



UNIVERSITA' DEGLI STUDI DI ROMA TRE  
FACOLTA' DI SCIENZE M.F.N.

Sintesi della Tesi di Laurea in Matematica

di

Luca Monno

# **Analisi Bayesiana di Misture di Distribuzioni $\alpha$ -stabili**

Relatore

Prof. Lea Petrella

Il Candidato

Luca Monno

Il Correlatore

Dott. Andrea Tancredi

Il Relatore

Prof.ssa Lea Petrella

ANNO ACCADEMICO 2003 - 2004

Luglio 2004

Classificazione AMS: 62F15, 65C40

Parole chiave: Inferenza Bayesiana, Modelli Mistura,  $\alpha$ -stabile, Metodi MCMC.

# Sintesi

La statistica si propone di studiare fenomeni reali sulla base dei risultati di esperimenti condotti su di essi.

Il primo passo di tale studio è la costruzione di un modello statistico, che in ambito bayesiano parametrico, si può rappresentare con una quaterna

$$(\mathcal{X}, f(x|\theta), \pi(\theta), \theta \in \Theta),$$

dove  $\mathcal{X}$  è l'insieme di tutti i possibili risultati ottenuti dall'esperimento e  $f(x|\theta)$  è una famiglia di leggi di probabilità, indicizzata dal parametro  $\theta \in \Theta \subseteq \mathbb{R}^k$ , che regola l'incertezza sull'esito dell'esperimento.

Oltre all'incertezza sul risultato dell'esperimento, occorre quindi considerare anche l'incertezza sul valore  $\theta \in \Theta$  che, attraverso il modello, genererà un particolare risultato. Secondo l'approccio bayesiano, il parametro incognito  $\theta$  è considerato una variabile aleatoria e la legge di probabilità che regola la sua incertezza è rappresentata dalla distribuzione  $\pi(\theta)$ , che per semplicità considereremo essere una densità. Essa è la sintesi delle opinioni e delle informazioni che si hanno sul parametro prima di avere osservato il risultato dell'esperimento, ed è per questo motivo che prende il nome di *distribuzione a priori*.

La costruzione del modello statistico passa quindi attraverso la scelta di una densità  $f$  che rappresenti, in senso probabilistico, il fenomeno d'interesse e

una distribuzione a priori  $\pi(\theta)$ .

In questo lavoro focalizzeremo l'attenzione sulla scelta della densità  $f$ , in particolare prenderemo in considerazione una famiglia molto ampia di densità allo scopo di costruire un modello statistico notevolmente flessibile in grado di interpretare una vasta gamma di fenomeni con caratteristiche diverse fra loro. Tale famiglia è costituita dall'insieme delle densità  $f$  rappresentabili attraverso la mistura di distribuzioni. In particolare considereremo le misture di distribuzioni che prendono il nome di  $\alpha$ -*stabili*. L'approccio che utilizzeremo nell'inferenza sarà di tipo bayesiano così nel primo Capitolo verranno introdotti i concetti fondamentali della *statistica bayesiana* e i metodi *Monte Carlo Markov Chains* (MCMC) che rappresentano uno strumento essenziale per lavorare in tale ambito.

In ambito bayesiano le informazioni che si hanno a priori sul parametro di interesse  $\theta$  contenute in  $\pi(\theta)$ , vengono aggiornate attraverso quelle provenienti dal campione osservato  $\mathbf{x} = (x_1, \dots, x_n)$  contenute nella funzione  $f(\mathbf{x}|\theta)$ , che, osservata come funzione di  $\theta$  per  $\mathbf{x}$ , dato prende il nome di *funzione di verosimiglianza* e viene indicata con  $L(\theta; \mathbf{x})$ . L'aggiornamento delle informazioni avviene attraverso la formula di Bayes

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta) d\theta} \propto L(\theta; \mathbf{x})\pi(\theta), \quad \theta \in \Theta. \quad (1)$$

in cui  $\pi(\theta|\mathbf{x})$  prende il nome di *distribuzione a posteriori* e rappresenta lo strumento inferenziale in questo contesto.

Il denominatore del Teorema di Bayes (1), che costituisce la costante di normalizzazione, altro non è che la densità marginale dei dati  $m(\mathbf{x})$ :

$$m(\mathbf{x}) = \int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta) d\theta.$$

In ambito bayesiano la distribuzione a posteriori viene utilizzata per calco-

lare le principali quantità di interesse dell'inferenza, ad esempio le medie a posteriori di funzioni di  $\theta$ :

$$E^\pi[g(\theta)|\mathbf{x}] = \int_{\Theta} g(\theta)\pi(\theta|\mathbf{x}) d\theta. \quad (2)$$

Integrali come questo risultano, nella maggior parte dei casi, impossibili da risolvere analiticamente e, per questo motivo, si ricorre spesso a metodi numerici, in particolare a quei metodi Monte Carlo basati sulle proprietà delle catene di Markov (MCMC).

Tali metodi sfruttano la particolare forma dell'integrale (2), ovvero il fatto che la funzione  $\pi(\theta|\mathbf{x})$  sia una densità di probabilità.

Le tecniche MCMC permettono di costruire una catena di Markov,  $(\theta^{(m)})_m$ , con distribuzione limite  $\pi(\theta|\mathbf{x})$  e, grazie, al *Teorema Ergodico* permettono di approssimare integrali del tipo (2) con una sommatoria:

$$\lim_{K \rightarrow \infty} \frac{1}{K} \sum_{i=1}^K g(\theta^{(i)}) \stackrel{q.s.}{=} E^\pi[g(\theta)|\mathbf{x}].$$

In questo lavoro verranno presentati due metodi MCMC che prendono il nome di *Metropolis–Hastings* e di *Gibbs Sampler*.

Il primo, introdotto da Metropolis *e altri* (1953) e poi generalizzato Hastings (1970), si applica ad un grande varietà di problemi, anche quando la distribuzione di interesse,  $\pi(\theta|\mathbf{x})$ , è conosciuta a meno di una costante moltiplicativa; esso è quindi in grado di soddisfare le esigenze tipiche del contesto bayesiano. Data  $\pi(\theta|\mathbf{x})$ , nota a meno di un fattore normalizzante, e data una densità condizionata  $q(\theta'|\theta)$ , detta *distribuzione di proposta*, l'algoritmo, noto con il nome di Metropolis–Hastings (MH), genera la catena  $(\theta^{(m)})_m$  come segue:

1. Inizializza  $\theta^{(0)}$
2. Aggiorna  $\theta^{(m+1)}$  dato  $\theta^{(m)}$  nel seguente modo:

(i) Genera  $\xi \sim q(\xi|\theta^{(m)})$

(ii) Definisci

$$\rho(\xi, \theta^{(m)}) = \min \left( \frac{\pi(\xi|\mathbf{x})q(\theta^{(m)}|\xi)}{\pi(\theta^{(m)}|\mathbf{x})q(\xi|\theta^{(m)})}, 1 \right)$$

(iii) Prendi

$$\theta^{(m+1)} = \begin{cases} \xi & \text{con probabilità } \rho(\xi, \theta^{(m)}) \\ \theta^{(m)} & \text{altrimenti.} \end{cases}$$

In pratica, al passo  $m + 1$  viene generato un candidato  $\xi$  dalla distribuzione di proposta  $q(\cdot|\theta^{(m)})$  che viene accettato o rifiutato con probabilità  $\rho(\xi, \theta^{(m)})$ . Il metodo Gibbs Sampler invece, introdotto da Geman e Geman (1984), risulta particolarmente utile quando la densità a posteriori  $\pi(\theta|\mathbf{x})$  è definita su uno spazio di dimensione elevata e qualora si conoscano quelle che prendono il nome di *distribuzioni condizionate complete*. Infatti, quando il parametro  $\theta$  può essere scritto come  $\theta = (\theta_1, \dots, \theta_k)$  basterà simulare, ad ogni iterazione dell'algoritmo, il parametro  $\theta_j$  con  $j = 1, \dots, k$  dalla distribuzione condizionata completa:

$$\theta_j \sim \pi_j(\theta_j|\theta_{-j}, \mathbf{x}) \tag{3}$$

con  $\theta_{-j} = (\theta_1, \dots, \theta_{j-1}, \theta_{j+1}, \dots, \theta_k)$ , per ottenere una catena di Markov  $(\theta^{(m)})_m$  che converge in distribuzione alla variabile aleatoria  $\theta$  con distribuzione a posteriori  $\pi(\theta|\mathbf{x})$ .

I due algoritmi presentati possono anche essere utilizzati congiuntamente, qualora non si sia in grado di simulare dalle condizionate  $\pi_j(\theta_j|\theta_{-j}, \mathbf{x})$ ; infatti in questo caso è possibile sostituire ai passi del Gibbs Sampler un singolo passo Metropolis–Hastings con distribuzione obiettivo  $\pi_j(\theta_j|\theta_{-j}, \mathbf{x})$ ; questo metodo *ibrido* viene denominato, appunto, *Gibbs Sampler con passo Metropolis–Hastings*.

Come abbiamo accennato, in questo lavoro verranno impiegate le distribuzioni  $\alpha$ -stabili ed è per questo motivo che nel secondo Capitolo tali distribuzioni verranno introdotte mostrandone definizioni e proprietà. Esse sono identificate dalla loro funzione caratteristica:

$$E(e^{i\xi X}) = \begin{cases} \exp \left\{ -\gamma^\alpha |\xi|^\alpha \left( 1 - i\beta (\text{sign } \xi) \tan \frac{\pi\alpha}{2} \right) + i\delta\xi \right\} & \text{se } \alpha \neq 1, \\ \exp \left\{ -\gamma |\xi| \left( 1 + i\beta \frac{2}{\pi} (\text{sign } \xi) \ln |\xi| \right) + i\delta\xi \right\} & \text{se } \alpha = 1, \end{cases} \quad (4)$$

dove

$$\text{sign } \xi = \begin{cases} 1 & \text{se } \xi > 0, \\ 0 & \text{se } \xi = 0, \\ -1 & \text{se } \xi < 0. \end{cases}$$

Le variabili  $\alpha$ -stabili sono indicizzate da quattro parametri:  $\alpha \in (0, 2]$ , l'indice di stabilità, che regola l'andamento delle code;  $\beta \in [-1, 1]$ , il parametro di simmetria;  $\gamma \in \mathbb{R}^+$ , il parametro di scala e infine  $\delta \in \mathbb{R}$ , il parametro di locazione e vengono indicate con il simbolo  $S(\alpha, \beta, \gamma, \delta)$ . Esse costituiscono una classe ricca di distribuzioni che include al suo interno, come casi particolari, la distribuzione Gaussiana e la distribuzione di Cauchy.

Tale classe di distribuzioni, grazie alla loro flessibilità, ottenibile attraverso la combinazione di diversi valori dei parametri, è particolarmente adatta a modellare quei fenomeni che presentano asimmetria e code pesanti ed è per questo motivo che vengono impiegate in numerosi contesti che vanno dai sistemi economici (Buckle, 1995) ai sistemi fisici (Herranz *e altri*, 2004). La particolarità di tali distribuzioni, che per diversi anni non ne ha permesso un rapido impiego empirico, è che non può essere espressa in forma esplicita la funzione di densità, oltre a non avere finiti i momenti di ordine minore di  $\alpha$  (dove  $\alpha \leq 2$ ).

Per quanto riguarda la prima problematica, osserveremo nel Capitolo 2 come essa può essere superata in un contesto bayesiano introducendo una variabile latente ed implementando un algoritmo proposto da Buckle nel 1995 con

cui è possibile simulare una successione di parametri  $(\alpha^{(m)}, \beta^{(m)}, \gamma^{(m)}, \delta^{(m)})$  con distribuzione limite la distribuzione a posteriori di questi. In particolare osserveremo come sia possibile scrivere, quando  $\alpha \neq 1$ , la densità congiunta di una coppia di variabili aleatorie  $X$  e  $Y$  tale per cui la marginale della variabile  $X$  sia di tipo  $\alpha$ -stabile utilizzando l'espressione

$$f(x, y | \alpha, \beta, \gamma, \delta) = \frac{\alpha}{|\alpha - 1|} \exp \left\{ - \left| \frac{z}{t_{\alpha, \beta}(y)} \right|^{\frac{\alpha}{\alpha-1}} \right\} \left| \frac{z}{t_{\alpha, \beta}(y)} \right|^{\frac{\alpha}{\alpha-1}} \frac{1}{\gamma |z|}, \quad (5)$$

dove

$$\begin{aligned} t_{\alpha, \beta}(y) &= \left( \frac{\sin[\pi \alpha y + \eta_{\alpha, \beta}]}{\cos \pi y} \right) \left( \frac{\cos \pi y}{\cos[\pi(\alpha - 1)y + \eta_{\alpha, \beta}]} \right)^{\frac{\alpha-1}{\alpha}}, \\ \eta_{\alpha, \beta} &= \beta K(\alpha) \pi / 2, \\ l_{\alpha, \beta} &= -\eta_{\alpha, \beta} / \pi \alpha, \\ K(\alpha) &= \min(\alpha, 2 - \alpha) \\ z &= \frac{x - \delta}{\gamma}. \end{aligned}$$

e  $\alpha \in (0, 1) \cup (1, 2]$ ,  $\beta \in [-1, 1]$ ,  $\gamma \in (0, \infty)$ ,  $\delta \in \mathbb{R}$ .

L'analisi bayesiana finale avviene così attraverso la distribuzione a posteriori  $\pi(\alpha, \beta, \gamma, \delta | \mathbf{x})$  per la quale vengono implementati gli algoritmi MCMC. In particolare verrà utilizzato il metodo Gibbs Sampler con passo Metropolis-Hastings; per rendere più maneggevoli le distribuzioni condizionate complete che servono per il Gibbs Sampler porremo

$$v = t_{\alpha, \beta}(y) \quad \text{e} \quad \phi = \frac{t_{\alpha, \beta}(y)}{x - \delta}.$$

In questo modo, dato il campione  $\mathbf{x} = (x_1, \dots, x_n)$ , l'algoritmo Gibbs Sampler può essere presentato sotto questa forma:

- 1) **Inizializza**  $\alpha^{(0)}, \beta^{(0)}, \gamma^{(0)}, \delta^{(0)}$ .
- 2) **Genera**  $y_i \sim f(y | \alpha, \beta, \gamma, \delta, x_i)$  per  $i = 1, \dots, n \dots$

3) Poni  $v_i = t_{\alpha,\beta}(y_i)$  e  $\phi_i = \frac{v_i}{x_i - \delta}$ , per  $i = 1, \dots, n$ .

4) Genera  $\alpha \sim \pi(\alpha|\beta, \gamma, \delta, \mathbf{x}, \mathbf{v})$ .

5) Genera  $\beta \sim \pi(\beta|\alpha, \gamma, \delta, \mathbf{x}, \mathbf{v})$ .

6) Genera  $\gamma \sim \pi(\gamma|\alpha, \beta, \delta, \mathbf{x}, \mathbf{v})$ .

7) Genera  $\delta \sim \pi(\delta|\alpha, \beta, \gamma, \mathbf{x}, \boldsymbol{\phi})$ .

dove le distribuzioni condizionate complete sono tali che

$$\begin{aligned}
 f(y|\alpha, \beta, \gamma, \delta, x) &\propto \exp \left\{ 1 - \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\frac{\alpha}{\alpha-1}} \right\} \left| \frac{z}{t_{\alpha,\beta}(y)} \right|^{\frac{\alpha}{\alpha-1}} \\
 \pi(\alpha|\beta, \gamma, \delta, \mathbf{x}, \mathbf{v}) &\propto \left( \frac{\alpha}{|\alpha-1|} \right)^n \exp \left\{ - \sum_{i=1}^n \left| \frac{z_i}{v_i} \right|^{\frac{\alpha}{\alpha-1}} \right\} \\
 &\quad \times \prod_{i=1}^n \left( \left| \frac{z_i}{v_i} \right|^{\frac{\alpha}{\alpha-1}} \left| \frac{dt_{\alpha,\beta}}{dy} \right|_{y:t_{\alpha,\beta}(y)=v_i}^{-1} \right) \pi(\alpha) \\
 \pi(\beta|\alpha, \gamma, \delta, \mathbf{x}, \mathbf{v}) &\propto \prod_{i=1}^n \left| \frac{dt_{\alpha,\beta}}{dy} \right|_{y:t_{\alpha,\beta}(y)=v_i}^{-1} \pi(\beta). \\
 \pi(\delta|\alpha, \beta, \gamma, \mathbf{x}, \boldsymbol{\phi}) &\propto \prod_{i=1}^n \left| \frac{dt_{\alpha,\beta}}{dy} \right|_{y:t_{\alpha,\beta}(y)=\phi_i(x_i-\delta)}^{-1} \pi(\delta).
 \end{aligned}$$

mentre la distribuzione  $\pi(\gamma|\alpha, \beta, \delta, \mathbf{x}, \mathbf{v})$  è tale che

$$\gamma^{\frac{\alpha}{\alpha-1}} \sim GI \left( n + a + \frac{1}{\alpha}, b + \sum_{i=1}^n \left| \frac{x_i - \delta}{v_i} \right|^{\frac{\alpha}{\alpha-1}} \right)$$

dove con  $GI(a, b)$  indichiamo una distribuzione Gamma Inversa di parametri  $a, b$ . Nella Tesi abbiamo implementato quest'algoritmo e lo abbiamo provato su un insieme di dati reali già studiato da Buckle (1995), costituito dai profitti giornalieri delle azioni dell'Abbey National tra il 31/07/91 e il 08/10/91 e i risultati ottenuti sono sintetizzati dal grafico 1. Notiamo che l'istogramma dei profitti mostra un'evidente asimmetria e una coda destra sufficientemente

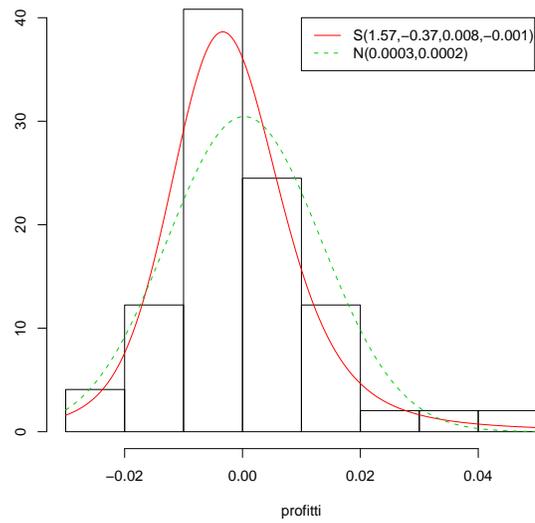


Figura 1: Istogramma dei profitti. La linea continua mostra una densità  $\alpha$ -stabile con parametri le medie a posteriori dei parametri simulati, mentre la linea tratteggiata mostra una densità normale con media la media empirica del campione e varianza la varianza empirica.

“pesante”, caratteristiche non associabili ad un modello, per esempio, di tipo normale. Osservando il grafico, infatti, è possibile osservare che la densità normale, rappresentata dalla linea tratteggiata, non è in grado di cogliere le particolarità del fenomeno, caratteristiche che invece risultano essere ben modellate dalla densità  $\alpha$ -stabile, rappresentata dalla linea continua.

Spesso i fenomeni oggetto di interesse sono caratterizzati dalla presenza, oltre che di asimmetria e code pesanti, di più picchi, ovvero di multimodalità: si pensi, ad esempio, alla distribuzione di serie storiche finanziarie in cui spesso si osservano diverse mode causate dall'evoluzione del processo sottostante che segue diversi regimi; oppure si pensi a tutti quei casi in cui si studiano popolazioni eterogenee, ovvero popolazioni che hanno al loro interno gruppi differenti, ognuno dei quali caratterizzato da una propria distribuzione. Per questo motivo introduciamo i modelli mistura. Le misture finite di distribuzioni forniscono un approccio matematico alla modellizzazione statistica di una vasta varietà di fenomeni (si veda per esempio Robert (1996), McLachlan e Peel (2000)). Grazie alla loro utilità in qualità di metodi estremamente flessibili, i modelli mistura continuano a ricevere una crescente attenzione attraverso gli anni, sia dal punto di vista pratico che da quello teorico, anche se le loro prime applicazioni risalgono a più di cento di anni fa (Pearson, 1894). In verità, nell'ultimo decennio la quantità e la potenzialità delle applicazioni delle misture sono cresciute enormemente. I campi in cui le misture sono state applicate con successo includono l'astronomia, la biologia, la genetica, la medicina, l'economia, l'ingegneria e numerosi altri campi all'interno delle scienze biologiche, fisiche e sociali. In letteratura, in particolare, per le analisi inferenziali si è fatto principalmente ricorso a misture finite di distribuzioni normali (Richardson e Green, 1997), soprattutto perché si è osservato che queste sono in grado di approssimare in modo sufficientemente preciso una

qualsiasi distribuzione continua. In questo lavoro intendiamo generalizzare questa modellistica facendo ricorso alle misture di distribuzioni  $\alpha$ -stabili, ricordando che le distribuzioni normali sono un caso particolare di queste ultime. In questo modo uniremo la flessibilità delle misture alla flessibilità propria delle distribuzioni  $\alpha$ -stabili.

Per definire una distribuzione mistura consideriamo  $k$  funzioni di densità  $f_j(x)$ , con  $j = 1, \dots, k$ , e  $k$  quantità  $p_j$  tali che

$$0 \leq p_j \leq 1 \quad \sum_{j=1}^k p_j = 1 \quad \text{con } j = 1, \dots, k$$

allora la funzione

$$f(x) = \sum_{j=1}^k p_j f_j(x) \tag{6}$$

è una densità che prende il nome di *mistura*, le quantità  $p_j$  sono dette *pesi* mentre le funzioni  $f_j$  sono dette *densità componenti*.

Per rendere più chiara tale definizione possiamo immaginare una mistura come la densità di una variabile aleatoria  $X$  estratta da una popolazione eterogenea formata da  $k$  gruppi con proporzioni  $p_j$ , ognuno caratterizzato da una densità  $f_j(x)$ . Nell'approccio che sviluppiamo in questo lavoro considereremo un numero finito e noto di densità componenti. Ovviamente generalizzazioni a tale approccio sono possibili ma non verranno prese in considerazione in questa sede (si veda per esempio Green (1995)). Inoltre supporremo che le densità  $f_j(x)$  appartengano alla stessa famiglia parametrica rappresentata, nel nostro caso, dalle distribuzioni  $\alpha$ -stabili e quindi  $f_j(x) = f(x|\theta_j)$  con  $\theta_j = (\alpha_j, \beta_j, \gamma_j, \delta_j)$ , per  $j = 1, \dots, k$ .

Come osservato precedentemente, lo schema bayesiano prevede l'utilizzo del Teorema di Bayes (1) per costruire la densità a posteriori dei parametri, rappresentati, nel nostro caso, dal vettore  $(\boldsymbol{\theta}, \mathbf{p})$ , dove  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  e

$\mathbf{p} = (p_1, \dots, p_k)$ , attraverso la formula

$$\pi(\boldsymbol{\theta}, \mathbf{p}|\mathbf{x}) \propto L(\boldsymbol{\theta}, \mathbf{p}; \mathbf{x}) \cdot \pi(\boldsymbol{\theta}, \mathbf{p}), \quad (7)$$

dove  $\pi(\boldsymbol{\theta}, \mathbf{p})$  è la densità a priori di  $(\boldsymbol{\theta}, \mathbf{p})$  e

$$L(\boldsymbol{\theta}, \mathbf{p}; \mathbf{x}) = f(\mathbf{x}|\boldsymbol{\theta}, \mathbf{p}) = \prod_{i=1}^n \sum_{j=1}^k p_j f(x_i|\theta_j) \quad (8)$$

Tale espressione, a causa principalmente della struttura della funzione di verosimiglianza (8), risulta essere molto complicata da maneggiare, anche in un contesto bayesiano in cui è possibile far ricorso ai metodi MCMC. Infatti la struttura moltiplicativa della funzione di verosimiglianza prevede l'espansione in  $k^n$  termini, che rappresenta un numero sufficientemente elevato anche nel caso in cui  $k$  ed  $n$  siono piccoli. Per ovviare a tale inconveniente introduciamo  $n$  variabili aleatorie  $\boldsymbol{\nu}_i$ , per  $i = 1, \dots, n$ , indipendenti che prendono il nome di *variabili di allocazione*. Ogni variabile aleatoria  $\boldsymbol{\nu}_i$  è rappresentata da un vettore di  $k$  componenti definite come 1 o 0 a seconda che la la variabile  $X_i$  sia stata estratta dal  $j$ -esimo gruppo o meno. Essendo  $p_j$  la probabilità di scegliere il  $j$ -esimo gruppo, possiamo scrivere

$$P(\boldsymbol{\nu}_i = (\nu_{i1}, \dots, \nu_{ik})|\mathbf{p}) = \prod_{j=1}^k p_j^{\nu_{ij}}, \quad (9)$$

ovvero  $P(\nu_{ij} = 1) = p_j$ , in quanto della produttoria sopravvive un unico termine: quello in cui  $\nu_{ij} = 1$ . In questo modo la densità di  $\mathbf{x}$  condizionatamente a  $\boldsymbol{\theta}, \mathbf{p}$  e  $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_n)$  diventa:

$$f(\mathbf{x}|\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\nu}) = \prod_{i=1}^n \prod_{j=1}^k [f(x_i|\theta_j)]^{\nu_{ij}} \quad (10)$$

Con l'introduzione della variabile di allocazione  $\boldsymbol{\nu}$  scompare la sommatoria della formula (8) rendendo tale densità più trattabile.

Per poter applicare tale ragionamento alle distribuzioni  $\alpha$ -stabili, a causa della mancanza di una forma esplicita per la densità di tali variabili, occorre utilizzare la variabile aleatoria  $Y$ , già introdotta in precedenza, che consente di scrivere la densità congiunta  $f(x, y)$ , data dalla formula (5), la cui marginale di  $x$  è  $\alpha$ -stabile. In questo modo, per poter calcolare ed utilizzare la distribuzione a posteriori  $\pi(\mathbf{y}, \boldsymbol{\nu}, \boldsymbol{\theta}, \mathbf{p}|\mathbf{x})$  occorre calcolare le distribuzioni condizionate complete:

$$\begin{aligned} &\pi(\mathbf{p}|\boldsymbol{\theta}, \boldsymbol{\nu}, \mathbf{x}, \mathbf{y}), \\ &\pi(\theta_j|\boldsymbol{\theta}_{-j}, \mathbf{p}, \boldsymbol{\nu}, \mathbf{x}, \mathbf{y}) \quad \text{per } j = 1, \dots, k, \\ &f(\mathbf{y}|\boldsymbol{\theta}, \mathbf{p}, \boldsymbol{\nu}, \mathbf{x}), \\ &\pi(\boldsymbol{\nu}|\boldsymbol{\theta}, \mathbf{p}, \mathbf{x}, \mathbf{y}). \end{aligned}$$

delle variabili aleatorie in gioco costruendo un algoritmo MCMC, basato sul metodo del Gibbs Sampler, per stimare i parametri della mistura, rappresentati sia dai parametri  $\theta_j$  delle singole componenti sia dai pesi  $p_j$ , per  $j = 1, \dots, k$ . Nel Capitolo 3 verranno calcolate dal punto di vista teorico le distribuzioni condizionate necessarie.

L'algoritmo implementato può essere quindi riassunto nei seguenti passi:

- 1) Inizializziamo  $\boldsymbol{\theta}^{(0)}$ ,  $\mathbf{p}^{(0)}$ ,  $\boldsymbol{\nu}^{(0)}$ .
- 2) Per  $i = 1, \dots, n$ , generiamo  $y_i$  da

$$f(y_i|\boldsymbol{\theta}, \boldsymbol{\nu}, x_i) \propto \prod_{j=1}^k [f(y_i|\theta_j, x_i)]^{\nu_{ij}}$$

- 3) Per  $i = 1, \dots, n$ , poni  $v_i = t_{\alpha_j, \beta_j}(y_i)$  e  $\phi_i = \frac{v_i}{x_i - \delta_j}$  con  $j$  tale che  $\nu_{ij} = 1$ .

4) Per  $i = 1, \dots, n$ , generiamo

$$\boldsymbol{\nu}_i \sim \text{Mult}_k(1, \hat{\mathbf{p}}_i)$$

con

$$\hat{\mathbf{p}}_i = (\hat{p}_{i1}, \dots, \hat{p}_{ik}) \quad \text{e} \quad \hat{p}_{ij} = \frac{p_j f(x_i, y_i | \theta_j)}{\sum_{j=1}^k p_j f(x_i, y_i | \theta_j)}.$$

5) Generiamo  $\mathbf{p} \sim \text{Diric}(n_1 + \delta_1, \dots, n_k + \delta_k)$  con

$$n_j = \sum_{i=1}^n \nu_{ij}.$$

6) Per  $j = 1, \dots, k$ , generiamo  $\theta_j$ :

a) Genera  $\alpha_j \sim \pi(\alpha | \beta_j, \gamma_j, \delta_j, \mathbf{x}, \mathbf{v}, \boldsymbol{\nu})$ .

b) Genera  $\beta_j \sim \pi(\beta | \alpha_j, \gamma_j, \delta_j, \mathbf{x}, \mathbf{v}, \boldsymbol{\nu})$ .

c) Genera  $\gamma_j \sim \pi(\gamma | \alpha_j, \beta_j, \delta_j, \mathbf{x}, \mathbf{v}, \boldsymbol{\nu})$ .

d) Genera  $\delta_j \sim \pi(\delta | \alpha_j, \beta_j, \gamma_j, \mathbf{x}, \boldsymbol{\nu})$ .

Per i passi 2, 3 e 6 dell'algorithmo per la simulazione dei parametri di una mistura è possibile ricorrere ai metodi illustrati nell'algorithmo per le distribuzioni  $\alpha$ -stabili, ovvero per la simulazione di  $y_i$  viene utilizzato il metodo di *Accettazione/Rifiuto* per i parametri  $\alpha_j$ ,  $\beta_j$  e  $\delta_j$  un passo *Metropolis-Hastings*, mentre per  $\gamma_j$  è possibile scegliere una distribuzione a priori in modo tale che  $\gamma^{\frac{\alpha}{\alpha-1}} \sim GI\left(n + a + \frac{1}{\alpha}, b + \sum_{i=1}^n \left| \frac{x_i - \delta}{v_i} \right|^{\frac{\alpha}{\alpha-1}}\right)$ . L'unica particolarità è data dal fatto che condizionando anche alla variabile aleatoria  $\boldsymbol{\nu}$ , i parametri  $\theta_j = (\alpha_j, \beta_j, \gamma_j, \delta_j)$  della  $j$ -esima componente vengono simulati esclusivamente in base alle osservazioni  $x_i$  che sono supposte provenire dalla  $j$ -esima componente della mistura e quindi non in base a tutto il vettore  $\mathbf{x} = (x_1, \dots, x_n)$ . Per simulare il vettore  $\boldsymbol{\nu} = (\boldsymbol{\nu}_1, \dots, \boldsymbol{\nu}_n)$  è possibile dimostrare, dal punto di vista teorico, che è sufficiente simulare separatamente

ogni  $\nu_i$  da una distribuzione Multinomiale con pesi  $\hat{p}_i$ . Scegliendo, inoltre, per il vettore dei pesi  $\mathbf{p}$  una distribuzione a priori Dirichlet di parametri  $(\delta_1, \dots, \delta_k)$  è possibile dimostrare che la distribuzione condizionata completa per  $\mathbf{p}$  è una distribuzione Dirichlet di parametri  $(n_1 + \delta_1, \dots, n_k + \delta_k)$  dove  $n_j = \sum_{i=1}^n \nu_{ij}$  rappresenta il numero di osservazioni allocate alla  $j$ -esima componente.

Per provare la bontà dell’algoritmo costruito abbiamo simulato un campione di numerosità 200 da una mistura di due distribuzioni  $\alpha$ -stabili con densità  $S(1.6, 0.3, 1, 0)$  e  $S(1.3, -0.7, 3, 40)$  e con pesi 0.3 e 0.7, ovvero da una densità del tipo:

$$0.3S(1.6, 0.3, 1, 0) + 0.7S(1.3, -0.7, 3, 40).$$

Per i parametri  $\alpha$  e  $\beta$  delle due componenti abbiamo scelto distribuzioni a priori uniformi rispettivamente nell’intervallo  $(1, 2]$  e  $[-1, 1]$ , per i parametri  $\gamma_1$  e  $\gamma_2$  una distribuzione coniugata, mentre per  $\delta_1$  e  $\delta_2$  una distribuzione normale con una varianza elevata. Per la simulazione dei parametri  $\alpha$ ,  $\beta$  e  $\delta$  delle densità componenti la mistura abbiamo utilizzato un passo *Random Walk Metropolis–Hastings*, ovvero un passo Metropolis–Hastings con distribuzione di proposta simmetrica centrata nel precedente valore del parametro simulato; in particolare la distribuzione utilizzata è una distribuzione normale. Come esempio riportiamo nella figura 2 i risultati ottenuti per gli indici di stabilità delle due componenti. Nella figura vengono mostrate le tracce delle simulazioni e gli istogrammi delle densità a posteriori dei parametri  $\alpha_1$  e  $\alpha_2$ .

Nel modello finora considerato abbiamo supposto che le osservazioni  $x_i$  provenienti dalle diverse densità della mistura siano indipendenti. Per questo approccio è possibile un ulteriore sviluppo che permette di tenere conto anche di una eventuale dipendenza tra le variabili aleatorie  $X_i$ . In particolare

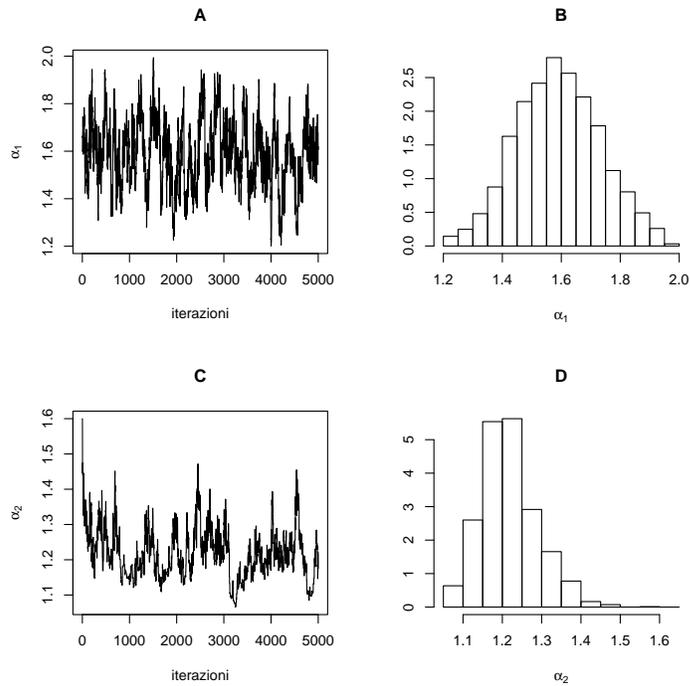


Figura 2: A. Simulazioni del parametro  $\alpha_1$ . B. Istogramma della distribuzione a posteriori condizionata di  $\alpha_1$ . C. Simulazioni del parametro  $\alpha_2$ . D. Istogramma della distribuzione a posteriori condizionata di  $\alpha_2$ .

introdurremo una dipendenza di tipo Markoviano nelle variabili di allocazione  $\nu_i$  considerate in precedenza, che si rifletterà in una dipendenza nelle osservazioni. Nell'ultimo capitolo di questa Tesi generalizzeremo il concetto di misture di distribuzioni  $\alpha$ -stabili introdotte nel Capitolo 3 al caso in cui si introduce la dipendenza suddetta. In particolare verranno poste le basi metodologiche per tale impostazione, conosciuta in letteratura con il nome di *Hidden Markov Model* (Robert e altri (1993), Robert e altri (2000)). Il nome "Hidden Markov Model" deriva dal fatto che le allocazioni  $\nu_i$  non sono osservabili e costituiscono una catena di Markov. Un esempio che rende chiara l'idea dei fenomeni che si intendono studiare con tale modellistica è dato dalla figura 3.

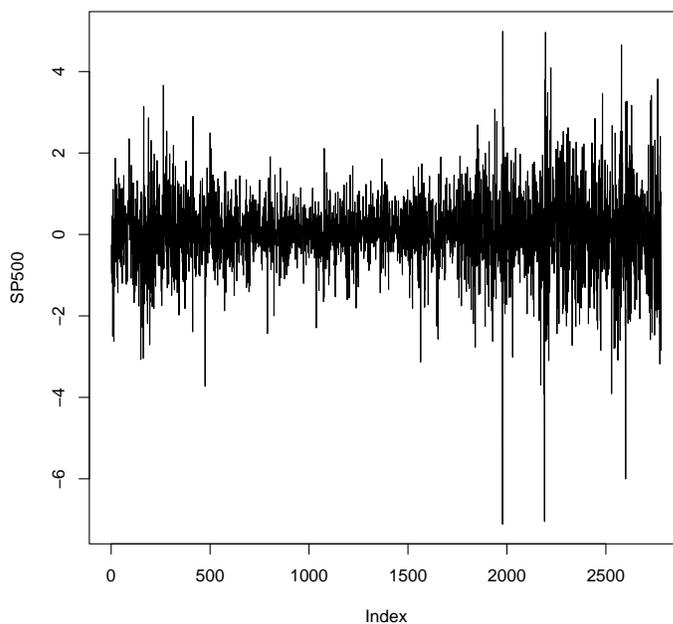


Figura 3:

Questi modelli ben si adattano ad interpretare dati osservati nel tempo che presentano volatilità o differenti regimi di volatilità; in particolare le distribuzioni  $\alpha$ -stabili sono particolarmente adatte per interpretare tale volatilità. Per formalizzare brevemente tali modelli introduciamo una nuova variabile aleatoria  $s_i$ , che, come  $\nu_i$ , indica il gruppo dal quale si estrae l'osservazione  $x_i$ :

$$s_i = \prod_{j=1}^k j^{\nu_{ij}}.$$

La variabile  $s_i$  assume valori nell'insieme  $\{1, \dots, k\}$ , in quanto

$$s_i = j \Leftrightarrow \nu_{ij} = 1.$$

Utilizzeremo, d'ora in poi,  $s_i$  in luogo di  $\nu_i$  e la chiameremo, comunque, variabile di allocazione. Per enfatizzare la dipendenza temporale delle allocazioni, e quindi delle osservazioni, utilizzeremo, al posto del solito indice  $i$ , l'indice  $t$ .

Assumendo quindi, come già accennato, che la successione  $s_t$  sia una catena di Markov con matrice di transizione  $\mathbb{P} = (p_{ij})$ , per  $i, j = 1, \dots, k$ , il modello si può scrivere sotto questa forma:

$$P(s_t = j | s_{t-1} = i, \mathbb{P}) = p_{ij}. \quad (11)$$

$$X_t | \boldsymbol{\theta}, s_t = j \sim f(x | \theta_j). \quad (12)$$

dove  $k$  rappresenta il numero di componenti o regimi,  $f(x | \theta_j)$  la densità di una variabile aleatoria  $\alpha$ -stabile,  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)$  e  $\theta_j = (\alpha_j, \beta_j, \gamma_j, \delta_j)$  il vettore dei parametri del  $j$ -esimo gruppo.

Per effettuare inferenza bayesiana sulle quantità incognite di questo modello rappresentate dall'insieme di tutti i parametri  $\boldsymbol{\theta}$  e dalla matrice di transizione  $\mathbb{P}$  è possibile implementare un algoritmo per costruire una successione

$(\boldsymbol{\theta}^{(m)}, \mathbb{P}^{(m)})_{m \geq 1}$  che converge in distribuzione alla densità  $\pi(\boldsymbol{\theta}, \mathbb{P} | \mathbf{x})$ . Per fare questo occorre costruire una catena di Markov  $(\boldsymbol{\theta}^{(m)}, \mathbb{P}^{(m)}, \mathbf{s}^{(m)}, \mathbf{y}^{(m)})_{m \geq 1}$  utilizzando il metodo Gibbs Sampler e quindi simulando iterativamente dalle distribuzioni condizionate complete:

$$\begin{aligned} \pi(s_t | \mathbf{s}_{-t}, \mathbb{P}, \boldsymbol{\theta}, \mathbf{y}, \mathbf{x}) & \text{ per } t = 2, \dots, n, \\ \pi(\mathbf{p}_i | \mathbf{p}_{-i}, \mathbf{s}, \boldsymbol{\theta}, \mathbf{y}, \mathbf{x}) & \text{ per } i = 1, \dots, k, \\ \pi(\theta_j | \boldsymbol{\theta}_{-j}, \mathbf{s}, \mathbb{P}, \mathbf{y}, \mathbf{x}) & \text{ per } j = 1, \dots, k, \\ f(y_t | \mathbf{y}_{-t}, \mathbf{s}, \mathbb{P}, \boldsymbol{\theta}, \mathbf{x}) & \text{ per } t = 1, \dots, n. \end{aligned}$$

Nell'ultimo capitolo verranno calcolate, dal punto di vista teorico, tali distribuzioni così l'algoritmo può essere riassunto dai seguenti passi:

1) Inizializziamo  $\boldsymbol{\theta}^{(0)}, \mathbb{P}^{(0)}, \mathbf{s}^{(0)}$ .

2) Per  $t = 1, \dots, n$ , generiamo  $y_t$  da

$$f(y_t | \boldsymbol{\theta}, \mathbf{s}, x_t) \propto f(y_t | \theta_{s_t}, x_t)$$

3) Per  $y = 1, \dots, n$ , poni  $v_t = t_{\alpha_{s_t}, \beta_{s_t}}(y_t)$  e  $\phi_t = \frac{v_t}{x_t - \delta_{s_t}}$ .

4) Per  $t = 2, \dots, n$ , generiamo

$$s_t \sim \text{Mult}_k(1, \hat{\mathbf{p}}_t)$$

con

$$\hat{\mathbf{p}}_t = (\hat{p}_{t1}, \dots, \hat{p}_{tk}) \quad \text{e} \quad \hat{p}_{tj} = \frac{p_{s_{t-1}j} f(x_t, y_t | \theta_{s_t}) p_{j s_{t+1}}}{\sum_{j=1}^k p_{s_{t-1}j} f(x_t, y_t | \theta_{s_t}) p_{j s_{t+1}}}$$

5) Generiamo  $\mathbf{p}_i \sim \text{Diric}(n_{i1} + \delta_1, \dots, n_{ik} + \delta_k)$  con

$$n_{ij} = \sum_{t=2}^n I_{\{s_{t-1}=i, s_t=j\}}$$

6) Per  $j = 1, \dots, k$ , generiamo  $\theta_j$ :

a) Genera  $\alpha_j \sim \pi(\alpha | \beta_j, \gamma_j, \delta_j, \mathbf{x}, \mathbf{v}, \mathbf{s})$ .

b) Genera  $\beta_j \sim \pi(\beta | \alpha_j, \gamma_j, \delta_j, \mathbf{x}, \mathbf{v}, \mathbf{s})$ .

c) Genera  $\gamma_j \sim \pi(\gamma | \alpha_j, \beta_j, \delta_j, \mathbf{x}, \mathbf{v}, \mathbf{s})$ .

d) Genera  $\delta_j \sim \pi(\delta | \alpha_j, \beta_j, \gamma_j, \mathbf{x}, \phi, \mathbf{s})$ .

Quest'algoritmo è ancora in fase di sviluppo e i risultati sono oggetto di indagine attuale.

# Bibliografia

- Buckle D. J. (1995). Bayesian inference for stable distributions. *J. Amer. Statist. Assoc.*, **90**(430), 605–613.
- Geman S.; Geman D. (1984). Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721–741.
- Green P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**(4), 711–732.
- Hastings W. (1970). Monte carlo smapling usings markov chains and their applications. *Biometrika*, **53**, 97–109.
- Herranz D.; Kuruoglu E.; Toffolatti L. (2004). An  $\alpha$ -stable approach to the study of the  $P(D)$  distribution of unresolved point sources in CMB sky map. *Astronomy & Astrophysics manuscript*, (0858).
- McLachlan G.; Peel D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York.
- Metropolis N.; Rosenbluth A.; Rosenbluth M.; Teller A.; Teller E. (1953). Equations of state calculations by fast computing machines. *J.Chem.Phys.*, **21**, 1087–1092.

- Pearson (1894). Contributions to the theory of mathematical evolution. *Phil. Trans. of the Roy. Soc. of London A*, **186**, 71–110.
- Richardson S.; Green P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *J. Roy. Statist. Soc. Ser. B*, **59**(4), 731–792.
- Robert C. P. (1996). Mixtures of distributions: inference and estimation. In *Markov chain Monte Carlo in practice*, Interdiscip. Statist., pp. 441–464. Chapman & Hall, London.
- Robert C. P.; Celeux G.; Diebolt J. (1993). Bayesian estimation of hidden Markov chains: a stochastic implementation. *Statist. Probab. Lett.*, **16**(1), 77–83.
- Robert C. P.; Rydén T.; Titterton D. M. (2000). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *J. R. Stat. Soc. Ser. B*, **62**(1), 57–75.