

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**Un modello gerarchico Bayesiano
per la stima della fecondità regionale**

Relatore Prof. Stefano Mazzuco
Dipartimento di Scienze Statistiche

Laureando: Federico Cauduro
Matricola N. 1013655

Anno Accademico 2014/2015

Indice

Introduzione	5
Capitolo 1 - Tavole di fecondità	7
1.1 Cenni metodologici	7
1.2 Determinazione delle nascite	9
Capitolo 2 - La distribuzione Normale Asimmetrica (SN)	13
2.1 Definizione	14
2.2 Inferenza	16
2.3 Normale Asimmetrica Unificate (SUN)	18
2.4 Skew-Normal e i tassi di fecondità	19
Capitolo 3 – Un approccio Bayesiano	23
3.1 Modelli Bayesiani	23
3.1.1 Modelli gerarchici Bayesiani	24
3.2 Markov Chain Monte Carlo	25
3.2.1 Gibbs Sampling	27
3.2.2 Metropolis Hasting	28

Capitolo 4 – Il modello	29
4.1 Distribuzione a priori e a posteriori di α	29
4.2 Distribuzioni a priori e a posteriori di ξ e ω	30
4.3 I dati e l'algoritmo	31
4.3.1 I dati	31
4.3.2 Costruzione dell'algoritmo	33
4.4 I risultati	34
Capitolo 5 – Conclusioni	37
Bibliografia	41
Appendice A - Grafici del 1964 senza Metropolis Hasting	43
Appendice B - Grafici del 1964 con Metropolis Hasting	51
Appendice C - Grafici del 1985 con Metropolis Hasting	60
Appendice D - Metropolis Hasting, il codice	68

Introduzione

Nella presente tesi vengono analizzati i tassi di fecondità per età e per regione dell'Italia degli anni 1964, ovvero da quando le regioni Abruzzo e Molise vennero riconosciute come due distinte regioni, 1985 e 2004. I tassi di fecondità sono stati presi in considerazione per svariati studi in tutte le loro forme, per anno, per età delle donne, per paesi, per ripartizioni geografiche e per regioni, e usando diversi modelli per poterli stimare.

In questo studio si è voluto far utilizzo della distribuzione normale asimmetrica, la quale sarà descritta brevemente nei capitoli seguenti. La scelta di questa distribuzione è dovuta dal fatto che i tassi di fecondità delle varie regioni italiane dal 1964 al 2004 possono assumere forme asimmetriche nei primi anni, ma via via che gli studi di sulla fecondità si avvicinano ai giorni nostri, il TFT tende ad avere una forma simmetrica simile a quella di una distribuzione normale. L'approccio scelto per l'analisi è di tipo bayesiano gerarchico per poter ricavare, dopo aver definito le distribuzioni a priori, le distribuzioni a posteriori degli iperparametri e aggiornando la stima della varianza del parametro di posizione della Skew-Normal tramite i dati in nostro possesso.

Per poter affrontare questo studio è stato creato un algoritmo di Gibbs sampling sfruttando alcuni risultati trovati nel lavoro proposto da Scarpa e Canale (2013) e utilizzati poi da Griggio Riccardo per la costruzione della sua tesi e introducendo una parte di Metropolis Hasting per lo studio gerarchico.

Capitolo 1 – Tavole di fecondità

Le tavole di fecondità rappresentano la base per la previsione dei modelli riproduttivi utilizzata nell'ambito delle previsioni demografiche diffuse e aggiornate periodicamente dall'Istat. Tramite esse si può analizzare la fecondità per ordine di nascita, con riferimento sia agli anni di calendario che alle coorti femminili, a livello regionale.

L'utilizzo di queste tavole fa sì di disporre di un sistema di misure di coorte e di periodo per permettere lo studio dell'andamento della fecondità dal passato fino ai giorni nostri. Questo non riscontra interesse solo in ambito demografico ma ha una rilevante importanza anche in ambito sociale ed economico. L'ordine di nascita rappresenta un punto saliente per lo studio dell'evoluzione e dei comportamenti di riproduzione delle varie generazioni, e con l'analisi regionale si ha l'opportunità di mettere in luce differenti modelli di fecondità e le loro variazioni nel tempo.

1.1 Cenni metodologici

Per poter ottenere una stima corretta dei parametri per lo studio di una popolazione, bisogna essere a conoscenza di quanto a lungo vivono i singoli individui. Tali informazioni vengono riassunte nelle tavole di mortalità, da dove si può trovare il numero di individui che sopravvivono per intervallo di età, e nelle tavole di fecondità che riportano il numero di figli che una donna può avere mediamente ad ogni intervallo di età.

Alla base della costruzione delle tavole di fecondità vi è il calcolo dei tassi di fecondità specifici per età della madre, a livello regionale per ogni anno del calendario.

Prendendo in considerazione l'anno t , si può calcolare il tasso di fecondità specifico per età attraverso le informazioni dette $N(x,t)$ e $\bar{F}(x,t)$ che rappresentano rispettivamente le nascite dell'anno t da madri di età compiuta $(x,x+1)$ e gli anni-persona vissuti dalla donne osservate tra le età x e $x+1$ nell'anno t .

$$f(x, t) = \frac{N(x,t)}{\bar{F}(x,t)} \quad (1.1)$$

Altri indici che sono importanti per ottenere nuove informazioni sono:

- a) Il tasso di fecondità totale, (TFT), usato in particolar modo per il confronto del numero di nascite tra popolazioni aventi caratteristiche diverse anche di tempi diversi dato che non risente della struttura per età delle popolazioni.

$$TFT(t) = \sum_{x=\alpha}^{\beta} f(x, t) = \begin{cases} > 2.1 & \text{ringiovanimento e aumento della pop.} \\ = 2.1 & \text{struttura costante della pop.} \\ < 2.1 & \text{invecchiamento e calo della pop.} \end{cases}$$

dove

α = età minima per la quale le donne possono avere figli

(solitamente posta a 14)

β = età massima per la quale le donne possono avere figli

(solitamente posta a 49)

f_x = tasso di fecondità specifico per l'età x , cioè il numero medio

di figli avuti a quell'età.

b) L'età media delle donne alla nascita dei figli al tempo t

$$\bar{x}(t) = \frac{\sum_{x=\alpha}^{\beta} x * f(x,t)}{TFT(t)} .$$

1.2 Determinazione delle nascite

Per calcolare i tassi specifici di fecondità si deve essere in possesso dei dati delle nascite dei nati ed età della madre. Tali informazioni sono state stimate negli anni attraverso differenti fonti e metodologie dalle quali possono essere distinti in tre tipi diversi di rilevamento dei dati, che sono (fonte demo.istat):

1. Rilevazione individuale delle nascite di fonte stato civile Questa metodologia è stata sfruttata dall'Istat per poter ottenere la struttura delle nascite per ordine ed età del periodo dal 1952 al 1998 tramite la selezione delle nascite viventi della popolazione presente e residente in un anno di calendario.

La procedura per l'acquisizione dei dati utilizzava due modelli, denominati ISTAT D1 e D2 a seconda se i nati erano rispettivamente di sesso maschile o femminile, suddivisi entrambi in tre sezioni:

- la prima sezione riguardante lo Stato Civile e l'Anagrafe;
- la seconda sezione contenente informazioni di tipo sanitario;
- la terza sezione, invece, conteneva notizie di natura socio-demografica (titolo di studio, condizione professionale, settore di attività economica dei genitori) acquisite direttamente dal dichiarante.

Tale approccio fu interrotto con l'introduzione della legge 127/97, detta Bassanini Bis, la quale permise di dichiarare la nascita del nato vivo non più solamente all'Ufficio di Stato Civile del Comune nel cui territorio dove si verificasse, ma anche presso l'Ufficio di Stato Civile del Comune di residenza dei genitori o direttamente alla direzione sanitaria del centro di nascita presso il quale era avvenuto tale evento.

2. Rilevazione degli iscritti in anagrafe per nascita.

Questo sistema fu introdotto nel 1° gennaio 1999 dove le informazioni erano di fonte Anagrafe e non più di Stato Civile. Permetteva, a livello territoriale comunale, di conoscere le nascite avvenute e di poterle rendere classificabili per informazioni riguardanti il nascituro (sesso, data di nascita, luogo di nascita, cittadinanza) e informazioni riguardanti entrambi i genitori come età e coorti, stati civili e cittadinanze e dal 2004 è stata introdotta un'ulteriore variabile di rilevante importanza che è servita a colmare una lacuna per il calcolo degli indicatori su cui si basa la costruzione delle tavole di fecondità regionali, ovvero il numero dei componenti minorenni presenti nella scheda anagrafica di famiglia.

Questo nuovo metodo fu adottato per consentire il calcolo dei principali indicatori di fecondità, come per esempio il TFT e l'età media al parto, per le donne a livelli comunali e ambiti territoriali sovracomunali amministrativi o meno.

3. Indagine campionaria sulle nascite.

Questo ultimo metodo di rilevazione è stato utilizzato negli anni 2002 e 2005 permettendo di ricostruire la struttura della fecondità per ordine di nascita a livello regionale fornendo, inoltre, un quadro accurato sulle opinioni e le aspettative di fecondità delle madri, sugli aspetti familiari e

sociali di contesto delle nascite, sulle strategie di cura adottate dalle famiglie e sull'impatto che la nascita di un figlio ha sulle scelte lavorative delle donne (cit. demo.istat).

L'edizione del 2002 prese in considerazione le donne aventi figli iscritti all'anagrafe dal secondo semestre del 2000 e del primo del 2001, mentre nell'edizione del 2005 sono state intervistate coloro che hanno registrato i figli all'anagrafe nell'anno 2003. Le interviste sono avvenute con tecnica C.A.T.I (Computer Assisted Telephone Interview) con due tipologie di questionari uno in versione ridotta somministrato all'intero campione per ottenere stime a livello regionale delle principali informazioni demografiche e socio-sanitarie sui genitori, il nato e il parto, e uno in versione esteso rivolto ad un terzo del campione con una rappresentatività a livello territoriale più ampio per approfondire temi riguardanti le dinamiche di fecondità in atto e per supporto informatico per possibili interventi di policy. L'obiettivo era quello di evidenziare i vincoli che si frappongono nel conciliare la dimensione familiare con quella lavorativa extra-domestica e le strategie di conciliazione messe in atto dalle neo-madri.

Capitolo 2 – La distribuzione Normale Asimmetrica (SN)

Azzalini (1985), riprendendo le caratteristiche assegnate alla funzione f_0 , la quale viene definita come la funzione di densità base di una densità probabilistica simmetrica attorno un punto x_0 se $f_0(x-x_0) = f_0(x_0-x)$ per tutti i valori di x , modulata grazie a un fattore che può essere scelto liberamente perché deve soddisfare delle semplici condizioni, introdusse il seguente lemma, il quale risulta essere molto importante per quanto riguarda lo sviluppo del nostro modello.

Proposizione 2.1 Siano f_0 una funzione di densità definita in \mathbb{R}^d , $G_0(\cdot)$ una funzione di ripartizione continua sull'asse reale, e $w(\cdot)$ una funzione a valori reali su \mathbb{R}^d , tali che

$$f_0(-x) = f_0(x), \quad w(-x) = -w(x), \quad G_0(-y) = 1 - G_0(y) \quad (2.1)$$

per tutte le $x \in \mathbb{R}^d, y \in \mathbb{R}$. Allora

$$f(x) = 2f_0(x)G_0\{w(x)\}, \quad (2.2)$$

è una funzione di densità su \mathbb{R}^d .

Impostando $G(x) = G_0\{w(x)\}$ in (2.2), questa può essere scritta come

$$f(x) = 2f_0(x)G(x) \quad (2.3)$$

dove

$$G(x) \geq 0, \quad G(x) + G(-x) = 1. \quad (2.4)$$

Le densità del tipo (2.2) o (2.3) descritte in precedenza, vengono chiamate '*skew-symmetric*', definite così dalla semplice forma della funzione $w(x)$, dove la caratteristica principale è l'asimmetria.

2.1 Definizione

Tra le tante famiglie di distribuzioni che possono essere generate dalla (2.2), la principale estensione che possiamo considerare è quella di una distribuzione normale-asimmetrica.

Se, sempre nella (2.2), si pongono $f_0 = \phi$, $G_0 = \Phi$ e $w(x) = \alpha x$, con α che assume valori reali, si ottiene la funzione di densità

$$\varphi(x; \alpha) = 2\phi(x)\Phi(\alpha; x) \quad (-\infty < x < \infty). \quad (2.5)$$

dove in Figura 2.1 si possono vedere i grafici della funzione di densità per alcuni valori di α scelti come esempio da Azzalini A. e Capitanio A. (2014).

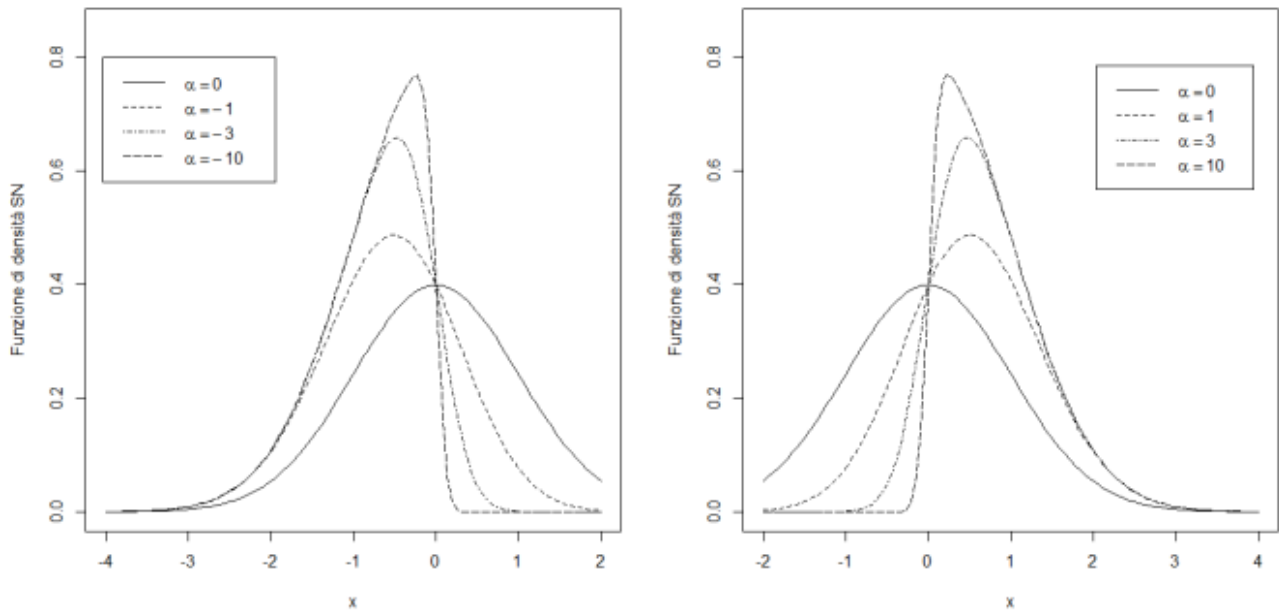


Figura 2.1 Funzione di densità della distribuzione normale asimmetrica con α posto uguale a 0,-1,-3,-10 nel grafico di sinistra, e α uguale a 0,1,3,10 in quello di destra.

Ponendo Z come una variabile casuale e continua avente densità pari alla (2.5), la variabile

$$Y = \xi + \omega Z \quad (\xi \in \mathbb{R}, \omega \in \mathbb{R}^+) \quad (2.6)$$

con parametro di posizione ξ , parametro di scala ω e α parametro di forma viene detta *skew-normal (SN)*. Questa viene scritta come $Y \sim SN(\xi, \omega^2, \alpha)$ con funzione di densità in $x \in \mathbb{R}$ pari a

$$f(x; \xi, \omega^2, \alpha) = 2\omega^{-1} \phi\left(\frac{x-\xi}{\omega}\right) \Phi\left\{\alpha \left(\frac{x-\xi}{\omega}\right)\right\} \equiv \frac{1}{\omega} \varphi\left(\frac{x-\xi}{\omega}; \alpha\right). \quad (2.7)$$

Nel caso in cui Z sia variabile casuale di una skew-normal con parametri $\xi = 0$, $\omega^2 = 1$ e α valgono le seguenti proprietà (Azzalini 2014):

1. $\varphi(x; 0) = \phi(x) \quad \forall x$;
2. $\varphi(0; \alpha) = \phi(0) \quad \forall \alpha$;
3. $-Z \sim SN(0, 1, -\alpha)$ è l'equivalente di $\varphi(-x; \alpha) = \varphi(x; -\alpha) \quad \forall x$;
4. $\lim_{\alpha \rightarrow \infty} \varphi(x; \alpha) = 2\phi(x)I_{[0, \infty)}(x) \quad \forall x$;
5. $Z^2 \sim \chi_1^2$ indipendentemente da α ;
6. se $Z' \sim SN(0, 1, \alpha')$ con $\alpha' < \alpha$ allora $Z' <_{st} Z$.

Questa classe di distribuzioni comprende modelli con diversi livelli di asimmetria e curtosi, oltre alla distribuzione normale stessa ($\alpha = 0$). Essa permetterebbe di partire dal modello normale, con l'introduzione del parametro α supplementare che controlla l'asimmetria, e il suo utilizzo garantirebbe una maggiore robustezza nei metodi inferenziali, e probabilmente, dei modelli migliori per l'adattamento dei dati.

2.2 Inferenza

Per quanto riguarda la stima dei parametri di scala e di posizione della distribuzione normale asimmetrica siamo in grado di ottenere dei valori numerici alla stima di massima verosimiglianza.

Considerando il campione Y_1, \dots, Y_n dove $Y \sim SN(\xi, \omega^2, \alpha)$, la funzione di massima verosimiglianza la si ottiene da

$$L_{SN}(\xi, \omega, \alpha) = \frac{2^n}{\omega^n} \prod_{i=1}^n \phi\left(\frac{y_i - \xi}{\omega}\right) \prod_{i=1}^n \Phi\left(\alpha \frac{y_i - \xi}{\omega}\right) \quad (2.8)$$

e la log-verosimiglianza

$$l_{SN}(\xi, \omega, \alpha) = 2 \ln 2 - n \ln \omega + \sum_{i=1}^n \ln \phi\left(\frac{y_i - \xi}{\omega}\right) + \sum_{i=1}^n \ln \Phi\left(\alpha \frac{y_i - \xi}{\omega}\right). \quad (2.9)$$

Tramite il calcolo delle derivate prime rispetto a ξ , ω^2 e α , e ponendole uguali a zero si ottengono le loro stime $(\hat{\xi}, \hat{\omega}^2, \hat{\alpha})$.

Nel caso i dati del campione fossero positivi/negativi, per dei valori fissati di ξ e ω^2 , la funzione di log-verosimiglianza sarebbe un funzione crescente/decrescente di α che produrrebbe delle stime marginali. Se α fosse posto uguale a 0 si avrebbe che la matrice d'informazione attesa di Fisher è singolare.

Per decidere se utilizzare la distribuzione normale o la distribuzione normale asimmetrica si può far ricorso all'utilizzo del test del rapporto di verosimiglianza aventi ipotesi

$$\begin{cases} H_0 = Y \sim SN(\xi, \omega^2, \alpha = 0) \\ H_1 = Y \sim SN(\xi, \omega^2, \alpha \neq 0) \end{cases}$$

$$T = \frac{L_{SN}(\hat{\xi}, \hat{\omega}^2, \alpha=0)}{L_{SN}(\hat{\xi}, \hat{\omega}^2, \hat{\alpha})} \quad (2.10)$$

dove

$$-2 \ln T \sim \chi_1^2 \quad \text{sotto } H_0.$$

Se $-2\ln T$ fosse elevato allora si rifiuta l'ipotesi nulla dato che la SN avente come parametri le stime di questi avrebbe un miglior adattamento dei dati rispetto alla distribuzione normale.

2.3 Normale Asimmetrica Unificata (SUN)

La normale asimmetrica unificata, scritta anche con l'acronimo SUN e introdotta da Arellano-Valle e Azzalini nel 2006, è un'estensione della skew-normal base e sarà successivamente utilizzata per il calcolo della stima a posteriori del parametro di forma α .

Prendendo in considerazione la trasformazione

$$Y = \xi + \omega(B_0 V_{0\gamma} + B_1 V_1) \quad (2.11)$$

con $V_{0\gamma} \sim LTN_m(-\gamma; 0, \Gamma)$, variabile normale multivariata con componenti troncati sotto $-\gamma$, e $V_1 \sim N_d(0, \Psi)$ variabili indipendenti (dove Γ e Ψ sono matrici di correlazione), $B_0 = \Delta\Gamma^{-1}$, B_1 è una matrice $d \times d$ tale che

$$B_1 \Psi B_1^T = \bar{\Omega} - \Delta\Gamma^{-1}\Delta^T \quad (2.12)$$

dove i termini vengono presi da una partizione della matrice di correlazione

$$\Omega^* = \begin{pmatrix} \Gamma & \Delta^T \\ \Delta & \bar{\Omega} \end{pmatrix}, \quad (2.13)$$

la funzione di densità di Y , che sarà la densità normale asimmetrica unificata che possiamo scriverla come $Y \sim SUN_{m,d}(\xi, \gamma, \omega, \bar{\Omega}, \Delta, \Gamma)$, è data dalla seguente espressione,

$$f(y) = \phi_d(y - \xi; \omega \bar{\Omega} \omega) \frac{\Phi_m(\gamma + \Delta^T \bar{\Omega}^{-1} \omega^{-1} (y - \xi); \Gamma - \Delta^T \bar{\Omega}^{-1} \Delta)}{\Phi_m(\gamma; \Gamma)} \quad (2.14)$$

dove per $y \in \mathbb{R}^d$, $\Phi_d(\cdot; \Sigma)$ è la funzione di distribuzione di una normale d -variata con matrice di varianze e covarianze Σ e ω è una matrice diagonale $d \times d$.

2.4 Skew-normal e i tassi di fecondità

La distribuzione skew-normal, o normale asimmetrica, è stata ben volentieri usata in alcune ricerche inerenti allo studio della fecondità. Il vantaggio di questa distribuzione per questo tipo di studi, come da nome, è dovuto all'asimmetria che fa sì che l'andamento dei dati riguardanti il numero di nati vivi, ricavati dalle tavole di fecondità, si adatti così in maniera più adeguata rispetto ad altre famiglie di distribuzione. Questo è dovuto al fatto che negli anni la natalità per motivi politici, guerre, epidemie, ripartizione geografica, etc., portando ad avere curve di fecondità con ampiezze diverse dovute alla velocità che hanno per raggiungere il punto massimo e la velocità per arrivare alla fine della campana che viene disegnata.

Hoem (1981) scrisse la curva della fecondità come

$$g(y; R, \vartheta_2, \dots, \vartheta_r) = Rh(x; \vartheta_2, \dots, \vartheta_r) \quad (2.15)$$

dove $g(y; R, \vartheta_2, \dots, \vartheta_r)$ è il tasso di fecondità per l'età y , R è l' r -esimo parametro che rappresenta il tasso di fecondità totale TFR, e $h(\cdot; \vartheta_2, \dots, \vartheta_r)$ è la funzione di densità sull'asse reale avente $r-1$ parametri. Di quest'ultima sono state date molte specificazioni attraverso l'utilizzo di distribuzioni come quella di Hadwinger (Gaussiana Inversa), Gamma e Beta. Hoem (1981) definì anche due modelli basati sulle funzioni di regressione Spline e polinomiale dai quali risultò che il modello creato su funzioni Spline forniva risultati migliore per i dati quando venivano utilizzati 10 o più parametri. Invece nel caso in cui fosse presente un numero inferiore di parametri i dati si adattavano meglio con l'utilizzo di distribuzioni gamma, Coale-Trussel e in alcuni casi quella di Hadwinger.

Ma recentemente si è visto che questi modelli non sempre sono adeguati per descrivere la fertilità di alcuni paesi perché utilizzano funzioni unimodali quando sarebbe più opportuno uno studio con funzioni bimodali. Chandola (1999) definì un 'modello Hadwinger misto'. Di seguito vediamo la funzione originale di Hadwinger (Hadwinger 1940),

$$g(y; a, b, c) = \frac{ab}{c} \left(\frac{c}{y}\right)^{3/2} \exp\left\{-b^2 \left(\frac{c}{y} + \frac{y}{c} - 2\right)\right\}. \quad (2.16)$$

Questa funzione venne definita da Hoem in modo leggermente diverso partendo dalla PDF di Hadwinger e aggiungendo il parametro R . Il modello (2.16) è equivalente al modello (2.15) nel caso in $R = a\sqrt{\pi}$ e $g(\cdot)$ sia la PDF di Hadwinger.

Mentre il modello misto di Chandola (1999) deriva dal modello (2.16), ed è

$$g(y; a, m, b_1, b_2, c_1, c_2) = m \frac{ab_1}{c_1} \left(\frac{c_1}{y}\right)^{3/2} \exp\left\{-b_1^2 \left(\frac{c_1}{y} + \frac{y}{c_1} - 2\right)\right\} \\ + (1 - m) \frac{ab_2}{c_2} \left(\frac{c_2}{y}\right)^{3/2} \exp\left\{-b_2^2 \left(\frac{c_2}{y} + \frac{y}{c_2} - 2\right)\right\} \quad (2.17)$$

dove $0 \leq m \leq 1$ è il parametro che determina la dimensione delle due popolazioni sottostanti.

Altre funzioni usate spesso nelle specificazioni demografiche sono la funzione Gamma (Hoem et al., 1981)

$$f(y) = R \frac{1}{\Gamma(b)c^b} (y - d)^{b-1} \exp\left\{-\left(\frac{y-d}{c}\right)\right\}, \quad \text{per } y > d \quad (2.18)$$

dove d rappresenta l'età di gravidanza più bassa mentre R determina il livello di fecondità. I parametri b e c sono stati riparametrizzati da Hoem con la moda m , la media μ e la varianza σ^2 , dove $c = \mu - m$ e

$$b = (\mu - d)/c = \sigma^2/c^2.$$

La funzione Beta utilizzata Hoem et al. (1981) è stata definita come segue,

$$f(y) = R \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)} (\beta - \alpha)^{-(A+B-1)} (y - \alpha)^{A-1} (\beta - y)^{B-1} \quad \text{per } \alpha < y < \beta \quad (2.19)$$

Dove i parametri sono legati alla media v e varianza τ^2 tramite le seguenti relazioni

$$B = \left\{ \frac{(v - \alpha)(\beta - v)}{\tau^2} - 1 \right\} \frac{\beta - v}{\beta - \alpha} \quad e \quad A = B \frac{v - \alpha}{\beta - v} .$$

Come ultima funzione ma non per questo la meno importante è la distribuzione Normale asimmetrica, (2.7) descritta nei precedenti paragrafi del capitolo 2,

utilizzata anche da R. Griggio nella sua tesi per lo studio dei tassi di fecondità del Costa Rica.

In questa ricerca, a differenza di quella di Griggio, è stata presa in considerazione la varianza del parametro ξ ponendola uguale per tutte le regioni ad ogni interazione, ma con la possibilità di variare all'interazione successiva dovuta all'algoritmo del Metropolis Hasting utilizzato per la creazioni di un modello bayesiano gerarchico.

Capitolo 3 - Un approccio Bayesiano

3.1 Modelli Bayesiani

Rappresentare l'incertezza dei parametri tramite distribuzioni di probabilità è il tratto distintivo dell'inferenza Bayesiana; il processo di apprendimento consiste nell'aggiornamento iniziale del parametro ϑ alla luce dei dati osservati ottenendo una distribuzione di probabilità per ϑ , $\pi(\vartheta|\mathbf{x})$, detta distribuzione a posteriori.

Tale procedimento utilizza il teorema di Bayes,

$$\pi(\vartheta|\mathbf{x}) = \frac{\pi(\vartheta)L(\mathbf{x}|\vartheta)}{\int_{\vartheta} \pi(\vartheta)L(\mathbf{x}|\vartheta) d\vartheta} \quad (3.1)$$

con $f(\mathbf{x}|\vartheta)$ funzione di densità di \mathbf{x} dato ϑ .

In presenza di alcune situazioni, l'utilizzo di un approccio Bayesiano rispetto a quello frequentista può risultare vantaggioso specialmente nel momento in cui ci si trova a dover svolgere un'analisi multilevel avente un numero ridotto di gruppi di secondo livello oppure come nel caso in cui sia presente un modello complicato da studiare evitando così di ottenere delle stime di variabilità imprecise dovute a mancanza d'informazione.

Una caratteristica fondamentale in ambito bayesiano è dovuta al fatto che tutti i parametri sono realizzazioni di variabili casuali aventi quindi una distribuzione a priori permettendoci di ottenere una distribuzione di probabilità e non una stima di un parametro, dando modo di ottenere degli intervalli di credibilità, deviazioni standard, scarti interquartili per tutti i parametri e le loro trasformate.

3.1.1 Modelli Bayesiani Gerarchici

Sia x_1, \dots, x_n un campione di dati i.i.d. per i quali si assume un modello parametrico avente legge $p(x|\vartheta)$; come accennato in precedenza l'inferenza bayesiana si basa sulla distribuzione a posteriori per il parametro ma volendo abbandonare l'ipotesi di identica distribuzione, si può supporre che ogni singolo dato derivi da una diversa distribuzione, cioè

$$x_i \sim p(x_i|\vartheta_i) \quad \forall i \quad (3.2)$$

Portando ad avere una distribuzione a posteriori per ogni parametro ϑ_i :

$$E(\vartheta_i|x_i) = \vartheta_i \quad (3.3)$$

Solitamente la distribuzione a priori del parametro dipende da parametri che non sono menzionati nella legge $p(\cdot|\cdot)$, quindi si può effettuare una sostituzione dell'a priori con una distribuzione avente parametri che non siano fissati ma che abbiano a loro volta una distribuzione di probabilità, detta iper-priori, ove i parametri vengono detti iper-parametri, φ . Otterremo così una nuova a posteriori,

$$\pi(\vartheta, \varphi|x) \propto L(x|\vartheta) \times \pi(\vartheta|\varphi) \times \pi(\varphi) \quad (3.4)$$

dove

$L(x|\vartheta)$ è la verosimiglianza per x ;

$\pi(\vartheta|\varphi)$ è la a priori per ϑ , avente parametro ignoto φ ;

$\pi(\varphi)$ è la a priori di secondo livello, con distribuzione “fissata”.

Con questo procedimento vengono stimati i parametri di secondo livello attraverso i dati ottenendo stime non più dipendenti e banali dovuto dal fatto che le a posteriori dei ϑ_i dipendono anche da φ .

3.2 Markov Chain Monte Carlo

Le distribuzioni a posteriori dei parametri possono essere trovate analiticamente solamente in certi casi. Il metodo Markov Chain Monte Carlo viene utilizzato per simulare dei campioni da una distribuzione di probabilità e, inoltre, permette di simulare anche distribuzioni multivariate e a meno di una costante di proporzionalità, rendendo così possibile determinare il valore dell'integrale posto a denominatore.

Il metodo si basa su due metodi congiunti, uno per l'approssimazione di integrali (Monte Carlo) e il secondo basato sulla costruzione di catene markoviana che abbiano come distribuzione limite la distribuzione desiderata, che in ambito bayesiano è $\pi(\vartheta|\mathbf{x})$.

In generale una catena di Markov si ottiene da una serie di variabili X_1, X_2, \dots, X_t aventi la stessa sigma-algebra A dove

$$P(X_{t+1} \in A | X_1, \dots, X_t) = P(X_{t+1} \in A | X_t) \quad \forall t \quad \forall A \in A \quad (3.5)$$

Se questa catena ha delle specifiche caratteristiche è dimostrato che esiste una distribuzione limite cioè che

$$\lim_{t \rightarrow \infty} P_t^j = \pi_j \quad (3.6)$$

Per poter avere una distribuzione limite la catena markoviana deve soddisfare determinate condizioni:

1. Aperiodicità: per ogni coppia $(i,j) \in S^2$ esiste un intero k^* tale per cui $\forall k > k^*$ vale che $P(X_{t+k}=j|X_t=i) > 0$;
2. Irriducibilità: per ogni coppia $(i,j) \in S^2$ esiste un intero k tale per cui $P(X_{t+k}=j|X_t=i) > 0$;
3. Ricorrenza positiva: se, indicando con T_{ii} il valore minimo di t tale per cui $X_t=i|X_0=i, \forall i \in S$ vale che $E(T_{ii}) < \infty$.

Se soddisfatte queste tre regole la catena trovata viene detta *ergodica*. Dove l'aperiodicità sancisce che non è possibile che la catena "oscilli" tra due o più distribuzioni limite, l'irriducibilità che è sempre possibile la transizione dallo stato i a qualsiasi stato j , la ricorrenza positiva che se il numero atteso di transizioni di ritorno allo stato originale è finito.

Tra gli algoritmi MCMC i più comuni sono tre e sono:

1. Campionamento Metropolis;
2. Campionamento Metropolis-Hastings;
3. Campionamento Gibbs Sampling.

Indipendentemente dall'algoritmo che viene utilizzato, bisogna controllare la convergenza e il "mixing" per verificare che effettivamente si stia utilizzando una distribuzione limite. Per poter far questo si può utilizzare il *trace plot* che mostra l'andamento delle simulazioni. Più l'autocorrelazione è elevata più lenta sarà la convergenza e maggiore sarà il numero di iterazioni necessario per raggiungerla e di conseguenza il "mixing" è buono se l'autocorrelazione è bassa e cattivo se alta.

Oltre al trace plot si può usare una statistica R, rapporto tra la varianza tra le catene markoviane e la varianza entro le catene, dove R deve essere pari a 1, mentre se è molto lontano da questo esito vuol dire che la convergenza non si è ancora raggiunta, oppure confrontare i valori della prima parte con quelli dell'ultima parte della catena, strategia di Geweke, o sfruttare il test di Rafttry-Lewis che indica quanto si deve specificare come burn-in e quante iterazioni fare per ottenere una maggior accuratezza della stima dei percentili desiderati, o infine si può osservare la numerosità campionaria effettiva (*Effective Sample Size - ESS*) dove più basso è il suo valore peggiore è il "mixing" della catena.

3.2.1 Gibbs Sampling

Questo algoritmo MCMC è usato come mezzo di inferenza bayesiana. E' un metodo randomizzato che di conseguenza produce risultati diversi ad ogni ciclo. Come primo passo l'algoritmo prende un valore iniziale ϑ , $\vartheta^0 = (\vartheta_1^0, \dots, \vartheta_p^0)$. Ad ogni avanzamento t, viene determinato un campione di ϑ_1^t della distribuzione

$$\pi(\vartheta_1 | \vartheta_2 = \vartheta_2^{t-1}, \dots, \vartheta_p = \vartheta_p^{t-1}, y)$$

e così via per gli altri campioni. Inoltre i campioni iniziali verranno scartati per il motivo che i primi campioni non provengono da $\pi(\vartheta|y)$ e quindi non

rappresentano accuratamente la distribuzione desiderata. Entra in gioco il *burn-in*, attraverso il quale si deve decidere a quanto far ammontare il numero di campioni iniziali. Questa quantità dipende dalla complessità del modello, per esempio in un semplice GLM la convergenza avviene in maniera abbastanza rapida e un burn-in di 100 iterazioni è più che sufficiente.

3.2.2 Metropolis-Hastings

Utilizzato nel caso che sia complicato trovare la distribuzione a posteriori con Gibbs sampling. Avendo un $t=0$ e ϑ^0 iniziali, viene generato un valore tramite una funzione di densità $q(\cdot|\vartheta_j^{t-1})$ per poi verificare se questo valore viene accettato o rifiutato con probabilità

$$\min \left\{ 1, \frac{\pi(\tilde{\vartheta}_j^t)L(y|\vartheta_j = \tilde{\vartheta}_j^t, -)q_j(\vartheta_j^{t-1}|\tilde{\vartheta}_j^t)}{\pi(\vartheta_j^{t-1})L(y|\vartheta_j = \vartheta_j^{t-1}, -)q_j(\tilde{\vartheta}_j^t|\vartheta_j^{t-1})} \right\} \quad (3.7)$$

dove $L(y|\cdot)$ è la verosimiglianza dato $\vartheta_j = \tilde{\vartheta}_j^t$, e il valore degli altri parametri e se $\tilde{\vartheta}_j^t$ viene “rifiutato” allora si tiene $\vartheta_j^t = \tilde{\vartheta}_j^{t-1}$.

Capitolo 4 – Il modello

In questo capitolo andremo a vedere le distribuzioni a priori e le distribuzioni a posteriori dei parametri di posizione ξ , di scala ω e di forma α della distribuzione Skew-Normal descritta nel capitolo 2 e studiati da B. Scarpa e A.Canale (2013), che ci permetteranno lo sviluppo del nostro modello, la preparazione dei dati per poter affrontare l'analisi e quella dell'algoritmo del quale sarà allegato parte del codice nell'Appendice D.

4.1 Distribuzione a priori e a posteriori di α

Al parametro di forma α viene assegnata come distribuzione a priori una Skew-Normal

$$\pi(\alpha) = \frac{2}{\psi_0} \phi\left(\frac{\alpha - \alpha_0}{\psi_0}\right) \Phi\left(\lambda_0 \frac{\alpha - \alpha_0}{\psi_0}\right), \quad (4.1)$$

dove gli iperparametri α_0 e ψ_0 rappresentano rispettivamente la posizione e la scala e λ_0 è l'iperparametro di forma.

La distribuzione a posteriori che si ottiene è la seguente (B.Scarpa, A.Canale),

$$\begin{aligned} \pi(\alpha; y) &\propto \phi\left(\frac{\alpha - \alpha_0}{\psi_0}\right) \Phi\left(\lambda_0 \frac{\alpha - \alpha_0}{\psi_0}\right) \prod_{i=1}^n \Phi(\alpha y_i) \\ &\propto \phi\left(\frac{\alpha - \alpha_0}{\psi_0}\right) \Phi_{n+1}\left(\begin{bmatrix} y\alpha_0 \\ 0 \end{bmatrix} + \begin{bmatrix} y \\ \lambda_0/\psi_0 \end{bmatrix} (\alpha - \alpha_0); I_{n+1}\right). \end{aligned} \quad (4.2)$$

La funzione di probabilità di densità (4.2) appartiene alla classe delle distribuzioni della SUN,

$$\alpha|y \sim SUN_{1,n|1}(\alpha_0, \gamma_1, \psi_0, 1, \Delta_1, \Gamma_1) \quad (4.3)$$

dove $\Delta_1 = [\delta_i]_{i=1, \dots, n+1}$ con $\delta_i = \psi_0 z_i (\psi_0^2 z_i^2 + 1)^{-1/2}$ e $z = (y^T, \lambda_0 \psi_0^{-1})^T$, $\gamma_1 = (\Delta_{1,1:n} \alpha_0 \psi_0^{-1}, 0)$, $\Gamma_1 = I - D(\Delta_1)^2 + \Delta_1 \Delta_1^T$.

La media e la varianza a posteriori risultano essere:

$$E[\alpha; y] = \zeta_1(0_n; \tilde{\Gamma}),$$

$$Var[\alpha; y] = \psi_0^2 + \zeta_2(0_n; \tilde{\Gamma}),$$

dove 0_n è un vettore di zeri di dimensione $n \times 1$ e $\tilde{\Gamma}$ è una matrice semidefinita positiva avente la diagonale pari a $1/\delta_i^2$ e il resto degli elementi, non appartenenti alla diagonale, pari a 1 ottenuti come $\tilde{\Gamma} = D(\Delta_1)^{-1} \Gamma_1 D(\Delta_1)^{-1}$.

4.2 Distribuzioni a priori e a posteriori di ξ e ω

Come per il parametro di forma α descritto in precedenza, anche ai parametri di posizione e di scala è stata assegnata una distribuzione a priori. ξ , parametro di posizione, si distribuisce come una Normale con media ξ_0 e varianza $k^* \omega^2$, mentre ω , parametro di scala, assume una distribuzione Gamma Inversa di parametri a e b ottenendo in questo modo la seguente distribuzione a priori:

$$\pi(\xi; \omega; \alpha) = N(\xi; \xi_0, k\omega^2) * InvGa(\omega^2; a, b) * SN(\alpha; \alpha_0, \psi_0, \lambda_0). \quad (4.4)$$

Nella seguente costruzione del Gibbs sampling saranno così campionati (B.Scarpa, A.Canale):

- $\xi \sim N(\hat{\mu}, \hat{k}\omega^2)$ con

$$\hat{\mu} = \frac{k \sum_{i=1}^n (y_i - \delta \eta_i) + (1 - \delta^2) \xi_0}{nk + (1 - \delta^2)}$$

$$\hat{k} = \frac{k(1 - \delta^2)}{nk + (1 - \delta^2)}$$

- $\omega \sim InvGam(a + (n + 1)/2, b + \hat{b})$ con

$$\hat{b} = \frac{1}{2(1 - \delta^2)} \left\{ \delta^2 \sum_{i=1}^n \eta_i^2 - 2\delta \sum_{i=1}^n \eta_i (y_i - \xi) + \sum_{i=1}^n (y_i - \xi)^2 + \frac{1 - \delta^2}{k} (\xi - \xi_0)^2 \right\}$$

Dove η_i sono delle variabili latenti normali standard indipendenti condizionatamente alle quali saranno prese in considerazione delle generiche i -esime osservazioni distribuite come delle normali aventi media pari a $\xi + \delta|\eta_i|$ e varianza $(1 - \delta^2)\omega^2$ che verranno aggiornate durante i primi passi del Gibbs.

4.3 I dati e l'algoritmo

4.3.1 I dati

Per poter iniziare a fare le prime analisi sui dati ci si è basati sulle tavole di fertilità per regione dell'anno 1964 prese dall'Istat (www.demo.istat.it).

Le tavole, come si può vedere in figura 4.1, sono formate da due colonne ognuna, dove nella prima troviamo l'età della madre al parto, dove il range va dall'età di 13 anni all'età di 50 anni, e sulla seconda i tassi di fecondità specifici per età (per mille abitanti).

L'Istat fornisce le tavole di fertilità dall'anno 1952 al 2004, ma la nostra scelta di partire dall'anno 1964 è dovuta dal fatto che da quell'anno, a seguito anche della nuova funzione che la Carta Costituzionale repubblicana assegnava alle regioni e l'approvazione della legge costituzionale, ci fu il distacco della regione Molise dall'Abruzzo, avendo così i dati dei tassi specifici per entrambe le regioni e permettendoci di mantenere un'analisi similare nello studio della variazione delle nascite negli anni presi in considerazione.

Età	TFT		
13	0	35	58
14	1	36	49
15	3	37	37
16	11	38	32
17	20	39	25
18	35	40	18
19	58	41	13
20	89	42	8
21	113	43	5
22	132	44	3
23	152	45	2
24	165	46	1
25	163	47	0
26	161	48	0
27	147	49	0
28	141	50	0
29	127		
30	120		
31	104		
32	90		
33	78		
34	65		

Figura 4.1 Tavola di fecondità della regione Piemonte del 1964

4.3.2 Costruzione dell'algoritmo

Dopo la creazione di un unico data set contenente tutti i valori dei tassi specifici delle tavole di fertilità di tutte le singole regioni appartenenti ad un anno, si è proseguito, come accennato nei capitoli precedenti, con la creazione di un Gibbs Sampling, il quale ci permette di poter trovare la convergenza dei nostri dati attraverso il giusto numero di iterazioni, calcolando durante il processo il valori dei parametri delle distribuzioni a posteriori di ξ , ω^2 e α descritti nei paragrafi 4.1 e 4.2.

L'algoritmo del gibbs sampling, utilizzato anche da Grigio Riccardo nella sua tesi, è stato arricchito con l'introduzione del Metropolis Hasting per effettuare l'aggiornamento della varianza di ξ , $k\omega^2$.

Come parametro iniziale da passare alla funzione gibbs, oltre ai parametri delle varie densità, è stato inserito σ^2 che non sarebbe altro che la varianza di ξ , attraverso il quale, una volta entro l'algoritmo, si calcola il valore di k che sarà dato dal rapporto tra σ^2 e ω^2 . E' stata effettuata questa scelta perché si vuole mantenere la stessa varianza per tutte le regioni in modo tale da avere una grande differenza del parametro di posizione.

Al termine di ogni ciclo del gibbs per tutte le regioni, si andrà a richiamare la funzione metropolis hasting alla quale verranno passati ξ , ω^2 , α di tutte le regioni italiane e σ^2 , parametro del quale vogliamo fare "l'aggiornamento".

Con questa seconda funzione si andrà a calcolare la funzione di densità della skew-normal totale di tutte le regioni sommata alla a priori di σ^2 proposta ottenuta da una distribuzione casuale normale e il risultato verrà sottratto al skew-normal di tutte le regioni sommata alla a priori di σ^2 passata come parametro che si distribuisce come una gamma.

Se l'esponenziale di questa differenza risulta essere superiore al valore casuale che assume una distribuzione uniforme di parametro pari a 1 allora il metropolis

hasting ritorna come valore di σ^2 il valore proposto altrimenti ritorna il valore del parametro inizialmente trovato e così via dopo ogni iterazione che eseguirà il gibbs per le 20 regioni.

4.4 I Risultati

Per curiosità e per verificare come cambia l'esito dello studio dei tassi di fecondità tra un modello bayesiano e un modello bayesiano gerarchico, inizialmente sono stati presi i dati del 1964 e dopo la pulizia del dataset sono stati utilizzati nel Gibbs Sampler prima senza e dopo con il passaggio attraverso il Metropolis Hasting (Risultati principali e grafici nell'Appendici B e C). Quindi nel caso in cui si utilizzi il modello bayesiano la varianza cambia per tutte le regioni, mentre nel caso in cui il modello fosse di tipo gerarchico bayesiano, la varianza ad ogni interazione potrebbe cambiare, grazie all'aggiornamento di essa attraverso il metropolis, a seconda se viene accettata l'ipotesi nulla o meno, ma rimane uguale per tutte le regioni fino alla fine dell'interazione stessa.

Successivamente l'attenzione è stata concentrata sui dati dell'anno 1985 studiati e analizzati tramite un modello bayesiano gerarchico e messi a confronto con i risultati ottenuti dal stesso tipo di modello ma dell'anno 1964.

Di seguito vengono riportati i risultati delle medie e degli intervalli di credibilità per le varie regioni dell'anno 1985 per vedere come cambia il TFT.

La tabella che segue è il risultato del gibbs sampling ottenuto da una quantità di iterazioni pari a 3000, con un burn in che vede interessati i primi 1500 esiti e con la varianza di ξ , ovvero $k*\sigma^2$, inizializzata a 40, valore deciso dopo molteplici tentativi fatti per poter raggiungere un risultato ottimale. Per prendere visione del grafico della densità della varianza di ξ e quelli delle densità che assume ξ per le varie regioni rimanda all'Appendice C.

Valori di ξ del 1985			
Regione	Mean	IC 0.10	IC 0.90
Piemonte	19,72	19,37	20,07
Valle D'Aosta	19,35	19,04	19,64
Liguria	20,19	19,84	20,53
Lombardia	20,04	19,72	20,36
Trentino A.A.	20,18	19,84	20,52
Veneto	20,03	19,70	20,33
Friuli V.G.	19,65	19,28	20,00
Emilia Romagna	19,36	19,02	19,67
Marche	19,90	19,54	20,24
Toscana	19,79	19,42	20,13
Umbria	19,60	19,25	19,94
Lazio	19,79	19,43	20,12
Campania	19,24	18,85	19,58
Abruzzo	19,46	19,17	19,78
Molise	19,24	18,93	19,51
Puglia	19,08	18,70	19,43
Basilicata	19,50	19,18	19,80
Calabria	19,08	18,76	19,38
Sicilia	18,54	18,20	18,87
Sardegna	19,72	19,38	20,06

Tabella 1. Principali valori di ξ per tutte le regioni dell'anno 1985 ottenuti dal modello bayesiano gerarchico

Capitolo 5. Conclusioni

Puntando l'attenzione specialmente su due anni, il 1964 e il 1985, in entrambi gli studi sviluppati sono state richieste 3000 iterazioni e la varianza del parametro di posizione è stata inizializzata a 40 ed è stato utilizzato un modello bayesiano gerarchico per l'analisi.

Mettendo a confronto gli esiti dei due modelli, dai grafici posti nelle Appendici B e C, per l'anno 1964, rispetto al 1985, risalta immediatamente all'occhio che è presente una sostanziale differenza nell'aggiornamento della varianza, infatti per il primo anno si nota come dal valore iniziale assegnatogli scenda verso lo zero per arrivare ad una convergenza, così sembrerebbe, dall'iterazione numero 2500 circa. Mentre per il 1985 la varianza di ξ , pur partendo da 40, prende una direzione opposta e dopo 3000 iterazioni sembrerebbe aver raggiunto la convergenza in maniera più difficoltosa.

Osservando i grafici delle varie regioni dei due anni, invece si osserva il fenomeno inverso, mentre per l'anno 1964 il parametro ξ fatica ad arrivare a convergenza, per il 1985 invece l'ha raggiunta. Questa differenza potrebbe esser data da una maggiore variabilità tra l'età media al parto delle madri nel 1985 rispetto al 1964? Per rispondere a questa domanda, nella Tabella 2, sono stati riportati i dati dell'età media al parto per poter fare questo confronto e i loro principali valori ottenuti dal summary presenti nella Tabella 3.

<i>Età Media al parto</i>		
Regione	1964	1985
Piemonte	27,22	27,65
Valle D'Aosta	27,26	27,37
Liguria	27,67	28,41
Lombardia	28,10	28,09
Trentino A.A.	29,57	28,45
Veneto	28,64	28,06
Friuli V.G.	27,71	27,77
Emilia Romagna	27,28	27,55
Marche	27,73	27,61
Toscana	27,57	27,80
Umbria	27,51	27,40
Lazio	27,90	27,81
Campania	29,37	27,44
Abruzzo	28,04	27,16
Molise	28,20	27,02
Puglia	28,96	27,32
Basilicata	28,84	27,36
Calabria	28,92	27,16
Sicilia	28,05	26,95
Sardegna	30,13	28,50

Tabella 2. Media dell'età al parto per regione degli anni 1964 e 1985

Anno	Min.	1st Qu.	Median	Mean	3rd Qu.	Max
1964	27,22	27,65	28,05	28,23	28,86	30,13
1985	26,95	27,35	27,58	27,64	27,87	28,5

Tabella 3. Risultati del summary ottenuti dall'età media al parto

In realtà dagli output ottenuti si può notare che la supposizione è errata, infatti l'anno 1964, conosciuto anche come l'anno del baby boom, l'età al parto è più variabile rispetto a quella del 1985 dovuta proprio al fatto che in quel periodo ci

fu un importante aumento delle nascite e il fatto che le donne abbiano più figli porta ad un aumento della variabilità.

In conclusione, la proposta dell'utilizzo di un modello bayesiano gerarchico per questa tipologia di analisi, dove la varianza di ξ è uguale per tutte le singole regioni, è dovuta al fatto che tale modello è un compromesso tra uno studio nel quale verrebbe analizzata la fecondità per tutta l'Italia, dove la variabilità non verrebbe ben spiegata, e uno studio che vedrebbe interessate le singole regioni, dal quale si otterrebbe una variabilità elevata.

Esaminando i due studi, quello del 1964 e quello del 1985, si può dire che il modello bayesiano gerarchico contribuisce in maniera significativa ad avere dei risultati interessanti ma che sicuramente non è facile arrivare a ottenere la perfezione e/o la convergenza delle stime dei parametri dovuta magari a qualche imperfezione dei dati e che forse potrebbe essere risolta aumentando il numero di iterazioni ottenendo dei risultati più accurati, ma che per mancanza di tempo non si è potuto accertare tale ipotesi dato che l'analisi per un singolo anno con 3000 iterazioni richiede al programma una durata di circa sette giorni.

Tuttavia, anche se ancora non si è potuta garantire la perfezione dell'algoritmo proposto, questo potrebbe essere un input per ulteriori studi di approfondimento di tale argomento.

Bibliografia

- Bruno Scarpa e Antonio Canale (2013), Informative Bayesian inference for the skew-normal distribution;
- Adelchi Azzalini with collaboration of Antonella Capitanio (2004), The Skew-Normal and Related Families;
- A. Azzalini (1985), A class of distribution which includes the normal ones;
- T.Chandola, D.A.Coleman and R.W.Hiorns (1999), Recent European fertility patterns fitting curves to 'distorted' distributions;
- Stefano Mazzuco and Bruno Scarpa (2013), Fitting age-specific fertility rates by a flexible generalized skew normal probability density function;
- Riccardo Griggio (2014), Tassi di fecondità per età del Costa Rica. Un'analisi Bayesiana;
- Hoem, Madsen, Nielsen, Ohlsen, Hansen, Rennermalm (1981), Experiments in modelling recent Danish fertility curves. *Demography*, 18: 231-244;
- Demo.istat.it .

Appendice A - Grafici del 1964 senza Metropolis Hasting

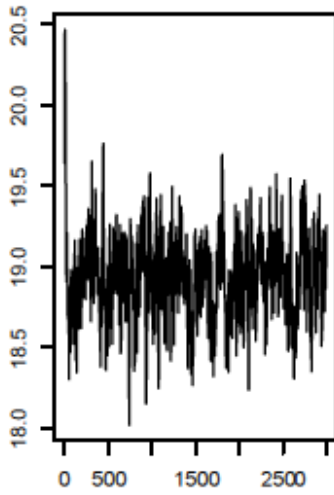
Valori di ξ del 1964 senza Metropolis Hasting			
Regione	Mean	IC 0.10	IC 0.90
Piemonte	18,94	18,60	19,25
Valle D'Aosta	18,90	18,55	19,23
Liguria	19,34	19,00	19,66
Lombardia	19,84	19,50	20,17
Trentino A.A.	20,63	20,25	20,99
Veneto	20,16	19,82	20,53
Friuli V.G.	19,31	18,94	19,67
Emilia Romagna	18,72	18,32	19,10
Marche	19,75	19,43	20,05
Toscana	19,28	18,91	19,65
Umbria	19,14	18,77	19,51
Lazio	19,38	19,03	19,73
Campania	20,19	19,77	20,58
Abruzzo	19,42	19,10	19,74
Molise	19,46	19,09	19,79
Puglia	19,74	19,38	20,11
Basilicata	19,92	19,51	20,28
Calabria	19,67	19,29	20,02
Sicilia	18,74	18,30	19,16
Sardegna	20,68	20,24	21,08

Tabella A.1 Principali valori di ξ per tutte le regioni dell'anno 1964 ottenuti dal modello bayesiano

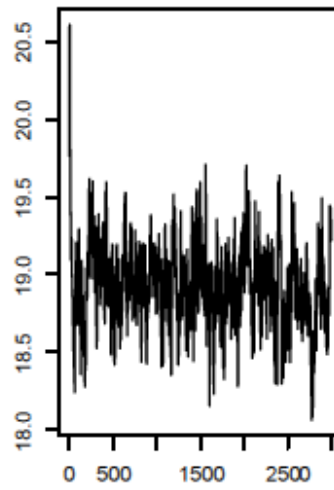
Questi primi grafici sono inerenti ad una prima verifica di analisi per il funzionamento del gibbs, e riguardano il modello bayesiano dell'anno 1964.

- Media dopo burn in
- IC 0.10 e 0.90 dopo burn in

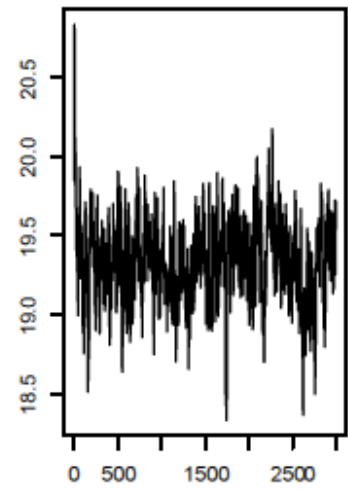
Piemonte



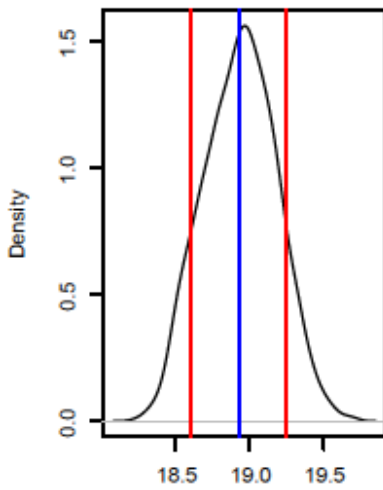
ValleDAosta



Liguria

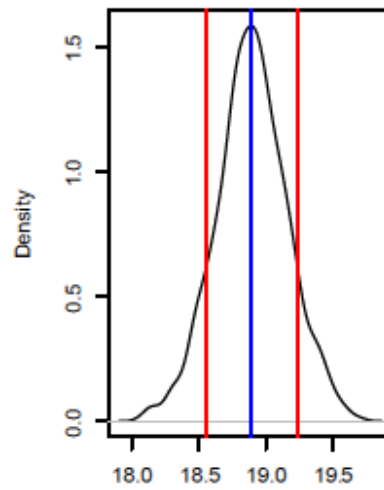


Xi Density



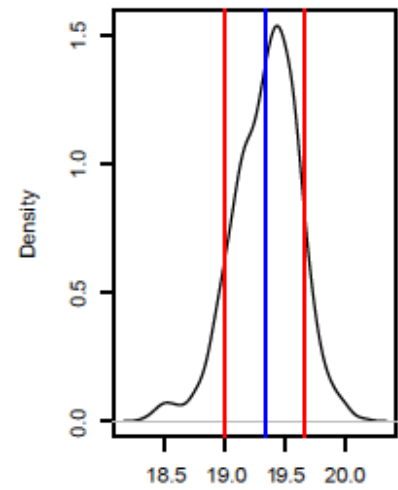
N = 1500 Bandwidth = 0.05119

Xi Density



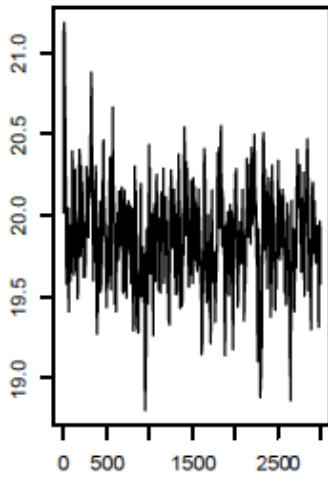
N = 1500 Bandwidth = 0.05183

Xi Density

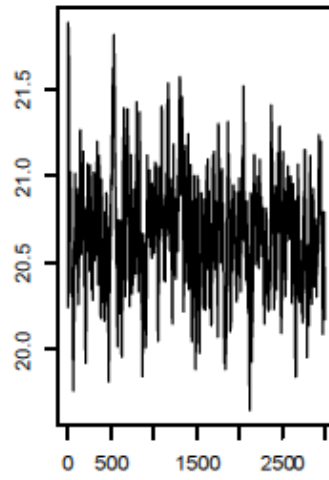


N = 1500 Bandwidth = 0.05722

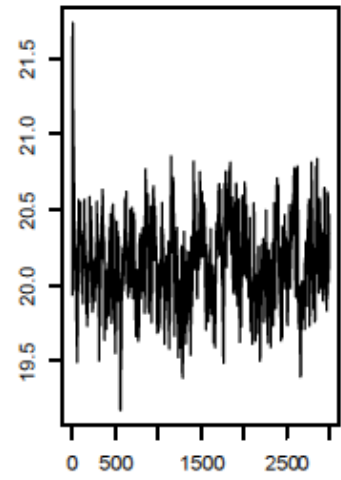
Lombardia



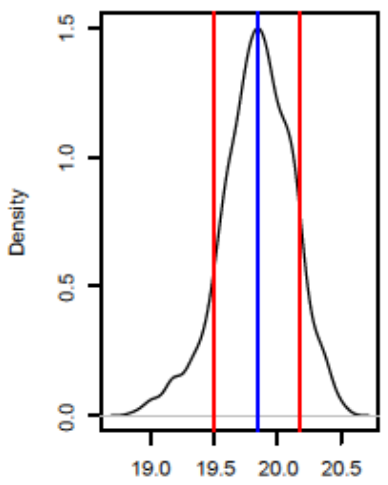
Trentino



Veneto

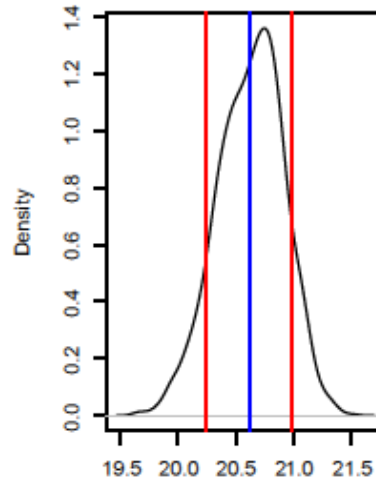


Xi Density



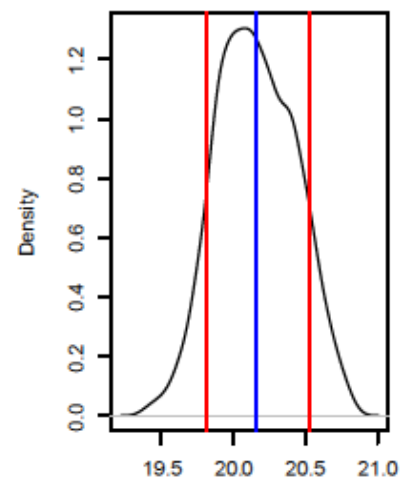
N = 1500 Bandwidth = 0.05702

Xi Density



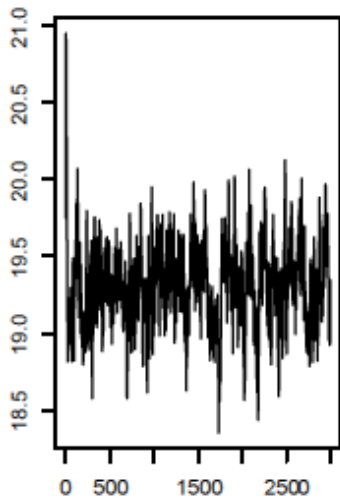
N = 1500 Bandwidth = 0.06003

Xi Density

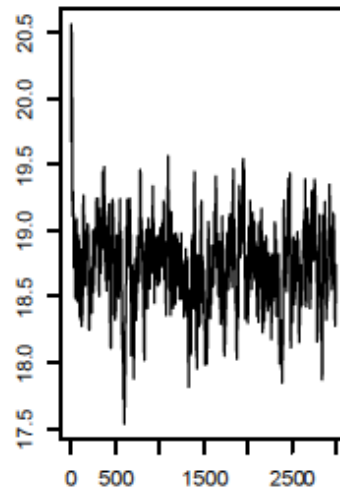


N = 1500 Bandwidth = 0.05669

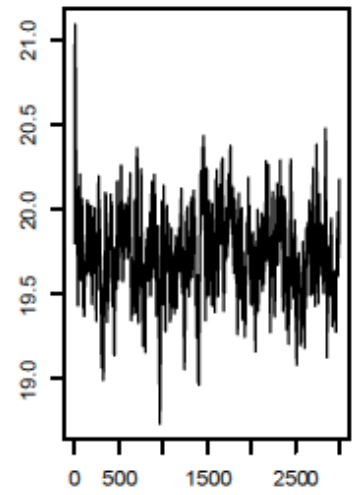
Friuli



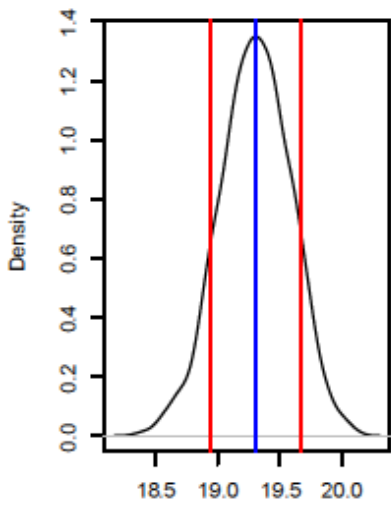
EmiliaRomagna



Marche

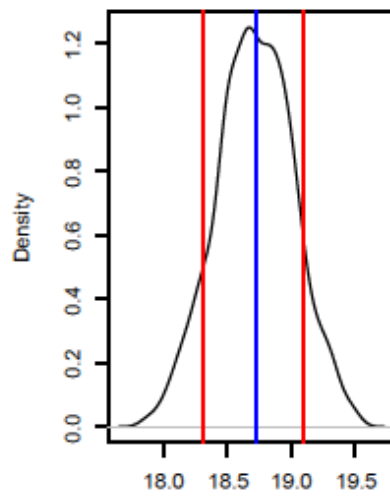


Xi Density



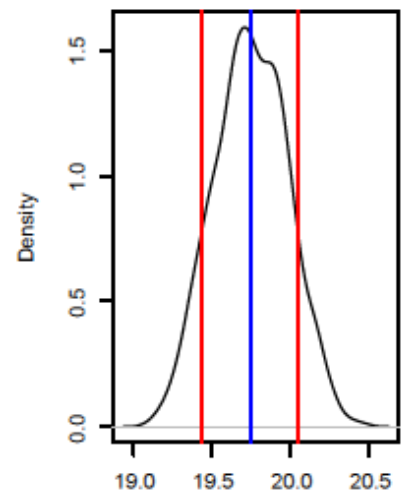
N = 1500 Bandwidth = 0.05952

Xi Density



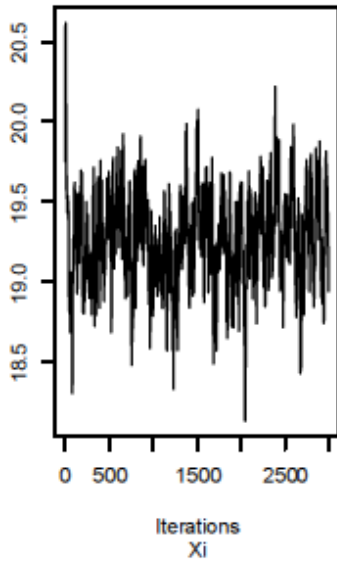
N = 1500 Bandwidth = 0.06304

Xi Density

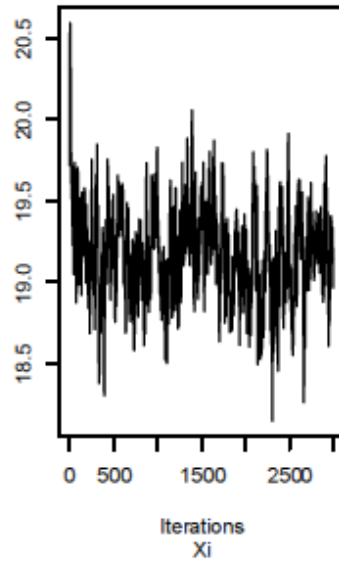


N = 1500 Bandwidth = 0.0491

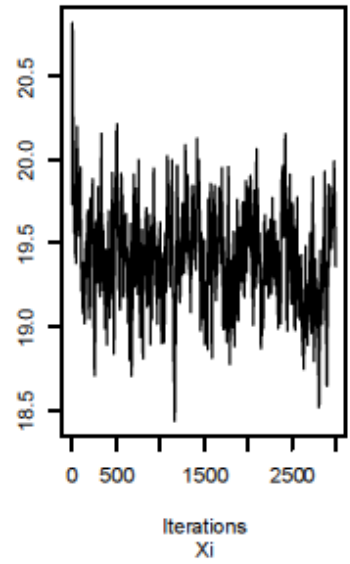
Toscana



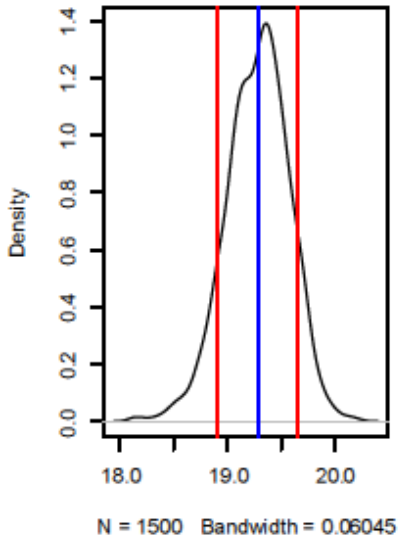
Umbria



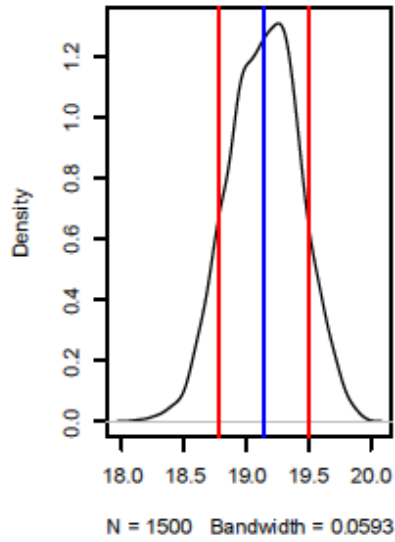
Lazio



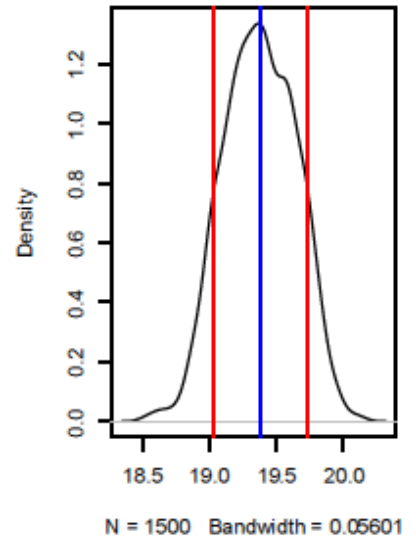
Xi Density

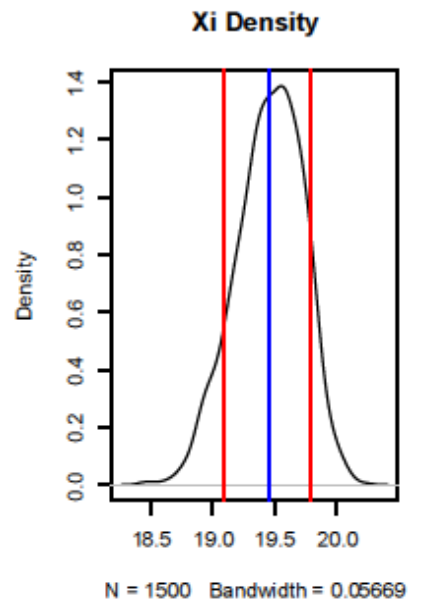
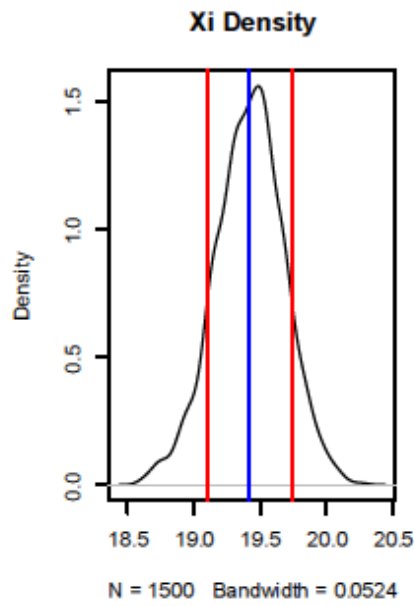
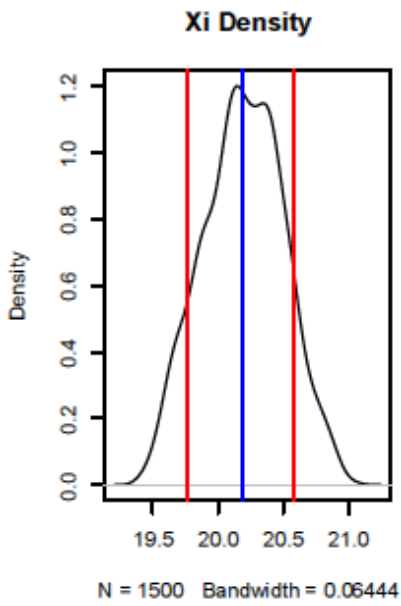
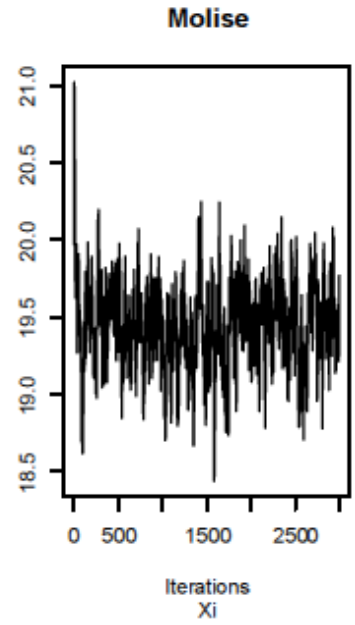
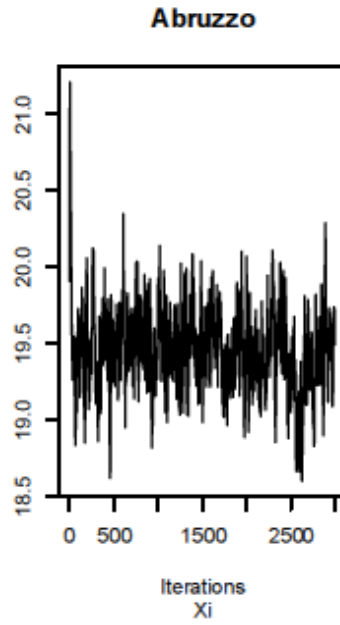
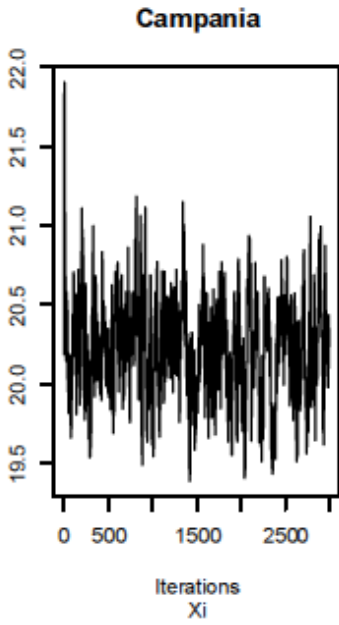


Xi Density

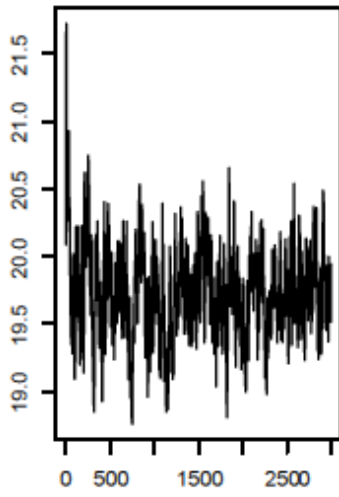


Xi Density



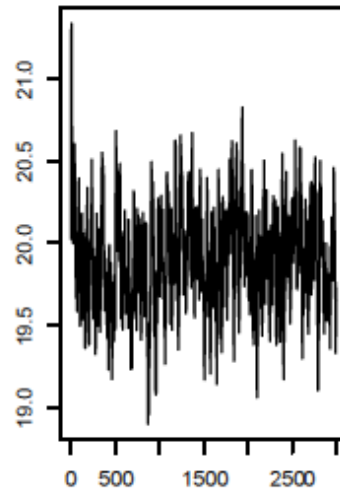


Puglia



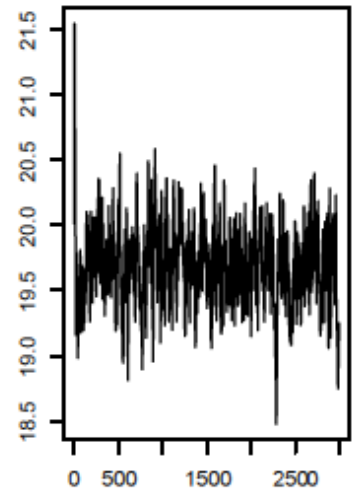
Iterations
 ξ_i

Basilicata



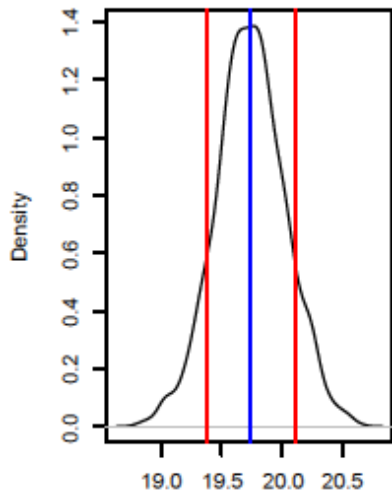
Iterations
 ξ_i

Calabria



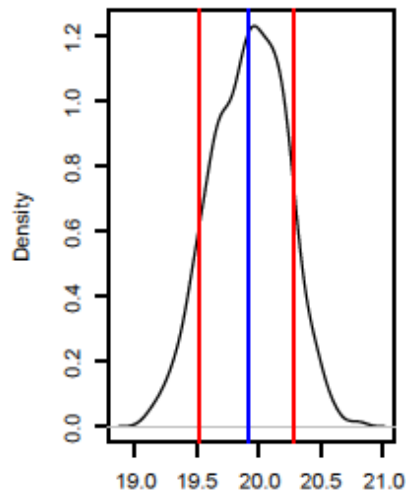
Iterations
 ξ_i

ξ_i Density



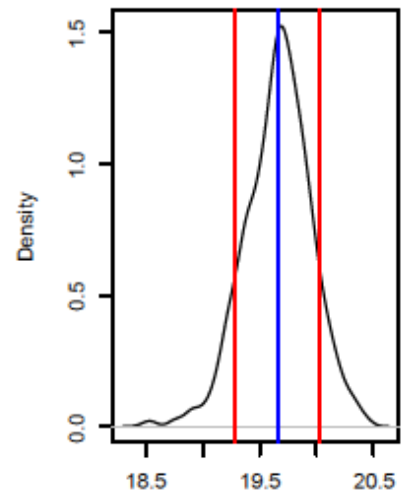
N = 1500 Bandwidth = 0.05815

ξ_i Density



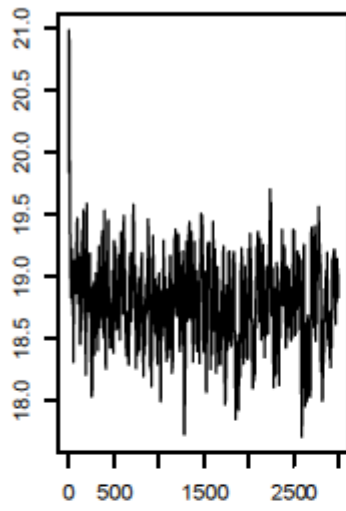
N = 1500 Bandwidth = 0.0624

ξ_i Density



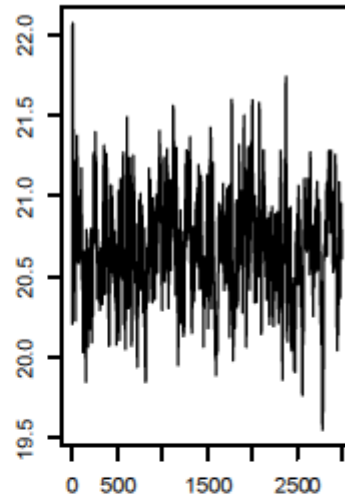
N = 1500 Bandwidth = 0.05779

Sicilia



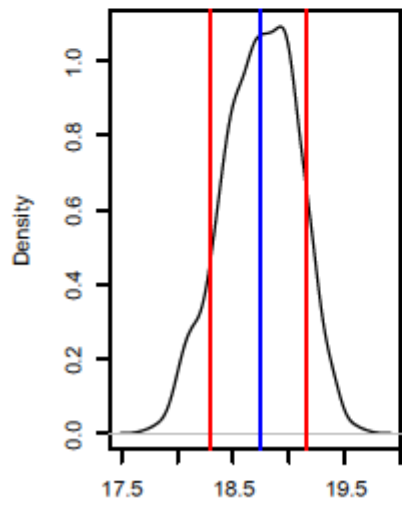
Iterations
Xi

Sardegna



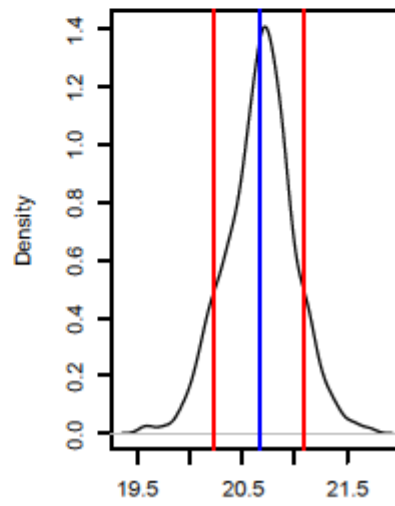
Iterations
Xi

Xi Density



N = 1500 Bandwidth = 0.06929

Xi Density



N = 1500 Bandwidth = 0.06301

Appendice B - Grafici del 1964 con Metropolis Hasting

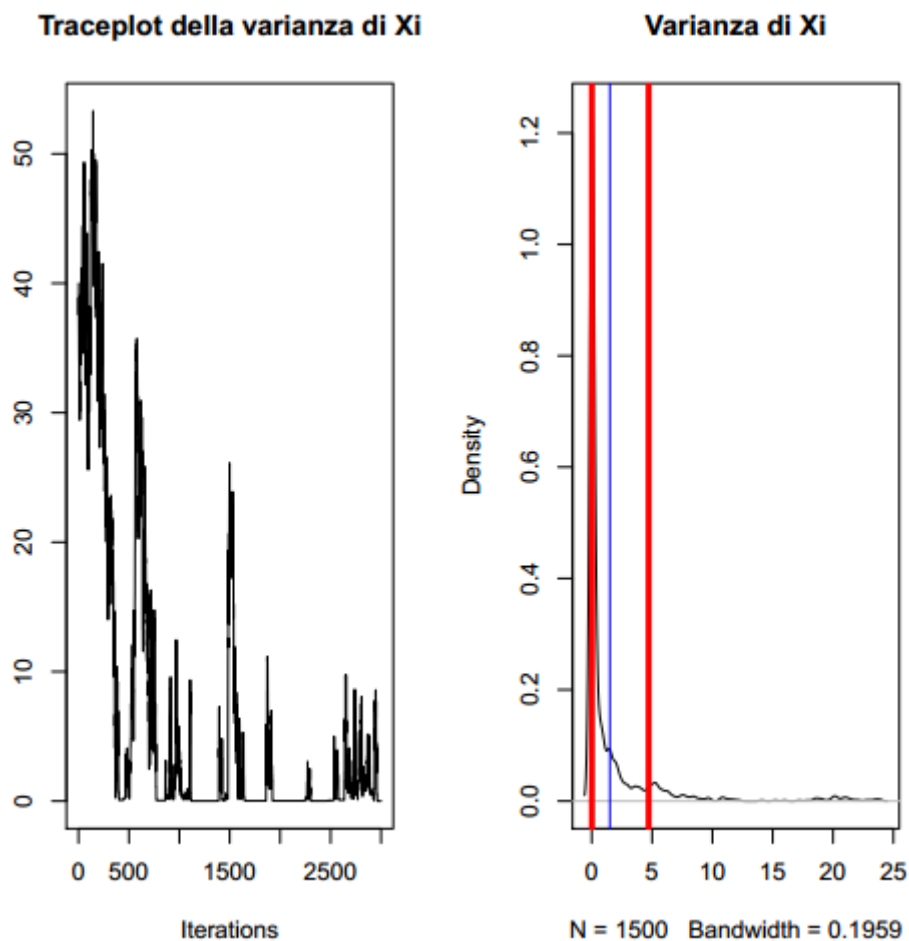
Nella tabella sottostante si possono vedere alcuni principali valori per ogni regione ottenuti dal Gibbs con 3000 iterazioni e varianza inizializzata a 40.

Valori di ξ del 1964			
Regione	Mean	IC 0.10	IC 0.90
Piemonte	18,56	18,15	19,10
Valle D'Aosta	18,56	18,18	19,03
Liguria	18,76	18,76	19,40
Lombardia	19,10	18,39	19,98
Trentino A.A.	19,50	18,48	20,80
Veneto	19,20	18,40	20,22
Friuli V.G.	18,81	18,26	19,51
Emilia Romagna	18,50	18,13	18,99
Marche	19,02	18,34	19,79
Toscana	18,77	18,28	19,43
Umbria	18,69	18,23	19,25
Lazio	18,82	18,31	19,47
Campania	19,24	18,42	20,27
Abruzzo	18,92	18,31	19,60
Molise	18,86	18,30	19,62
Puglia	19,02	18,35	19,90
Basilicata	19,11	18,37	20,04
Calabria	19,01	18,33	19,94
Sicilia	18,49	18,12	19,01
Sardegna	19,47	18,48	20,65

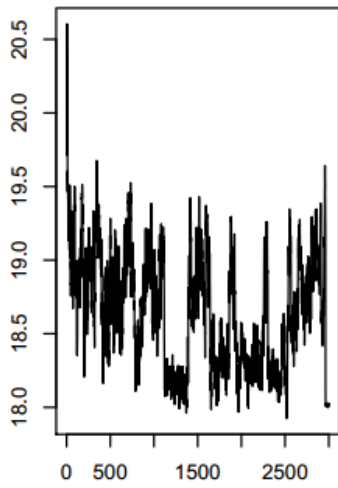
Tabella B.1 Principali valori di ξ per tutte le regioni dell'anno 1964 ottenuti dal modello bayesiano gerarchico

Mentre nell'Appendice A si vedono direttamente il traceplot e il density plot per ogni regione ottenuti dal gibbs, qui i primi due grafici rappresentano la convergenza della varianza di xi e la sua densità dopo il burn in, ottenuti dall'utilizzo del Metropolis Hasting.

- Media dopo burn in
- IC 0.10 e 0.90 dopo burn in

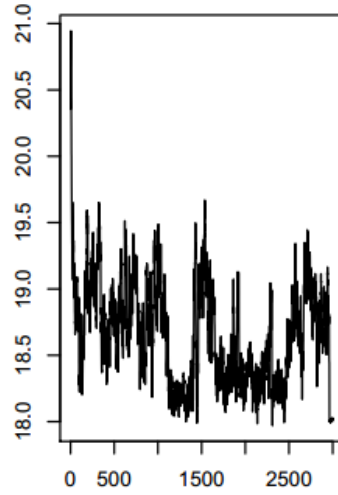


Piemonte



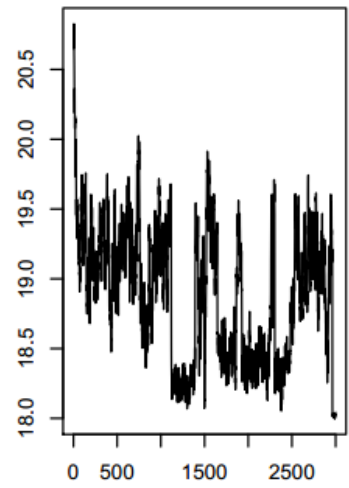
Iterations
Varianza di Xi

ValleDAosta



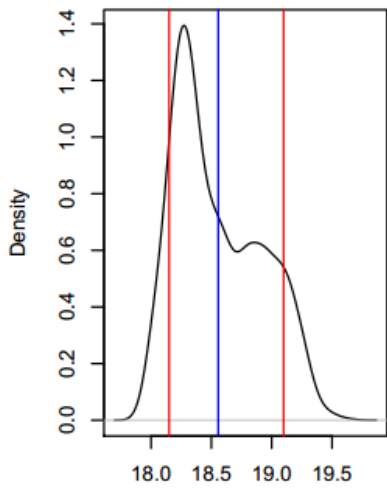
Iterations
Varianza di Xi

Liguria



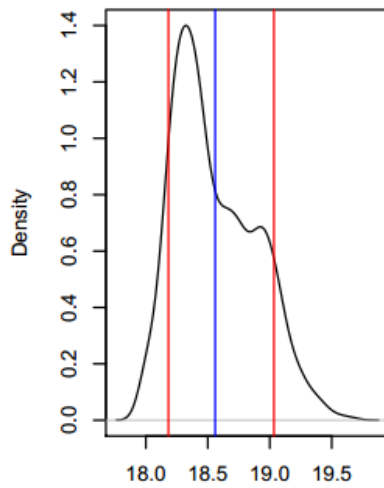
Iterations
Varianza di Xi

Densità di Xi



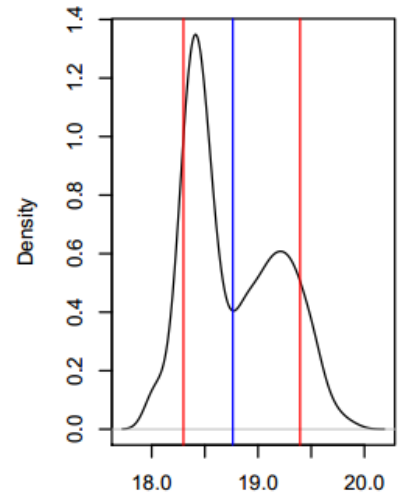
N = 1501 Bandwidth = 0.0755

Densità di Xi

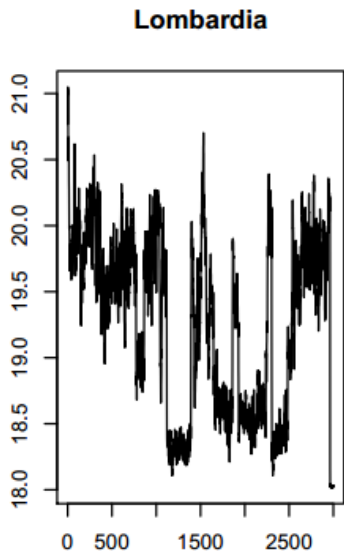


N = 1501 Bandwidth = 0.06929

Densità di Xi

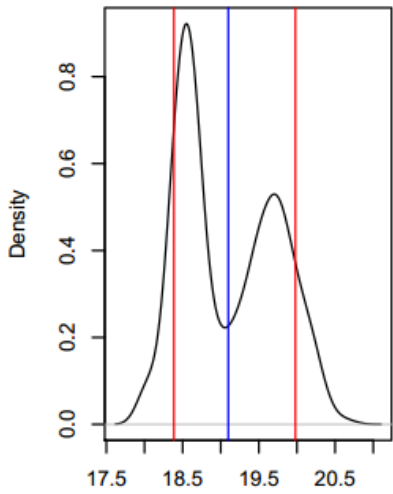


N = 1501 Bandwidth = 0.09139

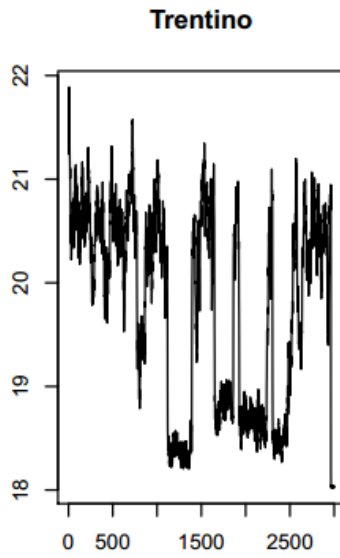


Iterations
Varianza di Xi

Densità di Xi

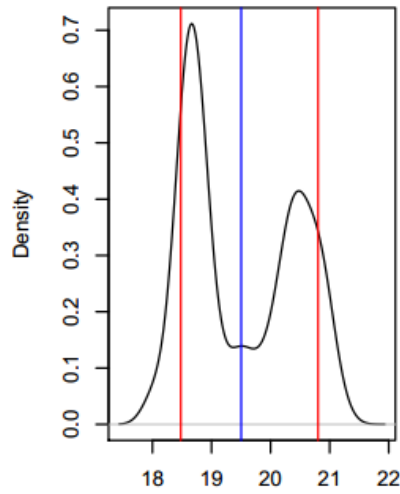


N = 1501 Bandwidth = 0.1329

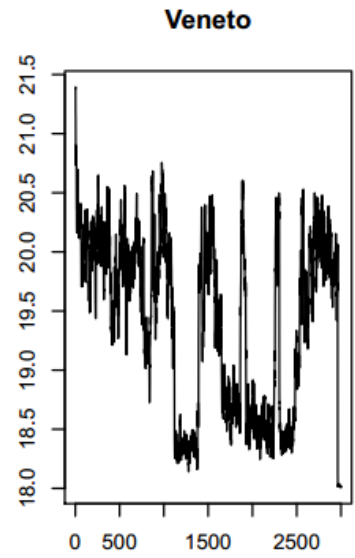


Iterations
Varianza di Xi

Densità di Xi

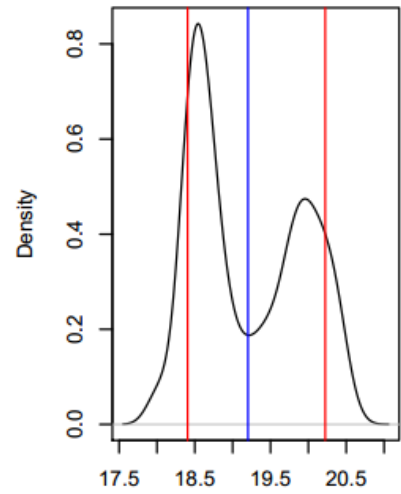


N = 1501 Bandwidth = 0.1955



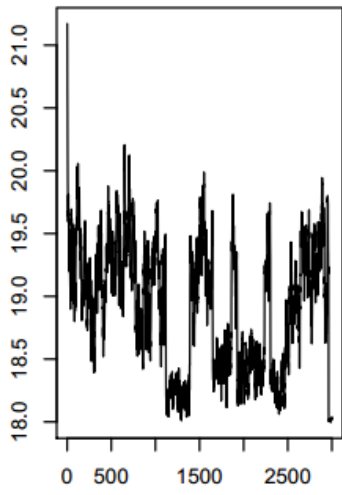
Iterations
Varianza di Xi

Densità di Xi



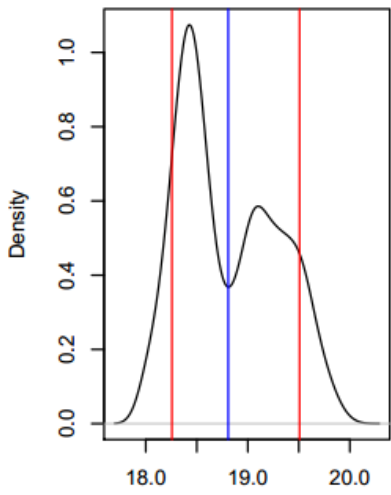
N = 1501 Bandwidth = 0.1515

Friuli



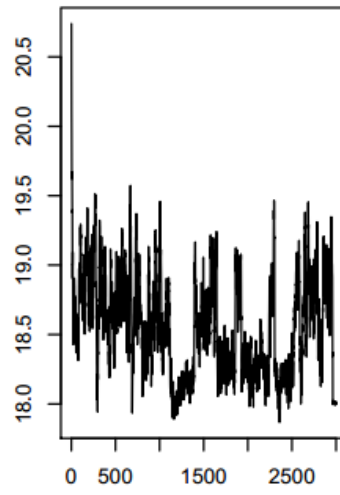
Iterations
Varianza di Xi

Densità di Xi



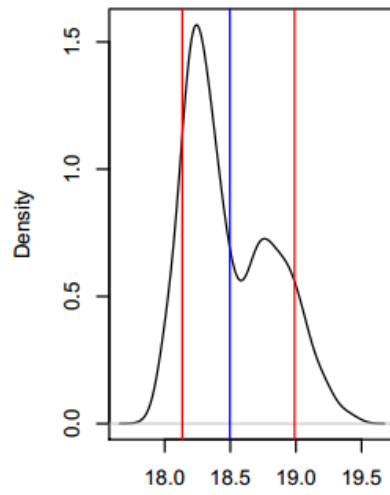
N = 1501 Bandwidth = 0.1004

EmiliaRomagna



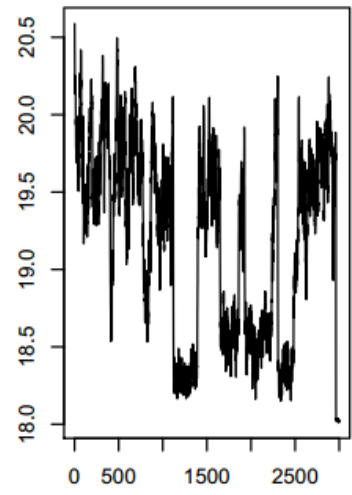
Iterations
Varianza di Xi

Densità di Xi



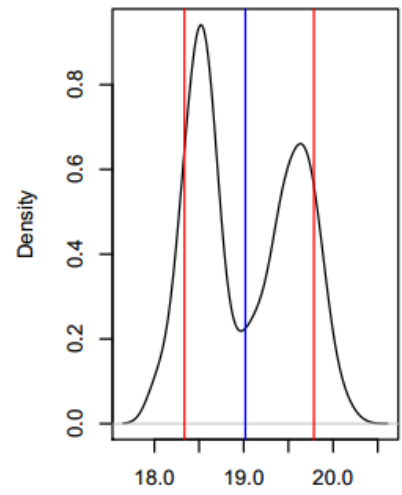
N = 1501 Bandwidth = 0.06993

Marche



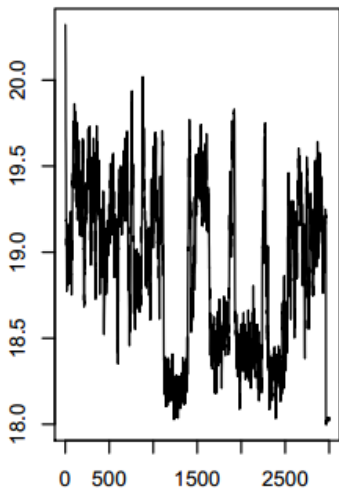
Iterations
Varianza di Xi

Densità di Xi

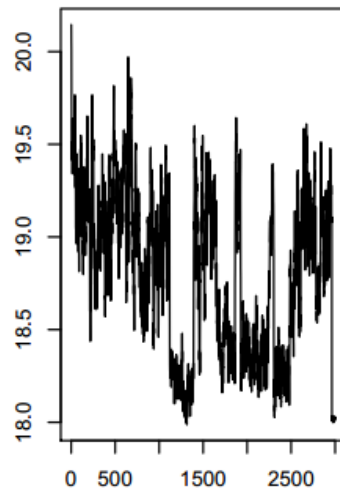


N = 1501 Bandwidth = 0.121

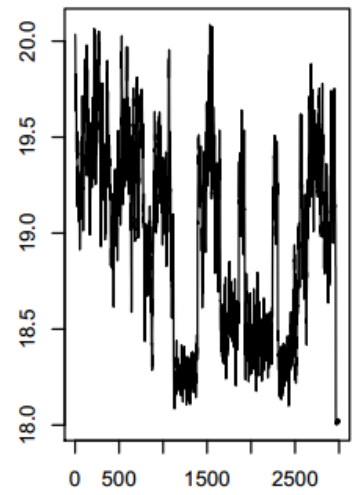
Toscana



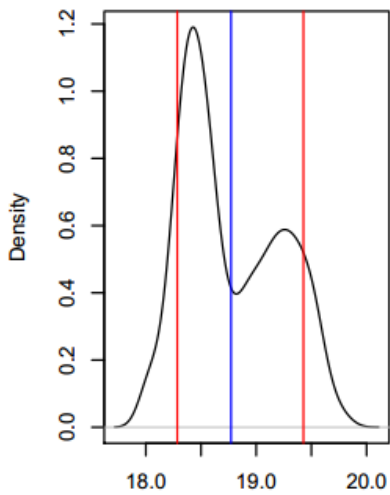
Umbria



Lazio

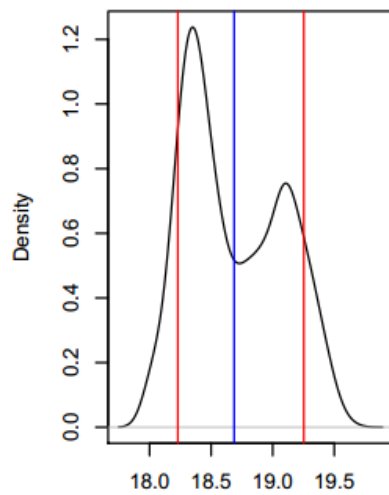


Iterations
Varianza di Xi
Densità di Xi



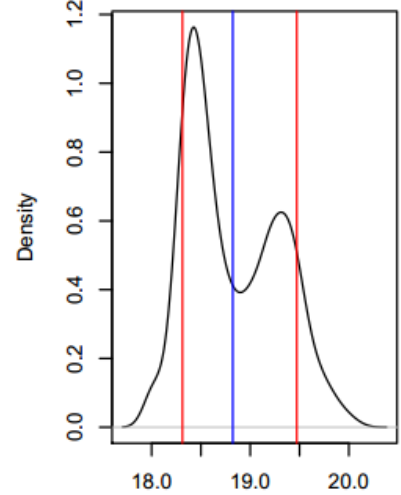
N = 1501 Bandwidth = 0.0931

Iterations
Varianza di Xi
Densità di Xi



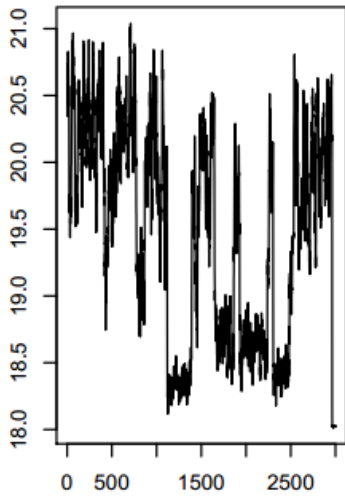
N = 1501 Bandwidth = 0.08313

Iterations
Varianza di Xi
Densità di Xi



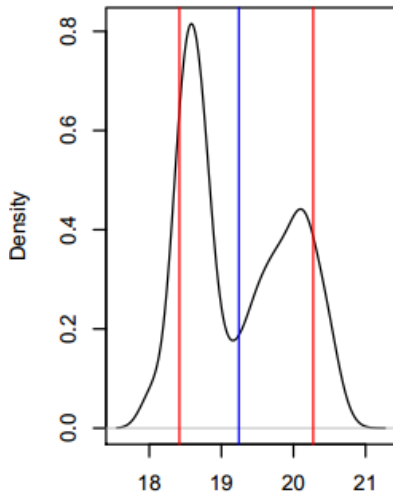
N = 1501 Bandwidth = 0.09871

Campania



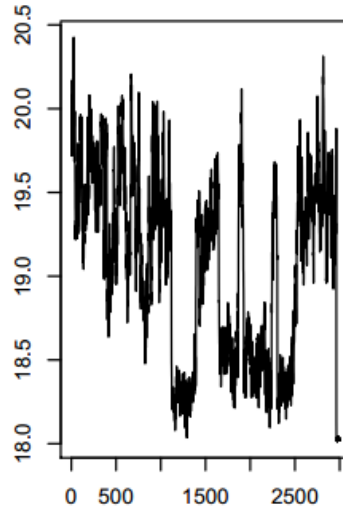
Iterations
Varianza di Xi

Densità di Xi



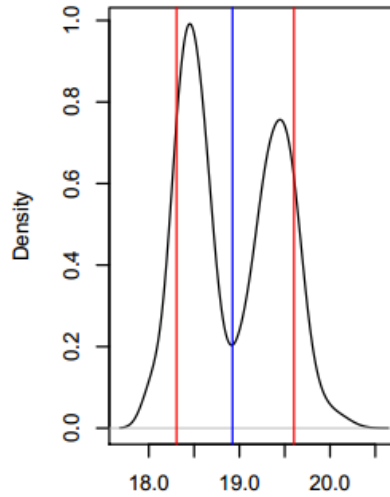
N = 1501 Bandwidth = 0.1548

Abruzzo



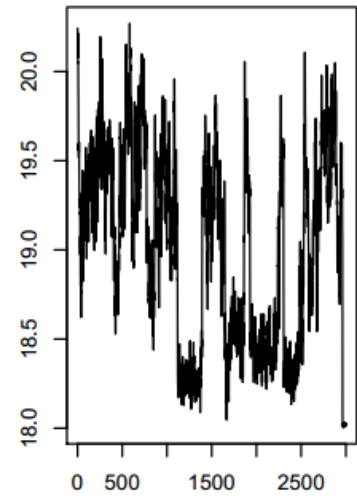
Iterations
Varianza di Xi

Densità di Xi



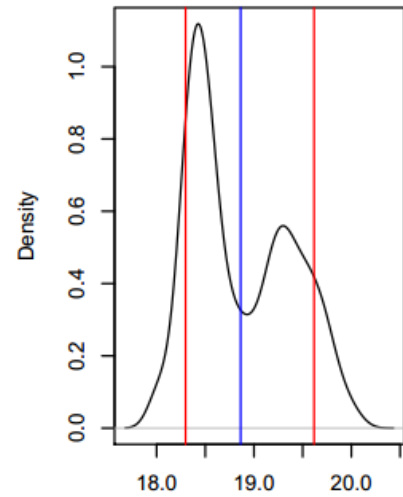
N = 1501 Bandwidth = 0.1099

Molise



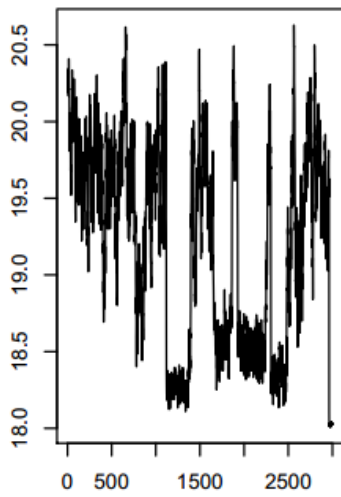
Iterations
Varianza di Xi

Densità di Xi

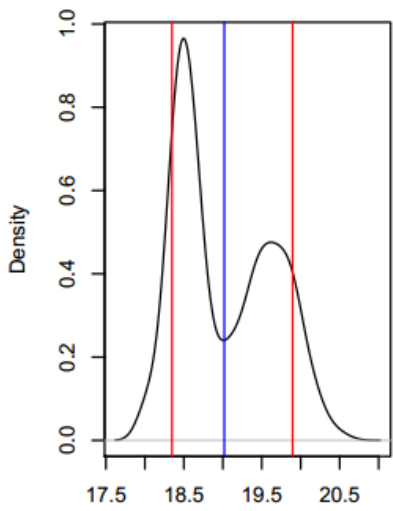


N = 1501 Bandwidth = 0.1084

Puglia

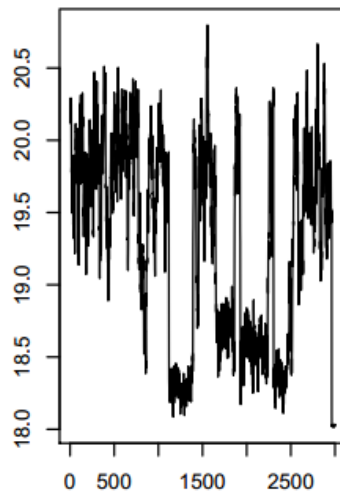


Iterations
Varianza di Xi
Densità di Xi

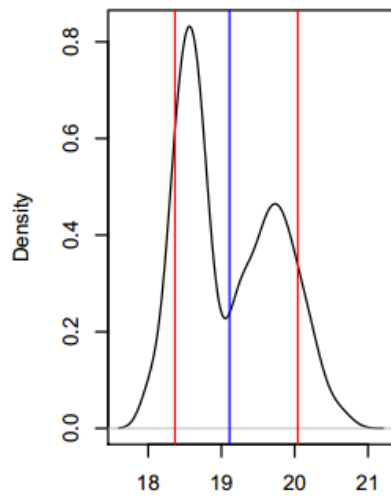


N = 1501 Bandwidth = 0.1297

Basilicata

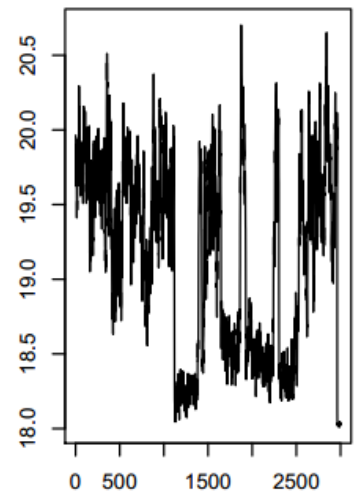


Iterations
Varianza di Xi
Densità di Xi

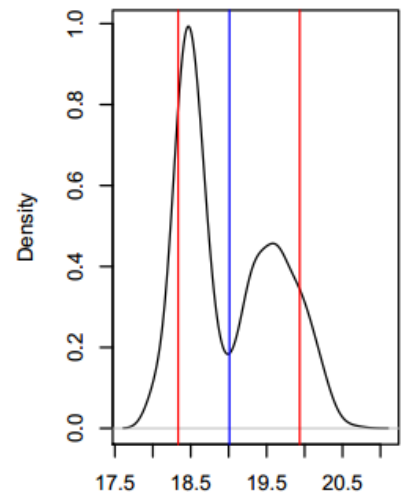


N = 1501 Bandwidth = 0.1388

Calabria

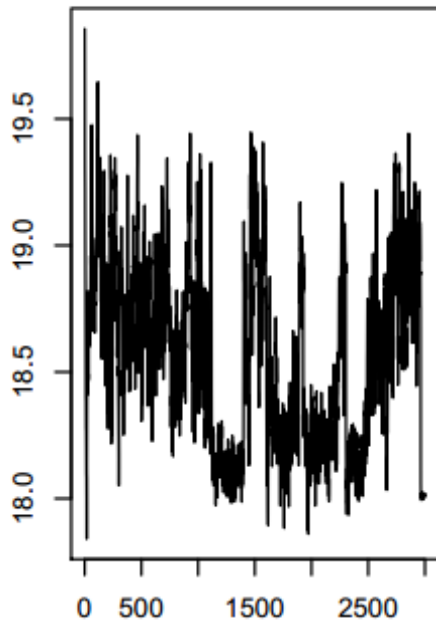


Iterations
Varianza di Xi
Densità di Xi



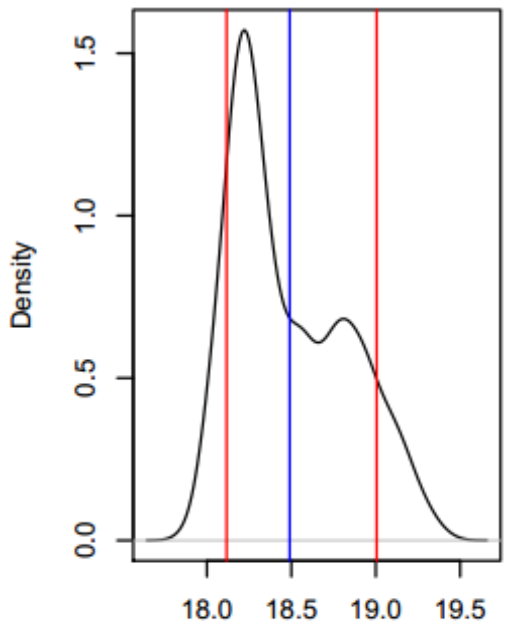
N = 1501 Bandwidth = 0.133

Sicilia



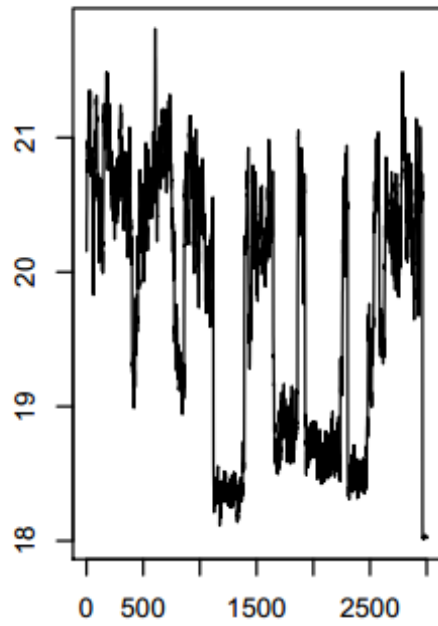
Iterations
Varianza di Xi

Densità di Xi



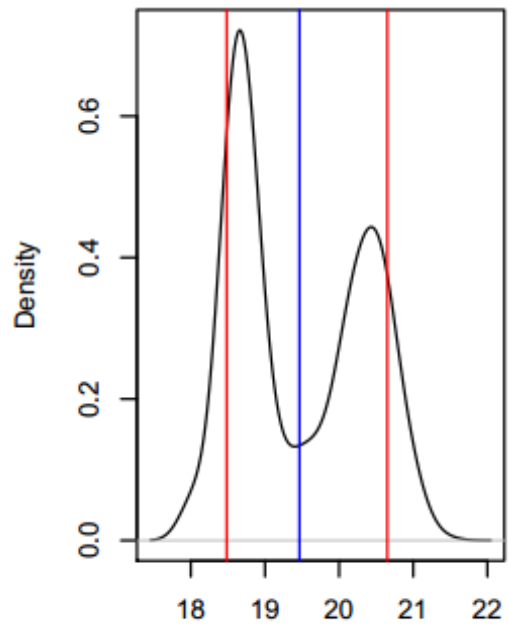
N = 1501 Bandwidth = 0.07212

Sardegna



Iterations
Varianza di Xi

Densità di Xi

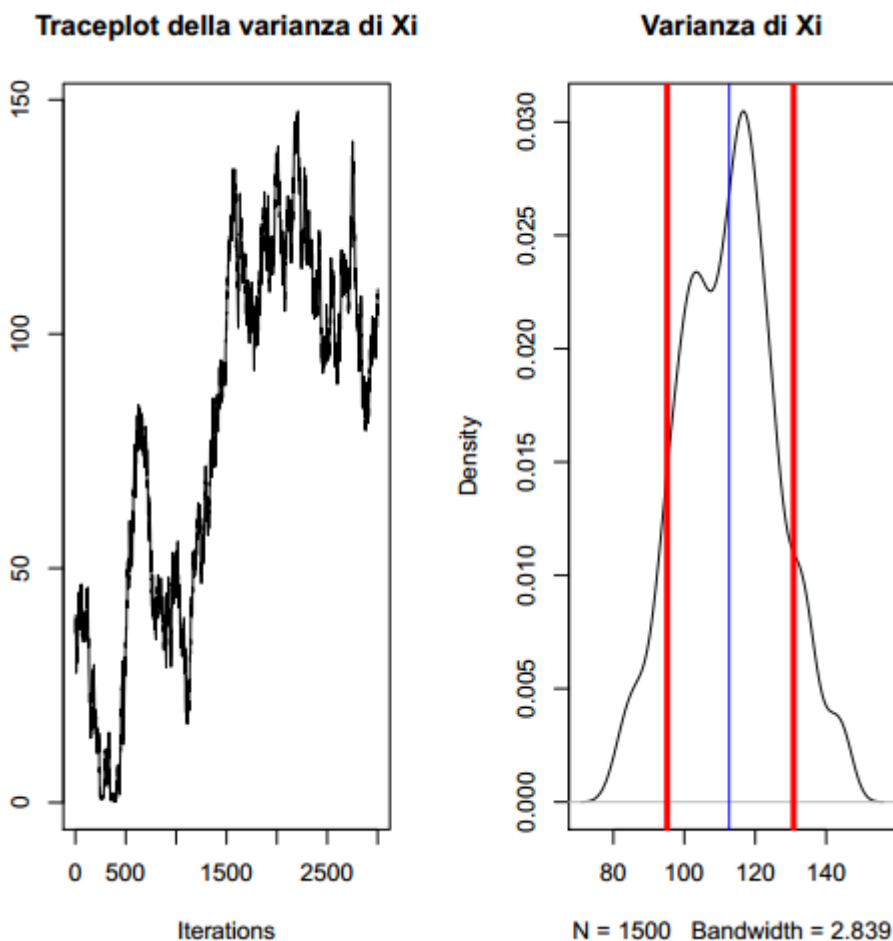


N = 1501 Bandwidth = 0.1859

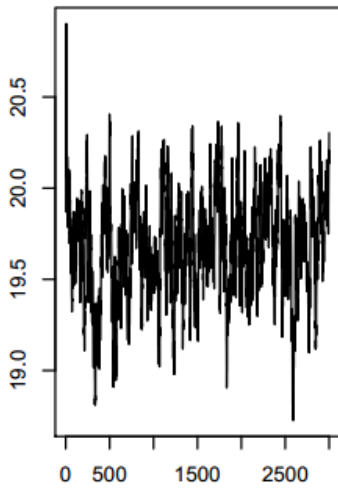
Appendice C - Grafici del 1985 con Metropolis Hasting

I grafici seguenti sono dati dal Gibbs e dal Metropolis trovati con 3000 iterazioni e varianza di ξ iniziale posta uguale a 40. Come in Appendice B i primi due grafici mostrano l'aggiornamento della varianza, mentre gli altri sono il traceplot e il density plot delle regioni del 1985.

- Media dopo burn in
- IC 0.10 e 0.90 dopo burn in

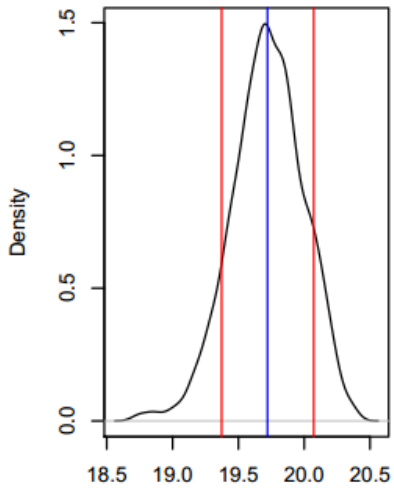


Piemonte



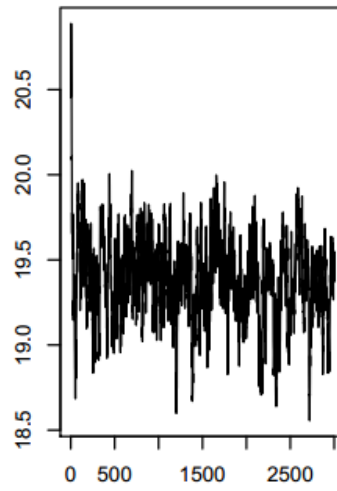
Iterations
Varianza di Xi

Densità di Xi



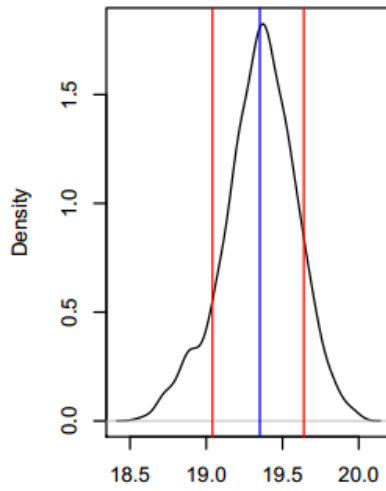
N = 1501 Bandwidth = 0.05482

ValleDAosta



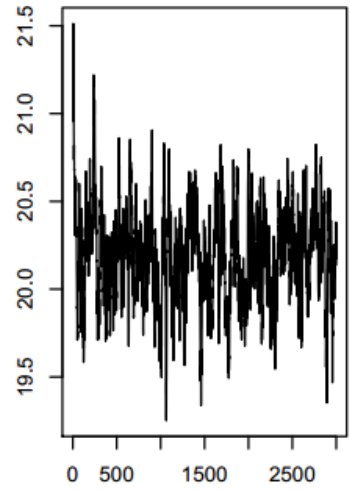
Iterations
Varianza di Xi

Densità di Xi



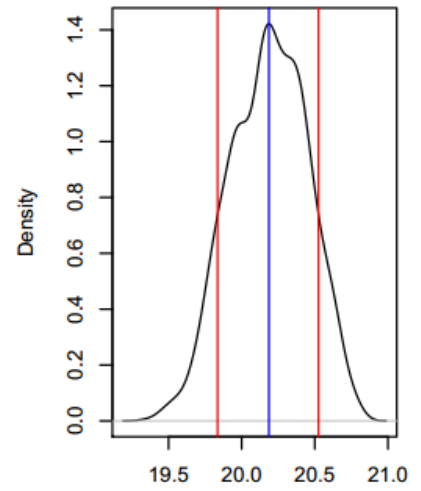
N = 1501 Bandwidth = 0.04744

Liguria

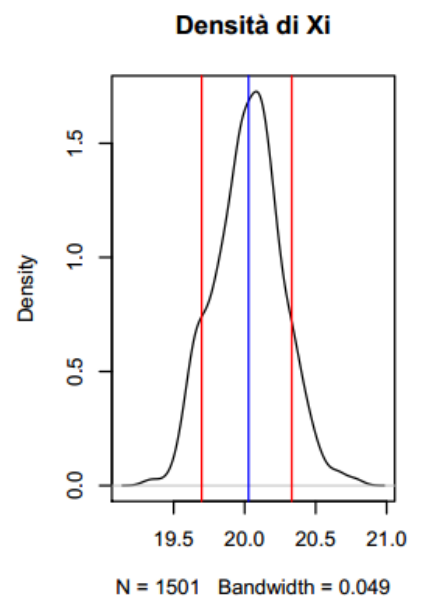
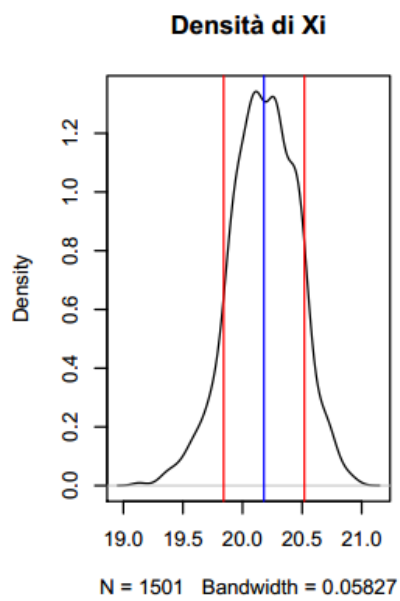
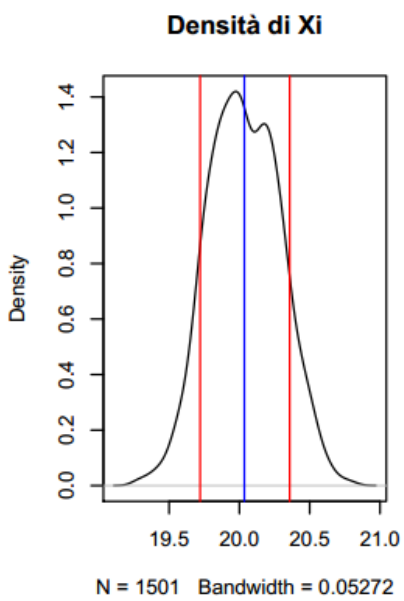
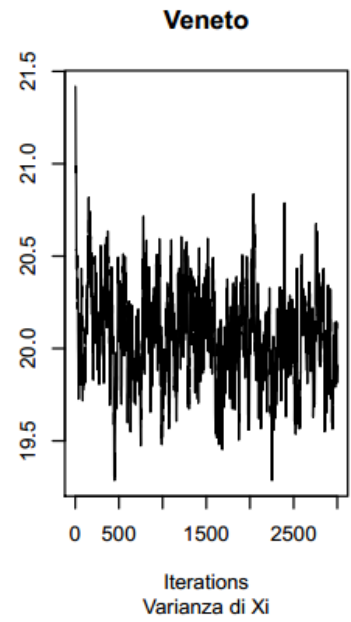
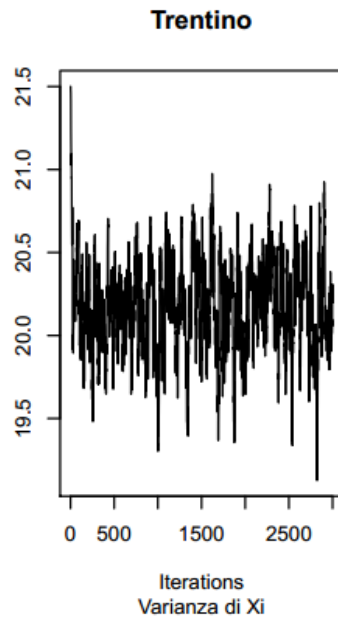
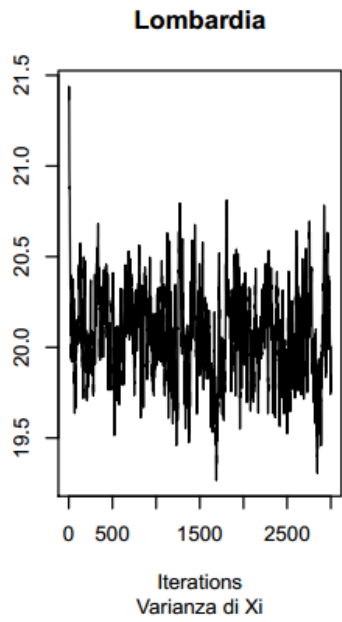


Iterations
Varianza di Xi

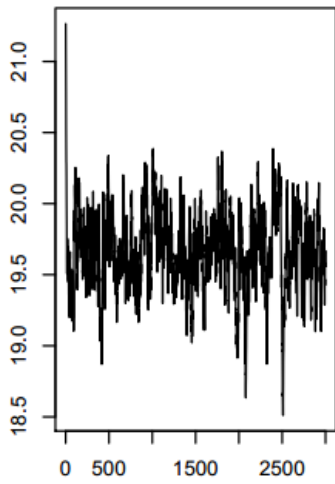
Densità di Xi



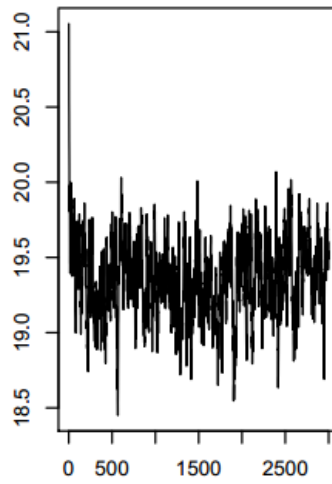
N = 1501 Bandwidth = 0.05492



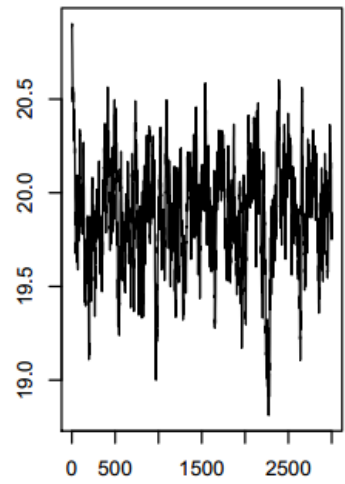
Friuli



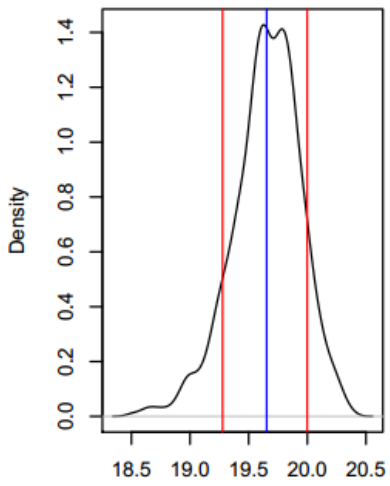
EmiliaRomagna



Marche

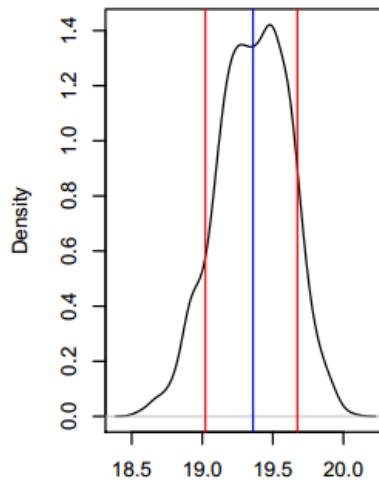


Densità di Xi



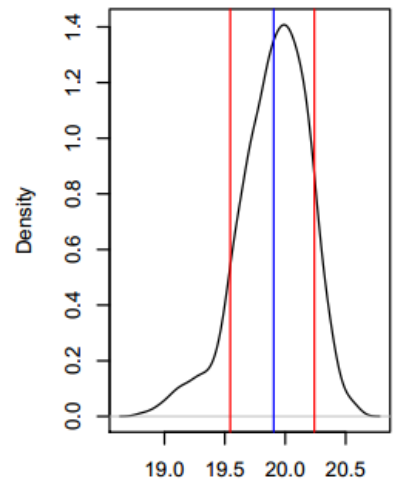
N = 1501 Bandwidth = 0.05642

Densità di Xi



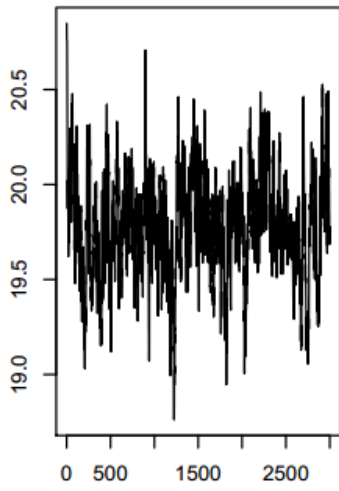
N = 1501 Bandwidth = 0.05303

Densità di Xi



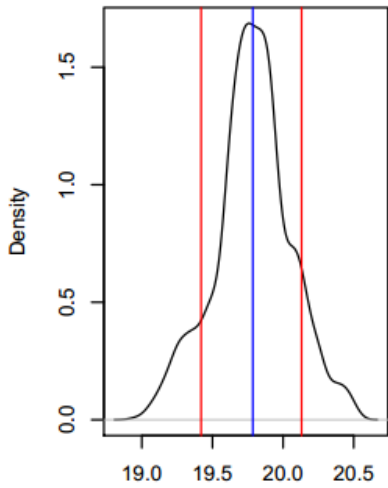
N = 1501 Bandwidth = 0.06001

Toscana



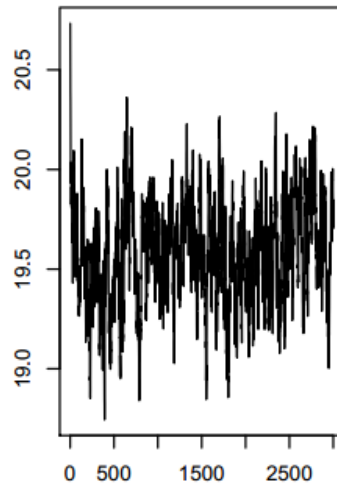
Iterations
Varianza di Xi

Densità di Xi



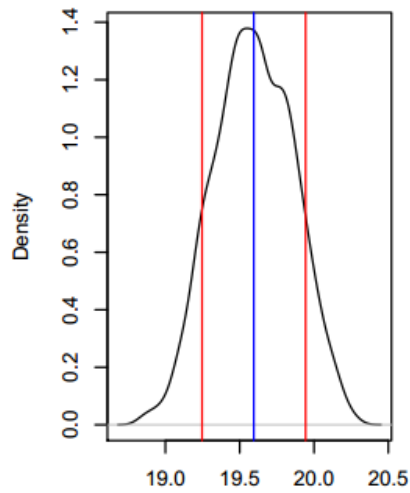
N = 1501 Bandwidth = 0.04687

Umbria



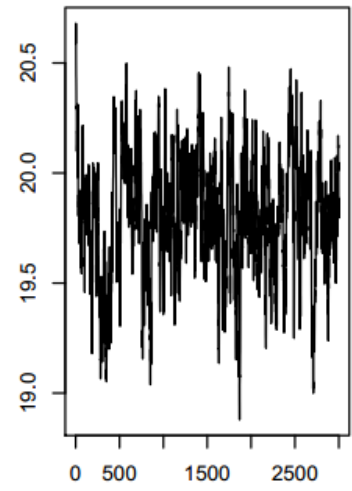
Iterations
Varianza di Xi

Densità di Xi



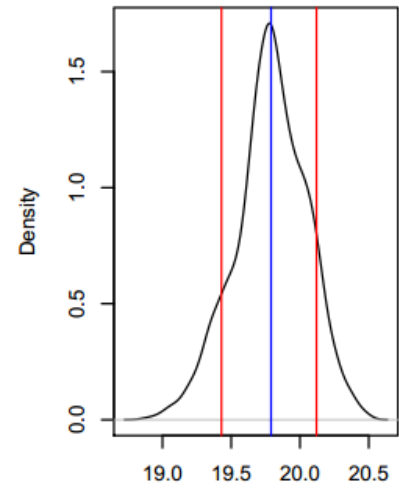
N = 1501 Bandwidth = 0.05516

Lazio



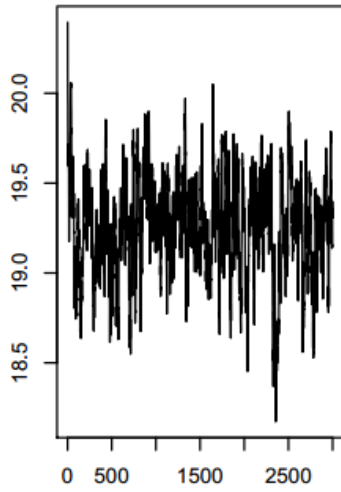
Iterations
Varianza di Xi

Densità di Xi



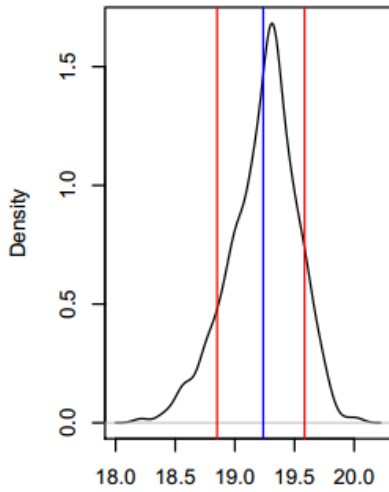
N = 1501 Bandwidth = 0.05194

Campania



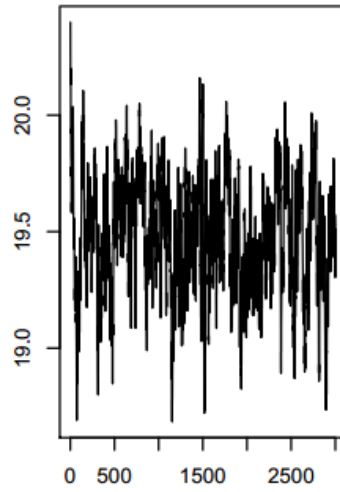
Iterations
Varianza di Xi

Densità di Xi



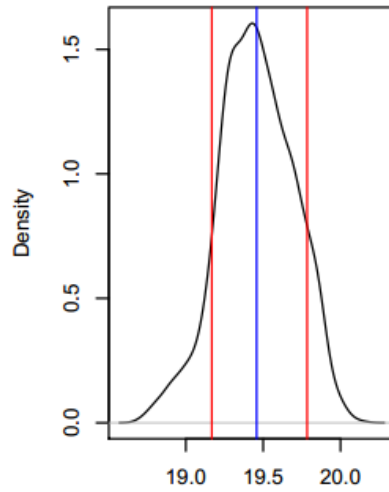
N = 1501 Bandwidth = 0.05729

Abruzzo



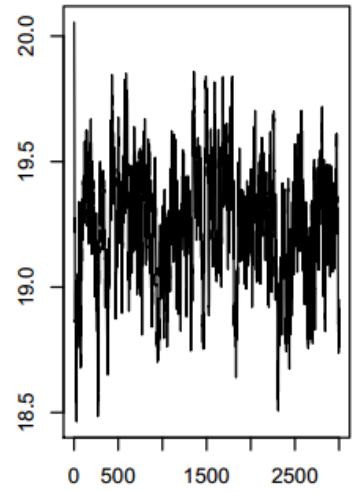
Iterations
Varianza di Xi

Densità di Xi



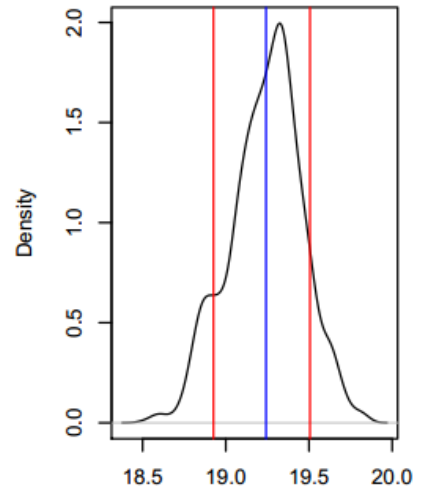
N = 1501 Bandwidth = 0.05008

Molise



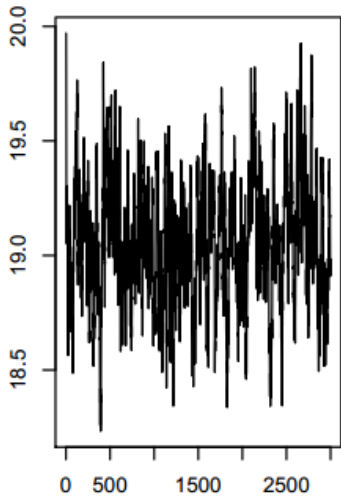
Iterations
Varianza di Xi

Densità di Xi



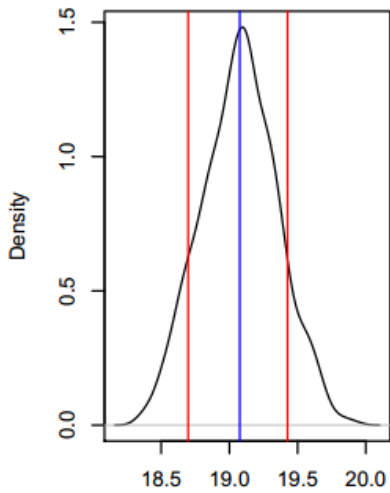
N = 1501 Bandwidth = 0.04359

Puglia



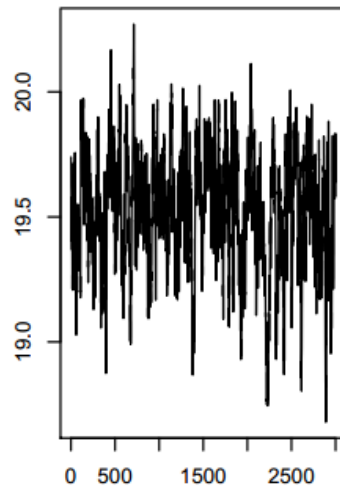
Iterations
Varianza di Xi

Densità di Xi



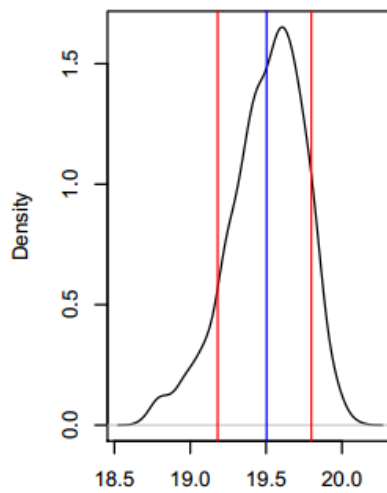
N = 1501 Bandwidth = 0.05818

Basilicata



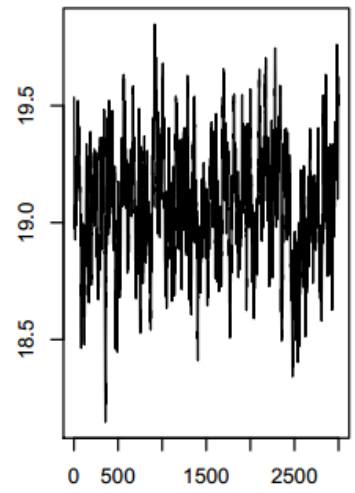
Iterations
Varianza di Xi

Densità di Xi



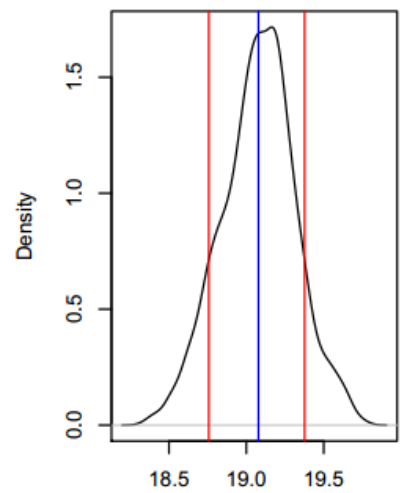
N = 1501 Bandwidth = 0.05168

Calabria



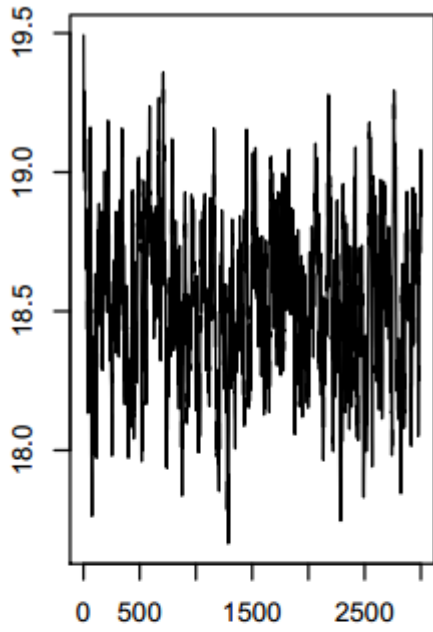
Iterations
Varianza di Xi

Densità di Xi



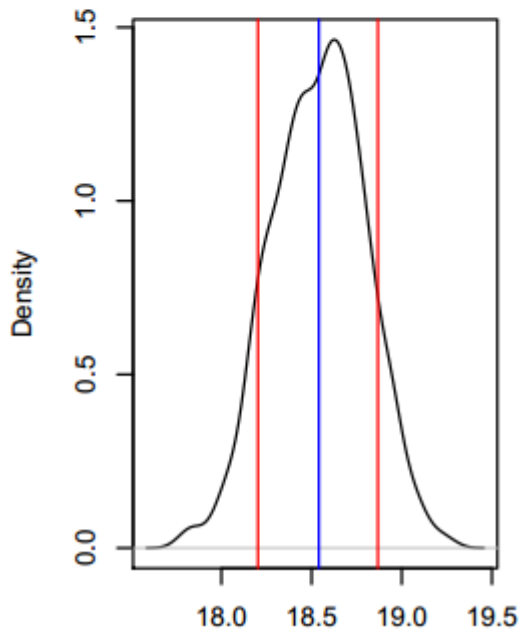
N = 1501 Bandwidth = 0.04827

Sicilia



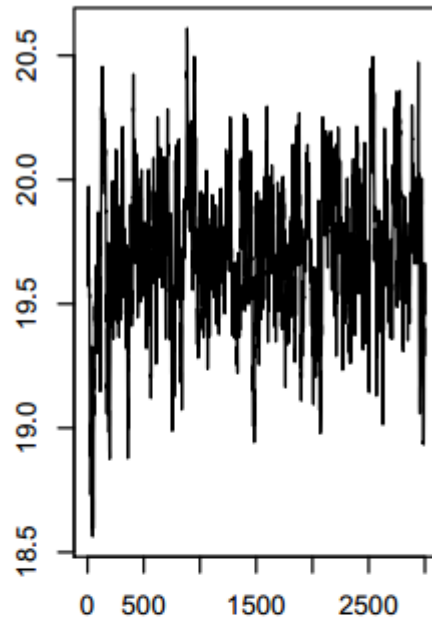
Iterations
Varianza di X_i

Densità di X_i



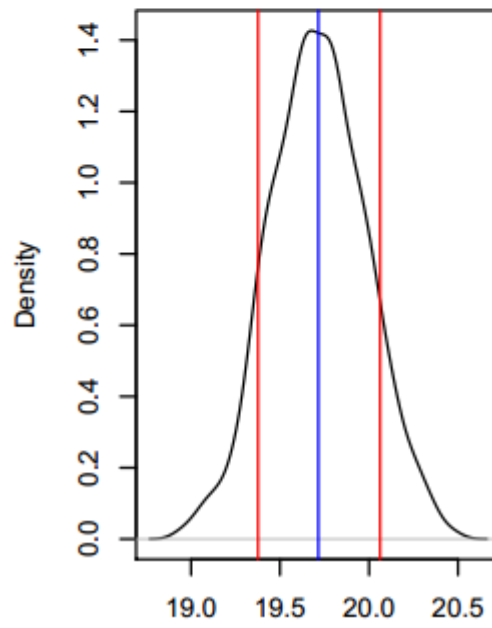
N = 1501 Bandwidth = 0.05403

Sardegna



Iterations
Varianza di X_i

Densità di X_i



N = 1501 Bandwidth = 0.05602

Appendice D – Metropolis Hasting, il codice

```
Metropolis_Hastings <- function(Outlist,sigma,it){  
  
  # Estraggo sigma dai par  
  
  sigma0.mh<-sigma  
  
  
  # Trovo il valore proposto di sigma da una distr. normale random  
  
  prop<- proposalfunction(sigma0.mh)  
  
  
  # Accettazione o Rifiuto dell'aggiornamento di sigma  
  
  probab <- exp(log.post(Outlist,prop,it) - log.post(Outlist,sigma,it))  
  
  RUnif<-runif(1)  
  
  
  if (RUnif < probab){  
  
    sigma.mh <- prop  
  
  }else{  
  
    sigma.mh <- sigma0.mh  
  
  }  
  
  
  return(sigma.mh)  
  
}
```