



POLITECNICO DI TORINO

Corso di Laurea  
in Ingegneria Matematica

## Homework 2 di Statistica computazionale

Cristina Baitan

343428

Anno Accademico 2024-2025

## Esercizio 1

Consideriamo un modello gerarchico Gaussian Process bidimensionale:

$$Y(s) | W(s) \sim \text{GP}(W(s), \tau^2), \quad W(s) \sim \text{GP}(m(s), C(\|s - s'\|; \phi, \sigma^2)),$$

con media lineare

$$m(s) = \beta_0 + \beta_1 s_1$$

e covarianza:

$$C(\|s - s'\|; \phi, \sigma^2) = \sigma^2 \exp(-\phi \|s - s'\|).$$

I parametri sono  $\tau^2 = 0.5$ ,  $\sigma^2 = 5$ ,  $\phi = 3$ ,  $\beta_0 = 2$ ,  $\beta_1 = 0.1$ .

Le attività svolte sono le seguenti:

1. (Punto 1) Simulazione di 100 osservazioni  $y_i$  in punti  $s_i \sim U([0, 10]^2)$ .
2. (Punto 2) Scatterplot dei punti colorati in quattro gruppi secondo i quartili di  $y$ .
3. (Punto 3) Selezione casuale di  $n \in [10, 90]$  osservazioni come set osservato  $y_o$  (coordinate  $D_o$ ) e il resto come non osservato  $y_u$  (coordinate  $D_u$ ). Si definiscono prior appropriate per  $(w, \beta_0, \beta_1, \tau^2, \sigma^2, \phi)$  e implementazione di un MCMC per ottenere campioni dal posteriori

$$f(w, \beta_0, \beta_1, \tau^2, \sigma^2, \phi | y_o),$$

(Punto 4) Successivamente si considerano i residui

$$\epsilon(s) = Y(s) - (\beta_0 + \beta_1 s_1), \quad s \in D_o,$$

e si ottengono campioni dalla loro distribuzione a posteriori. Successivamente, si mostra la distribuzione spaziale della media a posteriori dei residui.

(Punto 5) Infine si ottengono campioni dalla distribuzione a posteriori

$$f(Y(s) | y_o), \quad s \in D_u^* \subseteq D_u,$$

dove  $D_u^*$  contiene almeno 10 elementi. Si valuta, nella maniera ritenuta più appropriata, se i valori veri delle osservazioni ricadono all'interno degli intervalli di credibilità al 95%.

### Punto 1: costruzione del modello e simulazione

Il DAG del modello è rappresentato in Figura 1. Si tratta di un modello comprendente di processi gaussiani, gerarchico e con dati osservati  $\mathbf{Y}$ , variabile latente  $\mathbf{W}$  e parametri globali:  $\phi, \sigma^2, \tau^2, \beta_0, \beta_1$ . Il DAG rappresenta le relazioni tra queste grandezze; si osserva che le variabili  $w_i$  sono in generale dipendenti tra loro con covarianza fornita dalla matrice  $C(\|s - s'\|)$ . I dati osservati  $\mathbf{Y}$  sono indipendenti condizionatamente a  $\mathbf{W}$ .

Per la simulazione di osservazioni  $y(s_i)$  è stato sfruttato il condizionamento  $Y(s)|W(s)$  definito nel modello che permette di scrivere

$$Y(s_i) = W(s_i) + \epsilon(s_i) \quad \text{con } \epsilon(s_i) \sim N(0, \tau^2).$$

Quindi, sono stati prima simulati dei campioni  $w(s_i)$  e successivamente i campioni  $y(s_i) = w(s_i) + \epsilon(s_i)$ .  
 I campioni di  $\mathbf{W}$  vengono ottenuti da una distribuzione normale multivariata con media  $m(s)$  e matrice di covarianza  $C(||s - s'||)$  definita nel modello.

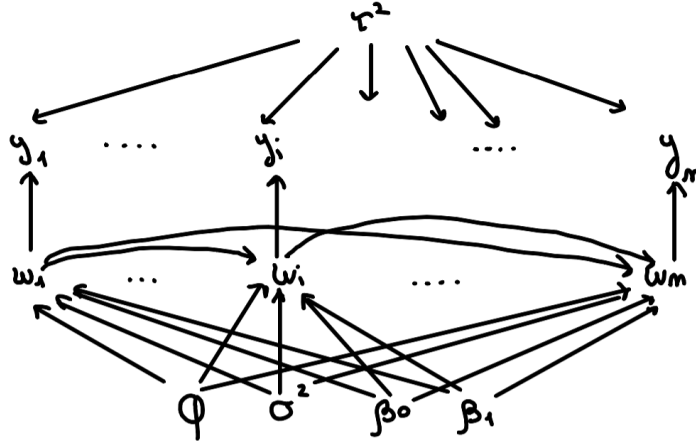


Figure 1: DAG modello esercizio 1

## Punto 2: studio dei campioni tramite scatterplot

Dalla Figura 2, scatterplot delle coordinate con colore in base al gruppo definito in base al quantile, si nota che, in generale, punti vicini spazialmente tendono ad appartenere allo stesso gruppo ovvero osservazioni vicine tendono ad assumere valori abbastanza simili. Questo denota una dipendenza spaziale dei dati osservati che in effetti viene modellizzata dalla variabile  $W(s)$  nel modello.

Per la realizzazione di tale scatterplot è stata usata la libreria *ggplot2*.

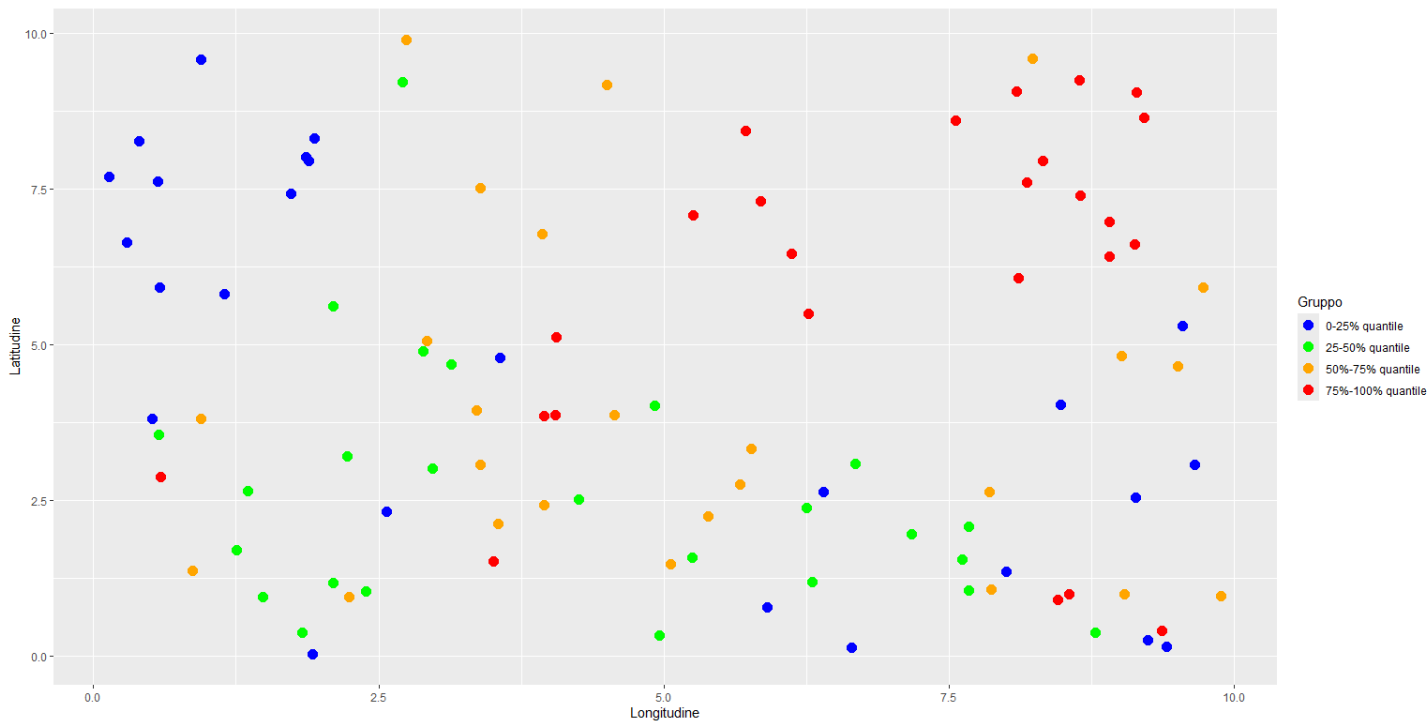


Figure 2: Rappresentazione dei campioni di Y in base ai quantili

### Punto 3: campionamento dalla a posteriori tramite MCMC

E' stato scelto  $n = 50$ .

#### Definizione delle prior

Sia per  $\beta_0$  che per  $\beta_1$  scelgo una prior poco informativa: normale con media 0 e varianza 10000. La scelta della prior di  $\tau^2$  e  $\sigma^2$  è una distribuzione Inverse-gamma in quanto ha supporto nei reali positivi. Il valore scelto per entrambi i parametri è 1 perchè permette di avere una distribuzione abbastanza diffusa e quindi poco informativa.

La prior di  $\phi$  è una uniforme che definisce il supporto di  $\phi$ . In particolare si impone che

$$\phi \in \left( \frac{3}{\text{maxDistance}}, \frac{3}{\text{minDistance}} \right)$$

con maxDistance e minDistance rispettivamente la massima e la minima distanze tra due coordinate nel dataset. Infatti la soglia inferiore indica il valore di  $\phi$  per cui tutte le  $w_i$  risultano correlate mentre la soglia superiore indica il valore di  $\phi$  per cui le  $w_i$  sono approssimativamente tutte indipendenti. Perciò non sarebbe sensato considerare dei valori di  $\phi$  fuori da questo range, soprattutto se  $\phi$  potesse assumere valori superiori a  $\frac{3}{\text{minDistance}}$  rischierei di incorrere nella non identificabilità del modello: esisterebbero più valori del parametro che porterebbero ad avere indipendenza tra tutte le osservazioni.

#### Full conditionals

Prima di applicare il Gibbs sampler, partiziono lo spazio dei parametri da stimare in:

- $\beta := (\beta_0, \beta_1)$
- $W := (W(s^{(1)}), W(s^{(2)}), W(s^{(3)}), \dots, W(s^{(n)}))$  con  $s^{(i)}$  coordinate spaziali in  $D_o$
- i restanti parametri formano ognuno un sottoinsieme.

Definisco inoltre alcune notazioni utilizzate nel calcolo delle full conditionals.

Definisco:

$$m := \begin{bmatrix} 1 & s_1^1 \\ 1 & s_1^2 \\ \dots & \\ 1 & s_1^n \end{bmatrix} \cdot \beta$$

E chiamo la matrice di covarianza

$$\Sigma := C(\|s - s'\|, \phi, \sigma^2)$$

Di seguito i calcoli per ricavare le full conditional per  $\beta$ ,  $\sigma^2$ ,  $\tau^2$  e  $W$ .

- Full conditional di  $\beta = (\beta_0, \beta_1)$ .

---


$$f(\beta \mid \dots) = \frac{f(W, \beta, \sigma^2, \tau^2, \phi, Y_0)}{f(W, \sigma^2, \tau^2, \phi, Y_0)} \propto f(W, \beta, \sigma^2, \tau^2, \phi, Y_0) \stackrel{(1)}{=}$$

$$f(\sigma^2) f(\beta) f(\tau^2) f(\phi^2) f(Y_0 \mid W, \tau^2) f(W \mid \beta, \sigma^2, \phi)$$

$$\propto f(W \mid \beta, \sigma^2, \phi) f(\beta)$$

(1) per la fattorizzazione della congiunta con le informazioni date dal DAG.

---

$f(\beta)$  è prior con normale  $(0, 0)$  e matrice di covarianza

$$\Sigma_p := \begin{bmatrix} 10000 & 0 \\ 0 & 10000 \end{bmatrix}$$

$$f(\beta \mid \dots) \propto f(W \mid \beta, \sigma^2, \phi) \cdot f(\beta) = \exp \left\{ -\frac{1}{2} [(W - m)^\top \Sigma^{-1} (W - m) + \beta^\top \Sigma_p^{-1} \beta] \right\}.$$

$$\text{con } (W - m)^\top \Sigma^{-1} (W - m) = W^\top \Sigma^{-1} W - 2m^\top \Sigma^{-1} W + m^\top \Sigma^{-1} m$$

$$\text{con } m = \begin{pmatrix} 1 & s_1 \\ 1 & s_2 \\ \vdots & \vdots \\ 1 & s_m \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = S \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

$$f(\beta \mid \dots) \propto \exp \left\{ -\frac{1}{2} \left[ -2m^\top \Sigma^{-1} W_o + m^\top \Sigma^{-1} m + \frac{\beta^\top \beta}{10000} \right] \right\}$$

$$= \exp \left\{ -\frac{1}{2} \left[ -2(S\beta)^\top \Sigma^{-1} W + (S\beta)^\top \Sigma^{-1} (S\beta) + \frac{\beta^\top \beta}{10000} \right] \right\}$$

$$= \exp \left\{ -\frac{1}{2} \left[ -2\beta^\top S^\top \Sigma^{-1} W + \beta^\top S^\top \Sigma^{-1} (S\beta) + \frac{\beta^\top \beta}{10000} \right] \right\}$$

$$= \exp \left\{ -\frac{1}{2} [\beta^\top (S^\top \Sigma^{-1} S + \Sigma_p^{-1}) \beta - 2\beta^\top (S^\top \Sigma^{-1} W)] \right\}$$

Riconosco il kernel di una normale con:

$$\text{Vettore delle medie: } M_{post} = V_{post} \cdot S^\top \Sigma^{-1} W_o$$

$$\text{Matrice di covarianza: } V_{post} = (S^\top \Sigma^{-1} S + \Sigma_p^{-1})^{-1}$$

- Full conditional di  $\sigma^2$

---


$$f(\sigma^2 \mid \dots) = \frac{f(W, \beta, \sigma^2, \tau^2, \phi, Y_0)}{f(\beta, W, \tau^2, \phi, Y_0)} \propto f(W, \beta, \sigma^2, \tau^2, \phi, Y_0) \stackrel{(1)}{=}$$

$$f(\sigma^2) f(\beta) f(\tau^2) f(\phi^2) f(Y_0 \mid W, \tau^2) f(W \mid \beta, \sigma^2, \phi)$$

$$\propto f(W \mid \beta, \sigma^2, \phi) f(\sigma^2)$$

(1) per la fattorizzazione della congiunta con le informazioni date dal DAG.

---

$f(\sigma^2)$  è la funzione di densità della prior  $IG(a=1, b=1)$ .

$$\begin{aligned} f(\sigma^2 \mid \dots) &\propto f(W \mid \sigma^2, \beta, \phi) \cdot f(\sigma^2) \propto \\ &(\det(\Sigma))^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [(W - m)' \Sigma^{-1} (W - m)] \right\} (\sigma^2)^{-(a+1)} \exp \left( -\frac{b}{\sigma^2} \right) = \\ &(\sigma^2)^{-(a+1)} (\det(\Sigma))^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ (W - m)' \Sigma^{-1} (W - m) + \frac{b}{\sigma^2} \right] \right\} \propto (*) \end{aligned}$$

Riscrivo considerando  $C := \sigma^2 \Sigma^{-1}$  e  $\det(\Sigma^{-1}) = \det(C) \sigma^{-2n}$

$$\begin{aligned} (*) \quad &(\sigma^2)^{-(a+1)} \sigma^{-n} \exp \left\{ -\frac{1}{2} \left[ (W - m)' \frac{C}{\sigma^2} (W - m) + \frac{b}{\sigma^2} \right] \right\} = \\ &(\sigma^2)^{-(a+n/2+1)} \exp \left\{ -\frac{1}{\sigma^2} \left[ b + \frac{1}{2} (W - m)' C (W - m) \right] \right\} \end{aligned}$$

Da cui  $\sigma^2 \mid \dots \sim IG(a_{post}, b_{post})$ , con

$$a_{post} = a + \frac{n}{2}, \quad b_{post} = b + \frac{1}{2} (W - m)' C (W - m)$$

- Full conditional di  $W$

---


$$f(W \mid \dots) = \frac{f(W, \beta, \sigma^2, \tau^2, \phi, Y_0)}{f(\beta, \sigma^2, \tau^2, \phi, Y_0)} \propto f(W, \beta, \sigma^2, \tau^2, \phi, Y_0) \stackrel{(1)}{=}$$

$$f(\sigma^2)f(\beta)f(\tau^2)f(\phi^2)f(Y_o \mid W, \tau^2)f(W \mid \beta_0, \beta_1, \sigma^2, \phi)$$

$$\propto f(Y_o \mid W, \tau^2)f(W \mid \beta, \sigma^2, \phi)$$

(1) per la fattorizzazione della congiunta con le informazioni date dal DAG.

---


$$f(Y_o \mid W, \tau^2)f(W \mid \beta, \sigma^2, \phi) \propto$$

$$\exp\left(-\frac{1}{2\tau^2}(Y_o - W)^T(Y_o - W)\right) \cdot \exp\left(-\frac{1}{2}(W - m)^T\Sigma^{-1}(W - m)\right) \propto (*)$$

dove:

$$(W - m)^T\Sigma^{-1}(W - m) = W^T\Sigma^{-1}W - 2m^T\Sigma^{-1}W + m^T\Sigma^{-1}m \propto W^T\Sigma^{-1}W - 2m^T\Sigma^{-1}W$$

$$(Y_o - W^T)(Y_o - W) = Y_o^TY_o - 2W^TY_o + W^TW \propto -2W^TY_o + W^TW$$

$$\begin{aligned} (*) \quad & \exp\left(-\frac{1}{2}\left(\frac{1}{\tau^2}(-2W^TY_o + W^TW) + W^T\Sigma^{-1}W - 2W^T\Sigma^{-1}m\right)\right) = \\ & \exp\left(-\frac{1}{2}\left[W^T\left(\frac{1}{\tau^2}I + \Sigma^{-1}\right)W - 2W^T\left(\frac{1}{\tau^2}Y_o + \Sigma^{-1}m\right)\right]\right) \end{aligned}$$

Da cui  $W \mid \dots \sim \mathcal{N}(M_{post}, V_{post})$ , con

$$V_{post} = \left(\frac{1}{\tau^2}I + \Sigma^{-1}\right)^{-1} \quad \text{e} \quad M_{post} = V_{post} \cdot \left(\frac{1}{\tau^2}Y_o + \Sigma^{-1}m\right)$$



- Full conditional di  $\tau^2$ .

---


$$f(\tau^2 \mid \dots) = \frac{f(W, \beta, \sigma^2, \tau^2, \phi, Y_o)}{f(\beta, \sigma^2, W, \phi, Y_o)} \propto f(W, \beta, \sigma^2, \tau^2, \phi, Y_o) \stackrel{(1)}{=}$$

$$f(\sigma^2) f(\beta) f(\tau^2) f(\phi) f(Y_o \mid W, \tau^2) f(W \mid \beta, \sigma^2, \phi)$$

$$\propto f(Y_o \mid W, \tau^2) f(\tau^2)$$

(1) per la fattorizzazione della congiunta con le informazioni date dal DAG.

---

$f(\tau^2)$  è la funzione di densità della prior  $IG(a=1, b=1)$

$$f(\tau^2 \mid \dots) \propto f(Y_o \mid \tau^2, W) \cdot f(\tau^2) =$$

$$\left( \frac{1}{\tau^2} \right)^{n/2} \exp \left( -\frac{1}{2\tau^2} (Y_o - W)^\top (Y_o - W) \right) \cdot \tau^{-(a+1)} \exp \left( -\frac{b}{\tau^2} \right)$$

$$\propto (\tau^2)^{-(a+n/2+1)} \exp \left( -\frac{1}{\tau^2} \left( \frac{(Y_o - W)^\top (Y_o - W)}{2} + b \right) \right)$$

Da cui  $\tau^2 \mid \dots \sim IG(a_{post}, b_{post})$ , con

$$a_{post} = a + \frac{n}{2} \quad \text{e} \quad b_{post} = b + \frac{1}{2} (Y_o - W)^\top (Y_o - W)$$

### Passo metropolis per $\phi$

Ricavo la full conditional:

---


$$f(\phi \mid \dots) = \frac{f(W, \beta, \sigma^2, \tau^2, \phi, Y_o)}{f(\beta, \sigma^2, W, \tau^2, Y_o)} \propto f(W, \beta, \sigma^2, \tau^2, \phi, Y_o) \stackrel{(1)}{=}$$

$$f(\sigma^2) f(\beta) f(\tau^2) f(\phi) f(Y_o \mid W, \tau^2) f(W \mid \beta, \sigma^2, \phi)$$

$$\propto f(W \mid \phi, \sigma^2, \beta^2) \cdot f_\phi(\phi)$$

(1) per la fattorizzazione della congiunta con le informazioni date dal DAG.

---

Non è possibile campionare direttamente dalla full conditional in quanto non ha la forma di una distribuzione nota.

Procedo con un algoritmo di tipo Metropolis-Hastings.

$\phi$  ha dominio in  $[a, b]$  con  $a = \frac{3}{\max \text{Distance}}$  e  $b = \frac{3}{\min \text{Distance}}$  per quanto visto in precedenza.

Applico quindi una trasformazione di  $\phi$  in modo da ottenere una v.a. con supporto in  $\mathbb{R}$  e quindi poter proporre da una normale.

Definisco  $\eta := \log \left( \frac{\phi - a}{b - \phi} \right)$  allora  $\phi(\eta) = \frac{a + be^\eta}{1 + e^\eta}$ .

Prior di  $\eta$ :

$$f_\eta(\eta) = f_\phi(\phi(\eta)) \left| \frac{d\phi}{d\eta} \right| = \frac{1}{b - a} \cdot \frac{be^\eta(1 + e^\eta) - e^\eta(a + be^\eta)}{(1 + e^\eta)^2} \propto \frac{e^\eta}{(1 + e^\eta)^2}$$

Viene quindi proposto un campione di  $\eta$  da una normale:  $\eta^* \sim \mathcal{N}(\eta_{\text{prec}}, \tau_\eta^2)$

- $\eta_{\text{prec}}$  ultimo campione accettato,
- $\tau_\eta^2$  varianza adattiva.

$$\alpha = \min \left\{ \frac{f(\eta^* | \dots) q(\eta_{\text{prec}} | \dots)}{f(\eta_{\text{prec}} | \dots) q(\eta^* | \dots)}, 1 \right\} = \min \left\{ \frac{f(\eta^* | \dots)}{f(\eta_{\text{prec}} | \dots)}, 1 \right\} \quad \text{perché proposta simmetrica.}$$

Dove la full conditional di  $\eta$ :

$$f(\eta | \dots) \propto f(\eta, \sigma^2, \tau^2, \beta, W) \propto f(\phi(\eta), \sigma^2, \tau^2, \beta, W) \left| \frac{d\phi}{d\eta} \right| =$$

$$\left( \frac{d\phi}{d\eta} \cdot f(\phi(\eta)) \right) f(\beta) f(\sigma^2) f(\tau^2) f(W | \beta, \sigma^2, \phi) f(Y_o | \tau^2, W) \propto f_\eta(\eta) f(W | \beta, \sigma^2, \phi(\eta)).$$

Nel codice viene calcolato il log-rapporto per evitare cancellazione numerica:

$$\log \left( \frac{f(\phi(\eta^*) | \dots)}{f(\phi(\eta_{\text{prec}}) | \dots)} \right) = \log(f(W | \phi(\eta^*))) + \log(f_\eta(\eta^*)) - \log(f(W | \phi(\eta_{\text{prec}}))) - \log(f_\eta(\eta_{\text{prec}}))$$

Viene accettato il campione proposto con probabilità pari ad  $\alpha$ .

## Burnin

Nell'algoritmo MCMC implementato è stato tenuto conto di eventuali burnin per ogni campionamento.

E' stato fatto burnin di 500 campioni per le catene di  $\sigma^2$ ,  $\tau^2$  e  $\phi$ . In Figura 3 la catena di  $\phi$  prima del burnin.

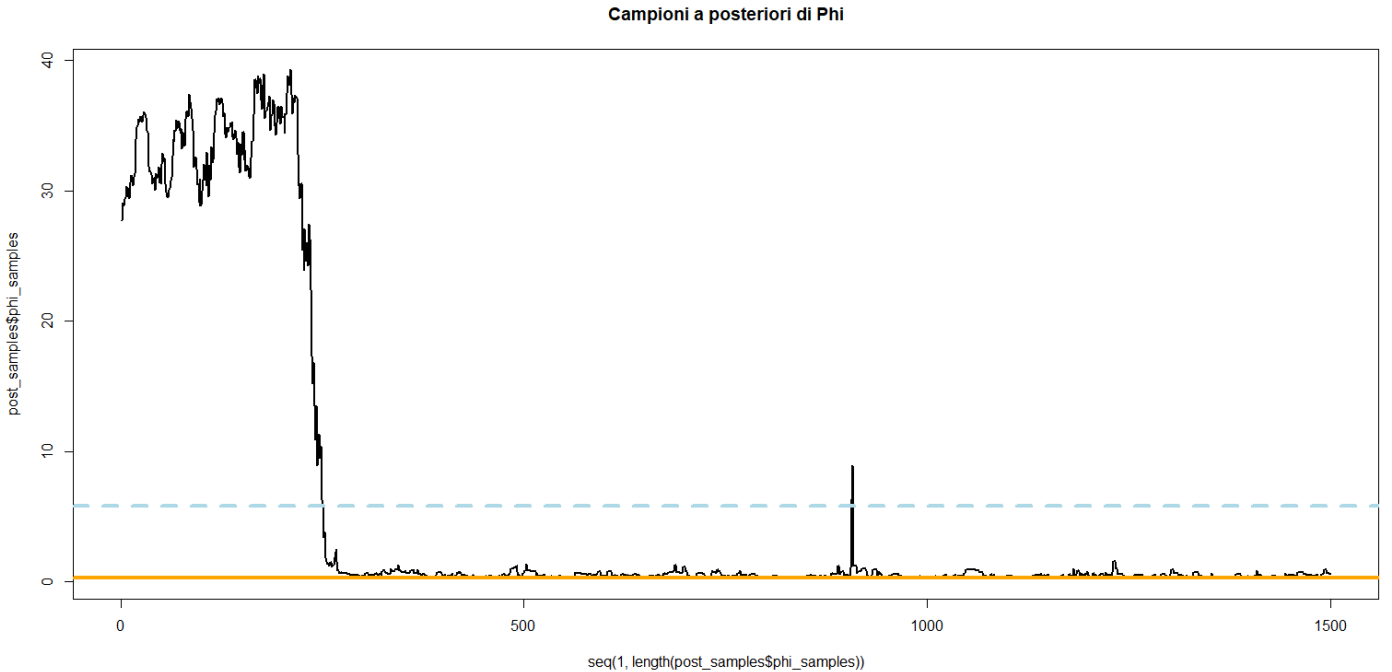


Figure 3: Catena di  $\phi$  prima del burnin

## Thinning

Nell'algoritmo MCMC implementato è stato tenuto conto di eventuali thin per ogni campionamento.

Sono stati scelti impostati valori di thin pari a 10 per i parametri  $\sigma^2$ ,  $\tau^2$  e  $\phi$ .

Tali valori sono stati scelti in modo da ridurre la correlazione tra i campioni.

Per calcolare tale correlazione è stata usata il comando `acf()` che mostra la funzione di autocorrelazione.

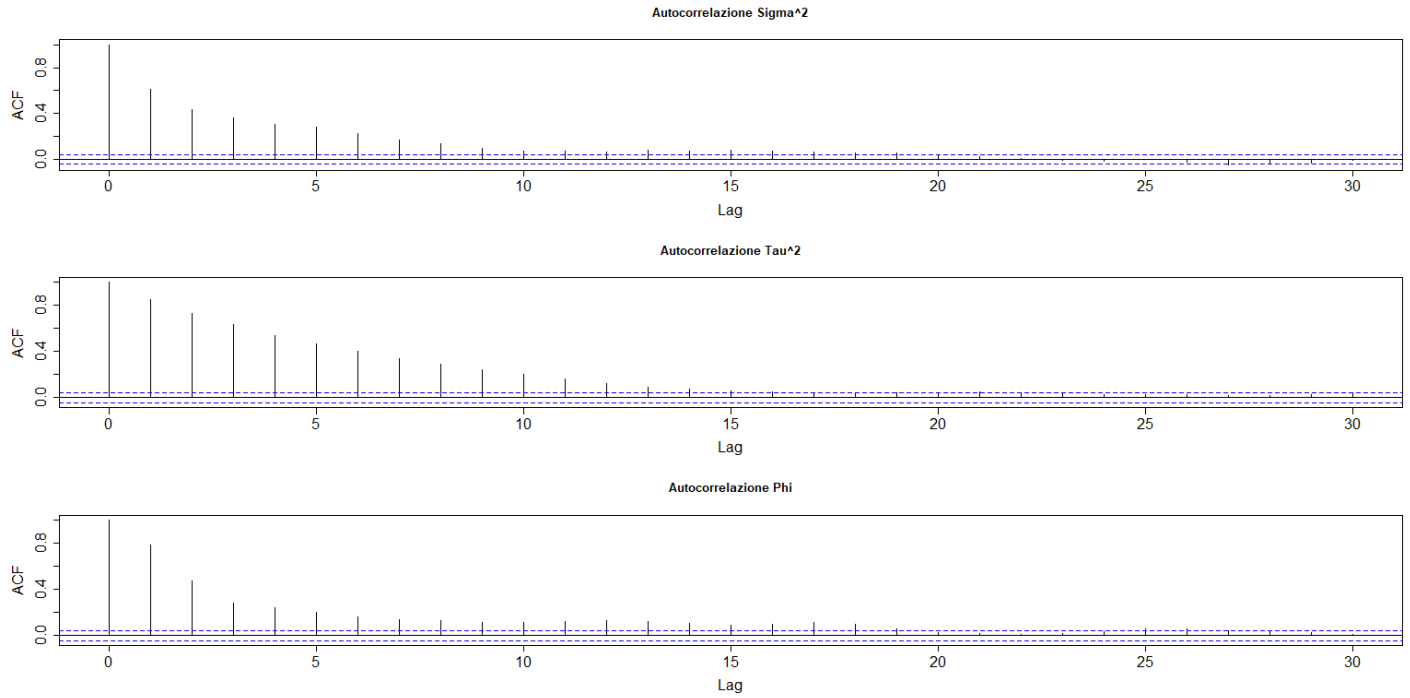


Figure 4: Autocorrelazione

## Risultati

Di seguito le catene a convergenza.

La linea tratteggiata in verde rappresenta la media campionaria dei campioni a posteriori, mentre la linea arancione il valore vero del parametro.

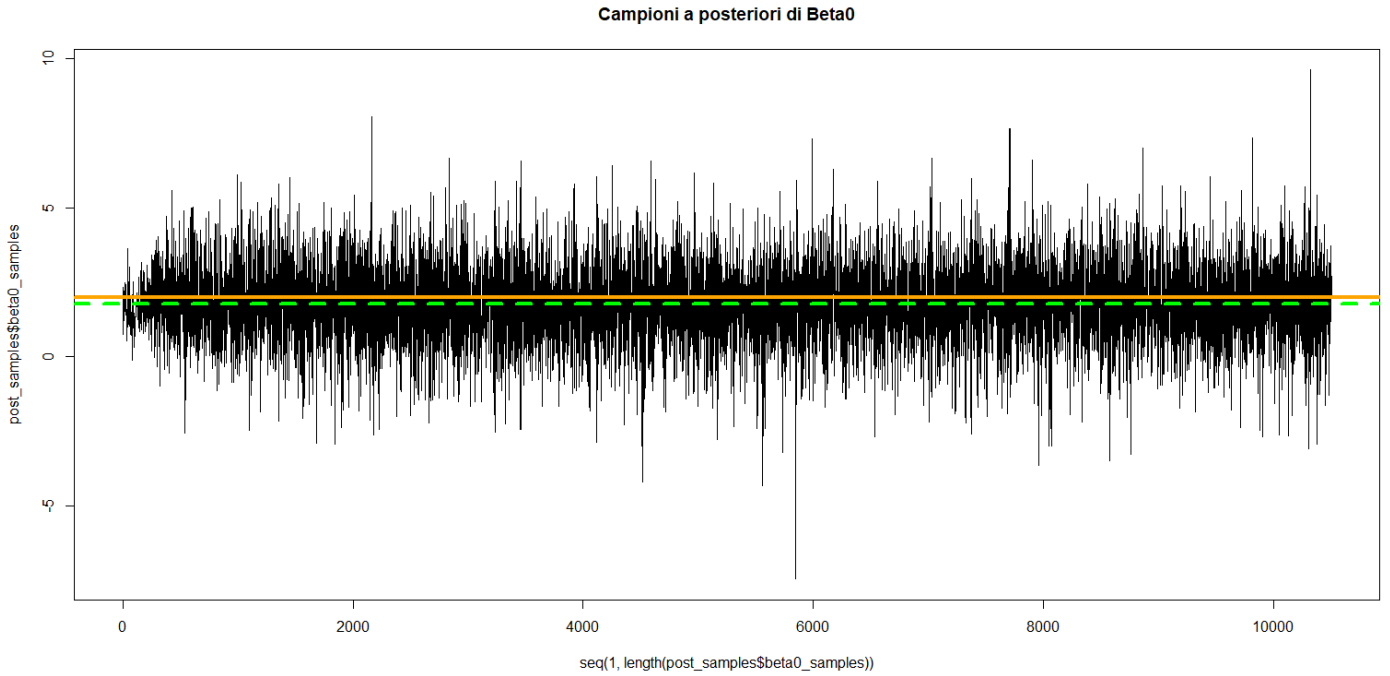


Figure 5: Campioni a posteriori di  $\beta_0$

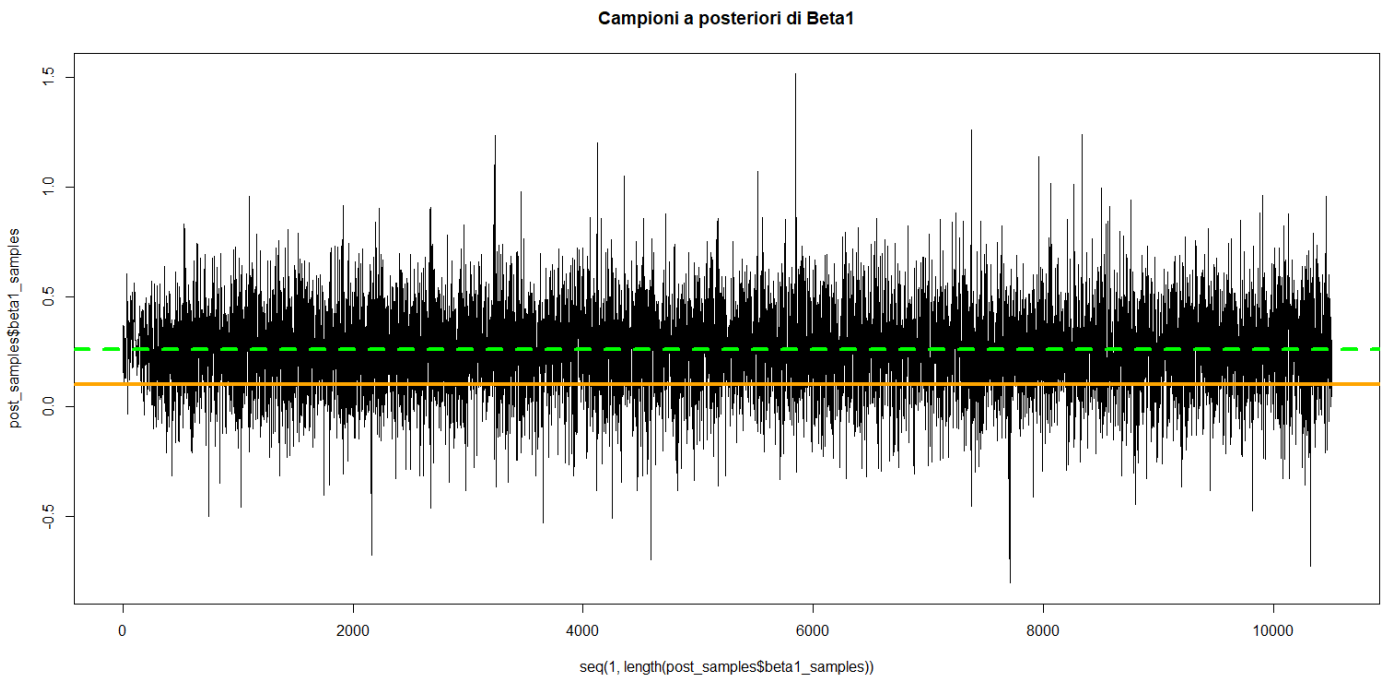


Figure 6: Campioni a posteriori di  $\beta_1$

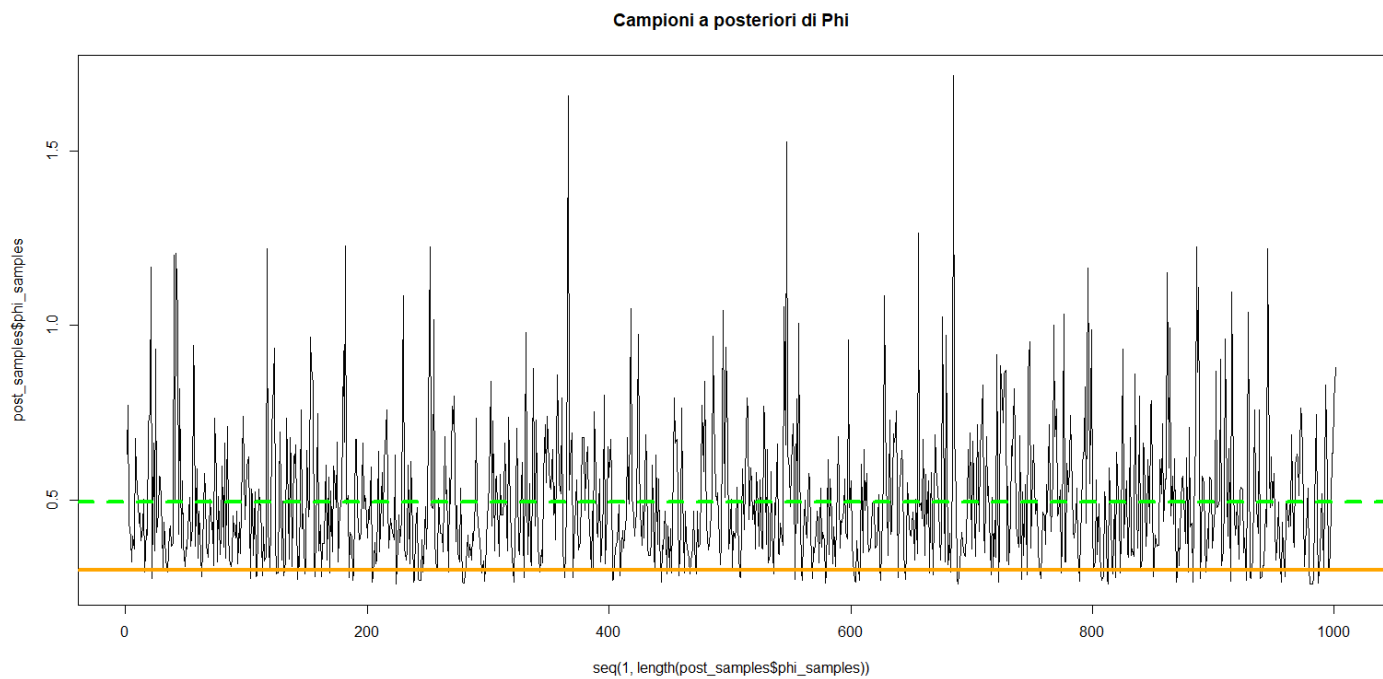


Figure 7: Campioni a posteriori di  $\phi$

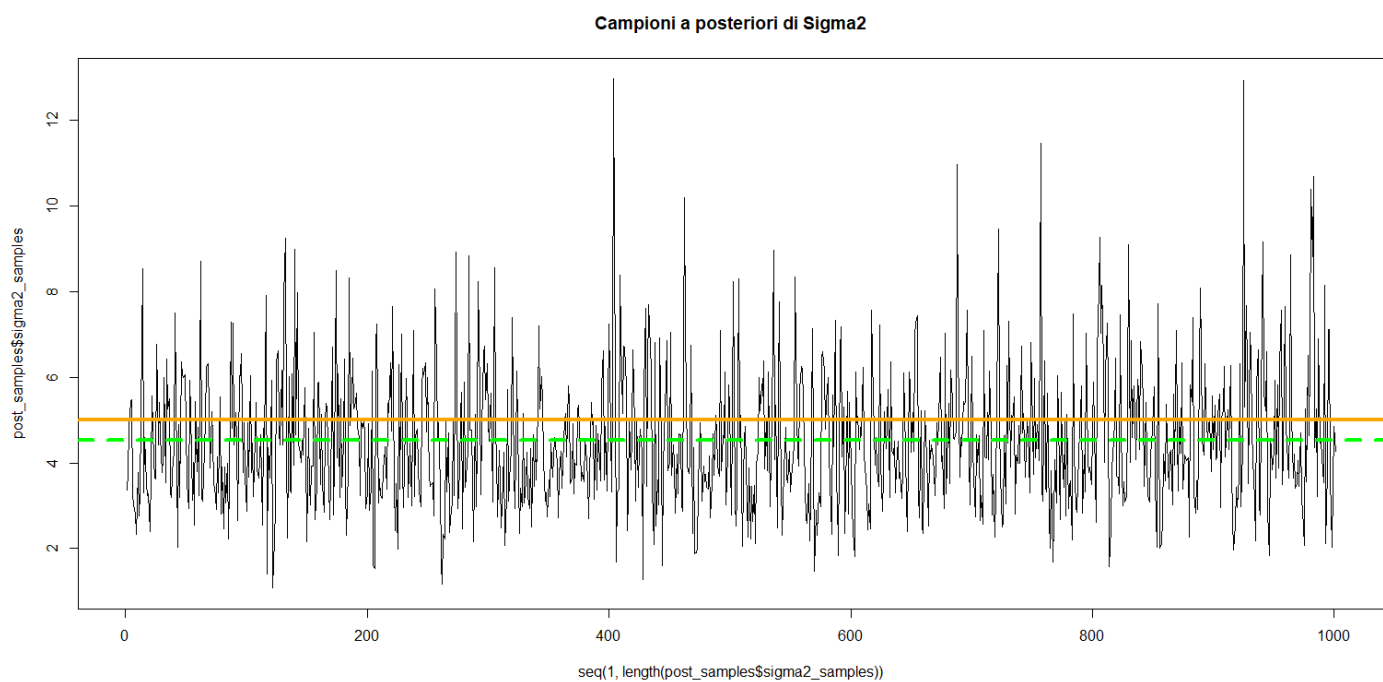


Figure 8: Campioni a posteriori di  $\sigma^2$

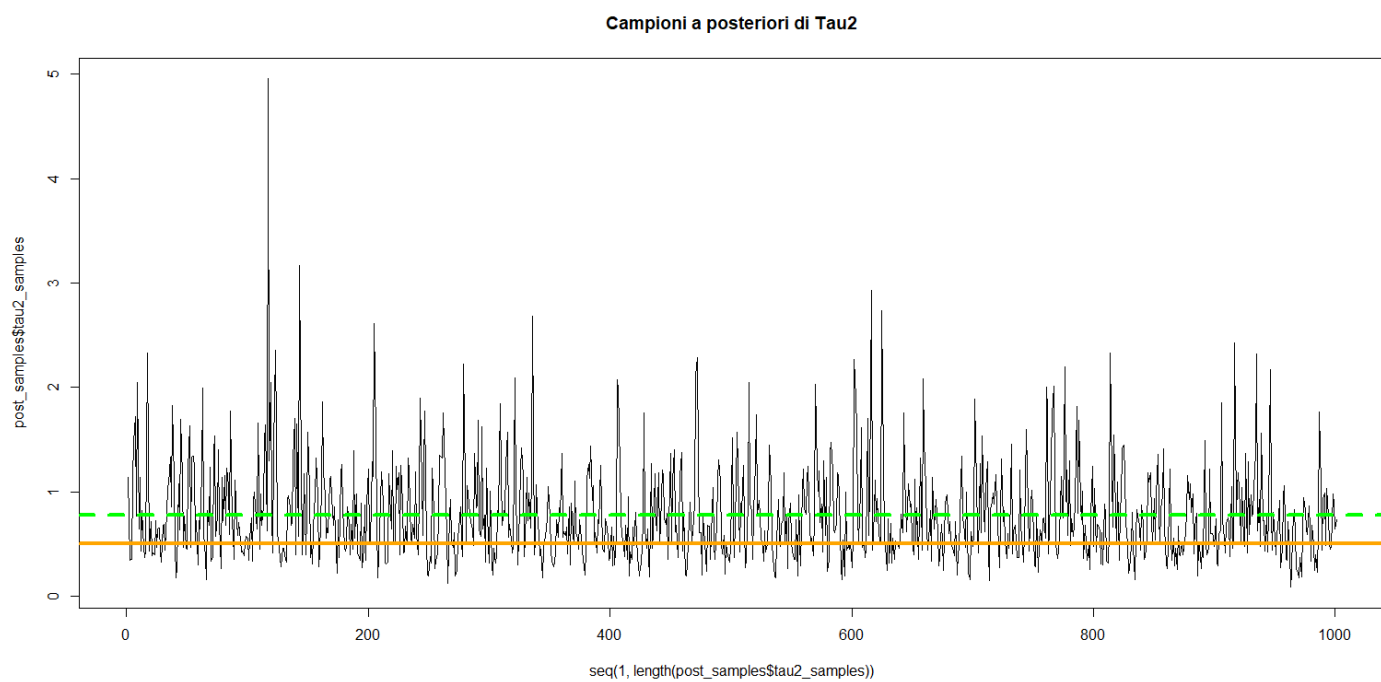


Figure 9: Campioni a posteriori di  $\tau^2$

#### Punto 4: campioni a posteriori da $\epsilon(s)$

Per definizione di  $\epsilon(s)$ , campioni di  $\epsilon(s_i) \forall s_i \in D_o$  si ottengono a partire dai campioni della distribuzione  $Y_o$  e dai campioni a posteriori di  $\beta_0$  e  $\beta_1$ .

$$\text{Campione a posteriore di } \epsilon(s_i) = y_i - (\beta_0^j + \beta_1^j s_i)$$

Fissato un punto in  $D_o$ :  $y_i$  è un campione di  $Y(s_i)$  (nel codice sono stati usati quelli campionati nel punto 1) e  $\beta_0^j, \beta_1^j$  sono campioni j-esimi ottenuti nel punto 3 con catena a convergenza.

Dei campioni ottenuti per ogni punto è stata calcolata la media campionaria.

La Figura 10 rappresenta la distribuzione della media campionaria nei punti: ogni punto è colorato in base al gruppo quantilico corrispondente.

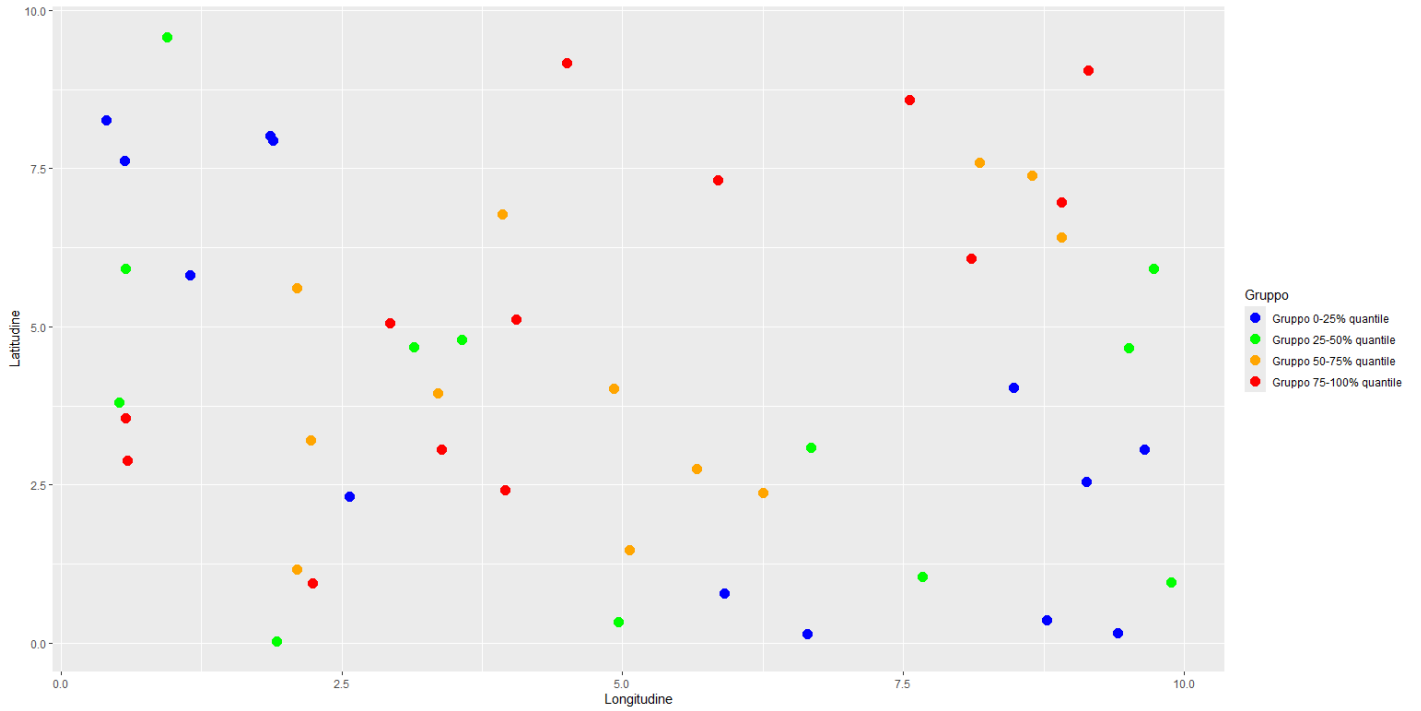


Figure 10: Campioni a posteriori di  $\epsilon(s)$

## Punto 5: previsione

L'intervallo di credibilità del 95% è quella regione circoscritta dal 0.25-quantile e dal 0.975-quantile della distribuzione a posteriori. I quantili sono approssimati dai quantili empirici calcolati dalla funzione *quantile()* a partire dai campioni a posteriori.

Quindi ricavo i campioni a posteriori di  $Y(s)$  con  $s \notin D_o$ , ovvero faccio previsione sfruttando i campioni a posteriori dei parametri ottenuti nel punto 3.

1. Campioni a posteriori di  $Y(s)$  con  $s \in D_u^*$  :

- Campione a posteriori da  $W(s)$ :

$W(s)$  è distribuito come una normale multivariata con media  $m(s) = \beta_0 + \beta_1 s_1$  e matrice di covarianza  $C(\|s - s'\|, \phi, \sigma^2)$ .

Ottingo molteplici campioni a posteriori dalla congiunta  $W(s)$ :

$$\text{campione a posteriori di } W(s) \text{ da } N_n \left( \begin{pmatrix} 1 & s_1 \\ 1 & s_2 \\ \vdots & \vdots \\ 1 & s_m \end{pmatrix} \cdot \begin{pmatrix} \beta_0^j \\ \beta_1^j \end{pmatrix}, C(\|s - s'\|, \phi^j, \sigma_j^2) \right)$$

Dove  $\beta_0^j, \beta_1^j, \sigma_j^2, \phi^j$  sono i j-esimi campioni a posteriori ottenuti nel punto 3 dell'esercizio.

Successivamente prelevo dai campioni multivariati solo le componenti marginali corrispondenti a  $s \in D_u^*$  (sfrutto il fatto un campione dalla congiunta posso vederlo come campioni dalle marginali).

- Campione da  $Y(s)$  :

Per ogni punto  $s \in D_u^*$  si ottiene campionando da:

$$Y(s) \sim N((W(s))^j, \tau_j^2)$$

Con  $W(s)^j$  campione a posteriori appena ottenuto e  $\tau_j^2$  campione a posteriori ottenuto nel punto 3 dell'esercizio.

2. Calcolo lower bound e upper bound intervallo di credibilità:

lower bound e upper bound dell'intervallo di credibilità della distribuzione a posteriori di  $Y(s)$  corrispondono rispettivamente al 0.25-quantile e 0.975 quantile. Vengono quindi calcolati lower bound e upper bound, per ogni punto, approssimativamente con la funzione *quantile()* applicato ai campioni a posteriori di  $Y(s)$ .

In Figura 11 sono rappresentati in rosso i punti in  $D_u^*$  laddove il valore vero (il valore ottenuto nel punto 1 dell'esercizio) non è compreso nell'intervallo di credibilità, in nero se invece è compreso.



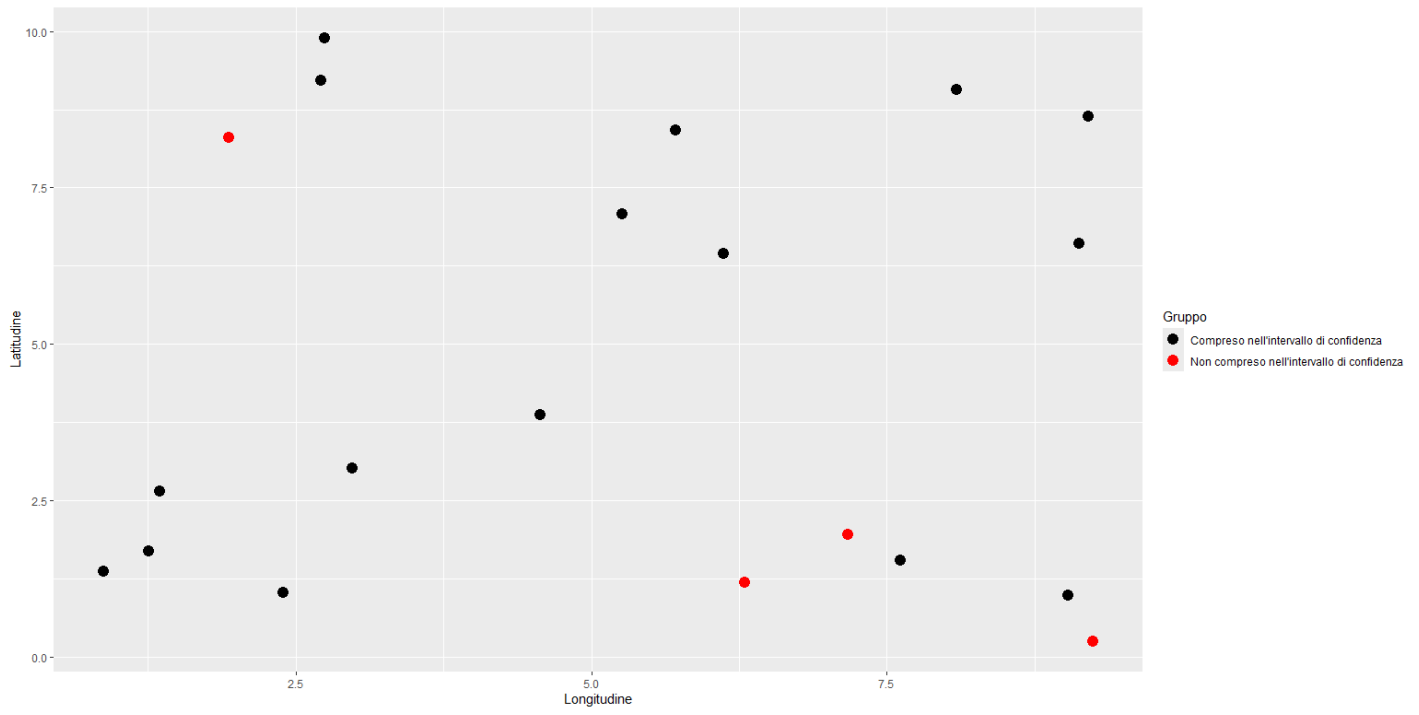


Figure 11: Intervallo di credibilità e valore vero di  $Y(s)$  in  $D_u^*$

## Esercizio 2

### Modello a mistura e simulazioni

Si considera un modello a mistura con

$$Y_i | z_i \sim P(\lambda_{z_i}), \quad P(z_i = k) = \pi_k, \quad z_i \in \{1, 2, 3\}, \quad i = 1, \dots, n,$$

con  $n = 200$ ,  $\pi_k = 1/3$  e  $\lambda_1 = 1, \lambda_2 = 10, \lambda_3 = 25$ .

Le attività svolte sono:

1. Simulazione dei dati dal modello e rappresentazione grafica dei gruppi
2. Assunzione di prior appropriate (per  $\lambda_k$  si suggeriscono distribuzioni Gamma) e implementazione di un MCMC per ottenere campioni a posteriori di

$$f(\lambda_1, \lambda_2, \lambda_3, \pi_1, \pi_2, \pi_3, z_1, \dots, z_n | y).$$

Il modello mistura ha DAG mostrato in Figura 12. La simulazione di 200 osservazioni  $Y$  avviene per campionamento da  $Z$  e successivamente da  $Y | z_i$ .

In Figura 13 l'istogramma di frequenze delle osservazioni  $y_i$  con suddivisione rispetto a  $z_i = k$ . Per il campionamento dalla a posteriori utilizzo il Gibbs Sampler e quindi il campionamento dalle full conditional.

Notazioni:

$$\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3),$$

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3),$$

$\lambda_{-k}$  è il vettore  $\boldsymbol{\lambda}$  senza la componente  $k$ -esima.

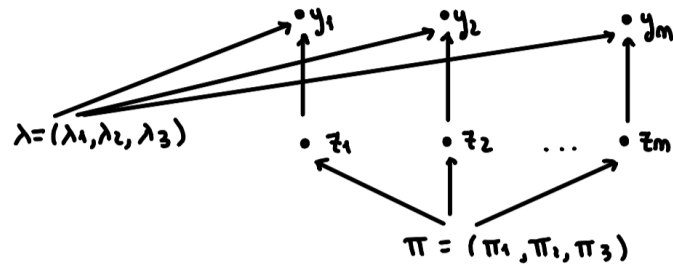


Figure 12: DAG modello esercizio 2

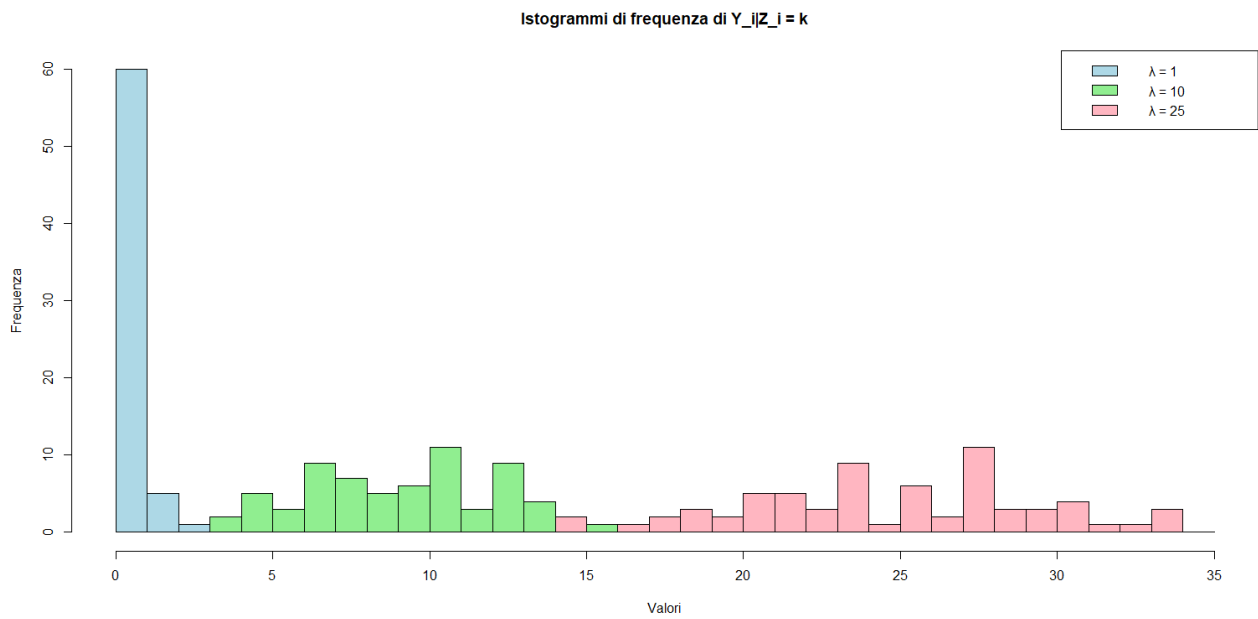


Figure 13: Rappresentazione dei dati osservati con suddivisione rispetto ai gruppi

### Full conditional di $\lambda_k$

---


$$f(\lambda_k \mid \dots) = \frac{f(\lambda_k, \lambda_{-k}, \boldsymbol{\pi}, \mathbf{z}, \mathbf{y})}{f(\lambda_{-k}, \boldsymbol{\pi}, \mathbf{z}, \mathbf{y})} \propto f(\lambda_k, \lambda_{-k}, \boldsymbol{\pi}, \mathbf{z}, \mathbf{y}) \stackrel{(1)}{=}$$

$$f(\lambda_k) f(\lambda_{-k}) f(\boldsymbol{\pi}) f(\mathbf{y} \mid \lambda_k, \lambda_{-k}, \mathbf{z}) f(\mathbf{z} \mid \boldsymbol{\pi})$$

$$\propto f(\mathbf{y} \mid \lambda_k, \lambda_{-k}, \mathbf{z}) f(\lambda_k)$$

(1) per la fattorizzazione della congiunta con le informazioni date dal DAG.

---

con  $f(\lambda_k)$  prior distribuita come una Gamma( $a, b$ ).

Graficamente variando i parametri  $a, b$  si nota come la distribuzione gamma sia poco informativa quando  $a = 1$  e  $b$  tende a 0. Nel codice è stato scelto  $a = 1$  e  $b = 0.01$ .

Definisco  $\mathcal{I}_k := \{i : z_i = k\}$ .

$$f(\mathbf{y} \mid \lambda_k, \lambda_{-k}, \mathbf{z}) f(\lambda_k) \propto \left( \prod_{i \in \mathcal{I}_k} \frac{\lambda_k^{y_i} e^{-\lambda_k}}{y_i!} \right) \cdot (\lambda_k^{a-1} e^{-b\lambda_k}) \propto \lambda_k^{\sum_{i \in \mathcal{I}_k} y_i + a - 1} e^{-\lambda_k(n_k + b)},$$

dove  $n_k = |\mathcal{I}_k|$ .

Quindi:

$$\lambda_k \mid \dots \sim \text{Gamma} \left( \sum_{i \in \mathcal{I}_k} y_i + a, n_k + b \right).$$

### Full conditional di $z_i$

---


$$f(z_i \mid \dots) \stackrel{(1)}{=} \frac{f(\boldsymbol{\lambda}, \boldsymbol{\pi}, y_i, z_i)}{f(\boldsymbol{\lambda}, \boldsymbol{\pi}, y_i)} \propto f(\boldsymbol{\lambda}, \boldsymbol{\pi}, y_i, z_i) \stackrel{(2)}{=}$$

$$f(\boldsymbol{\lambda}) f(\boldsymbol{\pi}) f(y_i \mid \boldsymbol{\lambda}, z_i) f(z_i \mid \boldsymbol{\pi})$$

$$\propto f(y_i \mid \boldsymbol{\lambda}, z_i) f(z_i \mid \boldsymbol{\pi})$$

(1) Mi soffermo a studiare la sola componente  $y_i$  poichè  $z_i$  e  $y_{-i}$  sono condizionatamente indipendenti dati  $\boldsymbol{\lambda}$  e  $\boldsymbol{\pi}$  per la struttura del DAG.

(2) per la fattorizzazione della congiunta con le informazioni date dal DAG.

---


$$f(y_i \mid \boldsymbol{\lambda}, z_i) f(z_i \mid \boldsymbol{\pi}) = \left[ \prod_{k=1}^3 \left( \frac{\lambda_k^{y_i} e^{-\lambda_k}}{y_i!} \right)^{\mathbf{1}(z_i=k)} \right] \left[ \prod_{k=1}^3 (\pi_k)^{\mathbf{1}(z_i=k)} \right]$$

Definendo:  $g_k = \frac{\lambda_k^{y_i} e^{-\lambda_k}}{y_i!}$ , allora

$$z_i \mid \dots \sim \text{discreta} \left( \frac{\pi_1 g_1}{\sum_k \pi_k g_k}, \frac{\pi_2 g_2}{\sum_k \pi_k g_k}, \frac{\pi_3 g_3}{\sum_k \pi_k g_k} \right).$$

I campioni vengono prelevati tramite il "Gumbel trick".

## Full conditional di $\pi$

$$f(\pi \mid \dots) = \frac{f(\lambda, \pi, \mathbf{y}, \mathbf{z})}{f(\lambda, \pi, \mathbf{y}, \mathbf{z})} \propto f(\lambda, \pi, \mathbf{y}, \mathbf{z}) \stackrel{(1)}{=}$$

$$f(\lambda)f(\pi)f(\mathbf{y} \mid \lambda, \mathbf{z})f(\mathbf{z} \mid \pi)$$

$$\propto f(\mathbf{z} \mid \pi)f(\pi)$$

(1) per la fattorizzazione della congiunta con le informazioni date dal DAG.

$f(\pi)$  funzione di densità di una Dirichlet con parametri  $\alpha_1 = \alpha_2 = \alpha_3 = 1$  in modo che non ci sia informazione nella prior (per parametri pari a 1 la distribuzione di Dirichlet è uniforme nel semplice in  $R^3$ ).

$$f(\mathbf{z} \mid \pi) \cdot f(\pi) = \left[ \prod_{i=1}^n \prod_{k=1}^3 (\pi_k)^{\mathbf{1}(z_i=k)} \right] \cdot \left[ \prod_{k=1}^3 \pi_k^{\alpha_k-1} \right] = \prod_{k=1}^3 \pi_k^{\alpha_k+n_k-1},$$

con  $n_k = |\mathcal{I}_k|$ .

Quindi:

$$\pi \mid \dots \sim \text{Dirichlet}(\alpha_1 + n_1, \alpha_2 + n_2, \alpha_3 + n_3).$$

Nelle figure che seguono sono rappresentate le catene.

La linea tratteggiata in verde rappresenta la media campionaria dei campioni a posteriori, mentre la linea arancione il valore vero del parametro.

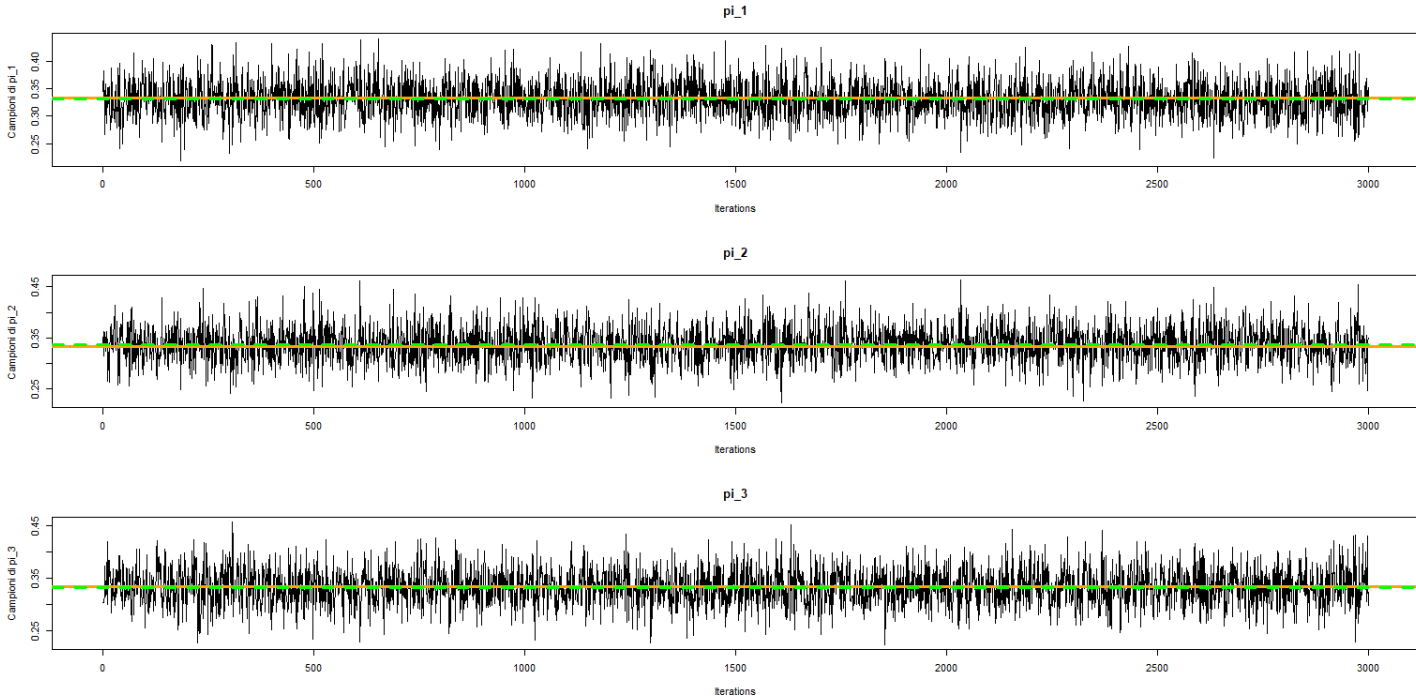


Figure 14: Catene di campioni a posteriori di  $\pi$

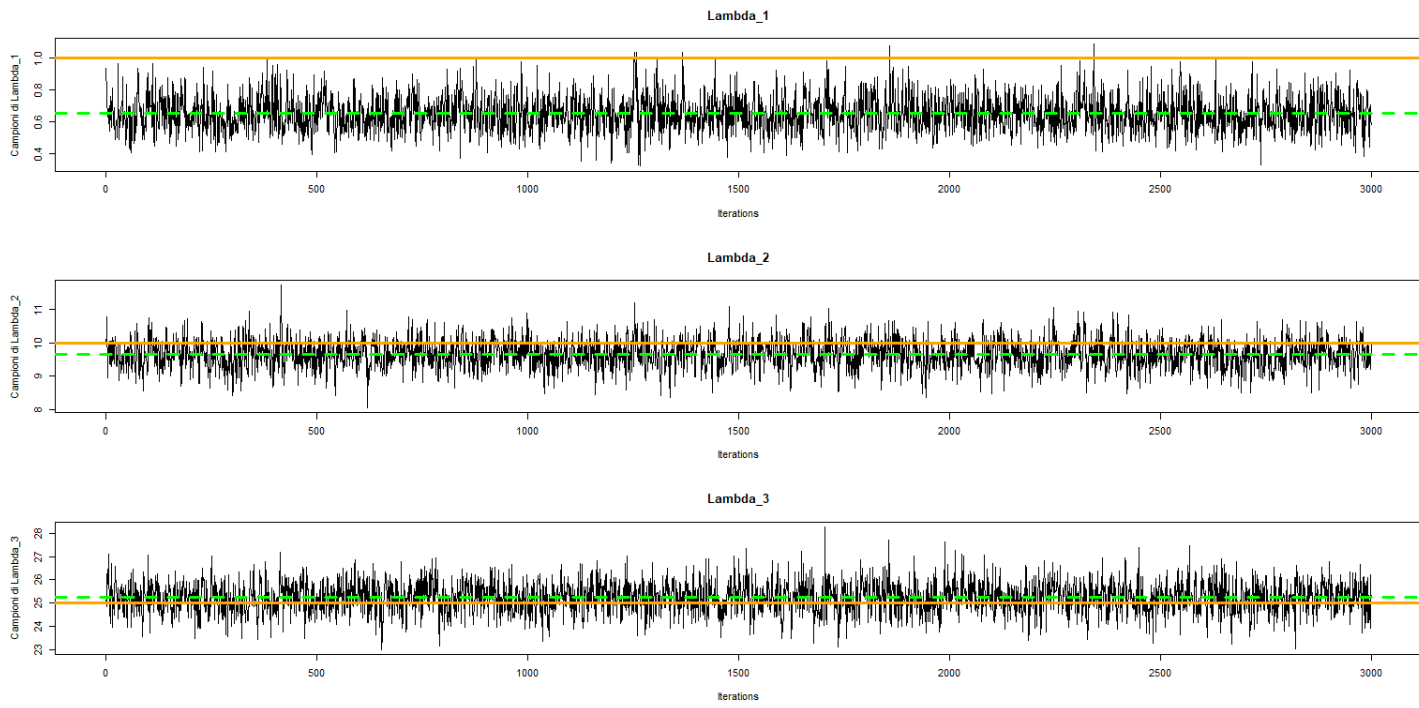


Figure 15: Catene di campioni a posteriori di  $\lambda$

Confrontato con Cecilia Bergamini, Rossana Ciarfaglia, Letizia Chiera.