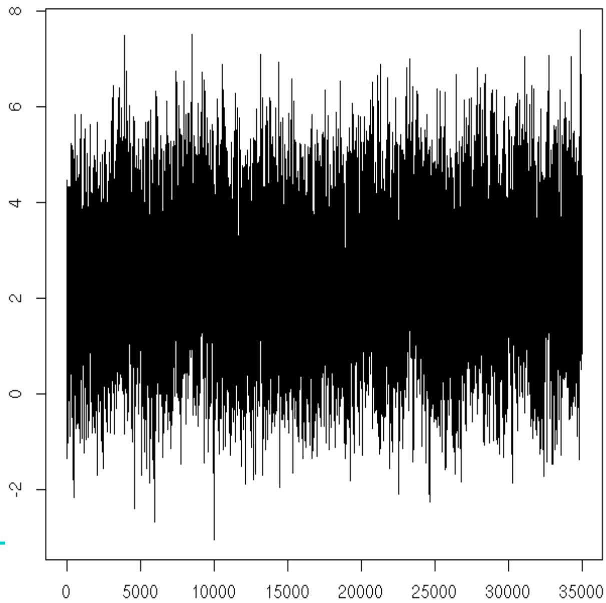


Figure 19: Sixth curve - Traceplot of parameter c.

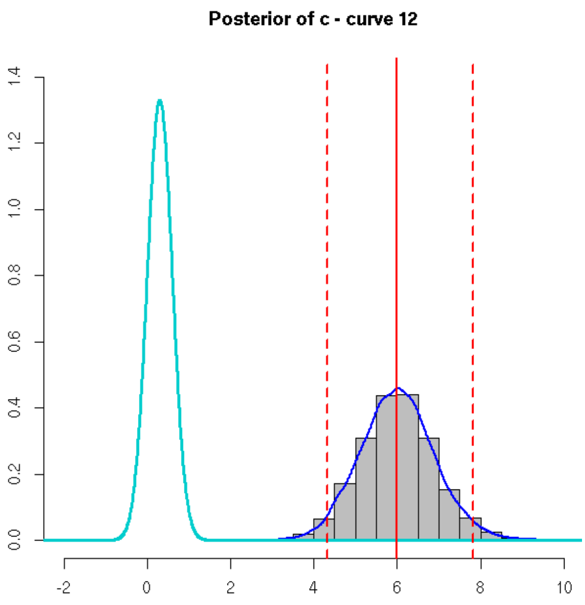


Figure 20: Twelfth curve - Prior (cyan) and posterior (blue) of parameter c.

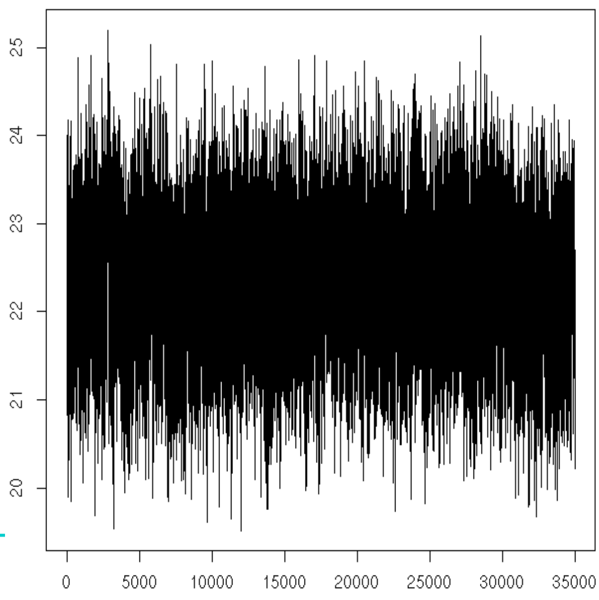


Figure 21: Twelfth curve - Traceplot of parameter c.

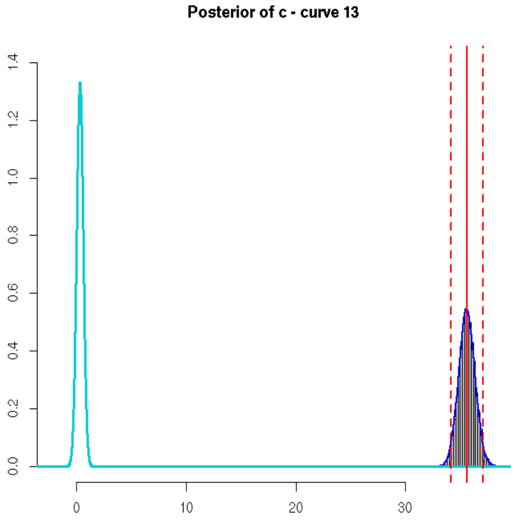


Figure 22: Thirtheenth curve - Prior (cyan) and posterior (blue) of parameter c.

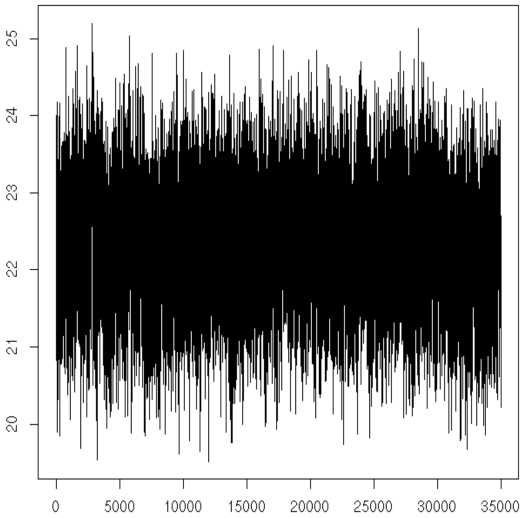


Figure 23: Thirtheenth curve - Traceplot of parameter c.

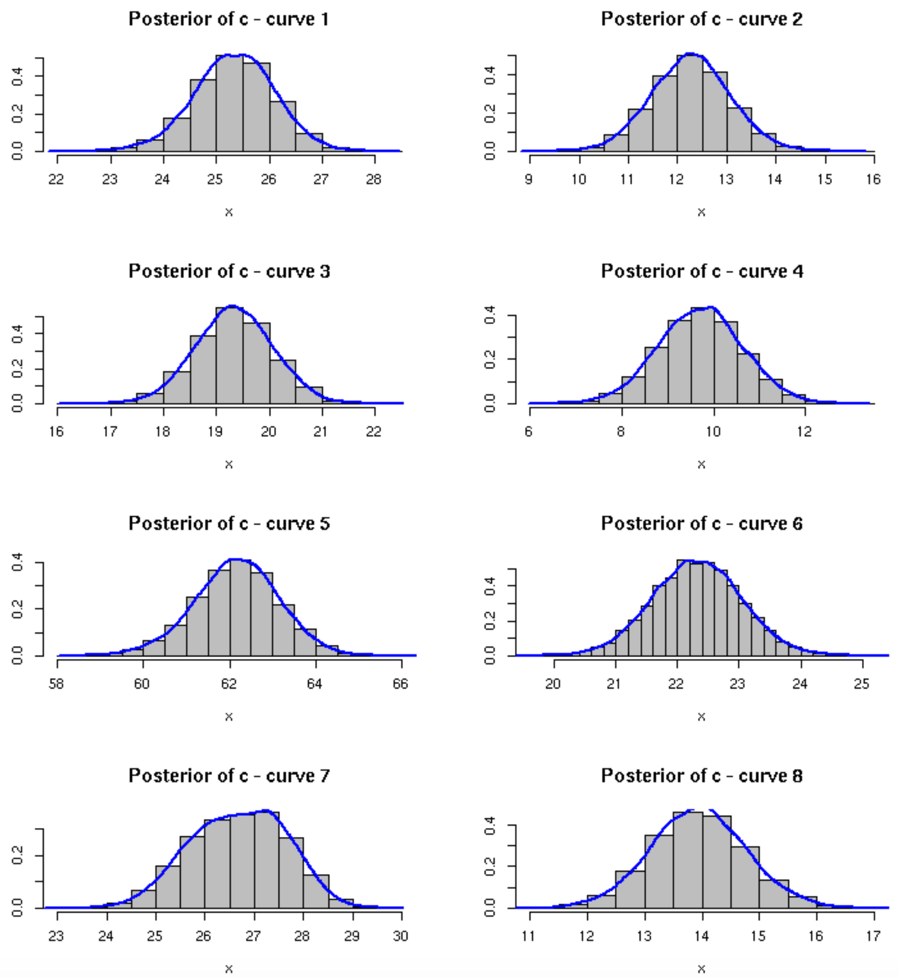


Figure 24: Some posterior of c

11.2 Parameter a

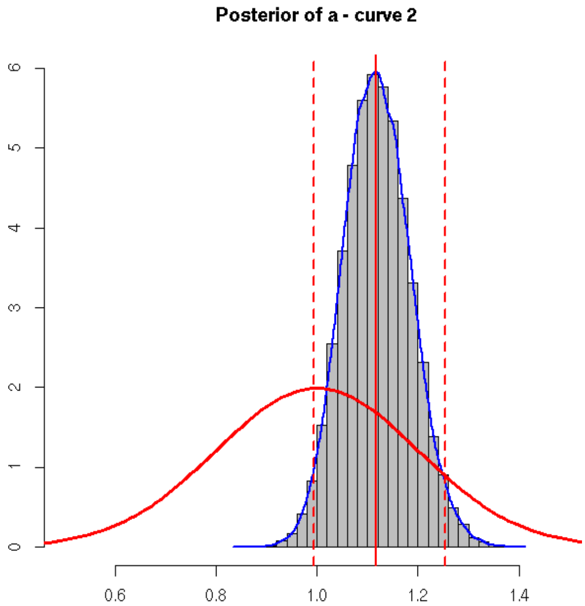


Figure 25: Second curve - Prior (red) and posterior (blue) of parameter a.

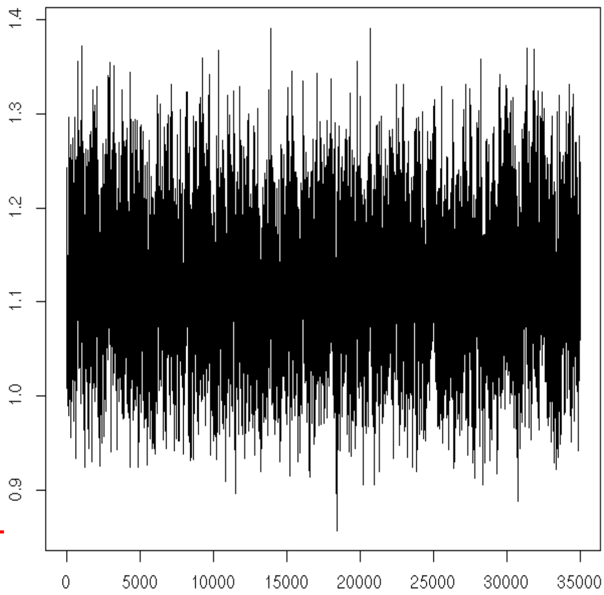


Figure 26: Second curve - Traceplot of parameter a.

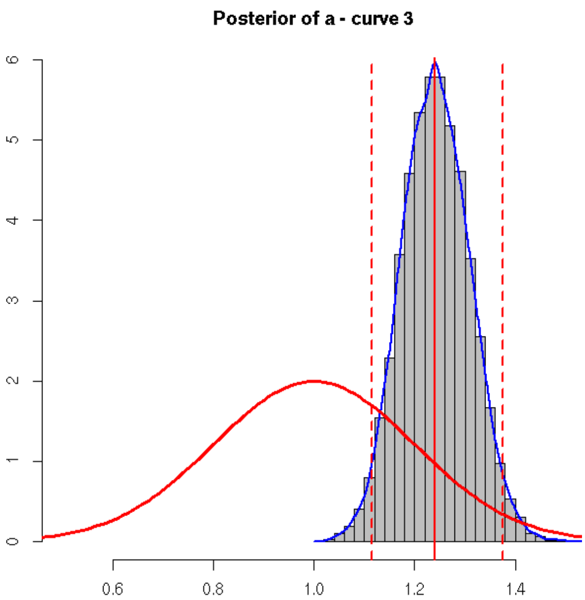


Figure 27: Third curve - Prior (red) and posterior (blue) of parameter a.

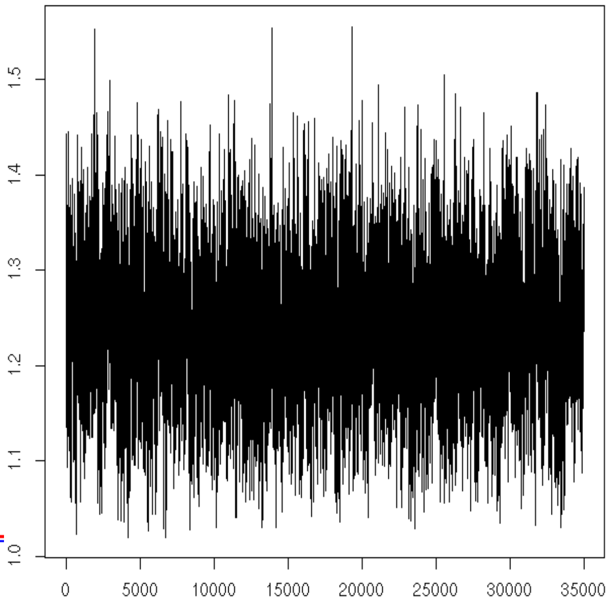


Figure 28: Third curve - Traceplot of parameter a.

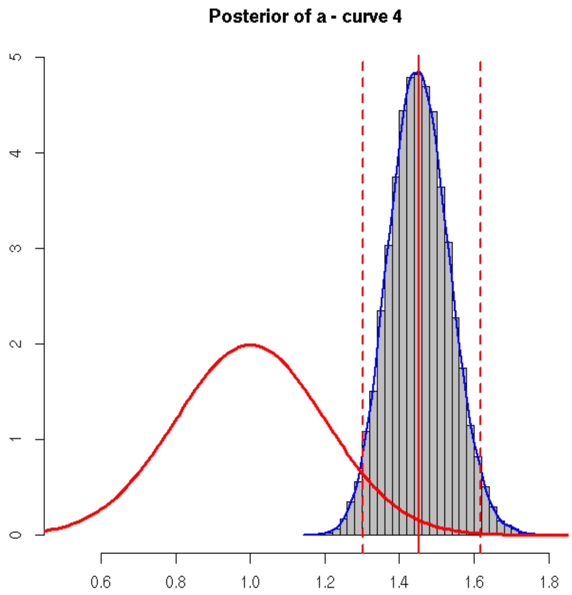


Figure 29: Fourth curve - Prior (red) and posterior (blue) of parameter a.

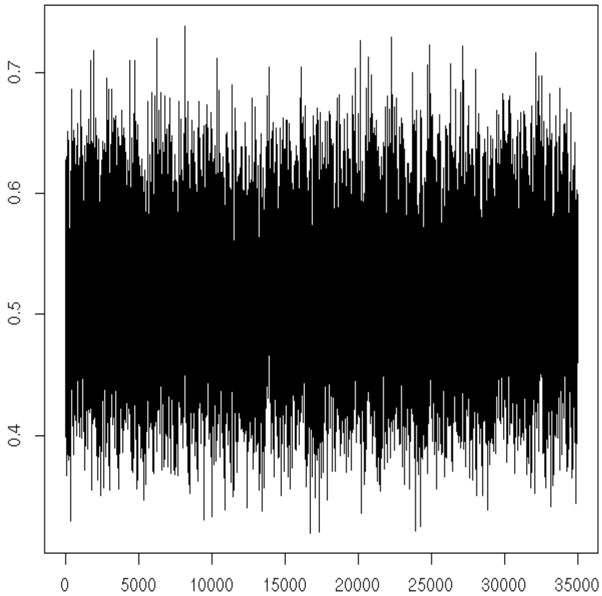


Figure 30: Fourth curve - Traceplot of parameter a.

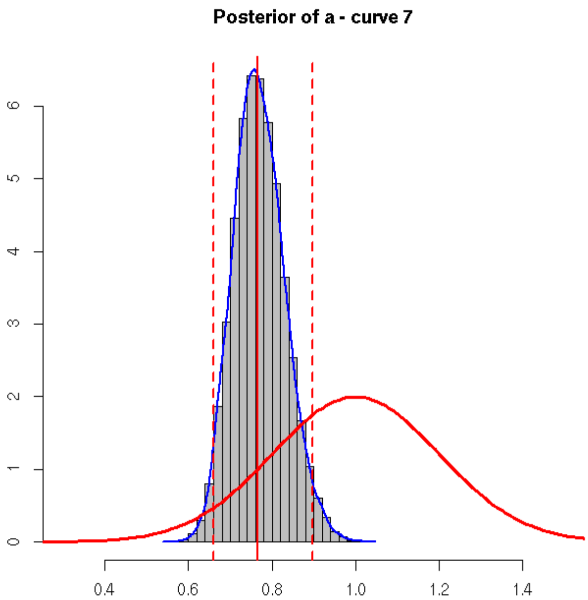


Figure 31: Seventh curve - Prior (red) and posterior (blue) of parameter a.

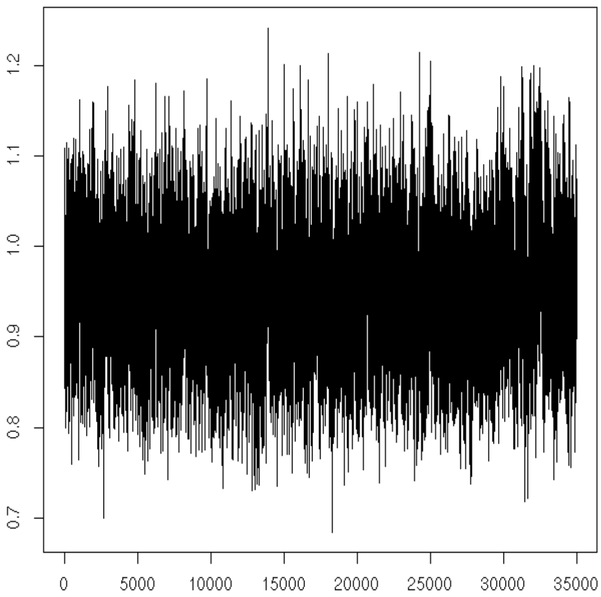


Figure 32: Seventh curve - Traceplot of parameter a.

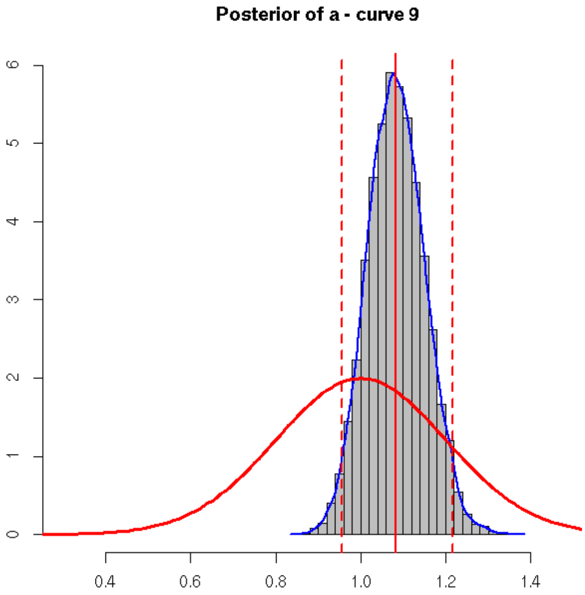


Figure 33: Nineth curve - Prior (red) and posterior (blue) of parameter a.

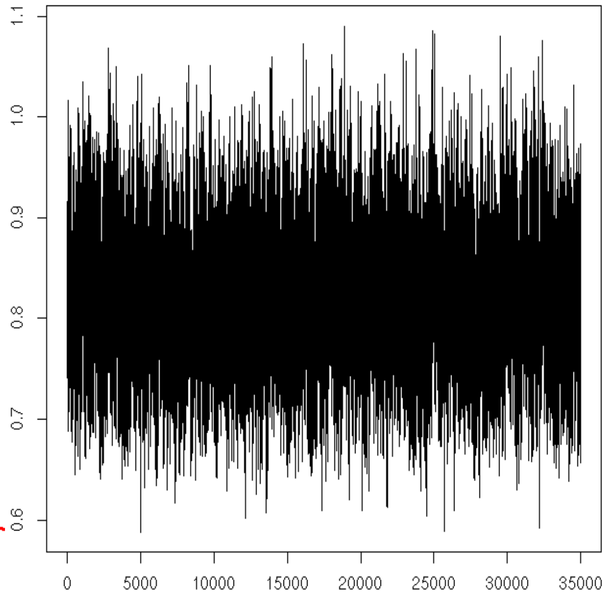


Figure 34: Nineth curve - Traceplot of parameter a.

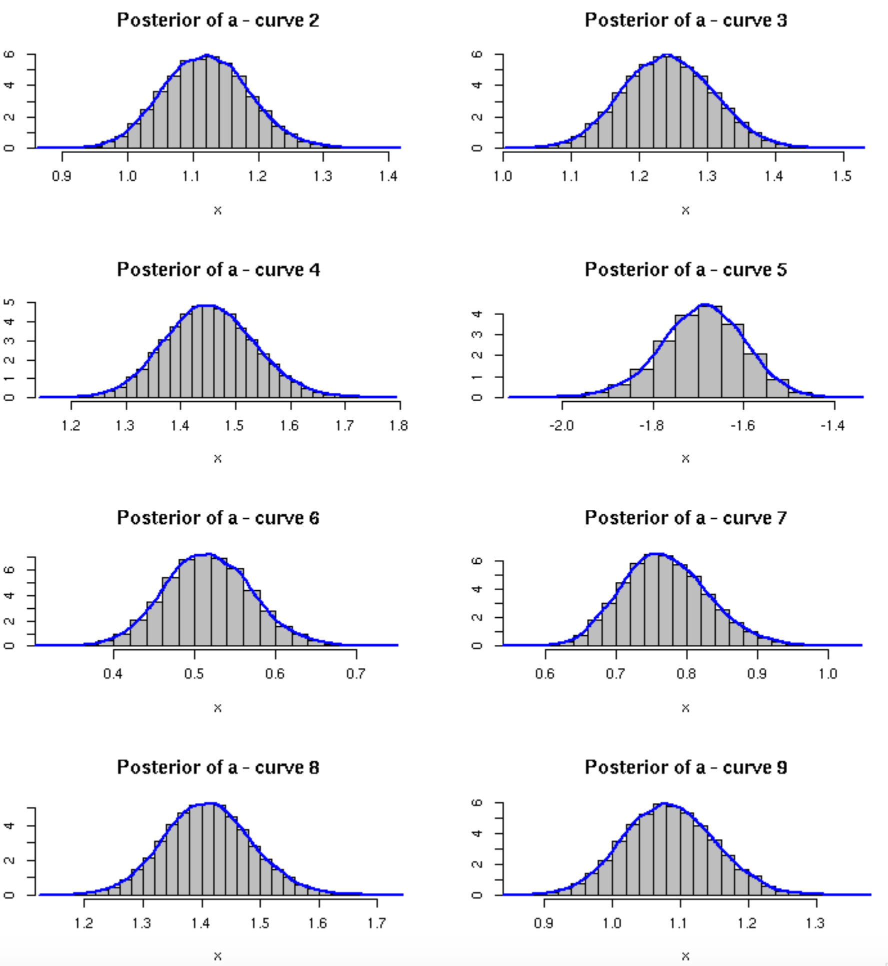


Figure 35: Some posterior of a

11.3 Parameter f

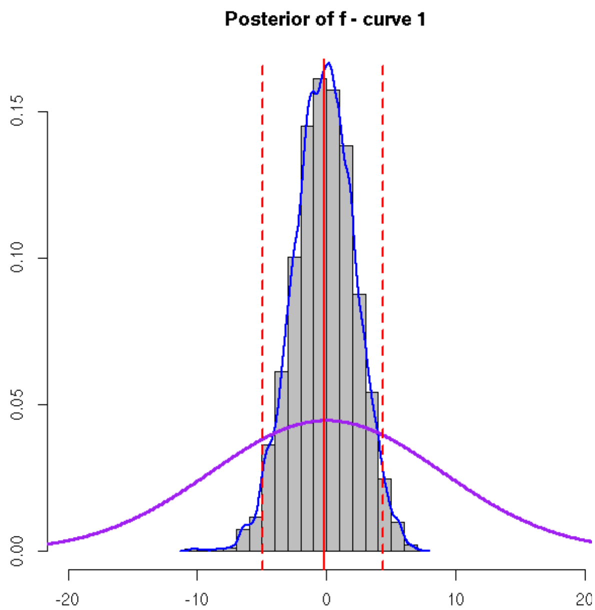


Figure 36: First curve - Prior (purple) and posterior (blue) of parameter f.

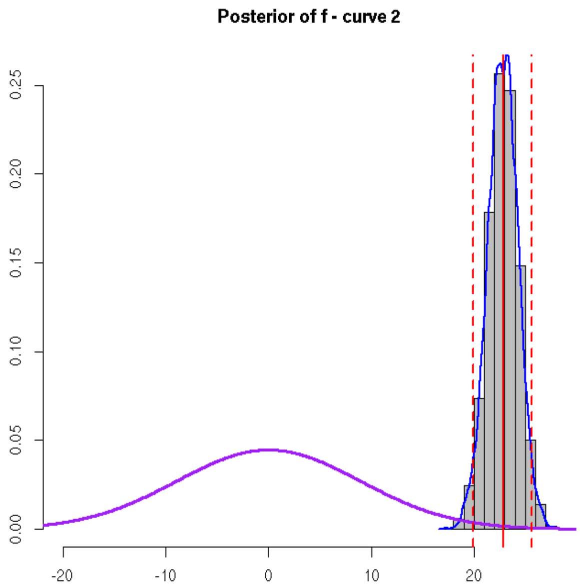


Figure 37: Second curve - Prior (purple) and posterior (blue) of parameter f.

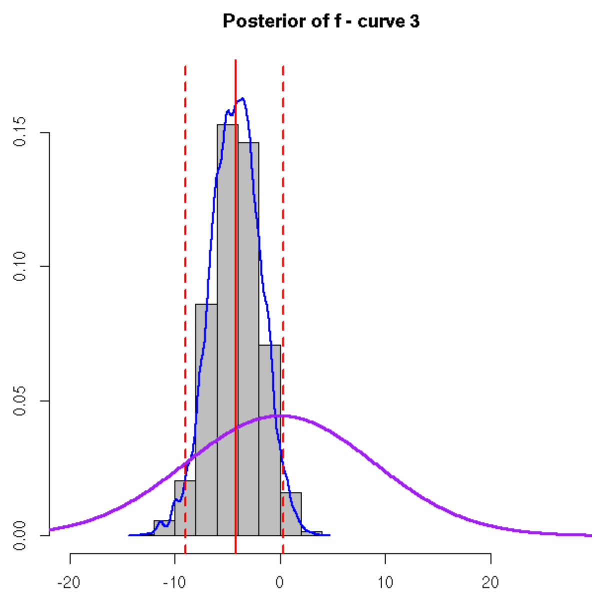


Figure 38: Third curve - Prior (purple) and posterior (blue) of parameter f.

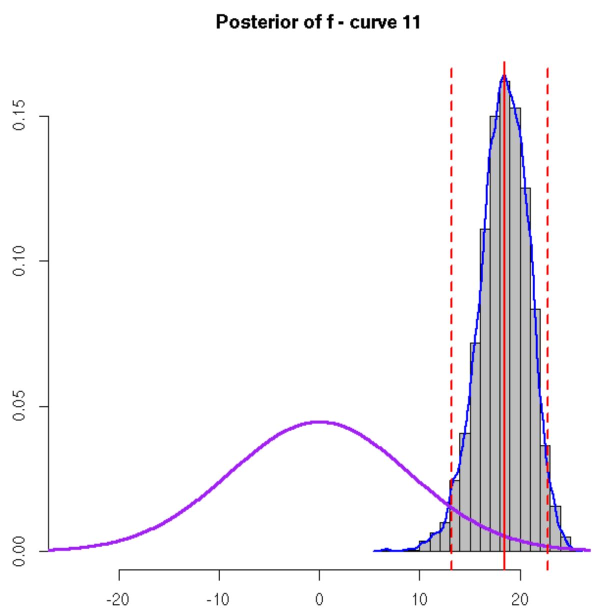


Figure 39: Eleventh curve - Prior (purple) and posterior (blue) of parameter f.

11.4 Parameter g

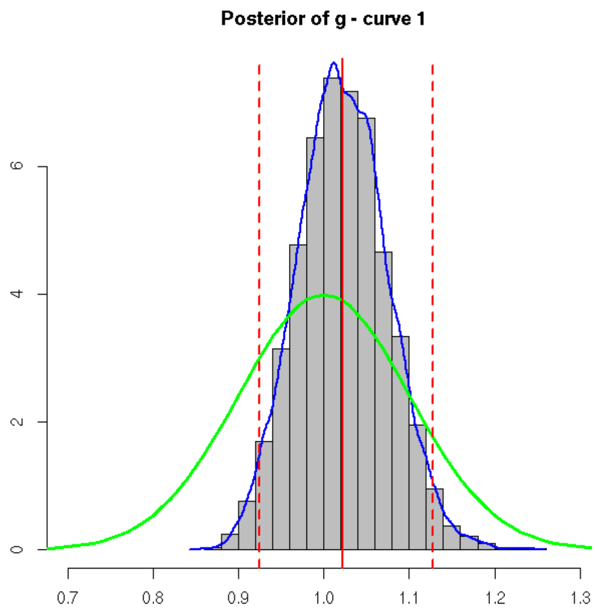


Figure 40: First cuve - Prior (green) and posterior (blue) of parameter g.

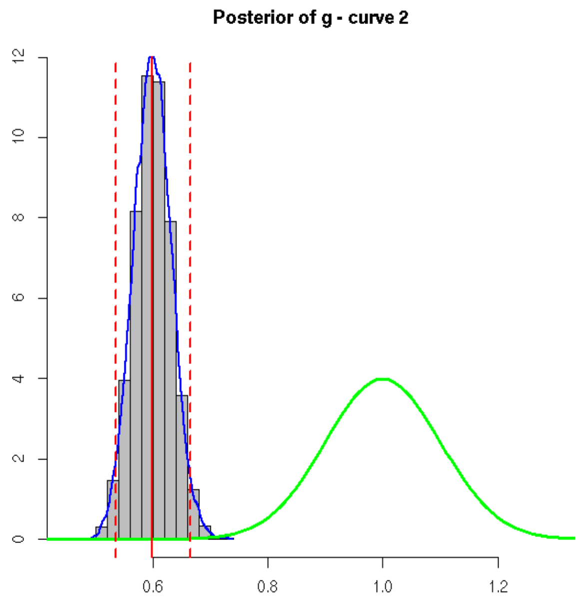


Figure 41: Second curve - Prior (green) and posterior (blue) of parameter g.

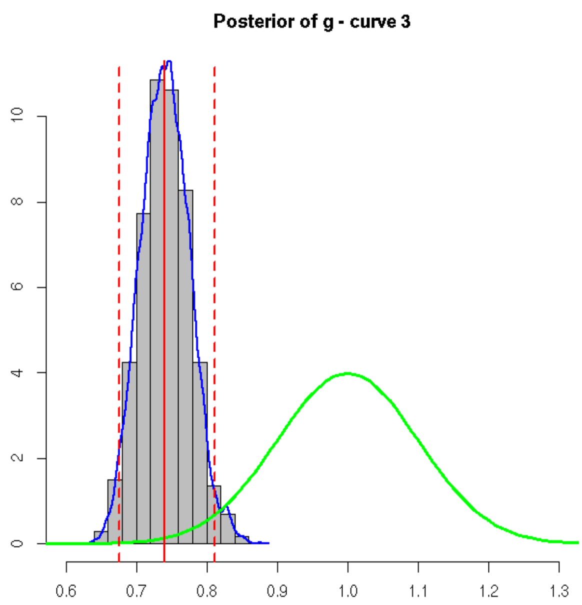


Figure 42: Third curve - Prior (green) and posterior (blue) of parameter g.

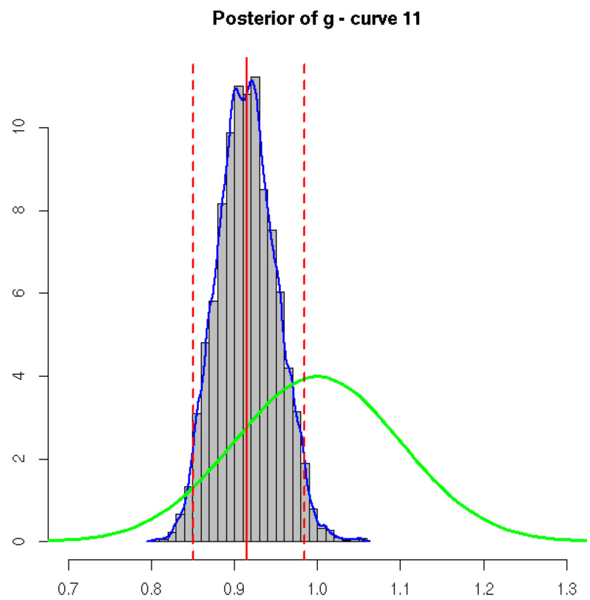


Figure 43: Eleventh curve - Prior (green) and posterior (blue) of parameter g.

11.5 Parameter c0

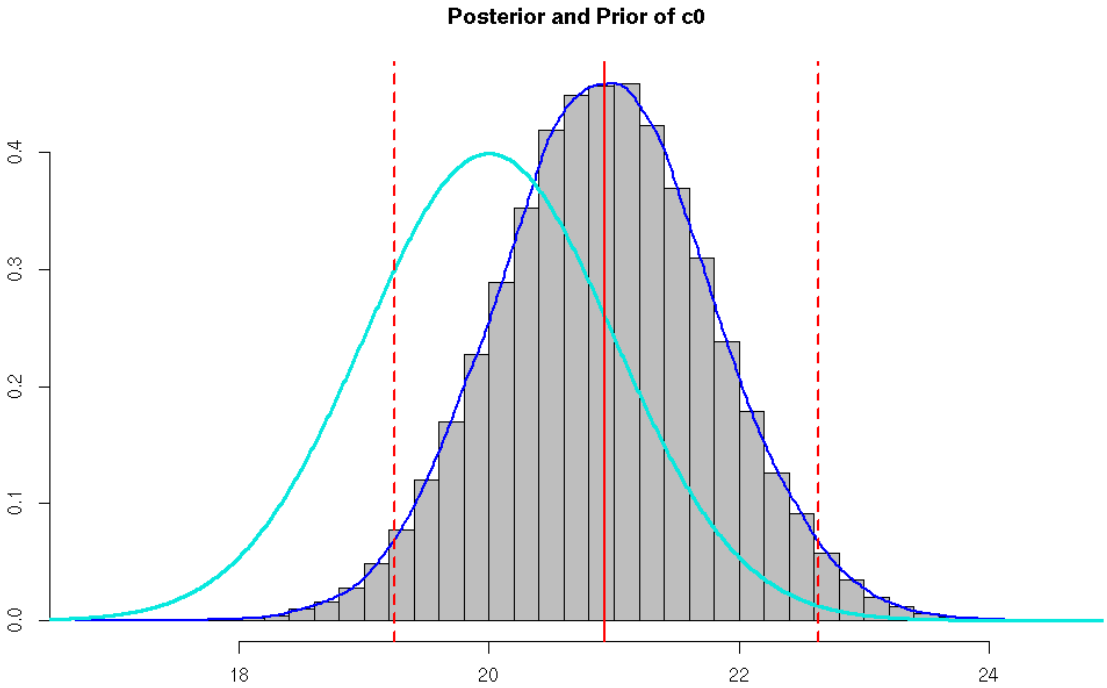


Figure 44: Prior (cyan) and posterior (blue) of parameter c_0.

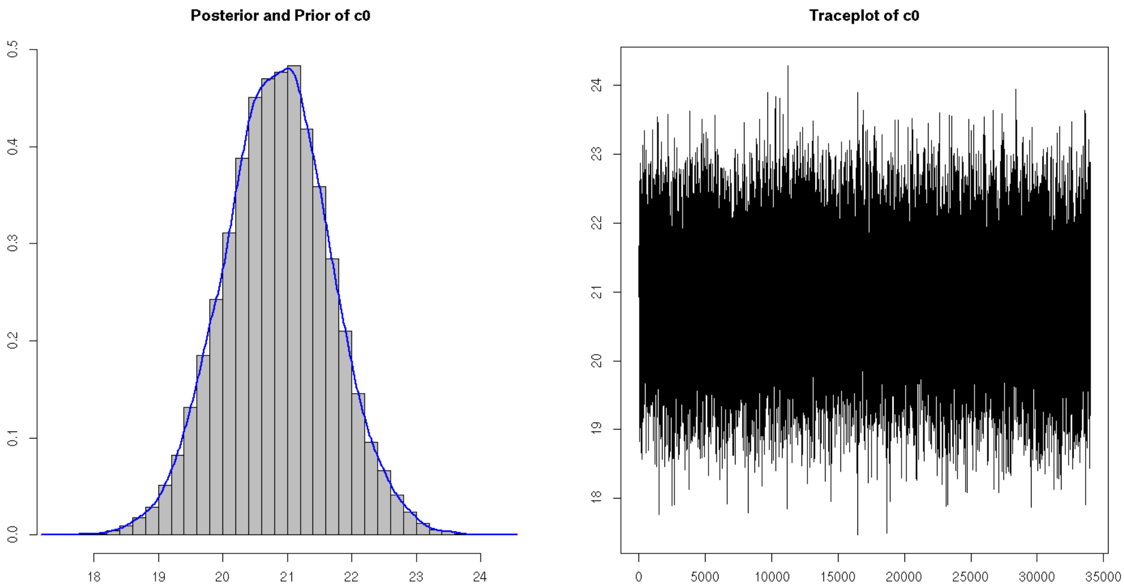


Figure 45: Posterior and Traceplot for parameter c_0

11.6 Parameter a0

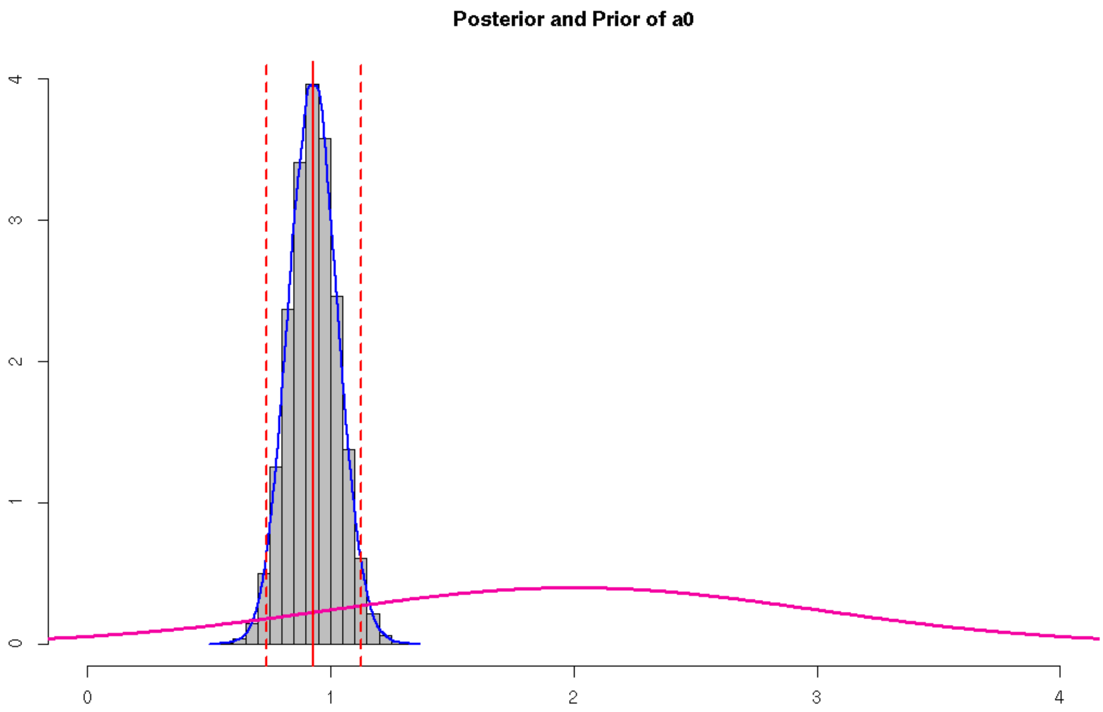


Figure 46: Prior (pink) and posterior (blue) of parameter a_0.

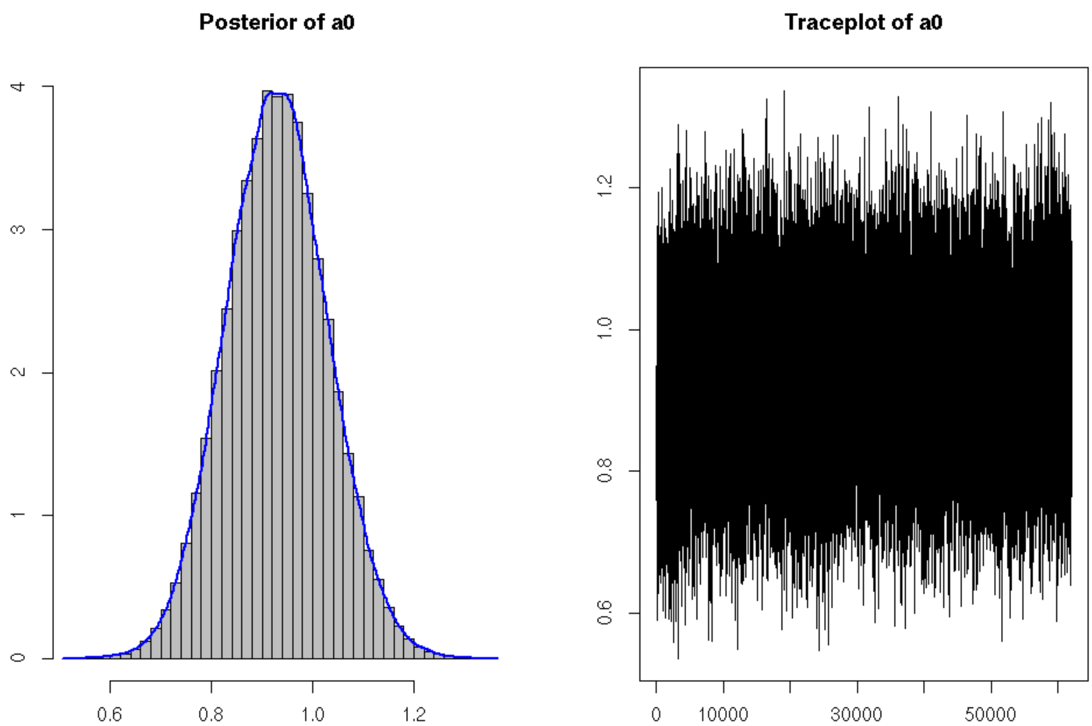


Figure 47: Posterior and traceplot for parameter a_0

12 Bibliography

References

Donatello Telesca, Lurdes Y. T. Inoue.
Bayesian Hierarchical Curve Registration.
Journal of the American Statistical Association (2008).

Simon Jackman.
Bayesian analysis for the social sciences.
Wiley, New York (2009).

Heikki Haario, Eero Saksman, Johanna Tamminen.
An Adaptive Metropolis Algorithm.
Bernoulli Journal (2001).

Eilers, P., and Marx, B. (1996)
Flexible Smoothing Using B–Splines and Penalized Likelihood (with discussion).
Statistical Science, 11, 1200–1224.

Lang, Stefan, and Andreas Brezger.
Bayesian P-splines.
Journal of computational and graphical statistics 13.1 (2004): 183-212.