

Università degli Studi di Padova

Corso di Laurea in

SCIENZE STATISTICHE



METODI MCMC PER L'ANALISI DI DATI FINANZIARI

MCMC METHODS FOR THE ANALYSIS OF FINANCIAL TIME SERIES

Relatore: Prof. Marco Ferrante

Dipartimento di Matematica

Laureando: Carlo Pianezzola, 623120-SST

ANNO ACCADEMICO 2012-2013

Indice

| | | |
|----------|----------------------------------------------------------------------------------------|-----------|
| 1 | Metodi MCMC | 5 |
| 1.1 | Introduzione | 5 |
| 1.2 | Campionamento di Gibbs | 8 |
| 1.3 | Catene di Markov e introduzione all’algoritmo di Metropolis-Hastings (MH) | 10 |
| 1.4 | L’algoritmo di Metropolis-Hastings | 13 |
| 1.4.1 | Random walk Metropolis | 15 |
| 1.4.2 | Independence sampler | 15 |
| 1.5 | La metodologia Griddy Gibbs | 16 |
| 2 | Applicazioni Teoriche | 19 |
| 2.1 | Regressione lineare con errori serialmente correlati | 19 |
| 2.2 | Valori mancanti e/o valori estremi | 23 |
| 2.2.1 | Valori Mancanti | 24 |
| 2.2.2 | Determinare i possibili Outlier | 27 |
| 2.3 | Applicazione ai modelli SV, Stochastic Volatility | 30 |
| 2.3.1 | Il modello SV | 30 |
| 2.3.2 | L’approccio tramite MCMC | 31 |
| 2.4 | Applicazione ai modelli ARMA-GARCH | 33 |
| 2.4.1 | Il modello ARMA-GARCH | 34 |
| 2.4.2 | L’approccio tramite MCMC | 34 |

| | | |
|----------|---------------------------------------------------------------|-----------|
| 3 | Applicazioni Pratiche | 37 |
| 3.1 | Introduzione | 37 |
| 3.2 | Software WinBugs | 38 |
| 3.3 | Stochastic Volatility Model | 39 |
| 3.3.1 | Implementazione in WinBugs | 41 |
| 3.3.2 | Risultati | 42 |
| 3.4 | (G)ARCH model | 45 |
| 3.4.1 | Implementazione in WinBugs | 45 |
| 3.4.2 | Risultati | 46 |
| 3.5 | Conclusioni | 49 |
| A | Prerequisiti teorici | 51 |
| A.1 | Processi stocastici e Catene di Markov | 51 |
| A.1.1 | Processi stocastici discreti | 51 |
| A.1.2 | Distribuzioni invarianti | 54 |
| A.1.3 | Time Reversal e Bilancio Dettagliato | 57 |
| A.2 | Generazione casuale di variabili aleatorie | 59 |
| A.2.1 | Metodi Monte Carlo | 59 |
| A.2.2 | Generare pseudo casualmente | 60 |
| A.2.3 | Combinazione di Generatori | 61 |
| A.3 | Statistica bayesiana | 63 |
| A.3.1 | Il teorema di Bayes | 63 |
| A.3.2 | Distribuzione a priori e distribuzione a posteriori | 64 |
| A.3.3 | Scelta della distribuzione iniziale | 65 |
| B | Complementi | 69 |
| B.1 | Catene di Markov, Teoremi e Formule | 69 |
| B.2 | Distribuzioni Utili | 70 |
| B.3 | Codice R, per l'analisi dei dati campionati | 75 |
| | Bibliografia | 79 |

INDICE

V

Ringraziamenti

81

Introduzione

Nel trattare dati appartenenti al mondo delle serie storiche e finanziarie, il primo problema da affrontare è la scelta di un approccio corretto, sia esso appartenente alla sfera della statistica inferenziale frequentista, Bayesiana o non parametrica (si veda ad esempio, Fan J., Yao Q. (2003) o Hardle W., Muller M., Sperlich S., Werwatz A. (2004)). Parallelamente a questa scelta iniziale si incorre spesso nel problema del campionamento, che a volte è ridotto e altre incompleto.

Se il dubbio che i dati non rappresentino in modo efficace la realtà è fondato, i metodi bayesiani possono venire in aiuto. Essi giungono a dei risultati attraverso la combinazione di conoscenze personali e informazioni presenti nei dati; il grosso vantaggio di poter inserire nell'analisi delle conoscenze pregresse del fenomeno si paga però con una crescente difficoltà computazionale. Nonostante la potenza sempre maggiore dei computer, i problemi, soprattutto nei casi multivariati, restavano in passato risolvibili solo in linea teorica e si scontravano nella pratica con difficoltà di calcolo analitico o numerico di integrali a più dimensioni, con i metodi MCMC è possibile superare questo scoglio.

Il lavoro di queste pagine verterà a dare una visione basilare dei metodi MCMC utili per risolvere i problemi di campionamento ridotto o incompleto utilizzando l'approccio bayesiano, in modo da ottenere stimatori più precisi e completi dal punto di vista statistico. La teoria MCMC (Markov Chain Monte Carlo) si basa sulla simulazione da distribuzioni note. Le simulazioni sono degli esperimenti che coinvolgono

l'utilizzo di numeri casuali¹, dove un insieme di numeri casuali è un vettore contenuto nell'intervallo $[0, 1)$; l'utilizzo di questi numeri si è rivelato utile in molte applicazioni che spaziano dalla medicina alla fisica, dalla matematica alla finanza.

Generalmente la simulazione si rileva molto utile quando ci si trova di fronte a situazioni che difficilmente si potrebbero risolvere analiticamente o numericamente; le statistiche derivanti dai campionamenti casuali (esempi sono media del campione, proporzione del campione ecc.) sono utilizzati per stimare alcuni parametri di interesse del problema originale, sistema o popolazione. D'altro canto generare casualmente tramite simulazione induce inevitabilmente a generare sequenze appropriate ma in qualche misura errate², di cruciale importanza è poter quantificare la precisione con la quale si giunge ai risultati, per poter dire qualcosa sulla possibile entità dell'errore.

Metodi Monte Carlo e integrali Esempio tra i più rappresentativi, e di importanza fondamentale per questo lavoro, per spiegare l'utilità delle simulazioni è la risoluzione di un integrale definito.

Si supponga di aver a che fare con il seguente integrale,

$$I_\alpha = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx ,$$

per $\alpha > 0$. Se si considera la variabile casuale X , con funzione di densità,

$$\mathcal{F}(x) = e^{-x} , \tag{1}$$

allora dalla definizione di valore atteso [Sheldon Ross, 2002], I_α altro non è che il valore atteso di $X^{\alpha-1}$ con X distribuita come in (1),

$$I_\alpha = E_{\mathcal{F}} [X^{\alpha-1}] .$$

¹ Simulazione e numero casuale non sono sinonimi, o per esser pignoli lo sono solo se si simula da una uniforme $(0, 1)$.

² Non potranno in nessun caso essere identiche alla realtà.

Per aver una stima di I_α , basterà allora ottenere una sequenza di valori indipendenti da $\mathcal{F}(x)$ e calcolarne la media,

$$\hat{I}_\alpha = \frac{1}{n} \sum_{i=1}^n X_i^{\alpha-1} .$$

Tenendo presente l'esempio fatto si intuisce facilmente che la metodologia Monte Carlo altro non è che approssimare un integrale; non importa quanto esso sia complicato. Le strade che si possono intraprendere per calcolare l'integrale comprendono da una parte la conoscenza della funzione (esatta) e la soluzione per via analitica, dall'altra l'approccio numerico tramite simulazione.

I termini *simulazione* e *Monte Carlo* a volte sono usati come sinonimi, ma sottendono a due concetti differenti: la simulazione è la scelta casuale di valori distribuiti in un certo modo, mentre con *Monte Carlo* si intende la soluzione di un integrale tramite simulazione. Questi due concetti legano la soluzione di un integrale (o di una somma nel caso discreto) al valore atteso di una funzione (o di una serie di variabili aleatorie).

Organizzazione del lavoro Il lavoro che si presenta è organizzato in tre capitoli ed in due appendici. La prima appendice presenta un breve excursus sulle conoscenze che sono richieste per comprendere al meglio i capitoli, all'interno di essa saranno sviluppati i quattro concetti che stanno alla base delle applicazioni MCMC (Processi stocastici, Catene di Markov, Simulazione, Statistica bayesiana).

Il primo capitolo presenterà la teoria alla base dei metodi MCMC ed il modo con cui essi interagiscono ed utilizzano la teoria esposta nella prima appendice.

Nel capitolo due si compie il passo successivo: lo sviluppo a casi teorici di ciò che si è esposto, quindi delle possibili applicazioni della teoria MCMC.

Il terzo ed ultimo serve da esempio pratico di applicazione, dato quanto visto nella teoria, per analizzare una serie finanziaria.

L'ultima appendice serve a raccogliere la restante teoria utile alla comprensione dell'elaborato (distribuzioni utilizzate, teoremi ecc.).

Alfabeto greco Nella tabella sottostante si riporta l'alfabeto greco.

| minus | MAIUS | Trascrizione | | minus | MAIUS | Trascrizione |
|-------|-------|--------------|--|-------|-------|--------------|
| α | A | Alfa | | ν | N | Ni |
| β | B | Beta | | ξ | Ξ | Xi |
| γ | Γ | Gamma | | ο | O | Omicron |
| δ | Δ | Delta | | π | Π | Pi, Pgreco |
| ε | E | Epsilon | | ρ | P | Rho |
| ζ | Z | Zeta | | σ | Σ | Sigma |
| η | H | Eta | | τ | T | Tau |
| θ | Θ | Theta | | υ | Υ | Upsilon |
| ι | I | Iota | | φ , φ | Φ | Phi |
| κ | K | Kappa | | χ | X | Chi |
| λ | Λ | Lambda | | ψ | Ψ | Psi |
| μ | M | Mi, Mu | | ω | Ω | Omega |

Tabella 1: Alfabeto greco

Metodi MCMC

1.1 Introduzione

Gli algoritmi MCMC trovano la loro più naturale applicazione nel contesto dell'inferenza Bayesiana. Si ricorda infatti che in tale paradigma, l'interesse è volto ad ottenere la densità a posteriori del vettore di parametri, θ , assunti essere in questo caso variabili casuali. La parte centrale viene quindi svolta dalla funzione di probabilità a posteriori $\pi(\theta|y)$, visto che $\pi(\theta)$ ¹ e $L(\theta; y)$ ² vengono o stimate tramite metodi standard o sono già date per via di esperimenti precedenti o in generale di esperienze passate. Da qui in avanti θ è un vettore d -dimensionale.

Tipicamente si è interessati a trovare il valore atteso, l'aspettativa, di una certa funzione dei parametri $h(\theta)$,

$$E_{\pi(\theta|y)} [h(\theta)] = \int_S h(\theta) \pi(\theta|y) d\theta_1 \dots d\theta_d , \quad (1.1)$$

per esempio, l'aspettativa marginale di θ_i , per $i \in \{1, 2, \dots, d\}$, si determina ponendo $h(\theta_i) = \theta_i$, da cui

$$E_{\pi(\theta|y)} [\theta_i] = \int_S h(\theta_i) \pi(\theta|y) d\theta_i .$$

¹La funzione di densità a priori.

²La funzione di verosimiglianza.

Spesso ci si ritrova a dover fare, letteralmente, i conti con gli integrali; lavorando con vettori di parametri gli integrali sono in più dimensioni e calcolarli analiticamente non è sempre agevole o addirittura possibile. Nonostante i progressivi miglioramenti in ambito informatico anche le soluzioni per via numerica all'aumentare della dimensione diventano via via più dispendiose fino a diventare inutilizzabili in tempi ragionevoli.

Dalle motivazioni sopra esposte risulta chiara l'esigenza di trovare soluzioni alternative: come visto nell'introduzione, un generico integrale può essere adeguatamente stimato tramite simulazione. Ad esempio, generando varie repliche di $\theta^{(i)}$ dalla distribuzione a posteriori³, si può ottenere $\{\theta^{(1)}, \theta^{(2)}, \dots, \theta^{(n)}\}$; allora uno stimatore non distorto di $h(\theta)$ sarà

$$\hat{\mu}_h = \frac{1}{n} \sum_{i=1}^n h(\theta^{(i)}) . \quad (1.2)$$

Si noti che le realizzazioni di $\theta^{(i)}$ sono identicamente distribuite, ma non necessariamente devono essere indipendenti. In statistica bayesiana $\pi(\theta|y)$ di solito è multivariata e spesso è molto difficile campionare valori indipendenti. Una possibilità è di usare le distribuzioni a posteriori marginali, cioè $\{\pi(\theta_1|y), \pi(\theta_2|\theta_1, y), \dots\}$; questo approccio⁴ non è sempre utilizzabile perché implica integrali in $d - 1$ dimensioni e difficilmente essi sono già noti. Nei casi in cui le marginali non sono reperibili, la condizione di indipendenza delle $\theta^{(i)}$ viene rilassata; ovviamente tutto ciò avviene ad un prezzo da scontarsi sulla varianza di (1.2):

$$VAR_{\pi(\theta|y)} [\hat{\mu}_h] = \frac{1}{n} VAR_{\pi(\theta|y)} [h(\theta)] + \frac{1}{n^2} \sum_{j=1}^n \sum_{i=1, i \neq j}^n COV [h(\theta^{(i)}), h(\theta^{(j)})] .$$

Come si intuisce se le covarianze sono tutte positive la varianza complessiva ne può risentire sensibilmente aumentando dal caso di indipendenza; d'altra parte avviene l'esatto contrario se si verificasse il caso delle covarianze tutte negative. Purtroppo i metodi MCMC sviluppati finora tendono a cadere nella prima categoria, si veda [J. S. Dagpunar, 2007] (Cap 8.1).

Questo appena descritto è il punto di vista partendo dalla teoria della statistica

³ $\pi(\theta|y)$

⁴Che verrà spiegato nella successiva sessione e che prende il nome di *campionamento di Gibbs*.

Bayesiana, dove si trovano difficoltà a calcolare le distribuzioni, perché risultato di integrali multidimensionali e quindi si ricorre alla simulazione; d'altro canto si può approcciare alla metodologia MCMC dalle catene di Markov.

Solitamente si è abituati ad avere una matrice di transizione (caso discreto) e da quella, tramite la verifica di certe proprietà, vedi (A.2), trovare la distribuzione stazionaria della catena. Con MCMC invece si ragiona al contrario: si ritiene nota la distribuzione stazionaria e si cerca di ottenere la probabilità di transizione tra gli stati. Tramite il Teorema A.3 si hanno le condizioni per cui una catena di Markov ha distribuzione stazionaria, se questo accade allora, per $t \rightarrow \infty$, la matrice di transizione a t passi converge alla distribuzione stazionaria. In pratica succede che la probabilità di transizione in t passi dallo stato i allo stato j , $p_{ij}^{(t)}$, tenda ad un limite π_j , indipendente dallo stato iniziale i .

Così, per generare una catena con distribuzione stazionaria π , si necessita solo di trovare un *kernel* di transizione⁵ \mathbf{P} che soddisfi le condizioni del Teorema in A.3, per la quale $\pi\mathbf{P} = \pi$ e tale che, date le osservazioni $\mathbf{x} \sim \pi(\mathbf{x})$, se $\mathbf{y} \sim \mathbf{P}(\mathbf{y}, \mathbf{x})$, allora $\mathbf{y} \sim \pi(\mathbf{y})$; si veda [Brooks, 1998] (Cap 2).

Tradizionalmente, la letteratura MCMC parla in termini di campionamenti MCMC e algoritmi. Tuttavia, questo limita inutilmente l'attenzione alle catene di Markov in base ad una sola forma/algoritmo di transizione. Si ritiene spesso più ragionevole combinare un certo numero di differenti *kernel* di transizione per la costruzione di una catena di Markov. In pratica, si suddivide il vettore di stato della catena di Markov in un numero di componenti distinti e vengono utilizzate matrici di transizione diverse per aggiornare ogni componente.

Facendo un piccolo passo indietro, prima viene introdotto quello che potrebbe essere considerato il tipo di transizione standard per aggiornare le componenti del vettore di stato, ovvero il *campionamento di Gibbs*.

⁵La matrice di transizione.

1.2 Campionamento di Gibbs

Il campionamento di Gibbs segue un'idea molto semplice, ma dalle grandi potenzialità visti i suoi risultati asintotici. Per esporla si ricorrerà ad un esempio teorico, ma adattabile ad una moltitudine di situazioni.

Si ponga $\theta_1, \theta_2, \theta_3$ come i rispettivi tre parametri di interesse e si indichi con X l'insieme dei dati disponibili e con M la classe di modelli. L'obiettivo è quello di stimare i parametri così che il modello adattato possa essere usato per fare inferenza. Si supponga che la funzione di verosimiglianza del modello sia difficile da ottenere, ma le tre distribuzioni condizionate di ogni singolo parametro dato gli altri siano disponibili. In altre parole le tre seguenti distribuzioni condizionate, a posteriori, sono conosciute:

$$\mathcal{F}(\theta_1|\theta_2, \theta_3, X, M) , \quad \mathcal{F}(\theta_2|\theta_1, \theta_3, X, M) , \quad \mathcal{F}(\theta_3|\theta_1, \theta_2, X, M) ,$$

ognuna delle \mathcal{F} è la distribuzione di un singolo parametro condizionata agli altri due e ai dati e al modello. Nelle applicazioni, non c'è bisogno di conoscere le forme esatte delle distribuzioni condizionali, ma sarà necessario invece estrarre una realizzazione casuale da ciascuna delle tre distribuzioni condizionate⁶.

Dato il valore, arbitrario, di due elementi del vettore dei parametri $\theta_{2,0}, \theta_{3,0}$, detto *start*, si possono creare i valori successivi, estraendo dalle tre distribuzioni condizionate:

1. Si estrae il valore per $\theta_{1,1}$, da $\mathcal{F}(\theta_{1,1}|\theta_{2,0}, \theta_{3,0}, X, M)$,
2. Si estrae il valore per $\theta_{2,1}$, da $\mathcal{F}(\theta_{2,1}|\theta_{1,1}, \theta_{3,0}, X, M)$,
3. Si estrae il valore per $\theta_{3,1}$, da $\mathcal{F}(\theta_{3,1}|\theta_{1,1}, \theta_{2,0}, X, M)$

così si completa una iterazione del campionamento di Gibbs, ottenendo Θ_1 . Il procedimento continua utilizzando i valori appena trovati come valori *start* per la nuova iterazione e così via fino ad ottenere m vettori,

$$(\theta_{1,1}, \theta_{2,1}, \theta_{3,1}) \dots (\theta_{1,m}, \theta_{2,m}, \theta_{3,m}) .$$

⁶Ciò significa che non è necessario normalizzarle.

Sotto alcune condizioni di regolarità, si può dimostrare che, per un m sufficientemente grande, la serie di vettori trovata approssima piuttosto bene la distribuzione congiunta, $\mathcal{F}(\theta_1, \theta_2, \theta_3 | X, M)$, dei tre parametri. Le condizioni sono abbastanza deboli e riguardano il fatto che per un valore di partenza arbitrario $(\theta_{1,0}, \theta_{2,0}, \theta_{3,0})$, le iterazioni Gibbs abbiano la possibilità di visitare tutto lo spazio dei parametri.

In pratica dopo un numero m , sufficientemente grande, di iterazioni si possono ritenere i successivi campionamenti convergenti alla distribuzione, $\mathcal{F}(\theta_1, \theta_2, \theta_3 | X, M)$; scartando perciò i primi m campioni si usano i successivi $(n - m)$ per fare inferenza sui parametri

$$(\theta_{1,m+1}, \theta_{2,m+1}, \theta_{3,m+1}) \dots (\theta_{1,n}, \theta_{2,n}, \theta_{3,n}) ,$$

e da queste nuove iterazioni si possono estrarre gli indicatori statistici come media varianza ecc, o fare analisi più complesse tipiche dell'inferenza e verifica d'ipotesi classica.

L'estensione al caso i cui la cardinalità del vettore dei parametri sia un generico numero appartenente ai $\mathbb{N} \setminus \{0\}$, non è difficile da quanto visto qui sopra.

Per ottenere una funzione proporzionale alla distribuzione a posteriori, $\mathcal{F}(\theta_i | \theta_{j \neq i}, \theta_{l \neq \{i,j\}}, X, M)$, è sufficiente impostare delle distribuzioni a priori dei parametri, anche non informative se non si è in possesso di dati utili a formalizzare una *a priori*, e della funzione di verosimiglianza del modello.

Nota 1.1. Alle prime m realizzazioni del campionamento di Gibbs, che sono scartate, ci si riferisce normalmente come realizzazioni *burn-in*. Queste prime realizzazioni sono usate per assicurare che le successive siano sufficientemente vicine alla distribuzione congiunta dei parametri.

◊

Nota 1.2. Il metodo così discusso consiste nell'eseguire una singola catena lunga e solo dopo il periodo di burn-in si ottiene il campionamento di Gibbs utile per fare inferenza. In alternativa, si possono eseguire molte catene relativamente brevi utilizzando diversi

valori di partenza. L'ultima iterazione⁷ in ciascuna catena viene quindi utilizzata per formare un campionamento di Gibbs.

◻

1.3 Catene di Markov e introduzione all'algoritmo di Metropolis-Hastings (MH)

L'idea venne inizialmente a Metropolis e fu poi estesa e migliorata da Hastings intorno agli anni '70. Si è ormai compreso che, in ottica bayesiana, l'attenzione è indirizzata verso la funzione multivariata di probabilità a posteriori, $\pi(\theta|y)$. In sostituzione a $\pi(\theta|y)$ si ponga $\pi(x)$ a valori in⁸ $S \subseteq \mathbb{R}^p$. Una sequenza di vettori $X^{(1)}, \dots, X^{(n)}$ verrà generata da una catena di Markov omogenea, con spazio degli stati S e matrice di transizione $P(z|x)$; questo implica

$$P[z|x] = \mathbb{P}[X^{(t)} \leq z | X^{(t-1)} = x^{(t-1)}, \dots, X^{(0)} = x^{(0)}] = \mathbb{P}[X^{(t)} \leq z | X^{(t-1)} = x^{(t-1)}] ;$$

e per il fatto che si tratta di una catena di Markov omogenea,

$$\mathbb{P}[X^{(t)} \leq z | X^{(t-1)} = x] = \mathbb{P}[X^{(1)} \leq z | X^{(0)} = x] , \quad \forall x, z \in S , \quad \forall t = 1, 2, \dots .$$

Essendo $z \in S$, $P(z|x)$ è una funzione di ripartizione, perciò la sua derivata rappresenta la funzione di densità,

$$p(z|x) = \frac{\partial P(z|x)}{\partial z} .$$

La catena è posta ergodica, quindi per il Teorema A.4, la distribuzione stazionaria π è unica. Questo deriva dalla convergenza alla distribuzione stazionaria dei vettori generati dopo un opportuno periodo di simulazione, nonostante le realizzazioni non sono indipendenti, vedi Teorema A.3.

⁷O le ultime iterazioni

⁸Questa generalizzazione oltre a semplificare la notazione serve a distaccarsi dall'idea di parametro facendo assumere a θ una connotazione più generica.

Dopo il periodo *burn in* con il vettore $\{X^{(m+1)}, \dots, X^{(n)}\}$, si può stimare μ_h ,

$$\mu_h = \int_S h(x)\pi(x) \partial x ,$$

con

$$\hat{\mu}_h = \frac{1}{n} \sum_{i=m+1}^n h(X^{(i)}) . \quad (1.3)$$

Scegliendo m sufficientemente grande da far sì che la varianza sia abbastanza piccola.

Per campionare il vettore necessario ad eseguire le analisi richieste si procede in questo modo: dato il generico stato $X^{(i)}$, $X^{(i+1)}$ diventerà z con probabilità di passaggio $\alpha(x, z)$ e resterà invariato con probabilità $1 - \alpha(x, z)$; z è campionato dalla densità $q(z|x)$, e la probabilità α^9 (detta anche *regola di salto*) è

$$\alpha(x, z) = \min\left[1, \frac{\pi(z)q(x|z)}{\pi(x)q(z|x)}\right] , \quad (1.4)$$

dove la quantità $(\pi(z)q(x|z))/(\pi(x)q(z|x))$ è detta *ratio*.

Sotto certe condizioni di regolarità $q(\cdot|\cdot)$ può prendere qualsiasi forma, cosa molto gradita per la grande flessibilità che ne deriva. La funzione $q(\cdot|\cdot)$ prende il nome di *Proposal distribution* o *Jumping distribution*.

I vettori $X^{(1)}, \dots, X^{(n)}$ formano una catena di Markov omogenea con

$$p(z|x) = \alpha(x, z)q(z|x) , \quad \forall x, z \in S \text{ con } x \neq z ,$$

Dalla formula (1.4) se $\alpha(x, z) < 1$, $\alpha(z, x) = 1$ ed ovviamente vice versa, allora guardando sempre la (1.4)

$$\alpha(x, z)\pi(x)q(z|x) = \alpha(z, x)\pi(z)q(x|z) \Rightarrow \frac{\pi(z)q(x|z)}{\pi(x)q(z|x)}\pi(x)q(z|x) = 1\pi(z)q(x|z);$$

⁹Questa forma di probabilità di accettazione non è unica, ci sono altre funzioni di accettazione che forniscono una catena con le proprietà desiderate.

questo mostra che la catena è *Time Reversal* o in bilancio dettagliato, vedi A.6, quindi

$$\pi_x p_{z|x} = \pi_z p_{x|z} .$$

Partendo da quest'ultima equazione si dimostra che π è una distribuzione stazionaria per la catena, che la catena è ergodica, π è unica infine π è il limite in distribuzione della catena¹⁰.

Per stimare μ_h si replica la catena più volte, in modo da far svanire l'effetto del valore iniziale, dopo il periodo già nominato di *burn in* si possono usare i campionamenti per estrarre indicatori statistici di interesse, principalmente la stima del parametro attraverso la media campionaria.

Resta la questione di come valutare quando una realizzazione è in equilibrio. Questo può essere un problema difficile, soprattutto con spazi alto-dimensionali. Se si vogliono generare più catene corte, piuttosto di una singola catena lunga, una soluzione è quella di plottare alcuni dei componenti delle sequenza $X^{(t)}$, più generalmente una funzione degli $X^{(t)}$; è appropriato pensare di valutare il grafico di $\{h(X_i^t), t = 1, 2, 3, \dots\}$ da ripetere per ciascuna delle repliche della catena. Partendo dal presupposto che ogni catena ha un valore iniziale diverso, l'equilibrio si raggiunge quando t è sufficientemente grande per far sì che i grafici siano simili, in altre parole devono fluttuare attorno al medesimo valore medio, visitano lo stesso spazio degli stati.

La seconda questione riguarda quante osservazioni in equilibrio n si generano in ogni esecuzione della catena; Se la catena ha una forte dipendenza, quindi le realizzazioni si muovono lentamente attraverso gli stati (lenta miscelazione), n dovrà essere grande in modo che lo spazio intero sia stato esplorato all'interno di una singola realizzazione della catena.

Ultima osservazione riguarda la (1.4) essendo presente la funzione π sia al numeratore che al denominatore allora non è necessario calcolare il valore della costante di normalizzazione (vedi A.7); questo vantaggio risulta determinante nei casi, non rari, in cui la distribuzione a posteriori è nota eccetto la costante di normalizzazione. Questo

¹⁰In altre parole la quantità $\alpha(\cdot, \cdot)$, formulata come visto, è introdotta per recuperare la condizione di reversibilità, necessaria per rendere π stazionaria

è il vantaggio maggiore che si ha nell'utilizzo di questa metodologia nel campo delle analisi multivariate.

1.4 L'algoritmo di Metropolis-Hastings

L'algoritmo di Metropolis-Hastings è quindi basato su una densità *proposal* $q(z|x)$ che fa muovere la catena di Markov da x a z , e sulla probabilità di accettazione del passaggio, $\alpha(x, z)$, necessaria per assicurare la proprietà di reversibilità.

La prima versione venne introdotta da *Metropolis*, mentre la seconda riguarda il perfezionamento ad opera di *Hastings*. L'idea che sta alla base di *Metropolis* è di prendere $q(z|x) = q(x|z)$ ¹¹ in modo da semplificare la 1.4 in

$$\alpha(x, z) = \min \left(1, \frac{\pi(z)}{\pi(x)} \right) .$$

Se la funzione q scelta è una distribuzione Normale,

$$q(z|x) \propto \exp \left[- (z - x)' \Sigma^{-1} (z - x) \right]$$

quindi¹² dato x , $Z \sim \mathcal{N}(x, \Sigma)$. La scelta di Σ deve gestire al meglio il *trade off* che esiste tra scegliere una varianza alta, per visitare in modo veloce e completo lo spazio dei parametri ma spesso rifiutata soprattutto se x è vicino alla moda della distribuzione¹³ e scegliere una varianza piccola spesso accettata ma con conseguenti problemi di velocità e di lunghezza della serie simulata.

L'algoritmo procede nel seguente modo:

- Per prima cosa trova un valore *start* per far partire la simulazione dalla distribuzione $\pi(x)$
- Per ogni $t = 1, 2, 3, \dots$ finché sarà necessario,

¹¹Prende quindi una distribuzione simmetrica.

¹²Si ricorda che in questa notazione x sta ad indicare il valore del parametro al tempo $t - 1$ ed z il valore del parametro al tempo t ; x e z sono vettori.

¹³Con questo si intende che la probabilità di accettazione α è piccola, si pensi a com'è costruita α .

- a. Simulare un nuovo valore del vettore dei parametri x_t^* , dalla distribuzione nota $q(\cdot|\cdot)$ (la *Proposal distribution*), dato il vettore precedente x_{t-1} ¹⁴. Come detto all'inizio la distribuzione di salto è simmetrica $q(x_j|x_i) = q(x_i|x_j)$ per ogni i, j in T .
- b. si calcola quello che viene anche detto *ratio*,

$$r = \frac{\pi(x_t^*)}{\pi(x_{t-1})} \quad (1.5)$$

- c. assegnare il nuovo valore $X_t = x_t$, vedi 1.4, come segue

$$X_t = \begin{cases} x_t^* & \text{Con probabilita' } \min(1, r) \\ x_{t-1} & \text{Con probabilita' } 1 - \min(1, r) \end{cases} .$$

Sotto alcune condizioni di regolarità la sequenza X_t converge in distribuzione a $\pi(x)$, cioè alla distribuzione a posteriori. L'implementazione dell'algoritmo di *Metropolis* richiede abilità di calcolo per il *ratio* α e di saper campionare i valori di x_t^* dalla distribuzione $q(\cdot|\cdot)$. Esempi di distribuzioni simmetriche sono la normale e la t di Student.

L'accettazione o rifiuto viene fatto tramite la simulazione di un valore u da una uniforme $(0, 1)$, se questo valore è minore di $\alpha(x_{t-1}, x_t^*)$ allora x al tempo t diventerà x_t^* altrimenti resterà x_{t-1} .

Hastings generalizza l'algoritmo di Metropolis in due modi; primo la *jumping distribution* $q(\cdot|\cdot)$ può non essere simmetrica e, secondo, modifica di conseguenza la regola di salto α in

$$r = \frac{\pi(x_t^*)/q(x_t^*|x_{t-1})}{\pi(x_{t-1})/q(x_{t-1}|x_t^*)} = \frac{\pi(x_t^*)q(x_{t-1}|x_t^*)}{\pi(x_{t-1})q(x_t^*|x_{t-1})} . \quad (1.6)$$

Si ricorda $\alpha = \min[1, r]$, vedi (1.4). A queste modifiche ci si riferisce come algoritmo di *Metropolis-Hastings*.

¹⁴Per semplificare la notazione e riferirsi in modo più evidente al tempo si è eliminato il simbolo z e si è inserito al simbolo x il pedice con il tempo t .

1.4.1 Random walk Metropolis

Una variante che si inserisce tra il caso più semplice formulato da Metropolis e il caso più generale teorizzato da Hastings è di scegliere $q(\cdot|\cdot)$ come un *random walk*

$$q(z|x) = g(z-x)$$

quindi

$$z = x + \sigma^{1/2}Z ,$$

dove Z è una realizzazione da una $\mathcal{N}_d(0, 1)$. La differenza è che essendo delle realizzazioni indipendenti si può porre $q(z|x) = q(z)$ allora la distribuzione del nuovo punto è indipendente dal punto attuale. In questa situazione una buona strategia è quella di scegliere q simile a π in modo tale da semplificare il calcolo del ratio, vedi (1.6).

1.4.2 Independence sampler

Come si può intuire dal nome, con questo campionatore si pone $q(z|x) = q(x)$, così la distribuzione del punto candidato alla prossima iterazione è indipendente dalla precedente¹⁵; di conseguenza la (1.4) diventa

$$\alpha(x, z) = \min\left[1, \frac{\pi(z)q(x)}{\pi(x)q(z)}\right] .$$

In questo caso una buona strategia può essere quella di scegliere $q(\cdot)$ simile ad π ; questo si traduce in una probabilità di accettazione vicina a 1, con successivi valori quasi indipendenti; la cosa, naturalmente, è molto positiva dal punto di vista della riduzione della varianza di uno stimatore. In un contesto Bayesiano $q(\cdot)$ potrebbe essere scelta come la distribuzione a priori dei parametri; ovviamente questa eventualità sussiste assieme alla prima se la distribuzione a posteriori differisce di poco dalla a priori.

¹⁵ è il caso più generale del *Random walk Metropolis*.

1.5 La metodologia Griddy Gibbs

Nelle applicazioni finanziarie, un modello può contenere alcuni parametri non lineari (ad esempio, la media mobile dei parametri in un modello ARMA o i parametri in un modello di volatilità GARCH). Dal momento che le distribuzioni condizionali a posteriori non hanno una forma chiusa, cioè difficilmente hanno una espressione che può essere espressa analiticamente, implementare un *campionamento di Gibbs* o l'algoritmo *Metropolis-Hasting* può diventare complicato. Tanner (1996) descrive una procedura per ottenere dei valori random in un campionamento di Gibbs quando le distribuzioni condizionate a posteriori sono univariate. Il metodo prende il nome di campionamento di *Griddy Gibbs*, e risulta ampiamente applicabile.

Si supponga che il parametro scalare θ_i abbia distribuzione a posteriori condizionata $\mathcal{F}(\theta_i|X, \boldsymbol{\theta}_{-i})$, dove $\boldsymbol{\theta}_{-i}$ è il vettore dei parametri a cui è stato tolto θ_i ; Griddy Gibbs si sviluppa nelle seguenti fasi:

- a. Creare una griglia di punti da un intervallo, selezionato correttamente, di θ_i ; ottenuta la griglia $\theta_{i1} < \theta_{i2} < \dots < \theta_{im}$, calcolare la distribuzione condizionata a posteriori per ogni punto per ottenere $w_j = \mathcal{F}(\theta_{ij}|X, \boldsymbol{\theta}_{-i})$ per $j \in \{1, 2, \dots, m\}$.
- b. Usare i valori trovati per calcolare un'approssimazione dell'inversa della funzione di ripartizione di $\mathcal{F}(\theta_i|X, \boldsymbol{\theta}_{-i})$.
- c. Campionare dei valori nell'intervallo $(0, 1)$ e ricavare il valore campionato di θ_i tramite la funzione di distribuzione cumulativa (FDC), calcolata al punto *b*.

Qualche osservazione per rendere più chiaro il metodo è doverosa. Intanto non si necessita della costante di normalizzazione della funzione $\mathcal{F}(\theta_i|X, \boldsymbol{\theta}_{-i})$, visto che l'inversa della funzione di distribuzione cumulativa la si ottiene direttamente dai w_i ; una semplice approssimazione della FDC è la distribuzione discreta per $(\theta_{ij})_{j=1}^m$ con probabilità $p_{ij} = w_j / \sum_{l=1}^m w_l$. In un'applicazione reale, la selezione dell'intervallo per il parametro θ_i deve essere controllato attentamente; una procedura semplice è quella di considerare l'istogramma del campionamento di Gibbs per θ_i : se l'istogramma indica una probabilità non prossima allo zero intorno θ_{i1} o θ_{im} allora l'intervallo deve essere

ampliato, diversamente se l'istogramma mostra una alta concentrazione di probabilità all'interno dell'intervallo scelto, l'intervallo è troppo largo e può essere accorciato. Se l'intervallo è troppo grande, il Gibbs Griddy diventa inefficiente perché la maggior parte delle w_j risulterebbe zero. Infine, Gibbs Griddy o Metropolis-Hasting possono essere utilizzati in un campionamento di Gibbs per ottenere delle realizzazioni casuali di alcuni parametri.

Applicazioni Teoriche

Si passa ora a valutare alcune applicazioni dei metodi MCMC. L'obiettivo è quello di evidenziare l'applicabilità e l'utilità dei metodi. La comprensione di questi esempi possono aiutare a esprimere la propria opinione in applicazioni di metodi MCMC soprattutto in questioni finanziarie.

L'organizzazione del capitolo è così strutturata: nelle prime due sezioni si seguirà [Tsay, 2010] (Cap 12), e saranno prettamente a scopo didattico illustrativo, non saranno poi implementati praticamente; in quella successiva si svilupperanno i metodi MCMC per il modello SV, seguendo l'approccio in [Kim, 1998] e [Meyer, 2000]. Infine, nella sezione 4, si svilupperà l'applicazione dei metodi MCMC ai modelli ARMA-GARCH, seguendo l'esposizione in [Nakatsuma, 2000]. Interessanti sviluppi per i modelli SV si hanno nell'articolo [Chib, 2002] che amplia i contenuti di [Kim, 1998] utilizzando due estensioni dei modelli SV: la prima con errori distribuiti come una t -Student, la seconda identica alla prima salvo che contiene un componente (parametro) di salto nell'equazione di osservazione per consentire grandi movimenti transitori.

2.1 Regressione lineare con errori serialmente correlati

Si tratta di valutare un modello del tipo

$$y_t = \beta_0 + \beta_1 x_{1t} + \cdots + \beta_p x_{pt} + z_t, \quad z_t = \phi z_{t-1} + a_t, \quad (2.1)$$

dove y è la variabile dipendente x sono le variabili esplicative che potrebbero contenere dei valori ritardati di y , infine z segue un semplice modello $AR(1)$ con a realizzazioni i.i.d. da una distribuzione normale a media zero e varianza σ^2 . Come si nota i parametri di interesse sono, $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \sigma^2)$; in notazione vettoriale

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + z_t, \quad z_t = \phi z_{t-1} + a_t, \quad t = 1, 2, \dots, n, \quad (2.2)$$

e n è la lunghezza del campione.

Un modo naturale per implementare il campionamento di Gibbs è di iterare alternativamente il modello di regressione e la serie storica degli errori correlati. Se il modello $AR(1)$ è noto si può facilmente, ricorrendo ai minimi quadrati, stimare il modello di regressione; d'altra parte conoscendo il modello di regressione si può ottenere facilmente la serie degli errori, $z_t = y_t - \mathbf{x}_t' \boldsymbol{\beta}$. Perciò si necessitano le seguenti distribuzioni condizionate,

$$\mathcal{F}(\boldsymbol{\beta}|Y, X, \phi, \sigma^2), \quad \mathcal{F}(\phi|Y, X, \boldsymbol{\beta}, \sigma^2), \quad \mathcal{F}(\sigma^2|Y, X, \boldsymbol{\beta}, \phi),$$

dove $Y = (y_1, \dots, y_n)$ è il vettore contenete tutte le variabili esplicative e X è la matrice con i vettori di tutte le variabili dipendenti.

Seguendo la teoria Bayesiana, vengono utilizzate delle distribuzioni coniugate a priori, in modo da ottenere una forma conosciuta per la distribuzione condizionata a posteriori. Le distribuzioni a priori dei parametri sono,

$$\boldsymbol{\beta} \sim \mathcal{N}(\boldsymbol{\beta}_0; \Sigma_0), \quad \phi \sim \mathcal{N}(\phi_0, \sigma_0^2), \quad \frac{\nu \lambda}{\sigma^2} \sim \chi_\nu^2, \quad (2.3)$$

dove tutte le quantità $\{\boldsymbol{\beta}_0, \Sigma_0, \phi_0, \sigma_0^2, \nu, \lambda\}$ sono quantità note, e vengono nominate *hyperparameter*. I loro valori dipendono dal problema in questione e dall'esperienza maturata da chi esegue l'analisi. Tipicamente si può scegliere, $\boldsymbol{\beta}_0 = 0, \phi_0 = 0$ e Σ_0 matrice diagonale con varianza piuttosto alta. In (2.3) si stanno usando delle distribuzioni a priori indipendenti partizionando il vettore dei parametri $\boldsymbol{\theta}$.

La distribuzione a posteriori per il vettore $\boldsymbol{\beta}$, $\mathcal{F}(\boldsymbol{\beta}|Y, X, \phi, \sigma^2)$ si può ottenere ricor-

rendo al campionamento di Gibbs. In particolare, dato ϕ , si definisce

$$y_{0,t} = y_t - \phi y_{t-1} \quad , \quad \mathbf{x}_{0,t} = \mathbf{x}_t - \phi \mathbf{x}_{t-1} \quad .$$

Usando (2.2)

$$y_{0,t} = \boldsymbol{\beta}' \mathbf{x}_{0,t} + a_t \quad , \quad (2.4)$$

sotto le assunzioni di a_t (2.4) è una regressione lineare multivariata. Perciò tutta l'informazione sul parametro $\boldsymbol{\beta}$ fornita dai dati la si può estrarre con in metodo dei minimi quadrati,

$$\hat{\boldsymbol{\beta}} = \left(\sum_{t=2}^n \mathbf{x}_{0,t} \mathbf{x}_{0,t}' \right)^{-1} \left(\sum_{t=2}^n \mathbf{x}_{0,t} y_{0,t} \right) \quad ,$$

che ha una distribuzione normale multivariata,

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N} \left[\boldsymbol{\beta}, \sigma^2 \left(\sum_{t=2}^n \mathbf{x}_{0,t} \mathbf{x}_{0,t}' \right)^{-1} \right] \quad .$$

Allora concludendo grazie al risultato in A.1, riguardo le distribuzioni normali multivariate coniugate, la distribuzione a posteriori per $\boldsymbol{\beta}$ dato X, Y, ϕ e σ^2 è una normale multivariata

$$(\boldsymbol{\beta} | X, Y, \phi, \sigma^2) \sim \mathcal{N} \left(\boldsymbol{\beta}_*, \Sigma_* \right) \quad , \quad (2.5)$$

con media $\boldsymbol{\beta}_*$ e varianza Σ_* pari a

$$\Sigma_* = \frac{\sum_{t=2}^n \mathbf{x}_{0,t} \mathbf{x}_{0,t}'}{\sigma^2} + \Sigma_0^{-1} \quad , \quad \boldsymbol{\beta}_* = \Sigma_* \left(\frac{\sum_{t=2}^n \mathbf{x}_{0,t} \mathbf{x}_{0,t}'}{\sigma^2} \hat{\boldsymbol{\beta}} + \Sigma_0^{-1} \boldsymbol{\beta}_0 \right) \quad .$$

Si procede analogamente, con le dovute cautele, per il parametro ϕ . Con $\boldsymbol{\beta}$ dato si possono calcolare gli errori del modello z_t , per ogni t ,

$$z_t = \phi z_{t-1} + a_t \quad ,$$

l'informazione data dalla verosimiglianza di ϕ è data ancora dalla stima dei minimi quadrati,

$$\hat{\phi} = \left(\sum_{t=2}^n z_{t-1}^2 \right)^{-1} \left(\sum_{t=2}^n z_t z_{t-1} \right) ,$$

normalmente distribuita con media ϕ e varianza $\sigma^2 (\sum_{t=2}^n z_{t-1}^2)^{-1}$. Rifacendosi sempre al risultato in A.1 la distribuzione condizionata a posteriori per il parametro ϕ è normale con media e varianza,

$$\sigma_*^{-2} = \frac{\sum_{t=2}^n z_{t-1}^2}{\sigma^2} + \sigma_0^{-2} , \quad \phi_* = \sigma_*^2 \left(\frac{\sum_{t=2}^n z_{t-1}^2}{\sigma^2} \hat{\phi} + \sigma_0^{-2} \phi_0 \right) . \quad (2.6)$$

Infine, si trova la distribuzione a posteriori di σ^2 dati β , ϕ e i dati. Conoscendo β , ϕ si ottiene

$$a_t = z_t - \phi z_{t-1} , \quad z_t = y_t - \beta' \mathbf{x}_t , \quad t = 2, \dots, n ,$$

dal risultato A.2 la distribuzione a posteriori di σ^2 è

$$\frac{v\lambda + \sum_{t=2}^n a_t^2}{\sigma^2} \sim \chi_{v+(n-1)}^2 . \quad (2.7)$$

Allora utilizzando le tre distribuzioni di probabilità a posteriori (2.5), (2.6) e (2.7) si può stimare con il campionamento di Gibbs la (2.2), con le seguenti fasi:

- (a) Specificare i parametri delle distribuzioni in (2.3), detti anche *hyperparameter*.
- (b) Specificare i valori di *start* per β , ϕ e σ^2 ; basta ad esempio la stima dei minimi quadrati ordinari di β senza la serie degli errori correlati.
- (c) Usare la distribuzione a posteriori normale multivariata in (2.5), per generare il nuovo valore di β .
- (d) Usare la distribuzione a posteriori normale univariata in (2.6), per generare il nuovo valore di ϕ .
- (e) Usare la distribuzione a posteriori *Chi-Quadro* in (2.5), per generare il nuovo valore di σ^2 .

Si ripetono gli ultimi tre passaggi, $\{c, d, e\}$, per molte iterazioni, in modo da ottenere un campionamento di Gibbs. Il campione così ottenuto viene utilizzato per ricavarne le stime puntuali ed altri indicatori descrittivi dei parametri del modello.

2.2 Valori mancanti e/o valori estremi

In questa sezione si analizza il problema di dati mancanti e degli *outliers* addittivi, cioè dei dati estremi (anomali) che sono più spesso dovuti ad errori nel campionamento o dipendenti dal verificarsi di condizioni anomale che non interessano l'analisi dei dati.

Soprattutto i valori anomali possono compromettere seriamente analisi di serie temporali, perché spesso inducono a distorsioni sostanziali sulla stima dei parametri e portano alla errata specificazione del modello.

Un *outliers* additivo, x_h , è definito come segue,

$$y_t = \begin{cases} x_h + \omega & \text{se } t = h \\ x_t & \text{altrove} \end{cases} .$$

Dove ω è la grandezza del valore anomalo, e x_t è una serie senza dati anomali. Gli errori di misurazione o di rilevazione sono i più tipici esempi di *outliers* addittivi.

Si consideri una serie storica x_t e un indice temporale fisso, h . Si può imparare molto su x_h trattandolo come valore mancante. Se il modello della serie x_t è conosciuto, si può derivare la distribuzione condizionata di x_h dato gli altri valori della serie. Confrontando y_h , valore osservato con la distribuzione derivata di x_h , è possibile determinare se y_h è un probabile *outlier* additivo. Nella fattispecie, se y_h è un valore probabile sotto la distribuzione derivata, allora non è un *outlier* additivo. Tuttavia, se la possibilità di osservare y_h è molto piccola sotto la distribuzione derivata, y_h può essere classificato come un valore anomalo. Pertanto, il rilevamento di valori anomali e il trattamento dei dati mancanti nell'analisi delle serie temporali sono basate sulla medesima idea.

Palesamente parlando di *outlier* additivo, implicitamente si deduce l'esistenza di

altre tipologie di *outlier*; in generale essi sono classificati in quattro categorie, si veda [Tsay, 2010] (Cap. 12.6), dipendenti dall'impatto che hanno sulla serie di dati. In questo elaborato l'attenzione sarà focalizzata sulla tipologia additiva.

2.2.1 Valori Mancanti

Si prende in considerazione un modello $AR(p)$,

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + a_t, \quad (2.8)$$

dove a_t è gaussiano con media zero e varianza σ^2 . In un set di valori lungo n si supponga che manchi l'osservazione h ($< n$). L'obiettivo è stimare il modello in presenza di valori mancanti. I parametri di interesse sono, $\boldsymbol{\theta} = (\boldsymbol{\phi}, x_h, \sigma^2)$, ed i dati della serie storica possono essere riassunti in una matrice \mathbf{X} . Le distribuzioni a priori dei parametri sono,

$$\boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\phi}_0; \boldsymbol{\Sigma}_0), \quad x_h \sim \mathcal{N}(\mu_0, \sigma_0^2), \quad \frac{v\lambda}{\sigma^2} \sim \chi_v^2, \quad (2.9)$$

prendendo come noti i parametri di inizializzazione, le distribuzioni a posteriori per $\boldsymbol{\phi}$ e σ^2 sono calcolate come nella sezione precedente, mentre la distribuzione a posteriori di x_h , $\mathcal{F}(x_h | \mathbf{X}, \boldsymbol{\phi}, \sigma^2)$, è normale univariata con medi μ_* e varianza σ_h^2 . Questi due parametri possono essere ricavati usando il modello di regressione lineare. Si noti che nel modello x_h interagisce con le osservazioni $\{x_{h-p}, \dots, x_{h-1}, x_{h+1}, \dots, x_{h+p}\}$; non si scordi che x_h è preso qui come parametro, considerando la seguente serie di passaggi:

- Per $t = h$, il modello è

$$x_h = \phi_1 x_{h-1} + \dots + \phi_p x_{h-p} + a_h,$$

prendendo $y_h = \phi_1 x_{h-1} + \dots + \phi_p x_{h-p}$ e ponendo $b_h = -a_h$ la prima equazione

può essere scritta come:

$$y_h = x_h + b_h = \phi_0 x_h + b_h ,$$

dove $\phi_0 = 1$.

- Per $t = h + 1$ si ha,

$$x_{h+1} = \phi_1 x_h + \dots + \phi_p x_{h-p+1} + a_{h+1} ,$$

prendendo $y_{h+1} = x_{h+1} - \phi_2 x_{h-1} - \dots - \phi_p x_{h-p+1}$ e ponendo $b_{h+1} = a_{h+1}$ l'equazione può essere riscritta come:

$$y_{h+1} = \phi_1 x_h + b_{h+1} ,$$

- Facilmente si ricava il caso generale per x_{h+j} , dove $j \in \{1, \dots, p\}$,

$$y_{h+j} = \phi_j x_h + b_{h+j} .$$

Si capisce che, per un modello $AR(p)$, il valore mancante x_h è influente, in $p + 1$ equazioni,

$$y_{h+j} = \phi_j x_h + b_{h+j} \quad , \quad j = 0, \dots, p . \quad (2.10)$$

Essendo la normale distribuzione simmetrica attorno alla sua media allora a_t e $-a_t$ hanno la stessa distribuzione $\forall t$ in $\{1, \dots, n\}$. Perciò l'equazione (2.10) è un semplice modello di regressione lineare con $p + 1$ dati. Lo stimatore dei minimi quadrati per x_h e la sua varianza sono rispettivamente:

$$\hat{x}_h = \frac{\sum_{j=0}^p \phi_j y_{h+j}}{\sum_{j=0}^p \phi_j^2} \quad , \quad Var[\hat{x}_h] = \frac{\sigma^2}{\sum_{j=0}^p \phi_j^2} .$$

Per esempio¹, se $p = 1$, allora $\hat{x}_h = [\phi_1 / (1 + \phi_1^2)](x_{h-1} + x_{h+1})$ viene indicato come il valore filtrato di x_h . Siccome il modello $AR(1)$ è reversibile, pesi simmetrici sono usati per un intorno di x_h .

Infine utilizzando il risultato in A.1, si ottiene la distribuzione a posteriori per x_h , normale con media μ_* e varianza σ_*^2 ,

$$\mu_* = \frac{\sigma^2 \mu_0 + \sigma_0^2 (\sum_{j=0}^p \phi_j^2) \hat{x}_h}{\sigma^2 + \sigma_0^2 \sum_{j=0}^p \phi_j^2}, \quad \sigma_*^2 = \frac{\sigma^2 \sigma_0^2}{\sigma^2 + \sigma_0^2 \sum_{j=0}^p \phi_j^2}. \quad (2.11)$$

I valori mancanti possono verificarsi a chiazze, con conseguente situazione di più valori mancanti consecutivi. Questo problema può essere gestito in due modi: primo, generalizzando il metodo precedente per ottenere direttamente una soluzione per molteplici valori filtrati; secondo è possibile applicare il risultato di un singolo valore mancante (2.11) più volte all'interno di una iterazione dell'algoritmo di Gibbs.

Nel primo caso si consideri di avere due valori mancanti consecutivi x_h, x_{h+1} , si definisce una variabile y_{h+j} in modo simile al caso di un singolo valore mancante, così da avere una regressione lineare multipla con parametri $\{x_h, x_{h+1}\}$. Combinando con le loro distribuzioni a priori si ottiene come a posteriori una normale bivariata per $\{x_h, x_{h+1}\}$; nel campionamento di Gibbs, questo approccio ricostruisce i valori consecutivi mancanti congiuntamente. Mentre nel secondo metodo essi vengono ricostruiti separatamente.

Tenendo conto che si tratta di serie storiche i valori, soprattutto se consecutivi, sono correlati perciò la prima metodologia è preferibile tanto più i valori mancanti sono in numero elevato; se la cardinalità è bassa allora anche la seconda opzione porta a buoni risultati. Ovviamente si è finora considerato il caso in cui, $h - p \geq 1$ e $h + p \leq n$, se h esce da questo intervallo è necessario riconsiderare il numero dei valori che entrano nel modello di regressione per stimare i valori mancanti.

¹Mettendo insieme gli stimatori con le informazioni su come creare le y , e tenendo presente $\{p = 1, \phi_0 = 1\}$ si ottiene:

$$y_h = \phi_1 x_{h-1}, \quad y_{h+1} = x_{h+1} \Rightarrow \hat{x}_h = \frac{\phi_0 \phi_1 y_h + \phi_1 y_{h+1}}{1 + \phi_1^2} = \frac{\phi_1 (x_{h-1} + x_{h+1})}{1 + \phi_1^2}.$$

Aumentando p si allungano solo i calcoli, ma la logica resta invariato.

2.2.2 Determinare i possibili Outlier

La ricerca di valori anomali addittivi, diventa centrale in un contesto di MCMC. Eccetto il caso di una sequenza di *outliers* con la stessa deviazione, il campionamento semplice di Gibbs sembra funzionare bene. Ancora si utilizzerà un modello $AR(1)$ per illustrare il problema; in ogni caso algoritmi alternativi si possono considerare nel caso si necessiti di campionare parametri non lineari.

Data una serie y_t con la presenza di *outliers* la cui posizione e segno dello scostamento sono ignoti, si può scrivere il modello come

$$y_t = \delta_t \beta_t + x_t \quad t = \{1, \dots, n\} , \quad (2.12)$$

dove δ_t è una variabile casuale bernulliana, con $\mathbb{P}[\delta_t = 1] = \varepsilon$ e $\mathbb{P}[\delta_t = 0] = 1 - \varepsilon$; mentre β_t è una sequenza di valori random da una distribuzione data; infine x_t è una serie *outliers-free*²,

$$x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + a_t ,$$

dove a_t è gaussiano con media zero e varianza σ^2 . Il modello può sembrare complicato ma così l'occorrenza di un valore estremo può accadere, con probabilità ε , ad ogni tempo t campionato.

Facendo riferimento alla (2.12), con n osservazioni, si hanno $2n + p + 3$ parametri; $\boldsymbol{\phi} = \{\phi_0, \dots, \phi_p\}'$, $\boldsymbol{\delta} = \{\delta_1, \dots, \delta_n\}$, $\boldsymbol{\beta} = \{\beta_1, \dots, \beta_n\}$, σ^2 , ε . I parametri binari δ sono governati dalla probabilità ε , e i β_i sono determinati da una specifica distribuzione. I parametri δ_t e β_t possono essere interpretati come la presenza o assenza di un *outlier* additivo e come la grandezza o importanza del valore anomalo.

Si assumano come a priori le seguenti distribuzioni,

$$\boldsymbol{\phi} \sim \mathcal{N}(\boldsymbol{\phi}_0, \boldsymbol{\Sigma}_0) , \quad \frac{\nu \lambda}{\sigma^2} \sim \chi_\nu^2 , \quad \varepsilon \sim \text{Beta}(\gamma_1, \gamma_2) , \quad \beta_t \sim \mathcal{N}(0, \zeta^2) ,$$

²Senza *outliers*.

i parametri delle distribuzioni a priori sono noti.

Per implementare un campionamento di Gibbs per una stima del modello in (2.12), devono essere considerate le a posteriori per le seguenti quantità,

$$\mathcal{F}(\boldsymbol{\phi}|Y, \boldsymbol{\delta}, \boldsymbol{\beta}, \sigma^2), \quad \mathcal{F}(\delta_h|Y, \delta_{-h}, \boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2), \quad \mathcal{F}(\beta_h|Y, \beta_{-h}, \boldsymbol{\phi}, \boldsymbol{\delta}, \sigma^2),$$

$$\mathcal{F}(\varepsilon|Y, \boldsymbol{\delta}), \quad \mathcal{F}(\sigma^2|Y, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\phi}),$$

dove nuovamente $\boldsymbol{\theta}_{-i}$ significa il vettore meno l'elemento i -esimo.

Condizionatamente alla conoscenza di $\boldsymbol{\delta}$ e $\boldsymbol{\beta}$, la serie libera da *outlier* x_t è data da $x_t = y_t - \delta_t \beta_t$. L'informazione dei dati riguardo $\boldsymbol{\phi}$ è contenuta nella stima dei minimi quadrati

$$\hat{\boldsymbol{\phi}} = \left(\sum_{t=p+1}^n x_{t-1} \mathbf{x}'_{t-1} \right)^{-1} \left(\sum_{t=p+1}^n x_{t-1} x_t \right)$$

dove in questo caso $x_{t-1} = \{1, x_{t-1}, \dots, x_{t-p}\}$ è un vettore di $p+1$ elementi; la stima è normalmente distribuita con media $\boldsymbol{\phi}$ e varianza,

$$\hat{\Sigma} = \sigma^2 \left(\sum_{t=p+1}^n x_{t-1} \mathbf{x}'_{t-1} \right)^{-1}.$$

Anche la distribuzione a posteriori per $\boldsymbol{\phi}$ è normale multivariata con media e varianza ottenuta come in, (2.5),

$$\Sigma_* = \frac{\sum_{t=2}^n \mathbf{x}_{t-1} \mathbf{x}'_{t-1}}{\sigma^2} + \Sigma_0^{-1}, \quad \boldsymbol{\phi}_* = \Sigma_* \left(\frac{\sum_{t=2}^n \mathbf{x}_{t-1} \mathbf{x}'_{t-1}}{\sigma^2} \hat{\boldsymbol{\phi}} + \Sigma_0^{-1} \boldsymbol{\phi}_0 \right).$$

Invece la a posteriori di σ^2 è una χ^2 inversa,

$$\frac{\nu \lambda + \sum_{t=2}^n a_t^2}{\sigma^2} \sim \chi_{\nu+(n-p)}^2,$$

dove $a_t = x_t - \boldsymbol{\phi}' \mathbf{x}_{t-1}$ e $x_t = y_t - \delta_t \beta_t$.

Proseguendo per il parametro δ_h si ottiene la a posteriori con il seguente processo. Primo δ_h è relativo ai valori di y_i, β_i, δ_i con $i \in \{h-p, \dots, h-1, h+1, \dots, h+p\} \neq h$, cioè $x_j = y_j - \delta_j \beta_j$ per $j \neq h$. Secondo x_h può avere due possibili valori a seconda di

δ_h :

$$x_h = \begin{cases} y_h - \beta_h & \text{se } \delta_h = 1 \\ y_t & \text{altrove} \end{cases} .$$

Si definisca

$$w_j = x_j^* - \phi_0 - \phi_1 x_{j-1}^* - \cdots - \phi_p x_{j-p}^* \quad j = \{h, \dots, h+p\} ,$$

dove $x_j^* = x_j$ se $j \neq h$ e $x_j^* = y_j$ se $j = h$. I due valori possibili di x_h danno luogo a due scenari:

- (Caso 1): $\delta_h = 0$, allora l'osservazione non è un *outlier* e $x_h^* = y_h = x_h$. Da qui $w_j = a_j$, cioè $w_j \sim \mathcal{N}(0, \sigma^2)$, per $j = h, \dots, h+p$.
- (Caso 2): $\delta_h = 1$, ora l'osservazione è un *outlier* e $x_h^* = y_h = x_h + \beta_h$. w_j risulta contaminato da β_h , infatti

$$w_h \sim \mathcal{N}(\beta_h, \sigma^2) \quad \text{e} \quad w_j \sim \mathcal{N}(-\phi_{j-h}\beta_h, \sigma^2) , \quad j = h+1, \dots, h+p .$$

definendo $\psi_0 = -1$ e $\psi_i = \phi_i$ si ha $w_j \sim \mathcal{N}(-\psi_{j-h}\beta_h, \sigma^2)$, $j = h, \dots, h+p$.

Ricordando che il Caso 1 si ha con probabilità $1 - \varepsilon$ ed il Caso 2 con probabilità ε si può scrivere la seguente distribuzione a posteriori per δ_h ,

$$\begin{aligned} \mathbb{P}[\delta_h = 1 | Y, \delta_{-h}, \boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2] = \\ \frac{\varepsilon \exp[-\sum_{j=h}^m (w_j + \psi_{j-h}\beta_h)^2 / (2\sigma^2)]}{\varepsilon \exp[-\sum_{j=h}^m (w_j + \psi_{j-h}\beta_h)^2 / (2\sigma^2)] + (1 - \varepsilon) \exp[-\sum_{j=h}^m w_j^2 / (2\sigma^2)]} \end{aligned} \quad (2.13)$$

Questa distribuzione a posteriori semplicemente confronta i valori pesati della funzione di verosimiglianza sotto le due situazioni, dove il peso diventa la probabilità di ciascuna situazione.

Infine la distribuzione a posteriori per β_h è come segue:

- Se $\delta_h = 0$, allora l'osservazione non è un *outlier* e $\beta_h \sim \mathcal{N}(0, \zeta^2)$.

- $\delta_h = 1$, ora l'osservazione è un *outlier* e l'osservazione è contaminata da β_h . Allora riprendendo la costruzione di $w_j \sim \mathcal{N}(-\Psi_{j-h}\beta_h, \sigma^2)$ per $j = h, h+1, \dots, \min(h+p, n)$; l'informazione relativa a β_h può essere estratta da

$$w_j = -\Psi_{j-h}\beta_h + a_t \quad j = h, h+1, \dots, \min(h+p, n) .$$

si passa alla stima dei minimi quadrati ordinari

$$\hat{\beta}_h = \frac{\sum_{j=h}^m -\Psi_{j-h}w_j}{\sum_{j=h}^m \Psi_{j-h}^2} , \quad m = \min(h+p, n) ,$$

essa è normalmente distribuita con media β_h e varianza $\sigma^2 / \sum_{j=h}^m \Psi_{j-h}^2$. Tramite il risultato delle distribuzioni normali coniugate in tabella A.1, si ottiene la a posteriori per β_h normale con media β_h^* e varianza $\sigma_{h^*}^2$,

$$\beta_h^* = \frac{-(\sum_{j=h}^m \Psi_{j-h}w_j)\zeta^2}{\sigma^2 + (\sum_{j=h}^m \Psi_{j-h}^2)\zeta^2} , \quad \sigma_{h^*}^2 = \frac{\sigma^2\zeta^2}{\sigma^2 + (\sum_{j=h}^m \Psi_{j-h}^2)\zeta^2} .$$

2.3 Applicazione ai modelli SV, Stochastic Volatility

2.3.1 Il modello SV

Un'altra importante applicazione, della teoria MCMC, riguarda i modelli a volatilità stocastica. Nei modelli SV la varianza segue un processo stocastico latente e tipicamente il modello SV è formulato per il tempo discreto. La rappresentazione canonica di questa classe di modelli è:

$$y_t = \beta \exp\left\{\frac{h_t}{2}\right\} \varepsilon_t \quad t \geq 1 , \quad (2.14)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t , \quad h_t \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{1 - \phi^2}\right) ,$$

dove y_t sono i rendimenti corretti per la media dell'asset di riferimento al tempo t , h_t è la log-volatilità e si assume che il processo sia stazionario ($|\phi| < 1$); ε_t e η_t sono

distribuzioni normali standard incorrelate, o altrimenti detti processi *white noise*. Il parametro $\beta = \exp\{\mu/2\}$ gioca il ruolo di fattore costante di scala, e può essere pensato come la volatilità istantanea, ϕ si interpreta come la persistenza della volatilità, infine σ_η come la varianza della log-volatilità. Per ragioni di identificabilità o β viene posto uguale ad uno o μ viene considerato nullo. Nelle applicazioni pratiche si preferisce lasciare $\beta = 1$ e non mettere restrizioni a μ quando si stima il modello, e poi si riportano le stime di $\beta = \exp\{\mu/2\}$, vista l'interpretazione economica più diretta.

2.3.2 L'approccio tramite MCMC

Come visto è necessario individuare le distribuzioni a priori, la verosimiglianza e quindi ottenere le distribuzioni a posteriori da cui campionare i valori dei parametri. L'approccio che segue prima [Kim, 1998] e poi con l'ausilio del software WinBugs [Meyer, 2000]³, si basa sulla stima delle a posteriori marginali⁴. Partendo da una densità multivariata $\pi(\psi|y)$ si provvederà a campionare i parametri del vettore ψ a blocchi associando ad ognuno la propria distribuzione condizionata, in modo da evitare la stima di integrali multidimensionali.

Facendo un passo indietro per impostare un'analisi bayesiana completa per il modello SV si devono scrivere le distribuzioni a priori dei parametri non osservabili μ, ϕ, σ_η , dell'insieme di stati della volatilità h_0, h_1, \dots, h_n ; e la verosimiglianza dei dati osservabili in funzione dei parametri non osservabili e del processo latente della variabilità h_t .

La distribuzione a priori congiunta sarà,

$$\mathcal{F}(\mu, \phi, \sigma_\eta^2, h_1, \dots, h_n) = \mathcal{F}(\mu, \phi, \sigma_\eta^2) \mathcal{F}(h_0 | \mu, \sigma_\eta^2) \prod_{i=1}^n \mathcal{F}(h_i | h_{i-1}, \mu, \phi, \sigma_\eta^2) .$$

Si assume l'indipendenza dei parametri μ, ϕ e σ_η^2 , e si utilizzano le a priori come in [Kim, 1998], e $\mathcal{F}(h_t | h_{t-1}, \mu, \phi, \sigma_\eta^2)$ sarà data da (2.14).

³Questo verrà esposto nel capitolo successivo.

⁴Questo approccio è stato in realtà suggerito in primis da [Jacquier, Polson, Rossi, 1994]

La verosimiglianza delle osservazioni

$$\ell(y_1, \dots, y_n | h_0, \dots, h_n, \mu, \phi, \sigma_\eta) = \prod_{i=1}^n p(y_i | h_i) ,$$

sarà estratta dall'equazione (2.14).

La distribuzione a posteriori congiunta ha forma

$$\begin{aligned} \mathcal{F}(\mu, \phi, \sigma_\eta^2, h_1, \dots, h_n | Y) &= \mathcal{F}(\psi | Y) \\ &= \mathcal{F}(\mu, \phi, \sigma_\eta^2) \mathcal{F}(h_0 | \mu, \sigma_\eta^2) \prod_{i=1}^n \mathcal{F}(h_i | h_{i-1}, \mu, \phi, \sigma_\eta^2) \prod_{i=1}^n p(y_i | h_i) , \end{aligned} \quad (2.15)$$

per campionare direttamente da essa si dovrebbero calcolare integrali a (n+4) dimensioni, viene quindi in aiuto la teoria del campionamento di Gibbs la quale divide lo spazio parametrico ψ .

Nel modello SV visto in (2.14) il vettore dei parametri ψ viene diviso in ogni sua componente. In ogni iterazione dell'algoritmo aggiornerà ogni componente uno alla volta, in pratica:

- (a) Si inizializza h e σ_η^2 , ϕ e μ .
- (b) Si campiona h_t da $h_t | h_{-t}, y, \theta$;
- (c) Si campiona da $\sigma_\eta^2 | h, y, \phi, \mu, \beta$;
- (d) Si campiona da $\phi | h, \sigma_\eta^2, \mu, \beta$;
- (e) Si campiona da $\mu | h, \sigma_\eta^2, \mu$;
- (f) Si torna al punto (b) per quante volte necessario.

Ogni ciclo da (b) a (e) è una iterazione, per raggiungere la convergenza a $\psi | y$ c'è bisogno di migliaia di iterazioni.

Il processo della volatilità h_t , ha distribuzione dipendente da (2.14), data l'ipotesi ε_t *white noise*, essa ha errori normali con varianza fissa σ_η^2 e media $\mu + \phi(h_t - \mu)$,

$$h_{t+1} \sim \mathcal{N}(\mu + \phi(h_t - \mu), \sigma_\eta^2) .$$

Seguendo la teoria in [Kim, 1998], la distribuzione a priori per σ_{η}^2 è una *inverse Gamma*, $\sigma_{\eta}^2 \sim IG(\sigma_r/2, S_{\sigma}/2)$. Per l'applicazione i valori della a priori vengono scelti come $\sigma_r = 5$ e $S_{\sigma} = 0.01 \cdot \sigma_r$. Prendendo il risultato nella nota A.3, la a posteriori per σ_{η}^2 sarà una *inverse Gamma* di parametri $[(\sigma_r + n)/2, (S_{\sigma} + \sum\{h_i - (\mu + \phi(h_{i-1} - \mu))\})/2]$.

Per campionare ϕ si parte dall'equazione $\phi = 2\phi^* - 1$ dove ϕ^* è distribuito come una variabile aleatoria Beta di parametri $(\phi_{(1)}, \phi_{(2)})$; perciò a meno del fattore di normalizzazione, la a priori per ϕ è

$$\pi(\phi) \propto \left\{ \frac{1+\phi}{2} \right\}^{\phi_{(1)}-1} \left\{ \frac{1-\phi}{2} \right\}^{\phi_{(2)}-1} \quad \phi_{(1)}, \phi_{(2)} > 0.5 ,$$

la riscrittura con ϕ^* permette a ϕ di avere un supporto nell'intervallo $(-1, 1)^5$. Rifacendosi allo stesso elaborato i valori dei parametri sono $\phi_{(1)} = 20$ e $\phi_{(2)} = 1.5$, il che implica una a priori con media 0.86.

Per μ verrà scelta una normale di media 0 e varianza 10, questa scelta è simile ad avere una distribuzione non informativa per il parametro: vista la media nulla si suppone che non ci sia una persistenza che prevale sia essa positiva o negativa, e con la varianza piuttosto alta si accetta l'ipotesi che esistano persistenze elevate. Una volta campionato μ si passa a calcolare $\beta = \exp\{\mu/2\}$, quantità di più immediata interpretazione economica.

2.4 Applicazione ai modelli ARMA-GARCH

In questa sezione verrà presentato l'approccio MCMC, per la stima e l'inferenza bayesiana del modello ARCH e GARCH. I modelli (ARCH) e (GARCH) sono stati teorizzati da Engle (1982) e da Bollerslev (1986); essi sono stati ampiamente studiati e applicati in molti campi di economici ma, specialmente, in finanza. Questo tipo di approccio non è solo applicabile alla classe di modelli per l'eteroschedasticità (GARCH)

⁵Vedi B.1, la distribuzione Beta ha supporto nell'intervallo (0,1).

ma anche per i modelli di regressione per la media con errori stimati tramite GARCH, questi modelli sono detti ARMA-GARCH, per la teoria si veda [Tsay, 2010] (Cap. 3).

2.4.1 Il modello ARMA-GARCH

Si è nel caso in cui la serie è autocorrelata ed i residui di una modellazione lineare della serie presentano effetti ARCH. Al fine di una più accurata analisi del fenomeno risulterà fondamentale la descrizione congiunta del comportamento del suo livello medio e della sua volatilità, considerazione, questa, che ha portato alla definizione di processi stocastici denominati ARMA-GARCH, che fondono le due strutture in un unico modello, il cui proposito sarà di spiegare il livello medio del fenomeno tramite un ARMA e la sua volatilità tramite un GARCH;

$$r_t = \mu + \sum_{i=1}^p \phi_i r_{t-i} + \sum_{j=1}^q \delta_j \varepsilon_{t-j} + \varepsilon_t = \mu_t + \varepsilon_t ,$$

$$\varepsilon_t = r_t - \mu_t = v_t \sigma_t \quad v_t \sim NID(0, 1) \quad \varepsilon_t | \Omega_{t-1} \sim \mathcal{N}(0, \sigma_t^2) ,$$

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_r \varepsilon_{t-r}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_s \sigma_{t-s}^2 . \quad (2.16)$$

In (2.16) r_t sono i rendimenti, ϕ_i sono i coefficienti della parte (AR) δ_i quelli della parte (MA), mentre α e β sono i vettori dei parametri del modello GARCH per gli errori. Le condizioni per la stazionarietà del modello si possono trovare in [Nakatsuma, 2000].

2.4.2 L'approccio tramite MCMC

Per applicare l'analisi proposta per prima cosa si delinea la distribuzione a posteriori,

$$\pi(\theta|R) = \frac{\ell(r_t|R, \theta)p(\theta)}{\int_S \ell(r_t|R, \theta)p(\theta)d\theta} , \quad (2.17)$$

dove θ è il vettore contenente tutti i parametri del modello, $\ell(r_t|R, \theta)$ è la verosimiglianza del modello ed infine $p(\theta)$ la distribuzione a priori dei parametri.

La funzione di verosimiglianza di un ARMA-GARCH è del tipo,

$$\ell(r_t|R, \theta) = \prod \frac{1}{\sqrt{2\pi\sigma_t^2}} \exp\left\{-\frac{\hat{\varepsilon}_t^2}{2\sigma_t^2}\right\},$$

dove $\hat{\varepsilon}_t = r_t - \mu_t$, si assume $r_0 = \varepsilon_0$ e $r_t = 0$ se $t < 0$, ε_0 si tratta come un parametro del modello. La distribuzione a priori ha una modellazione come segue

$$p(\varepsilon_0, \phi, \delta, \alpha, \beta) = \mathcal{N}(\mu_{\varepsilon_0}, \Sigma_{\varepsilon_0}) \times \mathcal{N}(\mu_\phi, \Sigma_\phi) \mathbb{I}_\phi \mathcal{N}(\mu_\delta, \Sigma_\delta) \mathbb{I}_\delta \times \mathcal{N}(\mu_\alpha, \Sigma_\alpha) \mathbb{I}_\alpha \mathcal{N}(\mu_\beta, \Sigma_\beta) \mathbb{I}_\beta.$$

le funzioni indicatrici si riferiscono ai vincoli di stazionarietà ed invertibilità del modello, se i parametri delle varie componenti del modello ARMA-GARCH sono dei vettori allora le distribuzioni saranno della normali multivariate. Certe tipologie di vincoli potrebbero essere inserite nelle distribuzioni a priori, ad esempio se un parametro dev'essere contenuto nell'intervallo $(0, 1)$ la distribuzione *Beta* può rivelarsi una scelta interessante.

Stimare direttamente da 2.17 necessita la conoscenza del fattore di normalizzazione. O si adopera il campionamento di Gibbs in modo da avere delle distribuzioni a posteriori univariate o, come visto nel metodo di *Metropolis-Hastings*, si sceglie una *proposal distribution* $q(\cdot)$ da cui si estrae un nuovo vettore dei parametri, $\hat{\theta}$, quindi si implementa la regola:

$$\alpha(\theta_t, \theta_{t-1}) = \min \left[\frac{\pi(\hat{\theta}|R)q(\theta|\hat{\theta})}{\pi(\theta|R)q(\hat{\theta}|\theta)} \right]. \quad (2.18)$$

Per costruire la procedura MCMC si divide il vettore dei parametri in due⁶: $\theta_1 = \{\varepsilon_0, \phi, \delta\}$ e $\theta_2 = \{\alpha, \beta\}$. Per i due gruppi di parametri si usano delle *proposal* diverse.

La *proposal distribution* per θ_1 è basata su

$$r_t = \mu + \sum_{i=1}^p \phi_i (r_{t-i} - \mu) + \varepsilon_t + \sum_{j=1}^q \delta_j \varepsilon_{t-j}, \quad (2.19)$$

⁶Anche prima con i modelli SV si divideva il vettore dei parametri.

supponendo che la varianza σ_t^2 sia conosciuta.

La *proposal distribution* per θ_2 è basata su una riscrittura del modello GARCH⁷,

$$\varepsilon_t^2 = \omega + \sum_{i=1}^l (\alpha_i - \beta_i) \varepsilon_{t-i}^2 + w_t - \sum_{i=1}^s (\beta_i) w_{t-i} , \quad w_t \sim \mathcal{N}(0, s\sigma_t^4) , \quad (2.21)$$

dove $l = \max\{r, s\}$ ed i coefficienti $\alpha_i = 0$ per $i > r$ e $\beta_i = 0$ per $i > s$.

La procedura suggerita da [Nakatsuma, 2000] consiste nel

- (a) Generare $\theta_1 = \{\varepsilon_0, \phi, \delta\}$ da (2.19) dato θ_2 ;
- (b) Generare $\theta_2 = \{\alpha, \beta\}$ da (2.21) dato θ_1 ;
- (c) Applicare la regola in (2.18), definita dall'algoritmo MH;
- (d) Ripetere i primi tre passi per quante iterazioni necessarie.

⁷Nominata *rappresentazione ARMA* del modello; definendo $\eta_t = \varepsilon_t^2 - \sigma_t^2$ quindi $\sigma_t^2 = \varepsilon_t^2 - \eta_t$ si può riscrivere l'espressione della varianza condizionata (2.16)

$$\varepsilon_t^2 = \omega + \sum_{i=1}^{\max(r,s)} (\alpha_i + \beta_i) \varepsilon_{t-i}^2 + w_t - \sum_{j=1}^s \beta_j w_{t-j} , \quad (2.20)$$

η_t è un processo martingala e quindi il modello in (2.20) è un ARMA per i quadrati di ε_t . In altre parole il modello GARCH può essere visto come un ARMA della serie ε_t^2 .

Applicazioni Pratiche

3.1 Introduzione

Delle varie applicazioni possibili in questo capitolo ne verranno viste due: una per stimare i coefficienti del modello SV (Stochastic Volatility) e l'altro per stimare quelli dei modelli GARCH. Questi modelli non sono di certo pochi in quanto a possibilità di scelta, comprendono vari (sotto)modelli diversi ognuno per caratteristiche peculiari. Per restare su una applicazione generale si intende stimare modello SV AR(0) e un modello GARCH(1,1). Volendo analizzare la volatilità non è previsto di implementare la modellistica ARMA-GARCH, che tra l'altro risulta troppo complessa per il pc in dotazione. La serie scelta è quella dei rendimenti dell'indice *S&P500* (Standard and Poor's 500)¹; l'indice è lo stesso degli esempi del testo [Tsay, 2010], così da poter avere un riferimento. Per interesse teorico verranno provate, per ogni modello, due set di distribuzioni a priori in modo da valutare quanto esse incidano sulle stime dei parametri.

Si chiarisce immediatamente che, in questo lavoro, non vi è alcuna intenzione di (ri)aprire la discussione tra modelli SV e GARCH, né tanto meno di dare indicazioni di qualsivoglia carattere sulla serie analizzata. L'obbiettivo resta fine a se stesso: esporre un metodo di stima per modelli alternativo all'approccio inferenziale con i metodi di

¹L'indice *S&P500* è stato realizzato da Standard & Poor's nel 1957 e segue l'andamento di un paniere azionario formato dalle 500 aziende statunitensi a maggiore capitalizzazione, [Wikipedia].

verosimiglianza (per questi ultimi metodi si può consultare [Tsay, 2010] Cap. 3).

Il maggior problema riscontrato nella stesura del capitolo è stato quello di poter reperire un software che potesse essere utilizzato per le analisi con metodi MCMC; si è iniziato con la ricerca di una libreria per il software *R-Gui* che ne implementasse il metodo, strada abbandonata nel momento in cui le librerie sull'argomento sono sembrate troppo specifiche e poco adattabili al lavoro che si voleva sviluppare. La scelta è infine caduta sul software *WinBugs* creato appositamente per il campionamento MCMC dal quale è possibile esportare le iterazioni dei singoli parametri ed analizzarle con *R-Gui*, sicuramente più completo e versatile dal punto di vista delle possibili analisi.

3.2 Software WinBugs

In questo capitolo si utilizza il programma *WinBugs* (per informazioni in merito si veda [WinBugs 1.4]) che come si intuisce dal nome per intero *Bayesian inference Using Gibbs Sampling*, usa principalmente il campionamento di *Gibbs*.

Alcuni cenni sono doverosi: il software in questione è sviluppato e distribuito dal MRC² Biostatistics Unit, al momento sembra che l'ultima versione risalga al 2007 e che per ora non ci siano in programma sviluppi ulteriori.

WinBugs è progettato esclusivamente per l'analisi bayesiana con metodi MCMC e rappresenta una valida alternativa a programmi come *R*, *Stata* o *S-plus*; esiste tuttavia la possibilità di far eseguire *WinBugs* tramite *R* e i vantaggi, se ne deduce, si esprimono sulle possibili analisi post simulazione senza dover salvare i dati e ricaricarli.

Il software si articola in due distinti approcci per creare il modello,

- Il primo è la classica programmazione, con le proprie parole chiave e cicli; questi ultimi tipici anche di altri programmi.
- Il secondo approccio è completamente differente: si può definire un modello con la rappresentazione detta *doodle*; si tratta di grafico in cui i vari nodi (che sono le variabili) sono collegati dalle rispettive relazioni che intercorrono tra di essi.

²Medical Research Council, è un'agenzia governativa finanziata con fondi pubblici con la responsabilità di coordinare e finanziare la ricerca medica nel Regno Unito; [Wikipedia].

Un esempio verrà inserito nell'analisi dei dati con il modello SV nella prossima sezione.

Un modello lo si può quindi scrivere sia con l'uno che con l'altro modo, sta alla scelta dell'utente.

Brevemente, per raggiungere la distribuzione a posteriori da cui campionare i parametri, si devono definire le distribuzioni a priori, scrivere quindi la verosimiglianza dei dati (in cui ovviamente compaiono i parametri delle distribuzioni a priori). A questo punto il modello può essere compilato, si esegue il caricamento dei dati (le osservazioni) per la verosimiglianza e si generano i valori iniziali dei parametri. Terza fase è la simulazione vera e propria della distribuzione a posteriori, da cui poi si estraggono gli indicatori di interesse e la densità. Il tutto si vedrà più nel dettaglio nella sezione successiva.

Il PC utilizzato per far girare le simulazioni è un ASUS con processore INTEL Atom N570 da 1.66GHz, questo incide sui tempi di esecuzione, perciò con PC dalle prestazioni superiori ci si augura una diminuzione sensibile dei tempi.

3.3 Stochastic Volatility Model

L'implementazione del modello SV verrà fatta seguendo l'articolo [Meyer, 2000], come spiegato nel Capitolo 2. Per questo modello viene esposto come lavora il software WinBugs. Il software ottiene delle distribuzioni univariate da cui campionare i parametri dalla a posteriori congiunta,

$$\begin{aligned} \mathcal{F}(\mu, \phi, \sigma_{\eta}^2, h_1, \dots, h_n | Y) &= \mathcal{F}(\psi | Y) \\ &= \mathcal{F}(\mu, \phi, \sigma_{\eta}^2) \mathcal{F}(h_0 | \mu, \sigma_{\eta}^2) \prod_{i=1}^n \mathcal{F}(h_i | h_{i-1}, \mu, \phi, \sigma_{\eta}^2) \prod_{i=1}^n p(y_i | h_i) . \end{aligned} \quad (3.1)$$

Si pensi ai vari parametri e funzioni come a dei nodi, ogni nodo ha genitori e/o figli ed esso può essere aleatorio o deterministico; questo modo di pensare sta alla base

della rappresentazione detta *Doodle* del software WinBugs:

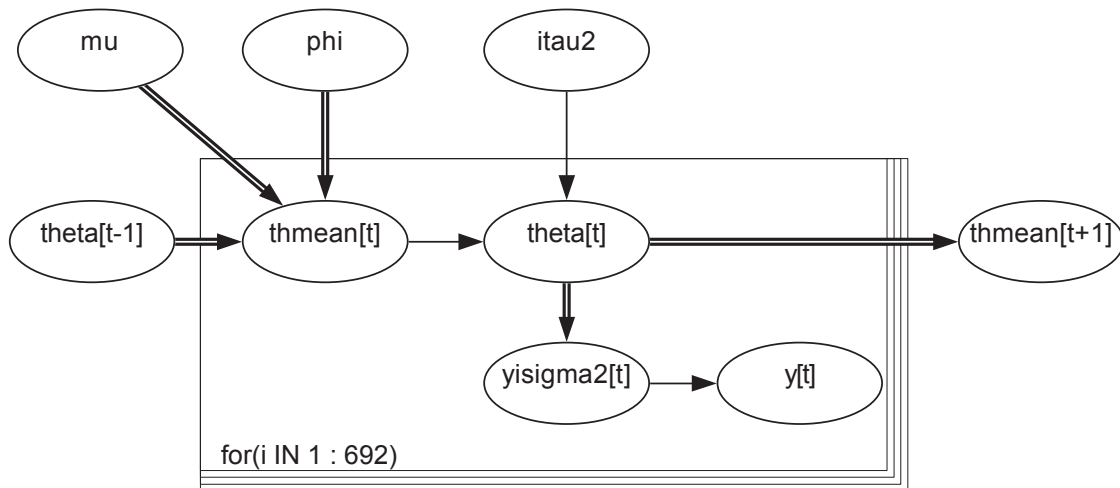


Figura 3.1: Rappresentazione Doodle per il modello SV.

I nodi *mu*, *phi* ecc... sono aleatori e si rifanno alle distribuzioni a priori dei parametri, mentre i nodi *thmean[t]* e *yisigma2[t]* sono deterministici³. La distribuzione univariata del generico nodo *v* è data da, dove *V* è l'insieme dei nodi,

$$p(v|V \setminus v) = p(v|\text{genitori}(v)) \prod_{w=\text{figli}(v)} p(w|\text{genitori}(w))$$

ad esempio, guardando la figura 3.1, la a posteriori per il nodo *mu* è data dalla sua a priori per la produttoria di *theta[t]* dati *theta[t-1]*, *phi* e *itau2* infine per la verosimiglianza di *y[t]*.

³Questo lo si nota anche nel codice sotto, se deterministico il nodo sarà preceduto dal simbolo $<-$, se aleatorio da \sim .

BUGS contiene un sistema per la scelta del miglior metodo di campionamento⁴. In breve esistono tre passaggi:

- Prima cerca una distribuzione coniugata alla a priori, nel caso la trovi campiona dalla a posteriori trovata. Questa eventualità è la più veloce, dato che la a posteriori è data dalla teoria.
- Se la densità non ha coniugata ma è una funzione log-concava, allora implementa l'algoritmo *Adaptive Rejection Sampling (ARS)*, per questo si veda l'articolo [Gilks, 1992].
- Se la densità non è nei primi due punti implementa l'algoritmo di *Metropolis-Hastings*, si veda [Jacquier, Polson, Rossi, 1994].

Per una più completa ed esaustiva spiegazione si rimanda a [WinBugs 1.4]; .

3.3.1 Implementazione in WinBugs

Di seguito viene riportato il codice del modello⁵

```
Model { for (t in 1:N) { yisigma2[t] <- 1/exp(theta[t]);
  y[t] ~ dnorm(0, yisigma2[t]);
}
# prior distributions
mu ~ dnorm(0, 0.1);
phistar ~ dbeta(20, 1.5);
itau2 ~ dgamma(2.5, 0.025);
beta <- exp(mu/2);
phi <- 2*phistar - 1;
tau <- sqrt(1/itau2);
theta0 ~ dnorm(mu, itau2);
thmean[1] <- mu + phi*(theta0 - mu);
theta[1] ~ dnorm(thmean[1], itau2);
for (t in 2:N) { thmean[t] <- mu + phi*(theta[t-1] - mu);
  theta[t] ~ dnorm(thmean[t], itau2);
}
}
```

Data list(N=692, y=c(0.0568907, -0.01014259, -0.03018903, ...))

⁴I metodi di campionamento cambiano, anche se di poco, di versione in versione.

⁵La funzione *dnorm* accetta come parametro di scala il reciproco della varianza.

Dell'indice *S&P500* si sono prese le chiusure giornaliere aggiustate e se n'è fatta la differenza logaritmica e corretti sottraendone la media⁶:

$$y_t = \log r_t - \log r_{t-1} - \frac{1}{n} \sum_{t=1}^n (\log r_t - \log r_{t-1}) .$$

Il primo ciclo *for* crea la verosimiglianza per y_t per la lunghezza del vettore delle osservazioni. Dopo di essa vengono definite le distribuzioni a priori dei parametri, e l'ultimo ciclo *for* serve a creare il vettore dei parametri h_t della volatilità.

A questo punto entrano in funzione le caratteristiche peculiari del programma *Win-Bugs*, per approfondimenti si rimanda a [WB User Manual 1.4]. In breve si compila il modello si creano i valori iniziali, *start*, dei parametri e si decide il numero di catene di markov da implementare, vedi (1.2); poi si scelgono i parametri da monitorare, si imposta quindi il numero di replicazioni per catena e si inizia la simulazione. Alla fine della simulazione il programma restituisce le statistiche per le distribuzioni a posteriori marginali dei singoli parametri, in questa fase conclusiva il periodo di *burn in* può essere scelto dall'utente.

In questo elaborato si prendono le serie risultate dalla simulazione e si utilizza il software *R* per le analisi proposte, a partire dalla valutazione del periodo di *burn in* per finire con le statistiche descrittive dei parametri, per il codice si veda l'appendice alla sessione B.3.

3.3.2 Risultati

Sono state stimate 5 catene di Markov da 30.000 estrazioni ciascuna. Tramite una funzione in *R*, creata ad hoc, si scartano le realizzazioni del periodo di *burn-in*; la funzione lavora in accordo all'idea espressa in [J. S. Dagpunar, 2007] (Cap 8, pag 161), confrontando le medie di una serie di finestre mobili. Il risultato grafico è visualizzato in figura 3.2.

⁶Il programma usato è *R - Gui* e la funzione per reperire i dati, presente nel pacchetto *TTR*, è nominata *getYahooData*

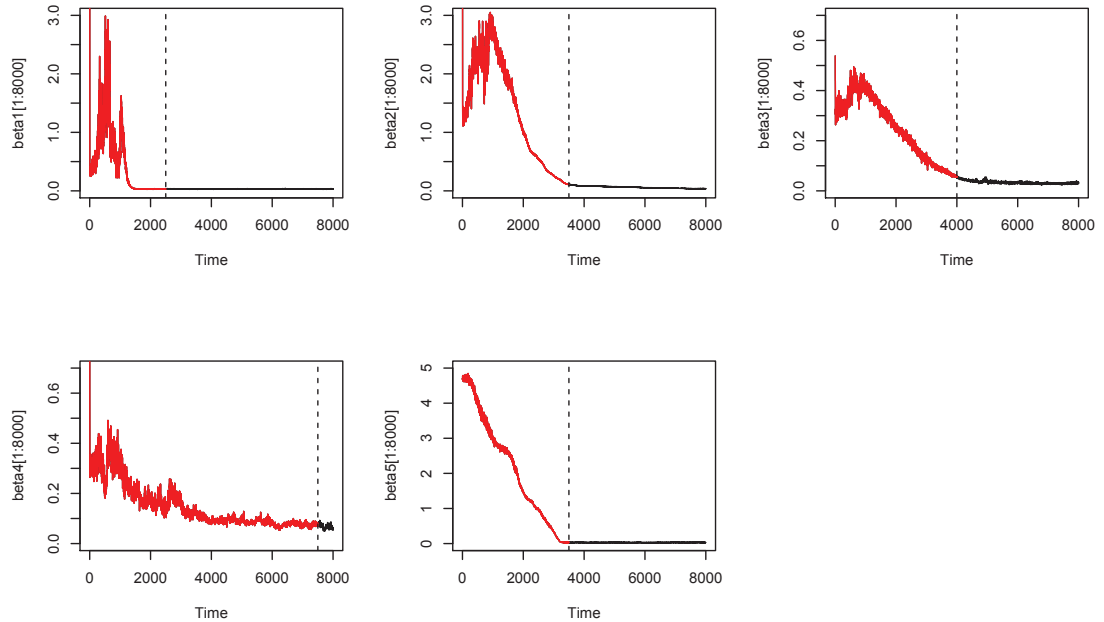


Figura 3.2: Periodo di *burn-in* per il parametro β .

Come si nota durante il periodo di *burn-in* la media cambia considerevolmente, per poi assestarsi intorno ad un dato valore, simile per tutte e 5 le serie simulate; la parte della serie in colore rosso⁷ è il periodo di *burn in* che verrà scartato. Così viene fatto per tutti e tre i parametri.

Scartato il periodo iniziale si ottengono le seguenti densità, esse sono fatte creando una serie unica dalle 5 serie in equilibrio⁸. Le densità stimate con un parametro di lisciamento di 0.01 sono in figura 3.3.

Ci si aspetta delle distribuzioni concentrate intorno ad un valore centrale sicuramente unimodali e si spera non troppo leptocurtiche.

⁷O a sinistra della linea tratteggiata verticale

⁸Ad esempio per il parametro β i periodi di *burn-in* sono stati di $\{2500, 3500, 4000, 7500, 3500\}$ avendo utilizzato finestre di 500 osservazioni, allora la serie con cui è calcolata la densità e gli indicatori descrittivi è composta da $30.000 \cdot 5 - (2500 + 3500 + 4000 + 7500 + 3500) = 129.000$ osservazioni.

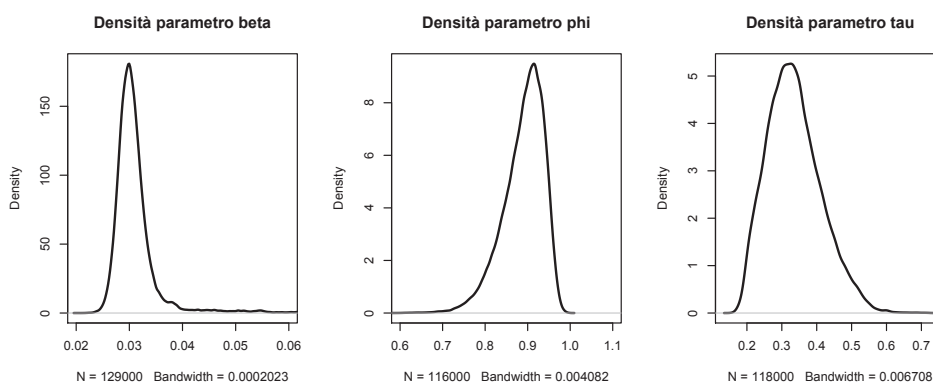


Figura 3.3: Densità lisciata dei parametri *beta phi e tau*.

Infine vengono riportate in tabella, 3.1 gli indicatori descrittivi per i tre parametri,

| Parametro | Qt 2.5% | Media | Mediana | Qt 97.5% | Scarto quad. |
|-----------|---------|---------|---------|----------|--------------|
| β | 0.02636 | 0.03189 | 0.03028 | 0.05493 | 0.00776 |
| ϕ | 0.776 | 0.8895 | 0.8977 | 0.9601 | 0.04834 |
| τ | 0.2071 | 0.3352 | 0.3283 | 0.5054 | 0.07704 |

Tabella 3.1: Statistiche descrittive dei parametri.

Modifica a priori

Si è eseguita una simulazione con distribuzioni non informative sui parametri, in sostanza le a priori sono uniformi e contengono solo i vincoli sui parametri per rendere stazionario il modello, questa volta si sono campionate serie da 40.000 estrazioni. I risultati sono riportati in tabella 3.2

| Parametro | Qt 2.5% | Media | Mediana | Qt 97.5% | Scarto quad. |
|-----------|---------|---------|---------|----------|--------------|
| β | 0.02818 | 0.03816 | 0.03304 | 0.09495 | 0.02943 |
| ϕ | 0.95704 | 0.97280 | 0.9743 | 0.9798 | 0.00620 |
| τ | 0.09426 | 0.1055 | 0.1049 | 0.1202 | 0.00667 |

Tabella 3.2: Statistiche descrittive dei parametri.

Ciò che più si nota, a parte il parametro β , è che le distribuzioni sono molto più concentrate, e il periodo di *burn in* si è rilevato molto più lungo del precedente, ad esempio per una catena del parametro ϕ ci sono volute più di 30.000 osservazioni prima che raggiungesse l'equilibrio, per questo le serie simulate sono più lunghe del caso precedente. Se come stima puntuale dei parametri si prendesse la media delle osservazioni in equilibrio ci sarebbero notevoli differenze ragionando in ambito statistico, diversamente in ambito economico β è abbastanza vicino nelle due parametrizzazioni, ed i valori di ϕ e τ non sono così drammaticamente distanti.

3.4 (G)ARCH model

Il modello GARCH viene visto ed implementato nella sua forma naturale. Anche in questo caso si utilizzano i dati con le differenze logaritmiche e corretti sottraendone la media.

Si ricorda che se il processo incorrelato a media zero ε_t segue un modello $GARCH(1, 1)$, allora esso ha varianza condizionata al passato σ_t^2 , con

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2 .$$

3.4.1 Implementazione in WinBugs

Di seguito viene riportato il codice

```

model {
  for (t in 1:N) { e[t]~dnorm(0,tau[t]) }
  for (t in 2:N) { h[t]<-omega+alpha*pow(e[t-1],2)+beta*h[t-1]
                  tau[t]<-pow(h[t],-1)
                }
  h[1]<-omega+alpha*pow(e0,2)+beta*h0
  tau[1]<-pow(h[1],-1)

  #inizializzazione dei dati
  e0~dt(0,0.01,10)
  h0~dnorm(1,0.1) I(0,10)

```

```
#A priori coefficienti GARCH
beta~dnorm(0.75,10) I(0.5,1)
alpha~dnorm(0.15,10) I(0,0.5)
omega~dnorm( 0.396,10) I(0,1)
}

Data list(N=692,e=c(0.05730085,-0.009732437,-0.02977888,....))
```

I dati analizzati sono gli stessi, si ricorda che il confronto tra i due modelli (SV e GARCH) esula dagli obiettivi dell'elaborato perciò ci si concentrerà sul semplice aspetto delle stime.

Come visto precedentemente i cicli *for* creano la verosimiglianza, che unita alle distribuzioni a priori dei parametri, formano la distribuzione a posteriori da cui verranno campionati i valori dei parametri, necessari al raggiungimento della densità a posteriori e delle statistiche descrittive.

Partendo dal presupposto che non sono stati trovati riferimenti di articoli che sostengano la scelta di a priori diverse, si inizia con delle distribuzioni a priori normali; vengono scelte con media in linea con la teoria classica dei modelli GARCH, che vuole un parametro α piuttosto piccolo ed un parametro β più grande, tali che la somma non superi l'unità, e con una varianza piccola in modo che non si discostino molto dal valore centrale.

3.4.2 Risultati

Sono state stimate 3 catene di Markov da 16.000 estrazioni ciascuna, questa diminuzione del campionamento è da imputarsi ai tempi maggiori che richiede il modello GARCH implementato⁹.

Per il resto si è proceduto in maniera analoga a quanto visto per il modello SV; sono state utilizzate le stesse funzioni in *R* per il calcolo del periodi di *burn in*, per la densità e per le statistiche descrittive, ponendo attenzione al fatto che per questo modello le serie di realizzazioni sono tre per parametro.

⁹Con queste dimensioni l'algoritmo ha richiesto circa 36-38 ore.

Il risultato grafico per il periodo di *burn in* scartato prima di considerare una realizzazione in equilibrio per il parametro β è in figura, 3.4

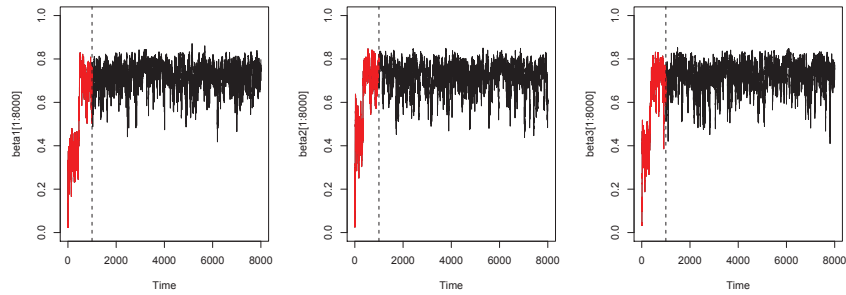


Figura 3.4: Periodo di *burn-in* per il parametro β .

Già dalla figura si nota che probabilmente la distribuzione sarà leptocurtica nella coda a sinistra.

Le densità marginali dei parametri sono in figura 3.6, la densità dei parametri α e β rispetta la loro costrizione all'interno dell'intervallo $(0, 1)$, ed essi presentano delle code pesanti in prossimità della zona più ampia del proprio spazio parametrico:

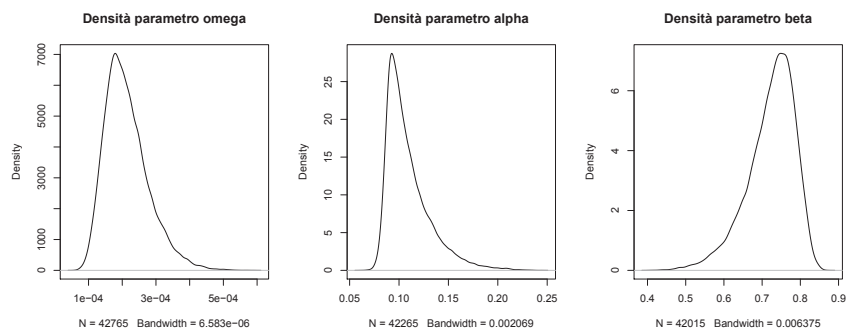


Figura 3.5: Densità lisciata dei parametri β , ϕ e τ .

Infine vengono riportate in tabella, 3.3 gli indicatori descrittivi per i tre parametri,

| Parametro | Qt 2.5% | Media | Mediana | Qt 97.5% | Scarto quad. |
|-----------|---------|----------|----------|----------|--------------|
| ω | 0.00021 | 0.000213 | 0.000203 | 0.000368 | 0.000065 |
| α | 0.08287 | 0.1093 | 0.1031 | 0.16980 | 0.022568 |
| β | 0.5708 | 0.7226 | 0.7332 | 0.8168 | 0.062855 |

Tabella 3.3: Statistiche descrittive dei parametri.

Modifica a priori

Anche in questo caso è stato fatto un confronto per valutare l'impatto della modifica della distribuzioni a priori; con il modello GARCH, per α e β , sono state scelte delle a priori che sicuramente non sfiorano le condizioni di stazionarietà per questa classe di modelli¹⁰, mentre per ω si è scelta ancora una normale ma con varianza elevata, quindi simile al concetto di distribuzione non informativa. Per soddisfare la condizione di stazionarietà si è estratto un singolo valore da una distribuzione Beta con media 0.92, che sarà il risultato della somma di α e β , poi sarà ripartito tra i due parametri tramite un uniforme; di seguito il codice usato,

```
#A propri coefficienti GARCH con vincoli di stazionarietà
x~dbeta(14,1.2)
p~dunif(0.7,0.9)
beta<-p*x
alpha<-(1-p)*x
omega~dnorm(0.396,0.001) I(0,1)
```

Questa scelta non prevedendo della a priori coniugate con la verosimiglianza del modello implica un tempo di esecuzione più alto¹¹.

Un confronto delle distribuzioni finali viene proposto nella figura 3.6

¹⁰Dato un processo GARCH(1,1), dove la varianza condizionata è scritta come

$$\sigma_t^2 = \omega + \sum_i \alpha_i \sigma_{t-i}^2 \vartheta_i^2 + \sum_j \beta_j \sigma_{t-j}^2 = \omega + \sum_i (\alpha_i \vartheta_i^2 + \beta_i) \sigma_{t-i}^2 ,$$

dove $\vartheta_i^2 \sigma_i^2 = \varepsilon_i^2$; è stazionario in senso forte se e solo se $E[\log(\alpha \vartheta^2 + \beta)] < 0$. Se $\alpha + \beta < 1$ allora la condizione è soddisfatta.

¹¹Per 15000 osservazioni ci sono volute approssimativamente 60 ore.

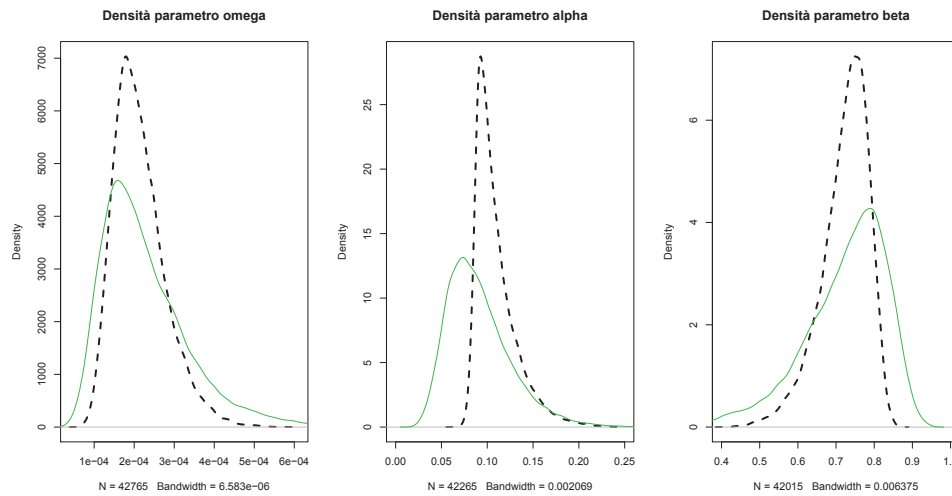


Figura 3.6: Confronto distribuzioni a priori: in nero tratteggiato a priori con vincoli di stazionarietà, in verde continuo a priori normali.

Come si può notare le differenze non sono poi molte, sempre con interpretazione più economica che statistica le stime sono simili. Interessante è notare che, in ogni caso la somma dei parametri $\alpha + \beta$ mediamente non supera uno, e che le distribuzioni con le a priori normali sono più leptocurtiche e meno concentrate. Mi sentirei quindi di concludere a favore delle a priori con i vincoli di stazionarietà: mantengono la stazionarietà, le distribuzioni sono migliori e di più facile interpretazione, anche se necessitano di un tempo di calcolo superiore.

3.5 Conclusioni

Effettuando l'analisi mi sono reso conto che avere delle precise conoscenze a priori facilitano di molto il lavoro portando a risultati più sicuri: buone distribuzioni a priori dovrebbero comportare periodi di *burn in* minori, e se esse sono delle coniugate, semplificare i calcoli. Sul modello SV mi sono basato su un articolo prendendo per buone le a priori suggerite, mentre per il modello GARCH mi sono basato sulla teoria discussa nei corsi universitari.

Tuttavia una non completa conoscenza di specifiche distribuzioni a priori sembra non compromettere del tutto le analisi fatte, anche se si tratta di due soli esempi è

conformante vedere come le distribuzioni stimate per il modello SV e GARCH si assestano su valori simili, anche cambiando le a priori.

Una possibile estensione al presente lavoro potrebbe indirizzarsi a sviluppare l'analisi o con altre distribuzioni a priori o con modelli leggermente diversi. Ad esempio: provare qualche modellistica GARCH particolare, inserire la parte ARMA in entrambi i modelli o infine togliersi dall'ipotesi di modelli gaussiani ed usare altre distribuzioni per gli errori, tipo la T-Student, come viene suggerito in [Chib, 2002].

A conclusione del mio ciclo di studi sono felice di aver letto ed approfondito un argomento per me nuovo, che fonda le proprie basi in materie da un lato interessanti e sviluppate come i processi stocastici, e dall'altro poco sviscerate come l'approccio bayesiano. Poter poi applicare il metodo studiato ad un ambito molto stimolante, come l'analisi dei dati in finanza, ha reso questo lavoro importante e a me caro.

Prerequisiti teorici

A.1 Processi stocastici e Catene di Markov

Si chiama processo stocastico una famiglia $\{X(t)\}$ di v.a. definite su uno spazio di probabilità Ω , dove $t \in T \subset \mathbb{R}^+$. I processi stocastici sono modelli matematici di fenomeni aleatori che si evolvono nel tempo.

Un processo stocastico viene detto *senza memoria* se il suo stato al tempo t dipende solo dallo stato all'istante immediatamente precedente. Dunque, la storia del sistema non è rilevante per la conoscenza dello stato attuale. Contrapposti ai processi senza memoria sono quelli *con memoria*, per i quali lo stato attuale del sistema dipende da tutta la storia del processo. Altra grande distinzione è tra i processi stocastici discreti e continui.

A.1.1 Processi stocastici discreti

Si consideri il caso in cui T è discreto e $X(t) \in E \subset \mathbb{R}$, dove $E = \{x_1, x_2, \dots, x_N\}$ è detto *spazio degli stati*; siccome E può essere messo in corrispondenza biunivoca con l'insieme $\{1, 2, \dots, N\}$, per semplicità si pone $E = \{1, 2, \dots, N\}$

Entrando nell'ambito delle catene di Markov si prenderanno in considerazione i processi stocastici Markoviani per i quali sussiste una particolare relazione per le pro-

babilità condizionali. Un processo stocastico markoviano (o di Markov) è un processo stocastico nel quale la probabilità di transizione che determina il passaggio ad uno stato di sistema dipende unicamente dallo stato di sistema immediatamente precedente e non dal come si è giunti a tale stato (in quest'ultima ipotesi si parla di processo non markoviano). Questa è detta proprietà di Markov.

Una catena di Markov è un processo stocastico discreto per cui ad ogni istante t si estrae una variabile casuale discreta. Nel caso di catena di Markov a tempo discreto si può assumere la notazione più semplice di *Catene Omogenee di Markov*, vedi B.1. Una catena omogenea di Markov è un processo markoviano nel quale la probabilità di transizione dipende unicamente dallo stato del sistema.

Definizione A.1. Si chiama *Catena di Markov (CM) discreta* con spazio degli stati discreto, una famiglia di v.a. $\{X_k\}$ indicizzate dal tempo discreto $k \in \mathbb{N}$; la cui distribuzione congiunta dipende da alcune regole elementari:

1. Il tempo T con cui sono scanditi gli eventi della catena di Markov, è l'insieme o un sottoinsieme dei numeri naturali.
2. $\forall n \in \{1, 2, \dots\}$ il supporto di X_n sia un insieme finito o al più numerabile $S = (s_1, s_2, \dots, s_j, \dots)$. S è detto insieme degli stati.
3. $\forall n$, e per ogni $(n+1)$ -pla (s_0, s_1, \dots, s_n) si assume che,

$$\mathbb{P}[X_n = s_n | X_{n-1} = s_{n-1}, X_{n-2} = s_{n-2}, \dots, X_0 = s_0] = \mathbb{P}[X_n = s_n | X_{n-1} = s_{n-1}] .$$

Si dice, che in una catena di Markov, il futuro è indipendente dal passato, condizionatamente al presente¹.

◻

¹Questo è alla base del concetto di indipendenza condizionata, $\mathbb{P}[A \cap B | C] = \mathbb{P}[A | C] \mathbb{P}[B | C]$, che non ha alcuna relazione con l'indipendenza semplice $\mathbb{P}[A \cap B] = \mathbb{P}[A] \mathbb{P}[B]$.

Si fissi con la notazione $\mathbf{p}(n) = (p_1(n), p_2(n), \dots, p_j(n), \dots)$ la distribuzione marginale della variabile aleatoria X_n dove,

$$p_j(n) = \mathbb{P}[X_n = j], \quad j \in S . \quad (\text{A.1})$$

per $n = 0$, $\mathbf{p}(0) = \lambda$ rappresenta la distribuzione iniziale della catena.

Con

$$\mathbf{P}_{ij}(m, n) = \mathbb{P}[X_n = j | X_m = i] \quad , \quad (i, j) \in S . \quad (\text{A.2})$$

si indicano le cosiddette probabilità di transizione. Risulta evidente, ed è una proprietà da verificare per confermare di essere in presenza di una catena di Markov, che

$$\sum_{j \in S} p_j(n) = 1 \quad \text{e} \quad \sum_{j \in S} p_{ij}(m, n) = 1 .$$

Riferendosi a catene di Markov omogenee nel tempo, vedi definizione in appendice, le precedenti $\mathbf{P}_{ij}(m, n)$ possono essere scritte, quando $m = n - 1$, semplicemente p_{ij} ; sarà $\mathbf{P} = \{p_{ij}, (i, j) \in S\}$ la matrice di transizione in un tempo unitario. Tale matrice è stocastica perché ogni riga rappresenta la distribuzione di probabilità con cui la catena si muove dallo stato generico i -esimo verso gli altri stati: di conseguenza la somma degli elementi per ogni riga è pari a 1². Sia inoltre $\mathbf{P}^{(k)}$ la matrice di transizione in k passi, ovvero l'elemento generico di $\mathbf{P}^{(k)}$ è $p_{ij}^{(k)} = \mathbb{P}[X_k = j | X_0 = i]$: è facile vedere, utilizzando le equazioni di Chapman e Kolmogorov che vale la relazione

$$\mathbf{P}^{(k)} = \mathbf{P}^{(k-1)}\mathbf{P} = \mathbf{P}^{(k-2)}\mathbf{P}\mathbf{P} = \dots = \mathbf{P}^k .$$

ovvero la matrice di transizione in k passi equivale alla k -esima potenza della matrice \mathbf{P} . Sempre ricorrendo all'equazione di Chapman e Kolmogorov si può vedere che la distribuzione della generica variabile aleatoria X_n di una catena di Markov omogenea dipende soltanto dalla distribuzione iniziale $\mathbf{p}(0) = \lambda$ e dalla matrice di transizione \mathbf{P} .

²Definizione di matrice stocastica.

Teorema A.1. : La legge di una catena di Markov è univocamente determinata da λ e \mathbf{P} . Due catene di Markov si dicono uguali se hanno la stessa legge finito dimensionale, cioè hanno la stessa distribuzione iniziale λ e la stessa matrice di transizione \mathbf{P} . Allora dati λ e \mathbf{P} viene definita una catena di Markov .

◻

A.1.2 Distribuzioni invarianti

Definizione A.2. DISTRIBUZIONE STAZIONARIA: Data una catena di Markov omogenea a stati discreti, una sua *distribuzione stazionaria* di probabilità (detta anche distribuzione di equilibrio) $\pi = \{\pi_1, \dots, \pi_n, \dots\}$ è una distribuzione discreta di probabilità che soddisfa le seguenti proprietà:

- $\forall i \in S, \pi_i \geq 0.$
- $\sum_{i \in S} \pi_i = 1.$
- $\forall j \in S, \sum_{i \in S} \pi_i P_{ij} = \pi_j .$

Euristicamente, una distribuzione stazionaria è una distribuzione di probabilità che si mantiene costante all'evolversi nel tempo della catena di Markov.

◻

L'importanza delle distribuzioni stazionarie per le catene di Markov omogenee a stati discreti è data dai seguenti teoremi:

Teorema A.2. ESISTENZA E UNICITÀ DELLA DISTRIBUZIONE STAZIONARIA: data una catena di Markov omogenea a stati discreti, con probabilità di transizione P_{ij} e spazio degli stati S , se la catena di Markov è irriducibile allora esiste un'unica distribuzione stazionaria π per la catena di Markov.

◻

Teorema A.3. CONVERGENZA ALLA DISTRIBUZIONE STAZIONARIA: data una catena di Markov omogenea a stati discreti, con probabilità di transizione P_{ij} e spazio degli stati S , se la catena di Markov è irriducibile ed aperiodica la distribuzione di probabilità $\tilde{\pi}_n$ al tempo t_n ³, converge alla distribuzione stazionaria π per ogni distribuzione iniziale di probabilità $\tilde{\pi}_0$ scelta. Si ha cioè che

$$\forall \tilde{\pi}_0, \forall (i, j) \in S, \lim_{n \rightarrow \infty} \sum_{i \in S} (\tilde{\pi}_0)_i (P^{(n)})_{ij} = \pi_j .$$

□

Definizione A.3. DISTRIBUZIONE INVARIANTE: Si supponga che esista una distribuzione di probabilità π su S tale che sussista la relazione,

$$\pi = \pi P \quad \text{ovvero} \quad \pi(\mathbb{I} - P) = 0 ,$$

Se inoltre accade che π rappresenti anche la distribuzione iniziale $p(0) = \lambda$, allora è facile vedere che $p(n) = \pi$ per ogni $n \in \mathbb{N}$: in tal caso la distribuzione π si dice *invariante* per la catena.

□

L'esistenza (ma non l'unicità) di una distribuzione invariante è garantita solo quando S è finito. Se esiste una tale distribuzione la catena ha un comportamento stazionario, nel senso che le distribuzioni marginali ai vari istanti sono identiche. Risulta importante determinare condizioni sufficienti a garantire la stazionarietà di una catena, ovvero l'esistenza di una distribuzione invariante.

Definizione A.4. MISURA INVARIANTE: La *misura invariante* è come la distribuzione invariante solo che non vengono posti vincoli sulla sua somma. Se la somma dei singoli elementi di una *misura invariante* è finita allora esiste una distribuzione invariante, viceversa se la somma è infinita allora non esiste la distribuzione invariante.

³Cioè $p_{ij}^{(n)} = \pi_{n,j}$ indipendentemente dallo stato i di partenza.

□

Definizione A.5. DISTRIBUZIONE REVERSIBILE: Una distribuzione di probabilità π sullo spazio degli stati S si dice *reversibile* se

$$\pi_i p_{ij} = \pi_j p_{ji}$$

per ogni coppia di stati $i, j \in S$.

□

Se una distribuzione π soddisfa la condizione di reversibilità allora essa è anche invariante per la catena. La condizione di reversibilità è più semplice da verificare rispetto all'invarianza: va però ricordato che essa rappresenta una condizione solo sufficiente, cosicché una distribuzione potrebbe essere stazionaria senza essere reversibile.

Teorema A.4. TEOREMA DI MARKOV o *Teorema Ergodico*: Una matrice di transizione regolare P , su un insieme di stati finito S ha un'unica distribuzione invariante π ; in più si ha per qualunque $j \in S$ la seguente proprietà:

$$\lim_{n \rightarrow +\infty} P_{i,j}^n = \pi_j .$$

◻

Si può sostituire la condizione di regolarità con quella di irriducibilità nel teorema di Markov, la sostituzione è lecita per il seguente lemma,

Lemma A.4.1. : Sia P una matrice di transizione, allora $Q = \sum_{n=1}^{+\infty} 1/2^{n+1} P^n$ è anch'essa matrice di transizione, per di più se P è irriducibile, allora Q ha tutti gli elementi positivi. Conseguenza se π è stazionaria per P lo è anche per Q .

□

Teorema A.5. TEOREMA DI MARKOV-KAKUTANI: Una matrice di transizione P , su un insieme di stati finito S ha almeno una distribuzione invariante.

◻

Con $\#S < +\infty$ per avere un'unica distribuzione invariante sono necessarie, *irriducibilità, positiva ricorrenza ed aperiodicità*.

A.1.3 Time Reversal e Bilancio Dettagliato

Viene definita come Time Reversal la possibilità di ribaltare nel senso temporale la catena di Markov senza che essa cambi le proprie caratteristiche; ovvero, data $(X_n)_{n>0}$, $0 < n < N$, catena di Markov (π, P) la catena $Y_n = X_{N-n}$, cioè rovesciando l'andamento del tempo, sia anch'essa catena di Markov (π, \hat{P}) .

Teorema A.6. TIME REVERSAL: Sia P irriducibile ed ammetta una distribuzione invariante π . Si supponga $(X_n)_{n>0}$, $0 < n < N$, catena di Markov (π, P) ; si definisca $Y_n = X_{N-n}$, allora anche $(Y_n)_{n>0}$ è catena di Markov (π, \hat{P}) , dove,

$$\forall i, j \in S \quad \pi_j P_{ji} = \pi_i P_{ij} \quad ,$$

inoltre \hat{P} è anch'essa irriducibile e ammette distribuzione invariante π ; $(Y_n)_{n>0}$ è detta Time Reversal di $(X_n)_{n>0}$.

◻

Visto il teorema precedente si evince che per poter rovesciare il tempo è necessaria una distribuzione invariante. La condizione è necessaria: se una catena di Markov gode del Time Reversal allora essa ha una distribuzione invariante.

Definizione A.6. EQUAZIONE DEL BILANCIO DETTAGLIATO: Una distribuzione λ e una matrice stocastica P soddisfano l'equazione del bilancio dettagliato se

$$\lambda_i P_{ij} = \lambda_j P_{ji} \quad \forall i, j \in S \quad .$$



La cosa più importante del bilancio dettagliato è che se esso è soddisfatto allora si ha sia che la distribuzione λ è invariante sia che la catena di Markov X_n è reversibile.

Una condizione sufficiente per il bilancio dettagliato è che la matrice di transizione P sia bistocastica e simmetrica, se questo avviene allora esiste una distribuzione invariante, la distribuzione uniforme sullo spazio degli stati, e la catena di Markov è reversibile.

A.2 Generazione casuale di variabili aleatorie

A.2.1 Metodi Monte Carlo

Per poter generare dei dati in accordo con le varie distribuzioni di probabilità è necessario partire da una sequenza di numeri generati casualmente; prima si genera un vettore di numeri casuali in un intervallo di interesse e poi si opera per ottenere dati distribuiti come si desidera.

Definizione A.7. Formalmente una sequenza di numeri a_1, a_2, \dots, a_n è detta sequenza casuale se $a_i \sim \mathbb{U}(0, 1)$, e se a_i è indipendente da a_j per ogni $i \neq j$.

◻

Per generare una sequenza di numeri casuali si può attingere da numeri che per definizione si conosce essere casualmente estratti da un intervallo conosciuto, come la lotteria alla TV, i secondi di un orologio nel momento in cui si guarda, il numero di un giro di roulette ecc. Questo approccio presenta dei grossi svantaggi: il primo è che nessuno di questi metodi è naturalmente interfacciato con un pc, secondo spesso si deve riprodurre la sequenza di numeri casuali molte volte, per aver vettori sufficientemente lunghi, rendendoli di fatto non casuali. Un altro approccio è quello di salvare la sequenza di numeri casuali all'interno di una periferica, e prendere quanti numeri necessari quando se ne ha bisogno; questo ha lo svantaggio che per rendere il tutto veramente casuale la stringa salvata dovrebbe essere molto lunga e questo presenta delle grosse problematiche in fase di raccolta dati⁴.

Pertanto, l'approccio preferito è quello generare numeri pseudo-casuali ogni volta che se ne ha bisogno, utilizzando una procedura deterministica. Questo permette la generazione rapida, elimina il problema di stoccaggio, e dà una sequenza riproducibile. Tuttavia, una grande cura è necessaria nella scelta di un appropriato algoritmo, per rendere la sequenza quasi-casuale.

⁴Basti pensare che per ottenere un milione di numeri sarebbero giocoforza necessarie un milione di estrazioni della lotteria.

A.2.2 Generare pseudo casualmente

Come visto sopra generare casualmente, nel senso letterale del termine, non è possibile e pertanto bisogna giungere a compromessi, utilizzando delle funzioni deterministiche. Una semplice funzione che permette di ordinare pseudo-casualmente una sequenza di numeri interi è la seguente:

$$X_i = (aX_{i-1} + c) \bmod m \quad i = \{1, 2, 3, \dots\} , \quad (\text{A.3})$$

questa equazione necessita di 4 valori dati dall'utente: a parametro moltiplicativo, c parametro additivo, m parametro di divisione⁵, e X_0 detto *seed* parametro iniziale. Sotto delle condizioni

- c e m sono relativamente primi, il più grande divisore comune è 1.
- $a - 1$ è un multiplo di ogni fattore primo di m .
- $a - 1$ è un multiplo di 4 se anche m lo è.

la procedura in (A.3) ordina casualmente i numeri da 0 a $m - 1$; quindi la si può interpretare come una estrazione senza reinserimento, di numeri interi in un intervallo prefissato. Ovviamente si cercherà il periodo più grande possibile in modo da aver una stringa di numeri casuali più lunga, così da non dover riutilizzare quelli già ottenuti. Sotto quest'ultimo aspetto è importante valutare la capacità massima di memorizzazione per un singolo numero di un computer, questo perché le operazioni aritmetiche eseguite da un pc sono limitate dalla memoria disponibile nel processore, il più delle volte la massima memoria allocabile riguarda i numero detti *double* che si spingono fino all'intero 2^{64} . Tendo presente che nel metodo di calcolo si incontrano operazioni che superano di gran lunga il periodo possono succedere problemi di *overflow*, esistono in tal senso procedure per evitare questa fastidiosa evenienza, si veda Schrage (1979).

In ultima analisi risulta evidente che dovendo campionare casualmente da una

⁵La scrittura $X \bmod m$ sta a significare il resto della divisione di X per m , inteso come il valore intero che separa X dall'ultimo multiplo di m minore o uguale a X . Esempio $14 \bmod 4 = 2$, la cosa è equivalente a $a - (b * \text{fix}(a/b))$ dove *fix* da il quoziente intero senza virgola.

$U(0,1)$ il risultato del vettore ottenuto in (A.3) deve essere diviso per il periodo, in modo da far coincidere il codominio con l'intervallo $[0,1)$, intuitivamente si capisce che più si aumenta il periodo più i problemi che derivano dall'estrazione senza reinserimento si affievoliscono; se il periodo è molto grande si potranno considerare due numeri uguali anche se diversi, perché differiranno di una quantità trascurabile. Di una certa importanza è il *seed*, visto che con la stessa partenza due vettori risulteranno uguali, se qualora questo rappresentasse un problema basta semplicemente cambiare il *seed* nelle successive generazioni; un'idea semplice è scegliere un *seme casuale* (per esempio usando l'orologio interno del computer) la speranza è che sezioni separate della sequenza non si sovrappongono. In tal caso, i risultati non saranno riproducibili.

A.2.3 Combinazione di Generatori

Combinando l'uscita da diversi generatori indipendenti si spera di aumentare il periodo ed inoltre di migliorare la casualità dell'uscita. Il primo aspetto è rilevante quando si deve lavorare con vettori di dimensioni elevate. Ad esempio un generatore sviluppato da Wichman e Hill (1982, 1984) combina l'uscita di tre generatori,

$$\begin{aligned} X_{1,i+1} &= (171X_i) \bmod 30269 \\ X_{2,i+1} &= (172X_i) \bmod 30307 \\ X_{3,i+1} &= (170X_i) \bmod 30323 \end{aligned} \quad (A.4)$$

si definisce

$$R_{i+1} = \left(\frac{X_{1,i+1}}{30269} + \frac{X_{2,i+1}}{30307} + \frac{X_{3,i+1}}{30323} \right) \bmod 1 \quad (A.5)$$

dove $X \bmod 1$ rappresenta la parte frazionaria di un reale positivo, cioè X a cui viene sottratta la parte intera di X .

La tipologia di generatori usati è di tipo moltiplicativo e hanno delle restrizioni leggermente diverse⁶, anche se comunque simili, da quelle viste in precedenza, si veda [J. S. Dagpunar, 2007] (Cap. 2.1.2). Quindi R_{i+1} rappresenta la parte frazionaria della somma di tre variabili aleatorie uniformi. Non è difficile dimostrare che

⁶Per aver un periodo completo di $m - 1$ termini.

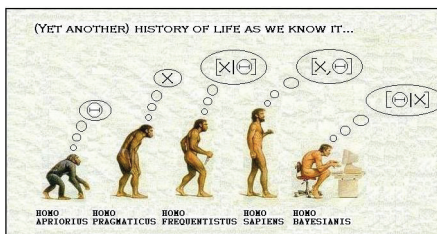
$$R_{i+1} \sim \mathbb{U}(0, 1).$$

Dal momento che i tre generatori rispettano le condizioni per aver il massimo periodo possibile, i loro periodi sono 30.268, 30.306 e 30.322 rispettivamente. Il periodo del generatore combinato è il minimo comune multiplo⁷ dei singoli periodi, che è $30268 \cdot 30306 \cdot 30324 / 4 \cong 695 \cdot 10^{12}$. Questo è un generatore affidabile anche se piuttosto lento. Viene utilizzato, per esempio, in Microsoft Excel 2003, si veda [J. S. Dagpunar, 2007] (Cap. 2.5).

⁷Si pone come divisore 4 visto che il massimo comune divisore dei tre periodi è 2.

A.3 Statistica bayesiana

Un simpatica sintesi dell'argomento:



A.3.1 Il teorema di Bayes

Dati due eventi qualsiasi F e E , la probabilità dell'intersezione $F \cap E$ si può scrivere,

$$\mathbb{P}(E \cap F) = \mathbb{P}(E|F)\mathbb{P}(F) \quad \text{oppure} \quad \mathbb{P}(E \cap F) = \mathbb{P}(F|E)\mathbb{P}(E) ,$$

facilmente si perviene alla forma più semplice del *teorema di Bayes*:

$$\mathbb{P}(F|E) = \frac{\mathbb{P}(E|F)\mathbb{P}(F)}{\mathbb{P}(E)} ,$$

essa suggerisce che la probabilità di un evento F non è una caratteristica unicamente intrinseca dell'evento, ma può essere calcolata sulla base delle informazioni a disposizione: il verificarsi di E , ad esempio, modifica la probabilità di F , e la trasforma in $\mathbb{P}(F|E)$, secondo la formula sopra.

Teorema A.7. TEOREMA DI BAYES (forma estesa): Sia E un evento contenuto in $F_1 \cup F_2 \cup \dots \cup F_k$, dove gli F_j , $j = 1, \dots, k$ sono eventi a due a due incompatibili (il verificarsi di uno esclude il verificarsi di un altro). Allora, per ognuno dei suddetti F_j vale la seguente formula

$$\mathbb{P}(F_j|E) = \frac{\mathbb{P}(E|F_j)\mathbb{P}(F_j)}{\sum_{i=1}^k \mathbb{P}(E|F_i)\mathbb{P}(F_i)} \quad (\text{A.6})$$

◻

Nella formula (A.6) il membro di sinistra prende il nome di *probabilità finale* o a *posteriori* dell'evento E ; il termine finale sta a significare: “dopo che è noto che si è verificato E ”. Il denominatore del membro di destra della (A.6), è un semplice fattore di normalizzazione; nel numeratore, invece, compaiono due quantità: la $\mathbb{P}(F_j)$ è la *probabilità a priori* dell'evento F_j e la $\mathbb{P}(E|F_j)$ rappresenta invece la verosimiglianza di F_j , ovvero la probabilità che si manifesti E quando si è verificato F_j . La formula (A.6) fornisce così un modo sintetico di valutare il grado di incertezza che si ha sul verificarsi di un evento, basandosi sia sulle informazioni a priori riguardo l'evento stesso, sia su ulteriori conoscenze sopraggiunte, magari mediante un apposito test.

A.3.2 Distribuzione a priori e distribuzione a posteriori

La base dell'inferenza bayesiana sta nell'inserire nell'indagine statistica le informazioni dei singoli soggetti che eseguono l'indagine, sul fenomeno di interesse. Quindi le informazioni pregresse possono aiutare ad ottenere conclusioni corrette.

Per inserire tali informazioni nell'analisi statistica si utilizza il teorema di Bayes: quello che serve è una formalizzazione di tutto ciò che è conosciuto sul parametro di interesse, in termini di una distribuzione di probabilità.

Definizione A.8. **PROBABILITÀ A PRIORI:** Dato θ parametro di interesse, se è possibile assumere, tramite informazioni precedenti sul parametro, che esso segua la legge di probabilità $\pi(\theta)$, allora, $\pi(\theta)$ è detta *distribuzione di probabilità a priori*, o *distribuzione iniziale*.

□

Matematicamente si tratta di trovare la distribuzione π per il parametro θ , sul dominio Θ di θ . Se Θ ha cardinalità numerabile allora si parla di legge di probabilità discreta differentemente se Θ ha cardinalità più che numerabile, è prassi utilizzare una legge di probabilità assolutamente continua. Ora θ passa da essere una quantità ignota ad essere una variabile aleatoria. La distribuzione $\pi(\theta)$ è fortemente soggettiva in base

alla conoscenza dell'individuo.

La conoscenza di θ dà la possibilità di scrivere la legge di probabilità $\pi(y|\theta)$ con una forma analoga alla (A.6),

$$\pi(\theta|y) = \frac{\pi(\theta)L(y;\theta)}{\int_{\Theta} \pi(\theta)L(y;\theta) d\theta} . \quad (\text{A.7})$$

Come nel caso della (A.6), il denominatore è un fattore di normalizzazione che non dipende da θ , indicato anche con $m(y) = \int_{\Theta} \pi(\theta)L(y;\theta) d\theta$, per sottolineare che si tratta della distribuzione marginale del vettore di variabili aleatorie Y .

La formula di Bayes produce la distribuzione finale $\pi(\theta|y)$, che rappresenta la distribuzione di probabilità del parametro θ oggetto di interesse, condizionata al risultato dell'esperimento. In essa è racchiusa tutta l'informazione su θ e su di essa ci si basa per produrre indicatori sintetici, esattamente come si è soliti fare in ambito descrittivo.

Si noti come ultimo appunto di questo paragrafo, che la formula (A.7) presentando un denominatore che non dipende da θ , si può anche scrivere nella forma:

$$\pi(\theta|y) \propto \pi(\theta)L(\theta;y) ,$$

in questo modo viene enfatizzato il fatto che essa dipende dal prodotto di due diverse fonti d'informazione.

A.3.3 Scelta della distribuzione iniziale

Di fatto, la scelta della distribuzione iniziale è il modo attraverso il quale le informazioni extrasperimentali vengono inserite nel procedimento, e l'adozione di una distribuzione iniziale rende l'analisi statistica, almeno sul piano formale, inequivocabilmente soggettiva.

L'uso di distribuzioni iniziali comporta inoltre problemi di tipo computazionale. A meno di non utilizzare forme analitiche particolari, le cosiddette distribuzioni coniugate al modello statistico utilizzato, risulta spesso difficile, se non impossibile, ottenere espressioni analitiche in forma esplicita per le distribuzioni finali.

Siano X_1, \dots, X_n , n variabili aleatorie i.i.d., da una distribuzione con vettore di parametri $\theta \in \Theta$. Si assume che la densità (caso continuo) o la distribuzione di probabilità (caso discreto), sia $p(x; \theta)$; ne segue che la funzione di verosimiglianza per θ associata a un vettore di osservazioni $(X_1 = x_1, \dots, X_n = x_n)$ è data da

$$L(\theta) \propto \prod_{i=1}^n p(x_i; \theta) .$$

Definizione A.9. DISTRIBUZIONE CONIUGATA: Una distribuzione di probabilità iniziale $\pi(\theta)$ si dice *coniugata* al modello utilizzato o, equivalentemente, alla verosimiglianza $L(\theta)$, se la forma funzionale della distribuzione iniziale e della distribuzione finale sono uguali.

□

Nota A.1. : *Normale Multivariata*. Si supponga di avere una serie di vettori $\mathbf{x}_1, \dots, \mathbf{x}_n$ da una normale multivariata con media $\boldsymbol{\mu}$ e varianza Σ ; inoltre il parametro $\boldsymbol{\mu}$ è distribuito anch'esso con una normale di media $\boldsymbol{\mu}_0$ e varianza Σ_0 . Allora la distribuzione a posteriori di $\boldsymbol{\mu}$ condizionata al vettore delle osservazioni \mathbf{x} è una normale multivariata con parametri

$$\Sigma_*^{-1} = \Sigma_0^{-1} + n\Sigma^{-1} \quad \text{e} \quad \boldsymbol{\mu}_* = \Sigma_*(\Sigma_0^{-1}\boldsymbol{\mu}_0 + n\Sigma^{-1}\bar{\mathbf{x}}) ,$$

dove $\bar{\mathbf{x}}$ è il vettore delle medie dei valori osservati, e ha distribuzione normale multivariata di media $\boldsymbol{\mu}$ e varianza Σ/n . Per la distribuzione \mathcal{N} , si veda B.2.

◇

Nota A.2. : *Chi-Quadro*. Si supponga che le realizzazioni a_1, \dots, a_n provengano da una normale con media $\mu = 0$ e varianza σ^2 ; inoltre il parametro σ^2 ha una distribuzione a priori *Chi-Quadro inversa* con ν gradi di libertà⁸. Allora la distribuzione a posteriori di σ^2 condizionata al vettore delle osservazioni a è una *Chi-Quadro inversa* con $\nu - n$ gradi di libertà:

$$\frac{\nu\lambda + \sum_{i=1}^n a_i^2}{\sigma^2} \sim \chi_{\nu+n}^2 .$$

⁸In altre parole $\lambda\nu/\sigma^2 \sim \chi_{\nu}^2$, con $\lambda > 0$.

| Modello X | Distrib. iniziale | Distr. finale | notazione |
|--------------------------------|----------------------------|---------------------------------------------------------------------------------------------------------------------|----------------------------|
| Bern(θ) | Beta(α, β) | Beta($\alpha + k, \beta + n - k$) | k=successi, n=tot prove |
| $\mathcal{N}(\mu, \sigma_0^2)$ | $\mathcal{N}(\mu, \tau^2)$ | $\mathcal{N}(\frac{\mu_0\sigma^2 + \bar{x}n\tau^2}{\sigma^2 + n\tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2})$ | σ_0^2 noto |
| Poiss(θ) | Ga(α, λ) | Ga($\alpha + n\bar{x}, \lambda + n$) | |
| Esp(θ) | Ga(α, λ) | Ga($\alpha + n, \lambda + n\bar{x}$) | |
| Unif($0, \theta$) | Pa(α, ξ) | Pa($\alpha + n, w$) | $w = \text{Max}(x_n, \xi)$ |
| Esp(θ) | IGamma(α, β) | IGamma($\alpha + 1, \beta + \sum x$) | |

Tabella A.1: Principali distribuzioni coniugate.

Per la distribuzione χ^2 si veda B.2.

◊

Nota A.3. : Si assuma $x_i | \mu, \sigma^2 \sim \mathcal{N}(\mu, \sigma^2)$ e che $\sigma^2 \sim \text{IGamma}(\alpha, \beta)$ allora la distribuzione a posteriori sarà

$$\sigma^2 | X \sim \text{IGamma}\left(\alpha + \frac{n}{2}, \beta + \frac{1}{2} \sum (x_i - \mu)\right) .$$

◊

Complementi

B.1 Catene di Markov, Teoremi e Formule

Teorema B.1. EQUAZIONE DI CHAPMAN E KOLMOGOROV: Le quantità p_j e P_{ij} , definite in (A.1) e (A.2), sono legate da una relazione importante detta equazione di Chapman e Kolmogorov: infatti si può scrivere

$$p_j(n) = p_{s_0}(m)\mathbf{P}_{s_0j}(m, n) + p_{s_1}(m)\mathbf{P}_{s_1j}(m, n) + p_{s_2}(m)\mathbf{P}_{s_2j}(m, n) + \dots$$

ovvero

$$p_j(n) = \sum_{k \in S} p_k(m)\mathbf{P}(m, n) .$$

Allo stesso modo si può dimostrare che, per un qualunque tempo u compreso tra m ed n , si ha

$$\mathbf{P}_{ij}(m, n) = \sum_{k \in S} \mathbf{P}_{ik}(m, u)\mathbf{P}_{kj}(u, n) .$$

◻

Definizione B.1. CATENA DI MARKOV OMOGENEA NEL TEMPO: Una catena di Markov si dice *omogenea nel tempo* se la quantità

$$\mathbb{P}[X_{n+1} = j | X_n = i]$$

non dipende dal tempo n .

□

Definizione B.2. **MATRICE REGOLARE:** Una matrice stocastica si dice *regolare* se $\exists n \in \mathbb{N} \setminus 0$ tale che $P_{ij}^n > 0 \forall i, j \in S$.

Se P irriducibile e $\#S < +\infty$ e se $\exists h \in S$ tale che $P_{hh} > 0$ allora P è regolare.

□

Definizione B.3. **CATENA DI MARKOV IRRIDUCIBILE:** Una catena di Markov è detta irriducibile se comunque fissati due stati i e j , esiste un intero $k = k(i, j)$ tale che dopo k unità di tempo sia possibile passare da i a j . Una catena di Markov irriducibile ammette un'unica distribuzione di probabilità stazionaria.

□

B.2 Distribuzioni Utili

Uniforme

La distribuzione uniforme di parametri α e β , ($\beta > \alpha$), ha densità,

$$f(x; \alpha, \beta) = \begin{cases} \frac{1}{\beta - \alpha} & \text{se } \alpha < x < \beta \\ 0 & \text{altrove} \end{cases} .$$

Essendo $F(a) = \int_{-\infty}^a f(x) \partial x$ ben definita, si ottiene dalla precedente la funzione di distribuzione nell'intervallo (α, β) :

$$F(a; \alpha, \beta) = \begin{cases} 0 & \text{se } a \leq \alpha \\ \frac{a - \alpha}{\beta - \alpha} & \text{se } \alpha < a < \beta \\ 1 & \text{se } a \geq \beta \end{cases} .$$

Si ricava facilmente che, se $X \sim \mathbb{U}(\alpha, \beta)$,

$$E[X] = \frac{\beta + \alpha}{2} \quad \text{e} \quad \text{Var}[X] = \frac{(\beta - \alpha)^2}{12} .$$

Normale

Se variabile casuale Y ha distribuzione normale con parametro di posizione $\mu \in \mathbb{R}$ e di scala $\sigma \in \mathbb{R}^+$, allora la sua densità vale,

$$p_Y(y; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mu)^2\right\} .$$

La funzione generatrice dei momenti è,

$$M_Y(t) = \exp\left\{t\mu + \frac{1}{2}\sigma^2 t^2\right\} .$$

Un'importante relazione collega la distribuzione normale con la distribuzione gamma: se $Z \sim \mathcal{N}(0, 1)$ e $T = Z^2$ allora $T \sim Ga(1/2, 1/2)$. Assieme alla proprietà additiva della distribuzione gamma si ha che se, Z_1, Z_2, \dots, Z_n sono variabili casuali normali *standard* indipendenti allora $\sum Z_i^2 \sim Ga(v/2, 1/2)$, quindi $\chi_v^2 \sim Ga(v/2, 1/2)$.

Le distribuzioni normali sono approssimazioni di altre distribuzioni, in particolare le distribuzioni che godono della proprietà additiva e hanno sia valore atteso che varianza finita sono approssimabili da normali. Ecco alcuni risultati:

- $Binom(n, \pi) \sim \mathcal{N}(n\pi, n\pi(1 - \pi))$ per n sufficientemente grande e $0 < \pi < 1$.
- $Poiss(\lambda) \sim \mathcal{N}(\lambda, \lambda)$ per λ sufficientemente grande, $\lambda > 10$.
- $Bineg(k, \pi) \sim \mathcal{N}(k/\pi, k(1 - \pi)/\pi^2)$ per k sufficientemente grande e $0 < \pi < 1$.
- $Ga(\alpha, \lambda) \sim \mathcal{N}(\alpha/\lambda, \alpha/\lambda^2)$ per α sufficientemente grande, $\alpha > 15$.
- $\chi_v^2 \sim \mathcal{N}(v, 2v)$ per v sufficientemente grande, $v > 30$.

Log-normale

Sia una variabile casuale Y da distribuzione normale di media μ e varianza σ^2 , definiamo poi una nuova variabile casuale X attraverso la relazione,

$$X = e^Y \quad \Leftrightarrow \quad Y = \ln X$$

Il dominio di definizione della X è il semi-asse $x > 0$ e la sua densità di probabilità è data da,

$$p_X(x; \mu, \sigma^2) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(\ln x - \mu)^2\right\}$$

questa funzione di densità si chiama *log-normale*.

La funzione generatrice dei momenti è,

$$M_X(t) = E[e^{ty}] = e^{t\mu + \frac{1}{2}t^2\sigma^2} .$$

I momenti primo e secondo sono rispettivamente: $E[X_i] = \exp\{\mu + \frac{\sigma^2}{2}\}$ e $E[X_i^2] = \exp\{2\mu + 2\sigma^2\}$. Quindi

$$E[X] = \exp\{\mu + \frac{\sigma^2}{2}\} \quad , \quad \text{Var}[X] = E[X^2] - E[X]^2 = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) .$$

Per finire si può dimostrare che il prodotto di variabili casuali log-normali ed indipendenti deve seguire una distribuzione log-normale.

t di Student

In teoria delle probabilità la distribuzione di Student, o *t di Student*, è una distribuzione di probabilità continua che governa il rapporto tra due variabili aleatorie, la prima con distribuzione normale e la seconda il cui quadrato ha distribuzione Chi-quadrato. La distribuzione di Student con n gradi di libertà è simmetrica, perché lo sono la distribuzione normale standard e la distribuzione Chi-quadrato.

Funzione di densità:

$$f(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} .$$

dove n sono i gradi di libertà.

Il valore atteso e la varianza sono rispettivamente:

$$E[T] = 0 \quad \text{se } n > 1 ; \quad \text{Var}[T] = \frac{n}{n-2} , \quad \text{se } n > 2 .$$

Si può far riferimento ad una *t di Student generalizzata* tramite una trasformazione si possono gestire il valore atteso e la varianza della distribuzione; sia $T \sim t_n$, *t di Student* con n gradi di libertà, e $X = \mu + \sigma T$, la variabile aleatoria trasformata con distribuzione *t di Student generalizzata*. La funzione di densità di X è data da

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{(n-2)\pi} \Gamma(\frac{n}{2})} \frac{1}{\sigma} \left(1 + \frac{(x-\mu)^2}{\sigma^2(n-2)}\right)^{-\frac{n+1}{2}} .$$

Il valore atteso e la varianza sono rispettivamente:

$$E[X] = \mu \text{ se } , n > 2 ; \quad \text{Var}[X] = \sigma^2 \frac{n}{n-2} , \text{ se } n > 2 .$$

Gamma

La distribuzione gamma con parametri (α, λ) , $\alpha > 0$ e $\lambda > 0$, con supporto $Y \in \mathbb{R}^+$, ha densità

$$p(y; \alpha, \lambda) = \frac{\lambda e^{-\lambda y} (\lambda y)^{\alpha-1}}{\Gamma(\alpha)} , \quad y > 0 ,$$

dove $\Gamma(\alpha)$, è la funzione gamma definita da

$$\Gamma(\alpha) = \int_0^{\infty} e^{-y} y^{\alpha-1} dy .$$

Vale la relazione ricorsiva $\Gamma(\alpha) = (\alpha - 1)!$ per valori interi di α .

Il valore atteso e la varianza della distribuzione gamma sono:

$$E[Y] = \frac{\alpha}{\lambda} ,$$

$$\text{Var}[Y] = \frac{\alpha}{\lambda^2} .$$

Teorema B.2. La distribuzione gamma con $\alpha = 1$ coincide con la distribuzione esponenziale. La distribuzione gamma di parametri $(n/2, 1/2)$ è chiamata anche distribuzione χ_n^2 , con n gradi di libertà. La distribuzione χ_n^2 è la distribuzione degli errori al quadrato in uno spazio di probabilità n -dimensionale in cui ogni dimensione gli errori hanno distribuzione normale. Se $\lambda = \frac{1}{2}$ e $\alpha = \frac{n}{2}$ allora la distribuzione gamma detta

anche χ_n^2 , coincide con la somma di n distribuzioni Z^2 con $Z \sim \mathcal{N}(0, 1)$

◻

Proprietà:

- Additività: se $X \sim Ga(\alpha_1, \lambda)$ e $Y \sim Ga(\alpha_2, \lambda)$ allora $X + Y \sim Ga(\alpha_1 + \alpha_2, \lambda)$.
- La variabile aleatoria $X \sim Ga(\alpha_1, \lambda)$ è la somma di α_1 variabili aleatorie $Ga(1, \lambda) \sim esp(\lambda)$.
- Se $X \sim Ga(\alpha, \lambda)$ e $T = Xb$ allora $T \sim Ga(\alpha, \frac{\lambda}{b})$

Inverse Gamma

La distribuzione gamma inversa a due parametri continua ha dominio sui reali positivi, ed è la distribuzione del reciproco di una variabile aleatoria distribuita secondo la distribuzione gamma. L'uso principale della distribuzione gamma inversa è in statistica bayesiana, dove serve come la a priori coniugata della varianza di una distribuzione normale¹.

La distribuzione Gamma inversa di parametri α, β ha densità:

$$p(y|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-\alpha-1} \exp^{-\frac{\beta}{y}},$$

dove α è il parametro di posizione e β quello di scala; essa ha come distribuzione cumulativa (CDF)

$$CDF = \frac{\Gamma(\alpha, \beta/y)}{\Gamma(\alpha)}.$$

I momenti della distribuzione *inverse Gamma* sono

$$E[Y] = \frac{\beta}{\alpha - 1} \quad E[Y^2] = \frac{\beta^2}{(\alpha - 1)(\alpha - 2)}$$

¹Tuttavia, è comune tra bayesiani considerare una parametrizzazione alternativa della distribuzione normale in termini di precisione, definita come il reciproco della varianza, che consente l'utilizzo diretto della distribuzione gamma come coniugata a priori. Come succede nel software WinBugs.

In pratica se $X \sim \text{Gamma}$ allora X^{-1} segue una distribuzione Gamma inversa.

Beta

La distribuzione beta, avendo supporto $[0,1]$, e grazie alla sua flessibilità, è facilmente adattabile ad un insieme molto grande di fenomeni casuali con risposte in $(0,1)$. Molti autori, indicano la distribuzione beta come la migliore scelta per la descrizione di dati proporzionali. La densità della distribuzione beta è data da

$$p(y; p, q) = \frac{1}{B(p, q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1 \quad (\text{B.1})$$

dove

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx, \quad p > 0, \quad q > 0 .$$

Al variare di p e q , la distribuzione beta assume forme differenti: se $p = q$ la distribuzione è simmetrica, se in più $p = q = 1$ la densità dà origine ad una distribuzione uniforme nell'intervallo $[0,1]$.

La funzione beta è legata alla funzione gamma, vedi sopra, tramite la seguente relazione,

$$B(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} .$$

Se Y_i è una variabile aleatoria beta di parametri p e q , allora

$$E[Y_i] = \frac{p}{p+q} ,$$

$$\text{Var}[Y_i] = \frac{pq}{(p+q)^2(p+q+1)} .$$

B.3 Codice R, per l'analisi dei dati campionati

Per recuperare i dati sono stati utilizzati i seguenti comandi,

```
rm(list=ls())
library(TTR)
```

```
simbolo="S&P500" # Standard and Poors
```

```

p0=getYahooData(symbol=simbolo , start=20100101,end=20121001 , adjust=TRUE, freq="daily")
chius<-ts(p0[, "Close"])
ret<-diff(log(chius))
n<-length(ret)
ret1<-c(ret[,1])
# Salvataggio per leggerli correttamente con WinBugs
write(ret1 , file="dati1.txt", sep = ", ", ncolumns=n)

```

Modello SV, per il modello GARCH il file è identico con i dovuti accorgimenti.

```

#####
### Lettura dei dati ###
#####
rm(list=rm())
beta1<-read.table("beta1.txt")[,2]
beta2<-read.table("beta2.txt")[,2]
beta3<-read.table("beta3.txt")[,2]
beta4<-read.table("beta4.txt")[,2]
beta5<-read.table("beta5.txt")[,2]
phi1<-read.table("phi1.txt")[,2]
phi2<-read.table("phi2.txt")[,2]
phi3<-read.table("phi3.txt")[,2]
phi4<-read.table("phi4.txt")[,2]
phi5<-read.table("phi5.txt")[,2]
tau1<-read.table("tau1.txt")[,2]
tau2<-read.table("tau2.txt")[,2]
tau3<-read.table("tau3.txt")[,2]
tau4<-read.table("tau4.txt")[,2]
tau5<-read.table("tau5.txt")[,2]

# Creazione matrice per i parametri
beta<-cbind(beta1 , beta2 , beta3 , beta4 , beta5)
phi<-cbind(phi1 , phi2 , phi3 , phi4 , phi5)
tau<-cbind(tau1 , tau2 , tau3 , tau4 , tau5)

#####
### Funzione per il calcolo del periodo di burn in ###
#####
burn_in<-function(parametro , n,m, plott=T) {
  colon<-dim(parametro)[2]
  burnin<-rep(0, colon)
  par(mfrow=c(2,3))
  for (vett in 1:colon) {
    cont<-1
    bet<-parametro[, vett]
    a<-trunc(length(bet)/n)

```

```

medie1<-rep(0,(a-1))
for (i in 0:(a-1)) {
  medie1[(i+1)]<-mean(bet[(i*n+1):((i+1)*n)]) }
media<-mean(bet)
if (mean(medie1[1:(1+m)])>media)
while (mean(medie1[cont:(cont+m)])>media) {
  cont<-cont+1 }
else
while (mean(medie1[cont:(cont+m)])<media) {
  cont<-cont+1 }
burnin[vett]<-cont
if (plott==T) {
  plot.ts(bet[1:(cont*n)]) }
return(burnin)
}
#####
### Esempio grafico del priodo di burn in per il parametro beta ###
#####
soglia<-500
buriin.beta<-burn_in(beta ,soglia ,5 ,T)
par(mfrow=c(2,3))
plot.ts(beta1[1:8000], ylim=c(0,3))
lines(beta1[1:((buriin.beta[1]+1)*soglia)], col=2)
abline(v=((buriin.beta[1]+1)*soglia), lty=2)
plot.ts(beta2[1:8000], ylim=c(0,3))
lines(beta2[1:((buriin.beta[2]+1)*soglia)], col=2)
abline(v=((buriin.beta[2]+1)*soglia), lty=2)
plot.ts(beta3[1:8000], ylim=c(0,0.7))
lines(beta3[1:((buriin.beta[3]+1)*soglia)], col=2)
abline(v=((buriin.beta[3]+1)*soglia), lty=2)
plot.ts(beta4[1:8000], ylim=c(0,0.7))
lines(beta4[1:((buriin.beta[4]+1)*soglia)], col=2)
abline(v=((buriin.beta[4]+1)*soglia), lty=2)
plot.ts(beta5[1:8000], ylim=c(0,5))
lines(beta5[1:((buriin.beta[5]+1)*soglia)], col=2)
abline(v=((buriin.beta[5]+1)*soglia), lty=2)

#####
### Calcolo periodi di burn, costruzione delle serie dei parametri in equilibrio ###
#####
buriin.phi<-burn_in(phi ,soglia ,5 ,F)
buriin.tau<-burn_in(tau ,soglia ,5 ,F)
nn<-length(beta1)
beta.serie<-c(beta1[((buriin.beta[1]+1)*soglia+1):nn],beta2[((buriin.beta[2]+1)*soglia+1):nn],
  beta3[((buriin.beta[3]+1)*soglia+1):nn],beta4[((buriin.beta[4]+1)*soglia+1):nn],

```

```

beta5 [(( buriin . beta [5]+1)* soglia +1):nn]
phi . serie <-c(phi1 [(( buriin . phi [1]+1)* soglia +1):nn] , phi2 [(( buriin . phi [2]+1)* soglia +1):nn] ,
  phi3 [(( buriin . phi [3]+1)* soglia +1):nn] , phi4 [(( buriin . phi [4]+1)* soglia +1):nn] ,
  phi5 [(( buriin . phi [5]+1)* soglia +1):nn]
tau . serie <-c(tau1 [(( buriin . tau [1]+1)* soglia +1):nn] , tau2 [(( buriin . tau [2]+1)* soglia +1):nn] ,
  tau3 [(( buriin . tau [3]+1)* soglia +1):nn] , tau4 [(( buriin . tau [4]+1)* soglia +1):nn] ,
  tau5 [(( buriin . tau [5]+1)* soglia +1):nn]

#####
### Grafico densita della distribuzione a posteriori dei parametri ###
#####
par(mfrow=c(1,3))
#par(mfrow=c(2,3))
plot(density(beta.serie),xlim=c(0.02,0.06),main="Densità parametro beta",lwd=2)
plot(density(phi.serie),xlim=c(0.6,1.1),main="Densità parametro phi",lwd=2)
plot(density(tau.serie),main="Densità parametro tau",lwd=2)
#plot.ts(beta.serie,main="Serie parametro beta")
#plot.ts(phi.serie,main="Serie parametro phi")
#plot.ts(tau.serie,main="Serie parametro tau")

#####
### Funzione per il calcolo delle statistiche descrittive ###
#####
stat<-function(serie,q1=0.025,q2=0.975){
med<-c(summary(serie)[3],summary(serie)[4])
qq1<-quantile(serie,q1)
qq2<-quantile(serie,q2)
stde<-sqrt(var(serie))
valori<-c(med[1],med[2],qq1,qq2,stde)
names(valori)<-c("Median","Mean","Qt 2.5%","Qt 97.5%","Std error")
return(valori)
}

stat(beta.serie)
stat(phi.serie)
stat(tau.serie)

```

Bibliografia

- [Andrieu, 2003] CHRISTOPHE ANDRIEU, NANDO DE FREITAS, ARNAUD DOUCET, MICHAEL I. JORDAN (2003). *An Introduction to MCMC for Machine Learning*, Kluwer Academic Publishers, 50 pages 5-43.
- [Gilks, 1992] W. R. GILKS AND P. WILD (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Journal of the Royal Statistical Society*, 41, No. 2, pp. 337-348.
- [J. S. Dagpunar, 2007] JOHN DAGPUNAR, (2007). *Simulation and Monte Carlo, With applications in finance and MCMC*, John Wiley & Sons Ltd, ISBN 13: 978-0-470-85494-5 (HB).
- [Jacquier, Polson, Rossi, 1994] ERIC JACQUIER; NICHOLAS G. POLSON; PETER E. ROSSI, (1994). Bayesian Analysis of Stochastic Volatility Models, *Journal of Business & Economic Statistics*, 12, No. 4, pp. 371-389.
- [Meyer, 2000] RENATE MEYER, JUN YU (2000). *BUGS for a Bayesian Analysis of Stochastic Volatility Models*. University of Auckland, Auckland, New Zealand.
- [Tsay, 2010] RUEY S. TSAY (2010). *Analysis of Financial Time Series*, 3rd Ed. Wiley, ISBN 987-0-470-41435-4.
- [Kim, 1998] SANGJOON KIM, NEIL SHEPHARD, SIDDHARTHA CHIB (1998). Stochastic Volatility: Likelihood inference and comparison with ARCH models. *The Review of Economics Studies Ltd*, 65, No 3, pages 361-393.

- [Sheldom Ross, 2002] SHELDON M. ROSS, (2002). Calcolo delle Probabilità, *Apogeo*, ISBN 88-503-2231-3.
- [Chib, 2002] SIDDHARTHA CHIB, FEDERICO NARDARI, NEIL SHEPHARD (2002). Markov chain Monte Carlo methods for Stochastic Volatility models. *Journal of Econometrics*, 108, pages 281-316.
- [Brooks, 1998] STEPHEN P. BROOKS (1998). Markov Chain Monte Carlo Method and Its Application, *Journal of the Royal Statistical Society*, 47, No. 1, pages 69-100.
- [Nakatsuma, 2000] TERUO NAKATSUMA (2000). Bayesian analysis of ARMA-GARCH models, A Markov chain sampling approach. *Journal of Econometrics*, 95, pages 57-69.
- [Wikipedia] WIKIPEDIA, JIMMY WALES <http://it.wikipedia.org>
- [WB User Manual 1.4] DAVID SPIEGELHALTER, ANDREW THOMAS, NICKY BEST, DAVE LUNN WinBUGS User Manual. January 2003, internet: <http://www.mrc-bsu.cam.ac.uk/bugs>
- [WinBugs 1.4] WINBUGS, BAYESIAN INFERENCE USING GIBBS SAMPLING. <http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/contents.shtml>, *MRC Biostatistics Unit, Cambridge, UK*.

Scritto in L^AT_EX.

Ringraziamenti

Desidero innanzitutto ringraziare il Prof. Marco Ferrante, per la disponibilità, competenza e serietà dimostrata nel seguirmi lungo la stesura di questo lavoro.

Ringrazio di cuore i miei genitori e mia sorella Benedetta che nell'arco di questo tempo in cui spesso mi sono assentato, dal consueto svolgersi della vita familiare, mi hanno sostenuto ed apprezzato per gli sforzi compiuti.

Un grazie anche agli amici di sempre con cui ho trascorso momenti felici e spensierati in cui condividevamo esperienze universitarie. Un grazie molto particolare alla mia ragazza Mara, ha avuto pazienza specialmente nell'ultimo periodo di stesura quando il tempo passato assieme si assottigliava sempre più.

Infine, ringrazio le cantine Soldà che mi hanno offerto possibilità di lavorare e quindi di mantenermi, almeno in parte, gli studi. Sono stati pazienti e gentili ad adattarsi alle mie esigenze di studio dimostrandosi affidabili e comprensivi.