

Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in  
Statistica per le Tecnologie e le Scienze



RELAZIONE FINALE

**INFERENZA BAYESIANA PER LA STIMA DEL MODELLO  
DI BASS**

Relatore Prof. Bruno Scarpa  
Dipartimento di Scienze Statistiche

Laureanda: Laura Masiero  
Matricola N. 1043323

Anno Accademico 2016/2017



*A mia Mamma.*



# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Approccio bayesiano e metodologie computazionali</b>	<b>3</b>
1.1 Inferenza statistica bayesiana . . . . .	3
1.2 Distribuzioni <i>a priori</i> . . . . .	5
1.2.1 <i>A priori</i> coniugate . . . . .	6
1.2.2 <i>A priori</i> non informative . . . . .	7
1.3 Teoria della stima . . . . .	8
1.3.1 Stimatore di Bayes . . . . .	8
1.3.2 Stima puntuale . . . . .	9
1.3.3 Intervalli di credibilità . . . . .	9
1.4 Aspetti computazionali . . . . .	11
1.4.1 Catene di Markov . . . . .	11
1.4.2 Algoritmo di Metropolis-Hastings . . . . .	12
1.4.3 Campionamento di Gibbs . . . . .	14
1.4.4 Analisi dell'output MCMC . . . . .	15
<b>2 Modelli di diffusione:</b>	
<b>il Modello di Bass</b>	<b>17</b>
2.1 Ciclo di vita di un prodotto . . . . .	17
2.1.1 Innovazioni . . . . .	18
2.2 Modelli di diffusione delle innovazioni . . . . .	19
2.2.1 Assunzioni fondamentali . . . . .	20
2.2.2 Modello di diffusione base . . . . .	21
2.3 Il Modello di Bass . . . . .	27

2.3.1	Formulazione del modello . . . . .	28
2.4	Confronto fra modelli . . . . .	32
<b>3</b>	<b>Stima bayesiana dei parametri del Modello di Bass</b>	<b>35</b>
3.1	I dati . . . . .	35
3.2	Il modello . . . . .	37
3.2.1	Distribuzioni <i>a priori</i> dei parametri . . . . .	37
3.2.2	Determinazione delle distribuzioni <i>a posteriori</i> . . . . .	39
3.3	Applicazione ai dati e risultati ottenuti . . . . .	41
<b>4</b>	<b>Diverse distribuzioni <i>a priori</i> portano a diverse previsioni?</b>	<b>53</b>
4.1	Densità a priori basate sul periodo storico . . . . .	54
4.2	"A priori" basate sulla valutazione del mercato . . . . .	63
<b>5</b>	<b>Conclusioni</b>	<b>71</b>
<b>A</b>	<b>Codice R</b>	<b>73</b>
A.1	Codice utilizzato nel capitolo 3 . . . . .	73
A.1.1	Stima dei parametri iniziali . . . . .	73
A.1.2	Stima dei parametri con Metropolis-Hastings e Gibbs Sampling . . . . .	74
A.1.3	Diagnostica e rappresentazioni grafiche . . . . .	75
<b>B</b>	<b>Distribuzione SBeta2</b>	<b>81</b>
B.1	Proprietà della SBeta2 . . . . .	82
	<b>Bibliografia</b>	<b>86</b>

# Elenco delle figure

2.1	Modello di Fourt e Woodlock al variare di $M$ . . . . .	24
2.2	Modello di Fourt e Woodlock al variare di $p$ . . . . .	24
2.3	Modello di Mansfield al variare di $M$ . . . . .	26
2.4	Modello di Mansfield al variare di $b$ . . . . .	26
2.5	Modello di Bass al variare di $p$ . . . . .	31
2.6	Modello di Bass al variare di $q$ . . . . .	31
2.7	Modelli a confronto: vendite cumulate e istantanee. . . . .	32
3.1	Vendite istantanee e vendite cumulate di Apple iPhone. . . . .	36
3.2	<code>trace plots</code> dei parametri. . . . .	43
3.3	ACF delle simulazioni dei vari parametri. . . . .	44
3.4	<code>cumuplots</code> delle <i>catene</i> di $M$ , $\sigma^2$ , $p$ e $q$ . . . . .	45
3.5	Distribuzioni <i>a posteriori</i> dei parametri. . . . .	48
3.6	Distribuzioni <i>a posteriori</i> e funzioni di ripartizione di $M$ e $\sigma^2$ . . . . .	49
3.7	Distribuzioni <i>a posteriori</i> e funzioni di ripartizione di $p$ e $q$ . . . . .	50
3.8	Vendite cumulate effettive, stimate con i parametri iniziali e con le stime bayesiane . . . . .	52
3.9	Vendite cumulate stimate e I.C. . . . .	52
4.1	<code>trace plots</code> dei parametri. . . . .	56
4.2	ACF dei parametri. . . . .	57
4.3	<code>cumuplots</code> delle <i>catene</i> di $M$ , $\sigma^2$ , $p$ e $q$ . . . . .	57
4.4	Distribuzioni <i>a posteriori</i> dei parametri. . . . .	58
4.5	Distribuzioni <i>a posteriori</i> e funzioni di ripartizione di $M$ e $\sigma^2$ . . . . .	59
4.6	Distribuzioni <i>a posteriori</i> e funzioni di ripartizione di $p$ e $q$ . . . . .	60

4.7	Vendite cumulate effettive e stimate con il modello ottenuto . . .	62
4.8	trace plots dei parametri. . . . .	64
4.9	ACF dei parametri. . . . .	65
4.10	cumuplots delle catene di $M$ , $\sigma^2$ , $p$ e $q$ . . . . .	65
4.11	Distribuzioni <i>a posteriori</i> dei parametri. . . . .	66
4.12	Distribuzioni <i>a posteriori</i> e funzioni di ripartizione di $M$ e $\sigma^2$ . . . . .	67
4.13	Distribuzioni <i>a posteriori</i> e funzioni di ripartizione di $p$ e $q$ . . . . .	68
4.14	Vendite cumulate effettive e stimate con il modello ottenuto . . .	70
B.1	Densità e ripartizione della Sbeta2 al variare di $\alpha$ . . . . .	83
B.2	Densità e ripartizione della Sbeta2 al variare di $\beta$ . . . . .	83
B.3	Densità e ripartizione della Sbeta2 al variare di $\gamma$ . . . . .	83



# Elenco delle tabelle

3.1	Parametri iniziali. . . . .	41
3.2	Informazioni sulla distribuzione <i>a posteriori</i> di $M$ . . . . .	49
3.3	Informazioni sulla distribuzione <i>a posteriori</i> di $\sigma^2$ . . . . .	49
3.4	Informazioni sulla distribuzione <i>a posteriori</i> di $p$ . . . . .	50
3.5	Informazioni sulla distribuzione <i>a posteriori</i> di $q$ . . . . .	50
3.6	Stime ottenute tramite l'inferenza bayesiana . . . . .	51
4.1	Informazioni sulla distribuzione <i>a posteriori</i> di $M$ . . . . .	59
4.2	Informazioni sulla distribuzione <i>a posteriori</i> di $\sigma^2$ . . . . .	59
4.3	Informazioni sulla distribuzione <i>a posteriori</i> di $p$ . . . . .	60
4.4	Informazioni sulla distribuzione <i>a posteriori</i> di $q$ . . . . .	60
4.5	Stime dei parametri ottenute. . . . .	61
4.6	Informazioni sulla distribuzione <i>a posteriori</i> di $M$ . . . . .	67
4.7	Informazioni sulla distribuzione <i>a posteriori</i> di $\sigma^2$ . . . . .	67
4.8	Informazioni sulla distribuzione <i>a posteriori</i> di $p$ . . . . .	68
4.9	Informazioni sulla distribuzione <i>a posteriori</i> di $q$ . . . . .	68
4.10	Stime dei parametri ottenute. . . . .	69



# Introduzione

L'inserimento di un nuovo prodotto, detto anche *innovazione*, nel mercato avviene dopo attente valutazioni operate dal *management* sulle possibili vendite future, sui profitti derivanti da esse e sull'impatto che l'*innovazione* può avere sulla reputazione dell'azienda stessa.

Per questo motivo vengono utilizzati modelli di previsione adatti a descrivere il possibile *ciclo di vita* del nuovo prodotto, per appurare la convenienza o meno del suo lancio nel mercato: i modelli di diffusione delle innovazioni.

Il modello più utilizzato è il Modello di Bass (1969), che incorpora nella stima delle vendite dell'*innovazione* considerata il mercato potenziale, il tasso d'innovazione e il tasso d'imitazione, rendendolo un modello completo da tutti i punti di vista.

Solitamente i parametri di questo modello sono stimati con i minimi quadrati non lineari (*NLS*), ma, in questa relazione, verranno stimati mediante l'inferenza statistica bayesiana, in modo da poter inserire quando necessario delle informazioni utili alla previsione.

Per questo motivo, nel Capitolo 1 verranno descritte brevemente le principali tematiche della statistica bayesiana e dei metodi computazionali per simulare valori da distribuzioni *a posteriori* sia conosciute che sconosciute e nel Capitolo 2 verranno affrontati i concetti relativi al ciclo di vita di un prodotto e ad alcuni modelli di diffusione delle innovazioni utilizzati, per poi dare maggior peso alla descrizione del Modello di Bass.

Nel Capitolo 3 invece ci si concentrerà nel punto focale della presente relazione: la stima bayesiana dei parametri del Modello di Bass utilizzando i dati delle vendite di iPhone, noto *smartphone* di Apple Inc., ma adottando

distribuzioni *a priori* molto diffuse, in modo da non incorporare informazioni in esse.

Nel Capitolo 4 invece si cercherà di valutare diversi scenari corrispondenti a diverse densità *a priori* dei parametri, per verificare se le previsioni delle vendite variano al variare delle ipotesi fatte inizialmente.

Infine nel Capitolo 5, verranno confrontate le varie previsioni ottenute nei Capitoli 3 e 4 per valutare le differenze e l'utilità della stima bayesiana.

# Capitolo 1

## Approccio bayesiano e metodologie computazionali

Nell'approccio classico all'inferenza statistica, l'obiettivo è quello di modellare il fenomeno d'interesse sulla base delle informazioni raccolte, per ottenere una conoscenza maggiore su di esso.

Tuttavia, molte volte è utile introdurre delle informazioni pregresse nel modello considerato: ciò è possibile anche nell'inferenza bayesiana, metodologia che sta conquistando un ruolo sempre più importante negli ambiti applicativi della statistica, come per esempio in elaborazioni di dati medico-sanitari o economico-finanziari.

Mediante l'utilizzo di distribuzioni *a priori* è possibile includere queste conoscenze, adattandole al verificarsi di eventi che migliorano la comprensione del fenomeno studiato.

### 1.1 Inferenza statistica bayesiana

*Bayesian statistics is, at it's core, about changing your opinion.*<sup>1</sup>

L'inferenza bayesiana fornisce strumenti matematici che modellano il ragionamento tipico in condizioni d'incertezza di qualsiasi essere razionale, e

---

<sup>1</sup><https://vwo.com/knowledge/bayesian-statistics>

ciò permette la flessibilità del modello atto a rappresentare i dati al cambiamento di opinione.

Per introdurre le informazioni pregresse nel modello statistico si usa il teorema di Bayes, formalizzando tutto ciò attraverso una distribuzione di probabilità *a priori*: si trattano quindi anche i parametri del modello come variabili casuali.

In termini matematici, dato un campione di osservazioni  $\mathbf{x} = (x_1, x_2, \dots, x_n)$ , provenienti da una distribuzione di probabilità  $p(\mathbf{x}|\theta)$ , in cui il parametro  $\theta$  è incognito ed ha distribuzione  $p(\theta)$ , si calcola attraverso il Teorema di Bayes la distribuzione *a posteriori*, cioè la distribuzione di  $\theta$  condizionata rispetto ai dati:

$$p(\theta|\mathbf{x}) = \frac{p(\theta)p(\mathbf{x}|\theta)}{\int_{\Theta} p(\theta)p(\mathbf{x}|\theta) d\theta} = \frac{p(\theta)p(\mathbf{x}|\theta)}{p(\mathbf{x})} \quad (1.1)$$

Nella maggior parte dei casi, viene specificata una funzione di verosimiglianza dei dati  $L(\theta; \mathbf{x})$ , di conseguenza, sostituendo a  $p(\mathbf{x}|\theta)$  la funzione di verosimiglianza, è possibile riscrivere la (1.1) come:

$$p(\theta|\mathbf{x}) = \frac{p(\theta)L(\theta; \mathbf{x})}{\int_{\Theta} p(\theta)L(\theta; \mathbf{x}) d\theta} = \frac{p(\theta)L(\theta; \mathbf{x})}{L(\mathbf{x})} \quad (1.2)$$

dove  $L(\mathbf{x}) = \int_{\Theta} p(\theta)L(\theta; \mathbf{x}) d\theta$  è una costante di normalizzazione<sup>2</sup>.

Scelte distinte della distribuzione *a priori*  $p(\theta)$  possono rendere difficoltoso il calcolo della costante di normalizzazione  $L(\mathbf{x})$  da un punto di vista meramente computazionale.

Per semplificare il calcolo si considera solo il numeratore della (1.2) ottenendo che la distribuzione *a posteriori* del parametro è proporzionale al prodotto tra la sua distribuzione *a priori* e la funzione di verosimiglianza dei dati, cioè:

$$p(\theta|\mathbf{x}) \propto p(\theta)L(\theta; \mathbf{x}) \quad (1.3)$$

---

<sup>2</sup>Essa ha il compito di rendere  $p(\theta|\mathbf{x})$  una funzione di probabilità, cioè tale che  $\int_{\Theta} p(\theta|\mathbf{x}) d\theta = 1$ , con  $\theta \in \Theta$ .

La distribuzione *a posteriori* così ottenuta, contenendo tutta l'informazione sul parametro  $\theta$ , verrà utilizzata per ottenere indicatori sintetici, stime puntuali e/o intervallari e per la verifica d'ipotesi (Liseo, 2008).

## 1.2 Distribuzioni *a priori*

La scelta della distribuzione *a priori* è estremamente soggettiva e si basa sulla previa conoscenza del singolo individuo. Oltre ad introdurre una componente informativa in aggiunta a quella derivante dai dati, spesso può portare anche a problemi di tipo computazionale: se non vengono utilizzate forme analitiche particolari, le cosiddette distribuzioni *coniugate*, spesso risulta addirittura impossibile ottenere espressioni esplicite per la distribuzione *a posteriori*. Inizialmente, l'incapacità di ottenere una forma esplicita per la distribuzione finale ha costituito un ostacolo alla diffusione delle tecniche bayesiane, ora invece grazie a tecniche computazionali, come per esempio i Metodi Monte Carlo basati sulle catene di Markov (*MCMC - Markov Chain Monte Carlo*) è possibile calcolare qualsiasi tipo di distribuzione *a posteriori* (§1.4).

Oltre alle distribuzioni *a priori* soggettive, cioè quelle finora illustrate, esistono anche densità *a priori* oggettive o non informative, cioè distribuzioni iniziali che esprimono vaghezza o informazioni a carattere generale rispetto ad un parametro. In problemi di stima parametrica, l'uso di una distribuzione *a priori* non informativa fornisce risultati simili a quelli derivanti dall'inferenza classica, dato che la maggior parte d'informazione è prodotta dai dati osservati e non dalla distribuzione iniziale del parametro.

In seguito verranno descritte brevemente la classe delle distribuzioni *coniugate* e alcuni metodi per trovare una distribuzione *a priori* non informativa.

### 1.2.1 *A priori* coniugate

Per non incorrere in problemi nel calcolo della distribuzione *a posteriori*, come menzionato precedentemente, vengono comunemente usate le distribuzioni provenienti da famiglie coniugate.

Questo tipo di distribuzioni presentano un tipo di relazione tale che:

- il calcolo della costante di normalizzazione sia eseguibile in modo analitico in forma chiusa;
- le distribuzioni *a priori* sui parametri siano abbastanza flessibili da consentire la rappresentazione di differenti opinioni riguardo all'incertezza sui parametri stessi;
- le distribuzioni *a priori* e *a posteriori* appartengano alla stessa famiglia parametrica.

Analiticamente, sia  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  un vettore di variabili aleatorie identicamente distribuite e indipendenti condizionatamente ad un vettore di parametri  $\boldsymbol{\theta} \in \Omega$ , con funzione di verosimiglianza

$$L(\boldsymbol{\theta}; \mathbf{x}) \propto \prod_{i=1}^n p(x_i | \boldsymbol{\theta})$$

Una distribuzione iniziale  $p(\boldsymbol{\theta})$  si dice *coniugata* al modello usato se la forma funzionale della distribuzione *a priori* e della distribuzione *a posteriori* sono uguali. Se il vettore aleatorio  $\mathbf{X}$  ha una distribuzione appartenente alla *famiglia esponenziale*, ovvero se la sua verosimiglianza può essere riscritta in questa forma (Pace e Salvani, 2001)

$$L(\boldsymbol{\theta}; \mathbf{x}) \propto k(\boldsymbol{\theta}) \exp(h(\boldsymbol{\theta}) \sum_{i=1}^n x_i)$$

è garantita l'esistenza di una distribuzione coniugata.<sup>3</sup>

---

<sup>3</sup>L'appartenenza alla famiglia esponenziale è condizione necessaria e sufficiente.



Le informazioni provenienti dai dati verranno incluse negli iperparametri della distribuzione *a priori*, formando così i parametri della distribuzione *a posteriori*.

### 1.2.2 *A priori non informative*

Una distribuzione *iniziale* si dice *non informativa* quando non apporta ulteriore conoscenza al modello.

Il principale motivo del loro utilizzo è che consentono di ottenere risultati simili a quelli raggiungibili tramite l'inferenza frequentista senza però abbandonare l'impostazione bayesiana; in questo modo non vengono utilizzate informazioni o valutazioni soggettive e ciò viene percepito come più condivisibile tra diversi utilizzatori.

Inoltre, il mantenimento della struttura bayesiana del modello statistico, consente il successivo inserimento, quando necessario, di informazioni utili atte ad affinare i risultati precedenti.

Tra le distribuzioni non informative possiamo trovare la distribuzione a priori di Jeffreys e le distribuzioni individuate con il metodo delle *reference prior* (Liseo, 2008).

La distribuzione non informativa di Jeffreys è così formata:

$$p^J(\theta) \propto \sqrt{\det(I(\theta))}$$

dove  $I(\theta)$  è l'Informazione attesa di Fisher riguardo ad una singola osservazione.

Il metodo delle *reference prior* invece trova delle distribuzioni *a priori* non informative massimizzando la *Divergenza di Kullback-Leibler*<sup>4</sup>

$$D_{KL} = \int_{-\infty}^{+\infty} p(\theta|\mathbf{x}) \log \frac{p(\theta|\mathbf{x})}{p(\theta)} d\theta$$

tra la distribuzione *a priori* e quella *a posteriori* ricavata da essa rispetto alla *a priori*.

---

<sup>4</sup>In caso di distribuzioni discrete, la *Divergenza di Kullback-Leibler* è così specificata:  
 $D_{KL} = \sum_i p(\theta_i|\mathbf{x}) \log \frac{p(\theta_i|\mathbf{x})}{p(\theta_i)}$

É comunque possibile ottenere lo stesso effetto dato dalla distribuzione di Jeffreys e da quelle ottenute tramite le *reference priors* introducendo nel modello densità *a priori* informative ma poco diffuse: in questa relazione verrà utilizzato quest'approccio.

## 1.3 Teoria della stima

Anche nell'approccio bayesiano possiamo calcolare stime puntuali, interval-  
lare, porci problemi di previsione e di verifica d'ipotesi (Liseo, 2008).

Le stime ottenute, a differenza di quelle calcolate con l'inferenza classica, saranno comprensive delle informazioni aggiuntive contenute nelle distribuzioni *a priori*.

### 1.3.1 Stimatore di Bayes

Uno Stimatore di Bayes è uno stimatore che minimizza il valore atteso di una funzione di perdita; equivalentemente, massimizza il valore atteso a posteriori di una funzione di utilità.

Il valore ottenuto dalla minimizzazione appena citata costituisce il valore atteso della densità *a posteriori*.

Dato che il calcolo dello stimatore, come detto precedentemente, si fonda soprattutto sulla minimizzazione di una funzione di rischio definita come valore atteso di una funzione di perdita

$$E_p\{F(\theta, \delta)\}$$

dove  $p$  indica la distribuzione *a priori* di  $\theta$ ,  $\delta = \delta(\mathbf{x})$  è uno stimatore di  $\theta$  basato su un vettore di osservazioni  $\mathbf{x}$  e  $F(\theta, \delta)$  è una funzione di perdita, esso varia al variare della funzione scelta. Quella più utilizzata è l'errore quadratico medio, detto anche **errore quadratico di rischio** (MSE) così definito:

$$MSE = E[(\hat{\theta}(\mathbf{x}) - \theta(\mathbf{x}))^2]$$

dove il valore atteso è calcolato sulla distribuzione congiunta di  $\theta$  e  $\mathbf{x}$ .

### 1.3.2 Stima puntuale

Una volta trovata la distribuzione *a posteriori*, il modo naturale di procedere consiste nel calcolo di qualche indicatore di sintesi.

Quello più usato è il valore atteso della distribuzione *a posteriori*

$$\hat{\theta}(\mathbf{x}) = E[\theta|\mathbf{x}] = \int \theta p(\theta|\mathbf{x}) d\theta$$

Esso coincide con lo stimatore di Bayes ottenuto tramite l'uso dell'MSE come funzione di rischio ed è noto come *stimatore dell'errore quadratico medio minimo - MMSE*.

A volte, a seconda della tipologia del problema che si vuole analizzare, invece di utilizzare come indicatore di sintesi il valore atteso, si preferiscono stime puntuali alternative come la mediana, la moda o altri quantili della distribuzione *a posteriori*.

### 1.3.3 Intervalli di credibilità

Nell'ambito dell'inferenza bayesiana non si parla più di intervalli di confidenza, ma di intervalli di credibilità o intervalli *Highest Posterior Density*, più brevemente *HPD* (Liseo, 2008 e Corradi, 2014), dato che ci si focalizza sulla ricerca di un intervallo che contenga con una certa probabilità il vero valore del parametro.

Gli intervalli di credibilità sono intervalli  $[a, b]$  che rispettano il seguente vincolo:

$$\int_a^b p(\theta|\mathbf{x}) d\theta = 1 - \alpha$$

dove  $p(\theta|\mathbf{x})$  è la densità *a posteriori* di  $\theta$ .

Essi sono quindi intervalli che contengono un insieme di valori  $\theta \in \Theta$  la cui densità *a posteriori* risulta più elevata, fino al raggiungimento di una probabilità complessiva non inferiore al livello prescelto  $1 - \alpha$ .

Non è sempre possibile determinare per via analitica gli intervalli HPD, e

quindi bisogna ricorrere a metodi numerici, che solitamente consistono in questi tre passi:

1. per un valore  $k$  fissato, si determinano le due radici  $\theta_1(k)$  e  $\theta_2(k)$  dell'equazione  $p(\theta|\mathbf{x}) = k$ ;
2. si calcola  $C(k) = Pr(\theta \in (\theta_1(k), \theta_2(k)))$ ;
3. si determina il valore di  $k$ ,  $\tilde{k}$ , che risolve l'equazione  $C(k) = 1 - \alpha$ .

Da questa procedura si ottiene un intervallo HPD del tipo:

$$\left(\theta_1(\tilde{k}), \theta_2(\tilde{k})\right).$$

Invece, se la distribuzione finale è monotona decrescente, l'insieme HPD di livello  $1 - \alpha$  si può ottenere determinando il quantile di ordine  $1 - \alpha/2$  della distribuzione *a posteriori*.

La differenza tra intervalli di confidenza classici e intervalli di credibilità è sostanziale: per gli intervalli di confidenza è l'intervallo stesso che, prima di effettuare la rilevazione campionaria ha probabilità dell'  $(1 - \alpha)\%$  di contenere il vero valore del parametro  $\theta$ , mentre, nel caso degli intervalli di credibilità, la probabilità è riferita al parametro  $\theta$  (variabile casuale) che ha probabilità dell'  $(1 - \alpha)\%$  di essere contenuto nell'intervallo.

Per quanto riguarda il calcolo analitico degli intervalli di credibilità, la regione d'interesse è costruita utilizzando i quantili della distribuzione *a posteriori* di  $\theta$ : gli estremi dell'intervallo vengono individuati prendendo il quantile o i quantili tali per cui la probabilità della regione così individuata sia pari ad  $1 - \alpha$ .

## 1.4 Aspetti computazionali

Come visto in precedenza, a meno che non vengano scelte distribuzioni *a priori* coniugate, possono sorgere problemi di tipo computazionale nel calcolo della distribuzione *a posteriori*.

Per riuscire a ricavare una sua approssimazione, si usano i metodi *Markov Chain Monte Carlo - MCMC*, algoritmi basati sulle catene di Markov particolarmente attraenti per la loro immediatezza.

Prima di trattare in dettaglio gli algoritmi MCMC utilizzati in seguito, è necessaria una breve introduzione alle catene di Markov.

### 1.4.1 Catene di Markov

Una *catena di Markov*  $\mathbf{X}^{(t)}$  è una sequenza di variabili casuali dipendenti

$$X^{(0)}, X^{(1)}, X^{(2)}, \dots, X^{(t)}, \dots$$

tale che la distribuzione di probabilità di  $X^{(t)}$  date tutte le variabili passate, dipenda solo da  $X^{(t-1)}$ .

Essa è un processo stocastico Markoviano che assume valori in uno spazio discreto, caratterizzato da un parametro  $t_i \in T$ , da un insieme  $S$  di stati e da una funzione di probabilità di transizione  $P : S \times S \times T \rightarrow [0, 1]$ .

Le probabilità di transizione si possono riassumere tramite una *matrice di transizione*  $P \in \mathbb{R}^{N \times N}$  e un vettore di probabilità iniziali  $\tilde{\pi}_0 \in \mathbb{R}^N$ , dove  $N$  indica il numero degli elementi contenuti nell'insieme degli stati  $S$ .

Gli elementi contenuti nella matrice  $P$  indicano le probabilità di transizione tra gli stati della catena: una catena nello stato  $i$  ha probabilità  $\pi_{ij}$  di passare allo stato  $j$  in un passo immediatamente successivo; mentre gli elementi sulla diagonale,  $\pi_{ii}$  indicano la probabilità di rimanere sullo stesso stato  $i$ .

Il vettore  $\tilde{\pi}_0$  definisce invece le probabilità che la catena di Markov inizialmente si trovi in ciascuno degli  $N$  stati.

Le probabilità  $\tilde{\pi}_n$  che al tempo  $t_n$  il sistema si trovi in ognuno degli  $N$  stati sono date dal seguente vettore:

$$(\tilde{\pi}_n)^T = (\tilde{\pi}_0)^T \cdot P^n$$

Dalla definizione di probabilità derivano queste proprietà della matrice  $P$ :

- $\pi_{ij} \geq 0, \forall j \in S$ ;
- $\sum_{i=1}^N \pi_{ij} = 1$ .

Le catene di Markov hanno importanti proprietà, tra cui l'irriducibilità e la periodicità.

L'*irriducibilità* indica la possibilità di passare da uno stato ad un altro in uno o più passi, vale a dire che tutti gli stati possono essere raggiunti; la *periodicità* invece specifica la capacità di ritornare in uno stato precedente ad intervalli periodici.

Se è possibile trovare un vettore di probabilità  $w$  tale che

$$wP = w$$

allora  $w$  identifica la distribuzione *stazionaria*.

Se una catena di Markov è aperiodica e irriducibile, allora essa ha un'unica distribuzione stazionaria e la distribuzione limite della catena di Markov, all'avvicinarsi ad infinito del numero di passi compiuti, equivale alla sua distribuzione stazionaria (Albert, 2009).

### 1.4.2 Algoritmo di Metropolis-Hastings

Il metodo di campionamento utilizzato dagli algoritmi MCMC crea una catena di Markov irriducibile e aperiodica, la cui distribuzione stazionaria equivale alla distribuzione *a posteriori*  $p(\theta|\mathbf{x})$  cercata.

Un modo generale per ottenerla è usare *Metropolis-Hastings*, un algoritmo che, nelle sue varianti, risulta applicabile ad una grande varietà di problemi inferenziali di tipo bayesiano.

L'algoritmo inizia con un valore iniziale  $\theta^0$  e specifica una regola per simulare  $\theta^t$ , il  $t$ -esimo valore in sequenza, dato  $\theta^{t-1}$ . Questa procedura utilizza una densità candidata  $g(\theta^*|\theta^{t-1})$  (*proposal density*) per simulare un valore candidato  $\theta^*$  e una *probabilità di accettazione*  $\pi$ , che specifica la probabilità con cui il valore candidato appena generato venga accettato come valore successivo della catena markoviana.

Questi sono i passi dell'algoritmo:

- Simulare un valore candidato  $\theta^*$  dalla densità candidata  $g(\theta^*|\theta^{t-1})$ ;
- calcolare il rapporto  $R = \frac{p(\theta^*|\mathbf{x})g(\theta^{t-1}|\theta^*)}{p(\theta^{t-1}|\mathbf{x})g(\theta^*|\theta^{t-1})}$ ;
- calcolare la probabilità di accettazione  $\pi = \min\{R, 1\}$ ;
- campionare un valore  $\theta^t$  tale che  $\theta^t = \theta^*$  con probabilità  $\pi$ , altrimenti  $\theta^t = \theta^{t-1}$ .

Sotto condizioni di regolarità poste sulla densità candidata  $g(\theta^*|\theta^{t-1})$ , la sequenza di valori simulati converge ad una variabile casuale che è distribuita come la distribuzione a posteriori  $p(\theta|\mathbf{x})$ .

Si hanno diversi tipi di algoritmi Metropolis-Hastings a seconda della *proposal* adottata: se essa è indipendente dal valore attuale della catena

$$g(\theta^*|\theta^{t-1}) = g(\theta^*)$$

allora si avrà un algoritmo che genera catene *indipendenti*, mentre un algoritmo in cui la *proposal* ha una forma del tipo

$$g(\theta^*|\theta^{t-1}) = h(\theta^* - \theta^{t-1})$$

con  $h(\cdot)$  densità simmetrica rispetto all'origine, viene chiamato *Metropolis-Hastings con random walk*. In quest'ultimo, il rapporto  $R$  viene calcolato come segue:

$$R = \frac{p(\theta^*|\mathbf{x})}{p(\theta^{t-1}|\mathbf{x})}$$

In conclusione, l'algoritmo Metropolis-Hastings è adatto quando non si conosce la forma della distribuzione a posteriori, però è possibile utilizzare altri tipi di algoritmi di generazione come il *campionamento di Gibbs* descritto in seguito.

### 1.4.3 Campionamento di Gibbs

Gli algoritmi chiamati *Gibbs sampling* appartengono sempre alla classe dei metodi MCMC, ma a differenza di *Metropolis-Hastings* permettono di scomporre problemi complessi (come distribuzioni target multidimensionali, nel cui caso sarebbe molto difficile costruire un *random walk Metropolis-Hastings*) in una serie di problemi più semplici da affrontare.

Sia  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)$  il vettore dei parametri d'interesse.

La distribuzione congiunta *a posteriori* di  $\boldsymbol{\theta}$ , cioè  $p(\boldsymbol{\theta}|\mathbf{x})$ , può essere difficile da sintetizzare.

Definiamo l'insieme delle distribuzioni condizionate

$$\begin{aligned} & p(\theta_1|\theta_2, \dots, \theta_p, \mathbf{x}) \\ & p(\theta_2|\theta_1, \theta_3, \dots, \theta_p, \mathbf{x}) \\ & \dots \\ & p(\theta_p|\theta_1, \dots, \theta_{p-1}, \mathbf{x}) \end{aligned}$$

Un ciclo di simulazioni dal *campionamento di Gibbs* consiste nella generazione di un valore da ciascuna distribuzione condizionata; procedendo in questo modo, viene simulata una catena di Markov che ha come distribuzione limite la distribuzione *a posteriori* cercata.

In alcune situazioni non è conveniente generare valori direttamente dalle distribuzioni condizionate, quindi si ricorre ad un approccio ibrido chiamato *Metropolis within Gibbs* (Robert e Casella, 2010 e Albert, 2009).



### 1.4.4 Analisi dell'output MCMC

Si è detto finora che, da un punto di vista teorico, utilizzando gli algoritmi di tipo MCMC la distribuzione dei valori simulati converge alla distribuzione *a posteriori*  $p(\theta|\mathbf{x})$  cercata, ma non è possibile essere completamente sicuri della correttezza dell'approssimazione distributiva ottenuta.

Un problema nell'analisi dell'output ottenuto dall'algoritmo è capire se esso ha raggiunto la convergenza e l'individuazione della dimensione del periodo di *burn-in*: i valori di  $\theta$  simulati nelle iterazioni iniziali non provengono esattamente dalla distribuzione cercata, di conseguenza bisogna eliminarli. Tuttavia, dopo il periodo di *burn-in*, l'effetto dovuto ai valori iniziali svanisce e la distribuzione dei nuovi valori ottenuti si avvicina alla vera distribuzione *a posteriori*.

Un modo per individuare il sopracitato periodo di *burn-in* è attraverso il *trace plot* dei valori simulati rispetto al numero di iterazioni.

Un'altra cosa importante da controllare in fase di diagnostica è l'*autocorrelazione* dei valori simulati: è necessario tenere presente che i valori ottenuti dall'algoritmo fanno parte di una catena markoviana: il valore  $t$ -esimo  $\theta^t$  dipende solamente da quello precedente, cioè da  $\theta^{(t-1)}$  e questo deve riflettersi sul grafico di autocorrelazione dei valori<sup>5</sup>.

Per constatare la stazionarietà della catena di Markov ottenuta, oltre alle verifiche grafiche sopra citate, è utile condurre alcuni test non parametrici, come il test di *Heidelberger-Welch* (1983) e il test di *Geweke* (1992).

Il test di *Heidelberger-Welch* usa, per approssimare la distanza tra due distribuzioni, la statistica di Cramer-von-Mises:

$$C = \int (F(x) - F_n(x))^2 dF(x)$$

dove  $F$  e  $F_n$  indicano rispettivamente la funzione di ripartizione di riferimento e quella empirica; esso saggia l'ipotesi nulla di stazionarietà della *catena*.

---

<sup>5</sup>Una forte autocorrelazione anche a lag alti, può indicare la mancata esplorazione dell'intero spazio parametrico.

Il test di *Geweke* invece usa la *densità spettrale* (Geweke, 1992) della catena  $h(x^{(t)})$ :

$$S_h(w) = \frac{1}{2\pi} \sum_{t=-\infty}^{+\infty} \text{cov}(h(x^{(0)}), h(x^{(t)})) e^{itw}$$

dove  $i$  indica la radice complessa<sup>6</sup> di 1, per derivare la varianza asintotica della media della *catena*,  $\gamma_h^2 = S_h^2(0)$ , per una funzione arbitraria  $h(\cdot)$  che in questo caso è la distribuzione *a posteriori* cercata:

$$\frac{1}{T} \sum_{t=1}^T h(x^{(t)})$$

Stimando con metodi non parametrici  $S_h$ , viene in seguito costruito un *test*  $t$  per verificare l'uguaglianza delle medie della prima e dell'ultima metà della *catena* presa in considerazione.

Infine, anche la scelta della dimensione campionaria dei valori simulati e l'accuratezza degli indicatori di sintesi conseguentemente ottenute sono una questione da non sottovalutare.

Dato che i valori generati non sono indipendenti tra loro, per il calcolo degli *errori standard* per esempio, è necessario l'utilizzo di metodologie apposite per valori correlati, come il metodo delle *medie batch*.

Supponiamo di stimare la media a posteriori di  $\theta_i$  con la media campionaria

$$\bar{\theta}_i = \frac{\sum_{j=1}^m \theta_i^j}{m}$$

Nel metodo delle *medie batch*, i valori simulati  $\theta_j$  sono divisi in  $b$  gruppi, ciascuno di dimensione  $v$ , dove  $m = bv$ .

In ogni gruppo viene calcolata la media, ottenendo così un insieme di medie campionarie  $\bar{\theta}_i^1, \bar{\theta}_i^2, \dots, \bar{\theta}_i^b$ : la deviazione standard delle medie batch divisa per la radice quadrata del numero di gruppi approssima l'errore standard della stima  $\bar{\theta}_i$  solamente se il lag uno dell'autocorrelazione della sequenza delle medie batch è relativamente contenuto.

---

<sup>6</sup>Quindi  $e^{itw} = \cos(tw) + tsin(tw)$ .

# Capitolo 2

## Modelli di diffusione: il Modello di Bass

In ambito aziendale, ogni decisione è presa in condizioni di incertezza, per questo sono necessarie metodologie indirizzate a guidare la scelta della migliore strategia da adottare.

Le previsioni delle vendite quindi possono aiutare molto il management, di conseguenza sono stati sviluppati, nel corso del tempo, vari modelli di previsione, in particolare i modelli di diffusione.

Dopo una breve introduzione al ciclo di vita che caratterizza qualsiasi prodotto verrà introdotta l'idea sottostante questi tipi di modelli e più specificamente il *Modello di Bass standard*.

### 2.1 Ciclo di vita di un prodotto

Il mercato, la concorrenza ed anche l'offerta di prodotti sono in continuo cambiamento e ciò implica anche la modifica della strategia di posizionamento e di differenziazione di un'azienda attraverso il lancio di nuovi prodotti chiamati *innovazioni* (Kotler *et al.*, 2012).

La teoria del *ciclo di vita del prodotto* (*PLC - Product Life Cycle*) considera il prodotto come un'entità distinta e per questo dotata di un proprio ciclo

di vita. La curva che rappresenta il ciclo di vita di un prodotto è divisa principalmente in quattro sezioni: introduzione, crescita, maturità e declino, e la loro evoluzione è il risultato della distribuzione nel tempo della curva di adozione, la cosiddetta *Curva di Rogers*, che modella la dinamica di adozione ipotizzando che la distribuzione dei singoli "adottanti" nel tempo sia una Gaussiana (Kotler *et al.*, 2012).

Da ciò deriva che il volume di vendite cambia al variare delle fasi di vita del prodotto, suggerendo tecniche di previsione distinte. In seguito saranno approfonditi i modelli che si riferiscono alle previsioni nel primo periodo di vita del prodotto, quando cioè esso è ancora un'*innovazione*.

### 2.1.1 Innovazioni

*Le imprese vedono l'innovazione come lo strumento chiave per aumentare profitti e quote di mercato [...]. Ma cos'è precisamente l'innovazione è difficile dirlo, ancora di più lo è misurarla.  
(The Economist, 20 febbraio 1999)*

Precedentemente si è detto che un'*innovazione* è un nuovo prodotto lanciato da un'azienda. Ma in realtà con questo termine non ci si limita ad indicare solo questo.

Il carattere di novità non si riferisce solo al prodotto ma anche al mercato in cui viene immesso, quindi, se un prodotto con determinate caratteristiche, già presente in un segmento di mercato, viene introdotto in un'altra tipologia geo-demografica di mercato, esso sarà connotato come *innovazione*, benchè fosse già esistente. Inoltre, è possibile utilizzare questo termine anche per riferirsi ad una versione aggiornata o modificata di un prodotto, perchè essa può comportare la riformulazione delle tecniche di produzione, gestione e posizionamento adottate finora dall'azienda, ponendola in una condizione di rischio dato dall'investimento di nuovo capitale.

Novità significa quindi alto rischio, componente sempre più presente nel sistema economico attuale, in cui è possibile osservare l'intensificarsi progressivo della competizione.

Ciò porta alla descrizione dei principali fattori che descrivono il successo di un'*innovazione*:

- la soddisfazione del consumatore che trova nel nuovo prodotto considerato un valore aggiunto rispetto ai prodotti forniti dai concorrenti;
- la conoscenza del mercato da parte dell'azienda e quindi la predisposizione di una strategia idonea allo sviluppo ottimale ed alla diffusione del prodotto;
- l'adeguatezza del processo di sviluppo, in modo da avere un'adeguata efficienza ed efficacia di produzione.

Questi tre fattori possono essere gestiti totalmente dall'impresa, quindi il successo di un'*innovazione* dipende quasi completamente dalla sua capacità di gestirli e solo raramente sono cause esterne a decretare il fallimento di nuovi prodotti.

## 2.2 Modelli di diffusione delle innovazioni

A causa dell'esigua quantità di dati a disposizione riguardanti le innovazioni, è difficile prevedere il loro eventuale successo o insuccesso ed individuare il loro *trend* futuro.

Per cercare di ovviare questo problema, vengono adottati dei modelli che, sfruttando i primi risultati di vendita mirano ad individuare l'estensione di ogni fase del ciclo di vita del prodotto considerato, in particolare la velocità di diffusione, il periodo temporale in cui si raggiunge il massimo della possibile diffusione e quello in cui ci si aspetta il suo declino.

Dato che esistono numerosi modelli quantitativi adatti a questo tipo di situazioni, sono necessari dei criteri per decidere qual è il modello più adatto all'uso.

I modelli previsionali possono essere differenziati e, di conseguenza scelti, in base alle seguenti caratteristiche (Valdani e Busacca, 1987):

- gli obiettivi della previsione: si hanno modelli diversi a seconda che la previsione si concentri sulla domanda globale del prodotto, su quella di primo acquisto o sulla domanda di riacquisto;
- le caratteristiche del prodotto considerato;
- le variabili indipendenti a disposizione;
- la struttura intrinseca del modello: esistono due tipologie essenziali, i modelli di *adozione* e quelli di *diffusione*<sup>1</sup>.

In conclusione, il fine dei modelli di diffusione è quello di rappresentare il livello di propagazione dell'innovazione all'interno di un sistema sociale, in termini di una funzione matematica del tempo trascorso dall'introduzione dell'innovazione (Peschechera, 2004).

### 2.2.1 Assunzioni fondamentali

Il processo di diffusione di un prodotto può essere pensato come il flusso delle adozioni dovute ai potenziali consumatori attraverso due segmenti di mercato (Lesa, 2000). Supponendo che un consumatore possa fare soltanto un'adozione<sup>2</sup>, possiamo definire il *mercato totale*  $M(t)$  come il numero di adozioni effettuate prima del ritiro del prodotto dal mercato:

$$M(t) = M_p(t) + N(t)$$

dove  $M_p(t)$  indica il *mercato potenziale o in senso stretto* al tempo  $t$  e  $N(t)$  indica il *mercato effettivo*, cioè i consumatori effettivi al tempo  $t$ .

In letteratura il mercato totale  $M(t)$  viene posto pari ad  $M$ , quindi supposto costante per tutto l'intervallo temporale preso in considerazione.

---

<sup>1</sup>In questo contesto non verranno trattati i modelli di adozione.

<sup>2</sup>Quindi si rientra nel caso di beni durevoli, non di beni soggetti ad acquisti ripetuti.

Quest'ipotesi, com'è possibile intuire, è irrealistica, dato che si hanno continue "migrazioni" dei consumatori tra segmenti di mercato (causate dal *passaparola* o *word-of-mouth*, dai *mass-media* o da altri fattori esterni come la condizione economica, propensione al consumo ecc...), ma è necessaria per giungere ad una soluzione chiusa delle equazioni generatrici del processo di diffusione<sup>3</sup>.

### 2.2.2 Modello di diffusione base

È possibile esprimere il modello di diffusione base<sup>4</sup> tramite il seguente problema di Cauchy (Bertsch *et al.*, 2009)

$$\begin{cases} \frac{dN(t)}{dt} = g(t)(M - N(t)) \\ N(t_0) = N_0 \end{cases} \quad (2.1)$$

dove:

$n(t) = \frac{dN(t)}{dt}$  rappresenta il numero non cumulato di adottanti al tempo  $t$ ,  
 $g(t) \geq 0 \quad t > 0$  è una funzione continua rappresentante il tasso di diffusione,  
 $M$  è il mercato totale,  
 $N(t)$  indica le adozioni cumulate al tempo  $t$  e  
 $N_0$  sono le adozioni cumulate al tempo  $t = t_0$ .

In base a questa formulazione, si ha che il numero di adottanti in un generico tempo  $t$  è determinato dal tasso di diffusione dell'*innovazione*, rappresentato mediante una funzione  $g(t)$  e dalla quota del mercato che non ha ancora acquistato il prodotto,  $M - N(t)$ . Questo fa sì che, più il numero di precedenti adottanti  $N(t)$  si avvicina al mercato totale  $M$ , più gli adottanti al tempo  $t$  decrescono.

Il modello descritto nell'equazione (2.1) è deterministico e continuo sotto l'ipotesi di continuità di  $g(t)$ : il suo valore specifico dipende dalle caratteristiche

---

<sup>3</sup>Esistono dei modelli che considerano la dinamicità del mercato, ma qui non sono trattati.

<sup>4</sup>Introdotta da V. Mahajan e R. A. Peterson (1985) per finalità didattiche, quindi posteriore al Modello di Bass (1969) che sarà trattato in seguito.

del processo di diffusione come il grado di innovazione, i canali comunicativi utilizzati e le proprietà del sistema economico considerato (Lesa, 2000).

Dato che  $g(t)$  può essere interpretata anche come la "probabilità" che un consumatore potenziale adotti al tempo  $t$  l'innovazione, si ha che  $g(t)(M - N(t))$  rappresenta il numero atteso di adottanti al tempo  $t$ .

La soluzione generica al problema (2.1) è così definita:

$$N(t) = (M - N_0)(M - e^{\int g(t_0) dt - \int g(t) dt}) \quad t > 0 \quad (2.2)$$

e indica le adozioni cumulate; invece per le adozioni istantanee, derivando la (2.2), si ha:

$$n(t) = (M - N_0)(g(t) \cdot e^{\int g(t_0) dt - \int g(t) dt})$$

Per ottenere invece una soluzione esplicita è necessario specificare una forma funzionale per  $g(t)$ : una è data da Bass (§2.3), poi, per esempio, ne sono state introdotte altre due, una da Fourn e Woodlock e una da Mansfield, nei rispettivi modelli<sup>5</sup>.

**Fourn e Woodlock** suppongono che il tasso di diffusione  $g(t)$  sia costante, ottenendo questa formulazione del problema di Cauchy:

$$\begin{cases} \frac{dN(t)}{dt} = p(M - N(t)) \\ N(t_0) = 0 \end{cases} \quad (2.3)$$

dove  $p > 0$  è il coefficiente di diffusione e

$N(t_0) = 0$  indica che le vendite al momento del lancio del prodotto sono nulle.

---

<sup>5</sup>Questi due modelli verranno brevemente descritti, ma la loro trattazione esula dall'obiettivo di questa relazione.



La soluzione al problema (2.3) evidenzia che il calcolo delle vendite cumulate è attuato mediante un'esponenziale modificata<sup>6</sup>

$$N(t) = M(1 - e^{-pt}) \quad t > 0$$

con costante di proporzionalità pari a  $p$  e con il valore limite di crescita indicato da  $M$ .

Invece, le vendite istantanee possono essere ricavate derivando l'equazione precedente, ottenendo:

$$n(t) = pMe^{-pt}.$$

Questo modello solitamente è usato per prevedere le adozioni di prodotti che non incontrano una particolare resistenza iniziale da parte dei consumatori.

Da notare è il fatto che il modello non include l'influenza esercitata dai primi consumatori su quelli successivi e che sia  $M$  che  $p$  sono costanti nel tempo: ciò implica che il mercato totale raggiungibile e lo sforzo comunicativo dell'azienda siano costanti, ipotesi non molto realistiche.

Nelle figure 2.1 e 2.2 è possibile osservare l'evoluzione della curva al variare dei valori assegnati ai parametri.

**Mansfield** invece si è concentrato nello studio dell'adozione di nuovi beni d'investimento, cioè delle innovazioni industriali, in cui il processo di diffusione è principalmente governato dal *passaparola*.

Tutto ciò viene formalizzato attraverso la seguente equazione differenziale:

$$\frac{dN(t)}{dt} = bN(t)(M - N(t)) \quad (2.4)$$

con  $b > 0$  coefficiente di diffusione assoluto.

---

<sup>6</sup>Ricordiamo che la curva di crescita data dall'esponenziale modificata è così formata:  $F(t) = \alpha(1 - \beta e^{-kt})$ , dove  $\alpha > 0$  indica il valore limite di crescita,  $k > 0$  è la costante di proporzionalità e  $\beta > 0$ .

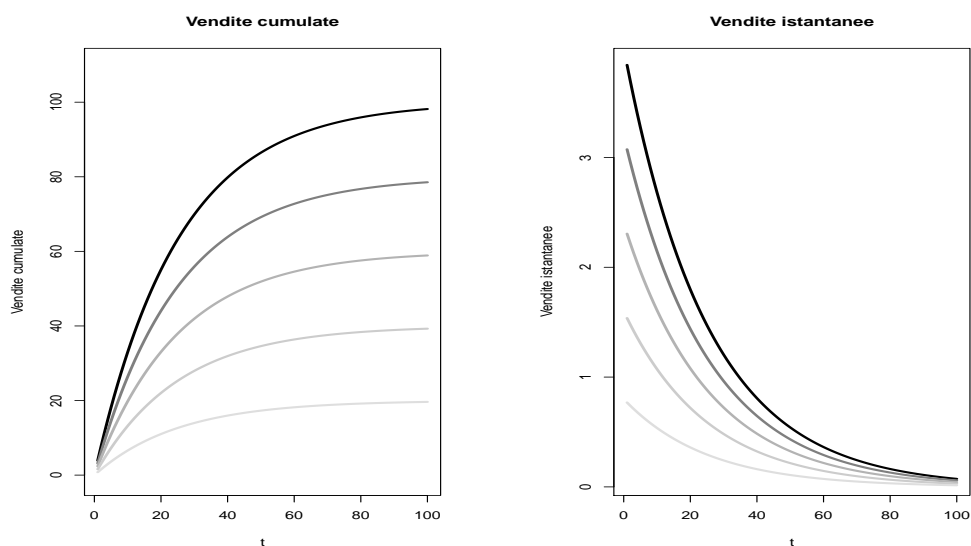


Figura 2.1: Modello di Fourt e Woodlock al variare di  $M$  ( $M = 100, 80, 60, 40, 20$ , in ordine di gradazione:  $M = 100$  nero,  $M = 20$  grigio chiarissimo),  $p = 0.04$  e  $t = 1, \dots, 100$ . Vendite cumulate e istantanee.

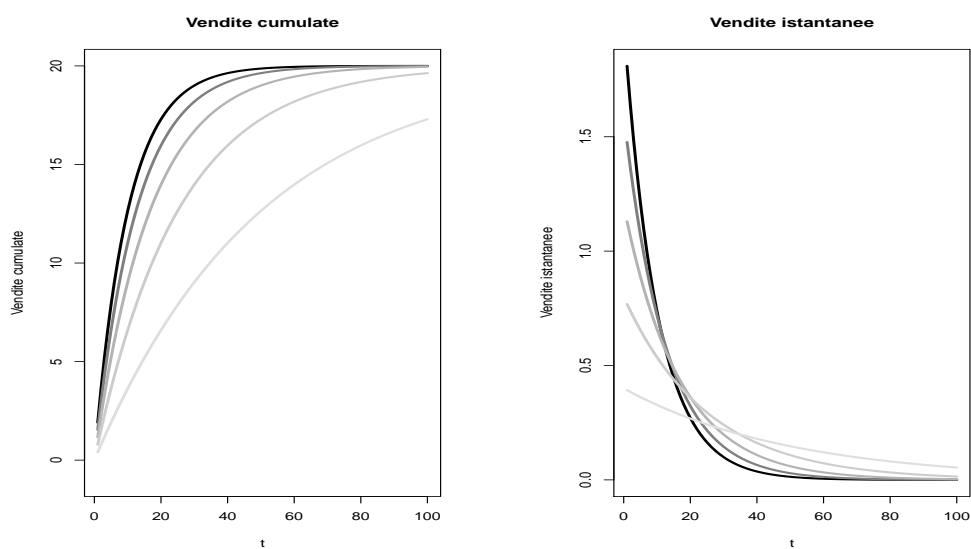


Figura 2.2: Modello di Fourt e Woodlock al variare di  $p$  ( $p = 0.1, 0.08, 0.06, 0.04, 0.02$ , in ordine di gradazione:  $p = 0.1$  nero,  $p = 0.02$  grigio chiarissimo),  $M = 20$  e  $t = 1, \dots, 100$ . Vendite cumulate e istantanee.

La soluzione della (2.4) è quindi la seguente:

$$N(t) = \frac{M}{1 + e^{a-ct}}$$

con  $a = \log \frac{M-N(0)}{N(0)}$  e  $c = bM$ .

Essa ha senso solamente per  $N(0) > 0$ , ciò implica che è possibile utilizzare questo modello solo dopo essere entrati a conoscenza dei primi dati di vendita.

È possibile riconoscere nella soluzione della (2.4) una curva di crescita logistica<sup>7</sup>, in cui il coefficiente di diffusione è  $bN(t)$ , in questo caso direttamente proporzionale all'ampiezza del mercato effettivo: ciò è coerente con l'ipotesi iniziale di adozione dell'innovazione avvenuta solo tramite *passaparola*, che agisce all'interno del mercato potenziale e si è sviluppata dai consumatori in  $N(t)$  e quindi è proporzionale a  $N(t)$  stesso (Lesa, 2000).

Per quanto riguarda le vendite istantanee, si ha che esse sono calcolate nel seguente modo:

$$n(t) = \frac{Mce^{a-ct}}{(1 + e^{a-ct})^2}$$

dove  $a$  e  $c$  sono definite come sopra.

Quest'ultima equazione è ottenuta dalla (2.4) mediante opportune sostituzioni.

Nelle figure 2.3 e 2.4 è possibile osservare il cambiamento delle curve date dal modello di Mansfield al variare dei parametri presenti nel modello, sia nel caso delle vendite cumulate che istantanee.

---

<sup>7</sup>La curva logistica è così definita:  $F(t) = \frac{\alpha}{1 + \beta e^{-kt}}$  con  $\beta > 0$ .

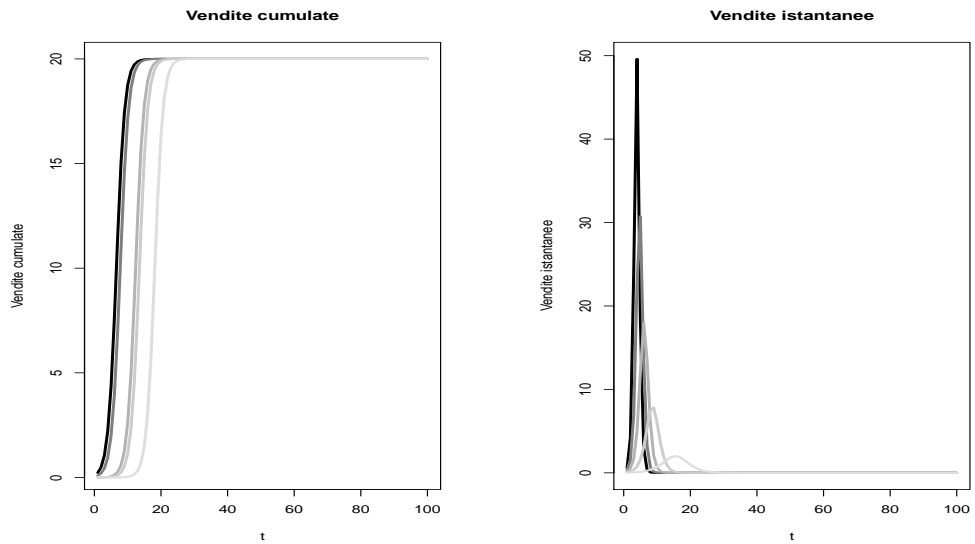


Figura 2.3: Modello di Mansfield al variare di  $M$  ( $M = 100, 80, 60, 40, 20$ , in ordine di gradazione:  $M = 100$  nero,  $M = 20$  grigio chiarissimo),  $b = 0.02$ ,  $N(0) = 0.04$  e  $t = 1, \dots, 100$ . Vendite cumulate e istantanee.

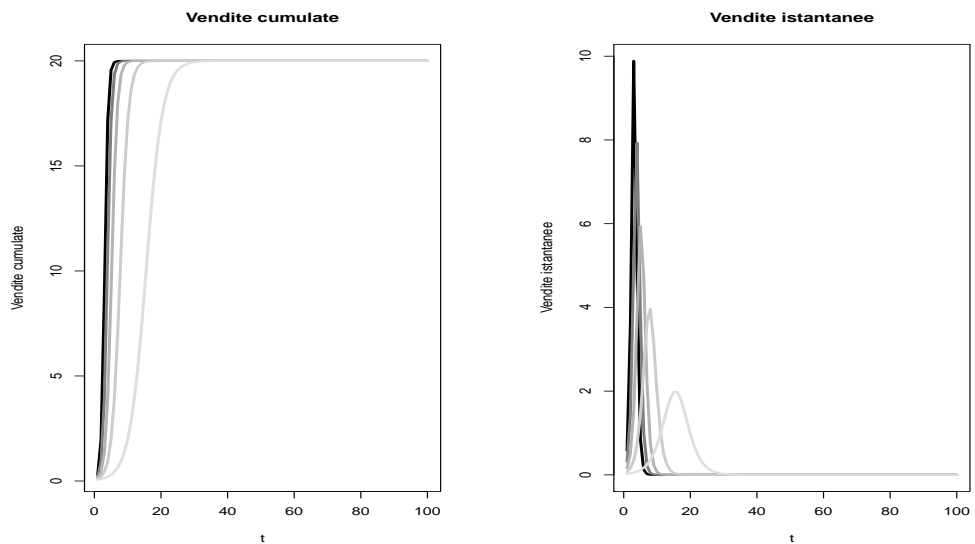


Figura 2.4: Modello di Mansfield al variare di  $b$  ( $b = 0.1, 0.08, 0.06, 0.04, 0.02$ , in ordine di gradazione:  $b = 0.1$  nero,  $b = 0.02$  grigio chiarissimo),  $M = 20$ ,  $N(0) = 0.04$  e  $t = 1, \dots, 100$ . Vendite cumulate e istantanee.

## 2.3 Il Modello di Bass

Il Modello di Bass (1969) è formato da un'equazione differenziale che descrive il processo di adozione delle innovazioni in un sistema economico.

Esso è uno dei primi modelli di diffusione che include entrambe le forme di comunicazione con le quali il comportamento degli adottanti può essere influenzato: il *passaparola* e i *mass media*, quindi cerca di modellare l'interazione tra i correnti e i possibili adottanti.

Bass sostiene l'esistenza di una distinzione tra gli acquirenti del nuovo prodotto, gli *innovatori* e gli *imitatori* e che la velocità di adozione dell'innovazione dipende dalla loro propensione verso la novità e il loro grado di imitazione.

**Innovatori e imitatori.** Gli *innovatori* sono coloro che hanno acquistato il nuovo prodotto solamente per l'effetto che hanno avuto sulle loro decisioni i *mass media* e la pubblicità, mentre gli *imitatori* sono stati influenzati dall'acquisto fatto dagli innovatori tramite il *passaparola*.

Inoltre, secondo il lavoro di Bass le due tipologie di adottanti non si distinguono per il periodo di acquisto, ma vengono distinti solo per il canale di informazione che li ha influenzati.

Come è possibile intuire, l'importanza<sup>8</sup> degli *innovatori* è maggiore rispetto a quella degli *imitatori* nel primo periodo di vita del prodotto e decresce con l'aumentare del tempo.

---

<sup>8</sup>Si intende da un punto di vista economico, dato che sono proprio gli *innovatori* ad influenzare gli *imitatori*.

### 2.3.1 Formulazione del modello

Il Modello di Bass standard è espresso dal seguente problema di Cauchy<sup>9</sup>:

$$\begin{cases} n(t) = \left(p + q \frac{N(t)}{M}\right)(M - N(t)) \\ N(0) = 0 \end{cases} \quad (2.5)$$

con  $p, q \in (0, 1)$ .

Da quest'equazione è possibile notare la proporzionalità delle vendite istantanee  $n(t)$  al mercato residuo o *mercato potenziale*  $M_p(t) = (M - N(t))$  e la dipendenza delle stesse da due fattori,  $p$  e  $q \frac{N(t)}{M}$ , che descrivono rispettivamente il comportamento degli *innovatori* e degli *imitatori*.

Più precisamente il parametro  $p$ , il *coefficiente d'innovazione* o *d'influenza esterna* cattura l'effetto delle influenze dei *mass media* e della pubblicità, mentre  $q$ , il *coefficiente d'imitazione* o *di influenza interna*, deriva la sua rilevanza dal rapporto  $\frac{N(t)}{M}$ , che rappresenta appunto la frazione del mercato composta dagli adottanti al tempo  $t$ : da  $q \frac{N(t)}{M}$  si deduce quindi la quota stimata degli *imitatori*, cioè di coloro che acquistano il nuovo prodotto in base all'esperienza degli *innovatori* e quindi in base al *passaparola*.

Al tempo  $t = 0$  si può infatti notare che la frazione  $\frac{N(t)}{M} = 0$  e che  $n(t) = pM$ , ciò significa che le vendite istantanee sono determinate solamente dagli *innovatori* e che non abbiamo ancora *imitatori*.

Ponendo  $\frac{N(t)}{M} = y(t)$  l'equazione differenziale contenuta nel problema (2.5) assume la seguente forma:

$$y'(t) = M(p + qy(t))(1 - y(t)) \quad (2.6)$$

È possibile vedere il problema anche da un punto di vista probabilistico, assumendo che  $y$  sia una funzione di ripartizione, rappresentante la probabilità cumulata di adozione dell'innovazione non oltre al tempo  $t$ , e  $y'$  sia la funzione densità corrispondente, atta ad indicare la variazione nel tempo delle adozioni (Bunea, 2014).

<sup>9</sup>Con riferimento al modello di diffusione base.

Sistemando la (2.6) possiamo interpretare il Modello di Bass standard come un particolare *tasso di rischio*:

$$\frac{y'(t)}{1-y(t)} = M(p + qy(t)) \quad p, q > 0 \quad (2.7)$$

La probabilità di adozione condizionata al fatto che l'innovazione non sia stata già adottata è uguale alla somma delle probabilità di due eventi indipendenti  $p$  e  $qy$ : non si può quindi essere sia *innovatori* che *imitatori* (Bass, 1969).

Per concludere, la soluzione in forma chiusa del problema di Cauchy (2.5) è definita come segue:

$$N(t) = M \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}} \quad t > 0 \quad (2.8)$$

Si nota che le vendite cumulate  $N(t)$  sono funzione dei parametri  $p$  e  $q$  e che  $M$  funge da parametro di scala del processo di diffusione, assunto costante durante tutto l'intervallo temporale.

**Utilizzo del modello.** Il Modello di Bass standard appena descritto è adatto a descrivere l'andamento della serie di vendite cumulate per la maggior parte delle innovazioni. A volte però, in presenza di particolari prodotti che non presentano un ciclo di vita *bell-shaped*, cioè che presentano, dopo la fase di crescita, vendite pressoché stazionarie, è possibile applicare il modello direttamente ai dati istantanei.

In questo caso, è necessario utilizzare la derivata della (2.8), vale a dire:

$$n(t) = M \frac{(p+q)^2}{p} \frac{e^{-(p+q)t}}{(1 + \frac{q}{p}e^{-(p+q)t})^2} \quad (2.9)$$

che modella appunto le vendite istantanee e non quelle cumulate.

Questo risultato è ottenibile anche sostituendo all'equazione differenziale contenuta nel problema di Cauchy (2.5) il risultato indicato nella (2.8).

È possibile notare il cambiamento delle curve delle vendite cumulate ed istantanee denotate dal modello di Bass, al variare di  $p$  e di  $q$ , rispettivamente nelle figure 2.5 e 2.6.

Dalla (2.9) è possibile derivare anche il punto di massima crescita della diffusione dell'innovazione, che rappresenta il raggiungimento della maturità del prodotto. Il tempo  $t^*$  in cui viene raggiunto il picco delle vendite è così calcolato:

$$t^* = \frac{\log\left(\frac{q}{p}\right)}{p + q}$$

di conseguenza otteniamo che le vendite istantanee al tempo  $t^*$  sono:

$$n(t^*) = \frac{M(p + q) \cdot 2}{4q}$$

e le vendite cumulate:

$$N(t^*) = \frac{M(q - p)}{2q}.$$

**Estensioni del modello.** Come è stato sottolineato, il Modello di Bass standard ipotizza il mercato totale  $M$  costante nel tempo e non tiene conto dell'effetto di variabili esogene<sup>10</sup> che possono modificare il processo di diffusione dell'innovazione.

Per ovviare a queste mancanze è possibile utilizzare una sua versione generalizzata, il *Modello di Bass generalizzato* (Bass *et al.* 1994) appunto, che introduce una *funzione d'intervento*  $x(t)$  che modella le variabili suddette<sup>11</sup>.

<sup>10</sup>Come per esempio la pubblicità, il prezzo, strategie di marketing...

<sup>11</sup>Essa può essere specificata in vari modi, si veda Guseo, 2004.

In questa relazione ci si limiterà all'utilizzo del modello standard.



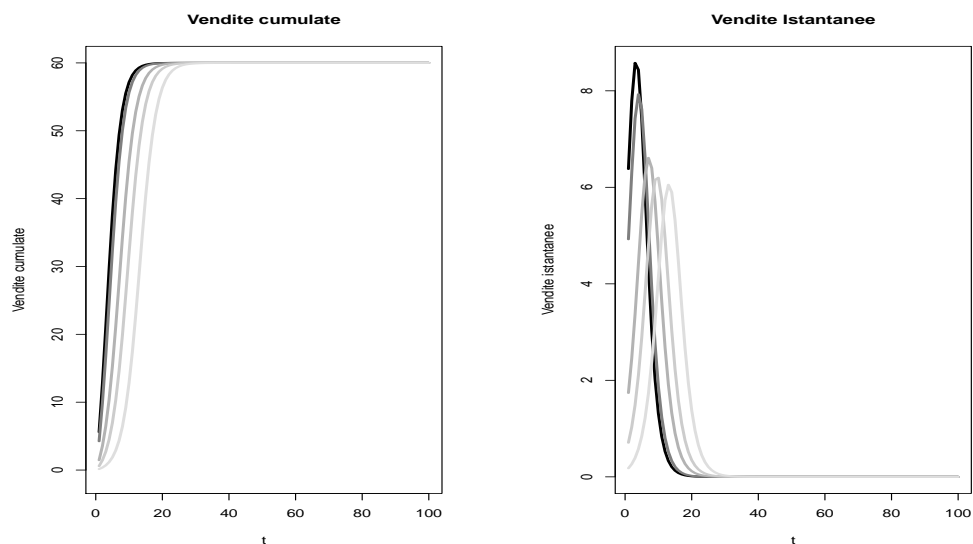


Figura 2.5: Modello di Bass al variare di  $p$  ( $p = 0.08, 0.06, 0.02, 0.008, 0.002$ , in ordine di gradazione:  $p = 0.08$  nero,  $p = 0.002$  grigio chiarissimo),  $M = 60$ ,  $q = 0.4$  e  $t = 1, \dots, 100$ . Vendite cumulate e istantanee.

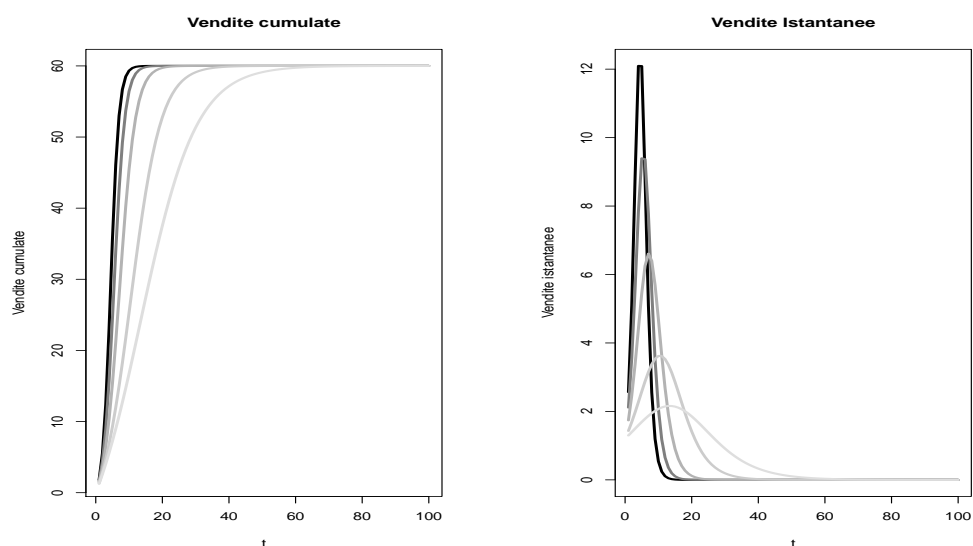


Figura 2.6: Modello di Bass al variare di  $q$  ( $q = 0.8, 0.6, 0.4, 0.2, 0.1$ , in ordine di gradazione:  $q = 0.8$  nero,  $q = 0.1$  grigio chiarissimo),  $M = 60$ ,  $p = 0.02$  e  $t = 1, \dots, 100$ . Vendite cumulate e istantanee.

## 2.4 Confronto fra modelli

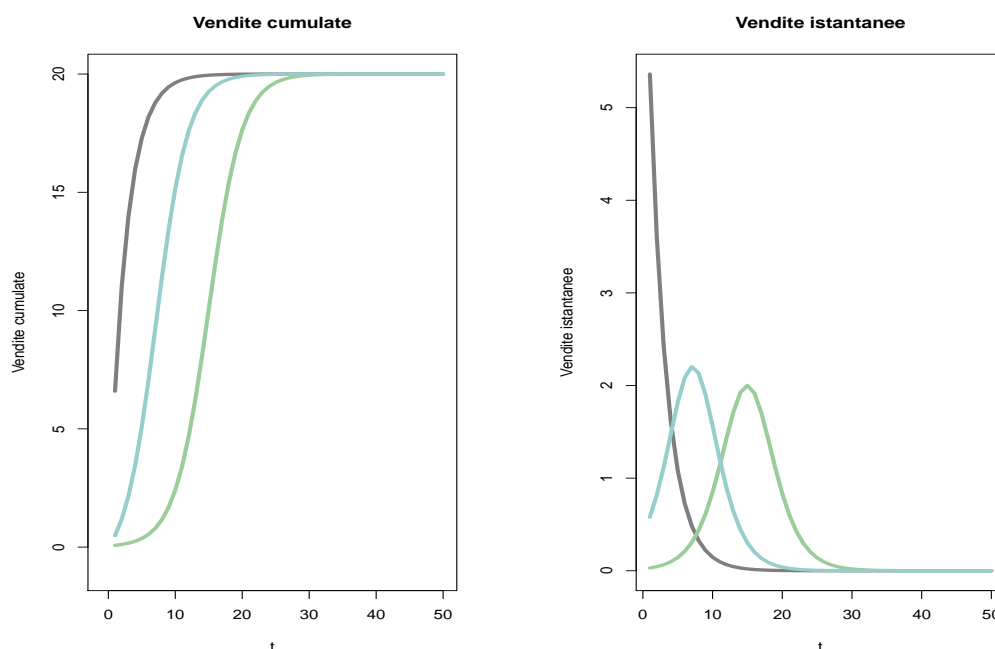


Figura 2.7: Vendite esemplificative cumulate (a sinistra) e istantanee (a destra) calcolate mediante il modello di Fourn e Woodlock (in grigio), di Mansfield (in verde) e di Bass (in azzurro).

In questo paragrafo si vuole confrontare i modelli citati nei due paragrafi precedenti: il modello di *Fourn e Woodlock*, quello di *Mansfield* ed il modello di *Bass*: per far ciò, nella figura 2.7 sono state tracciate le curve derivanti dai tre modelli, a parità di parametri, per vendite cumulate ed istantanee esemplificative.

Dalla figura si può notare che al variare della forma funzionale di  $g(t)$ , cioè del *tasso di diffusione*, è possibile ottenere modelli previsionali che differiscono anche molto l'uno dall'altro.

Si veda per esempio la differenza tra il modello di *Fourn e Woodlock* e quello di *Mansfield*: quest'ultimo è adatto a modellare i cicli di vita di prodotti che hanno una forma "a campana", a differenza di quello di *Fourn e Woodlock* che invece modella egregiamente le adozioni di prodotti le cui vendite esplodono all'entrata nel mercato e decrescono in rapporto al valore di  $p$ .

Non c'è molta differenza invece tra il modello di *Mansfield* e quello di *Bass*: entrambi assumono una forma "a campana", differendo l'uno dall'altro solo per il periodo del raggiungimento del picco delle vendite: il modello di *Mansfield* prevede un periodo di "crescita" nel ciclo di vita del prodotto più lungo di quello dato dal modello di *Bass*<sup>12</sup>.

In base alla funzione  $g(t)$  specificata quindi è possibile adattare il modello alle caratteristiche specifiche dell'*innovazione* e del processo di diffusione che la caratterizza.

Infatti, la forma di  $g(t)$  adottata da *Fourt e Woodlock*, costante pari a  $p$ , modella solamente l'influenza dei *mass media* sui consumatori e non prevede la presenza del *passaparola*: ciò è in linea con le ipotesi riguardo il suo utilizzo, che consiste nella previsione delle vendite di beni che non sono destinati ad avere una permanenza lunga nel mercato<sup>13</sup>; questo però costituisce anche un limite del modello, che non introduce nessuna informazione aggiuntiva per quanto riguarda l'influenza dei consumatori iniziali.

Nel modello ideato da *Mansfield*, invece, la forma funzionale di  $g(t)$ , cioè  $bN(t)$ , considera solamente l'influenza *interna*, data cioè dal *passaparola*: il tasso di diffusione è infatti direttamente proporzionale al mercato effettivo al tempo  $t$ . Questo fa sì che all'aumentare degli acquirenti effettivi aumenti anche la pressione all'acquisto esercitata sulla parte di mercato non ancora raggiunto. È una dinamica sensata in un contesto di *innovazioni tecnologiche* in cui è necessario un investimento per acquisirle, difatti il processo di diffusione si basa soprattutto sulle opinioni dei consumatori che sono già in possesso del bene.

---

<sup>12</sup>Si intende a parità di parametri.

<sup>13</sup>Come per esempio articoli di lusso, che fanno affidamento prevalentemente sulle campagne pubblicitarie per la loro diffusione.

Il modello di *Bass* incorpora tutti e due i canali d'informazione: sia quello *interno*, che quello *esterno*, attraverso i parametri  $p$  e  $q$  inseriti in  $g(t) = p + q \frac{N(t)}{M}$ , generalizzando gli altri due modelli appena trattati<sup>14</sup>; di conseguenza è adatto a modellare una vasta gamma di *innovazioni*, anche molto diverse tra loro. L'unica pecca di questo modello è che presuppone il mercato totale,  $M$ , costante al variare di  $t$ , ipotesi che comunque è presente anche negli altri due modelli.

In conclusione, cambiando  $g(t)$  possiamo adattare le previsioni di vendita a molti tipi di prodotti, avendo però ben chiaro il tipo di *innovazione* e il suo presunto processo di diffusione.

---

<sup>14</sup>Per  $q = 0$  si ottiene il modello di *Fourt e Woodlock*, mentre con  $p = 0$  si ottiene quello di *Mansfield*.

## Capitolo 3

# Stima bayesiana dei parametri del Modello di Bass

Punto centrale di questa relazione è la stima dei parametri del modello di Bass utilizzando l'inferenza statistica bayesiana.

Ciò implica l'imposizione di una distribuzione di probabilità da cui si ipotizza la provenienza dei dati: per questo motivo, oltre al parametro di *innovazione*  $p$ , al parametro di *imitazione*  $q$  e a quello rappresentante il *mercato potenziale*  $M$ , è necessaria la stima di un ulteriore parametro,  $\sigma^2$ , indicante la dispersione dei dati.

Prima di descrivere in dettaglio il procedimento di stima, verranno illustrati brevemente i dati utilizzati, la distribuzione di probabilità degli stessi ipotizzata e le distribuzioni *a priori* adottate per ogni parametro.

### 3.1 I dati

Per studiare l'andamento delle stime ottenute con quest'approccio innovativo sono stati utilizzati i dati delle vendite globali di *iPhone*, noto smartphone di Apple Inc.

La serie storica analizzata è trimestrale, a partire dal secondo trimestre del 2007 e fino al secondo trimestre del 2016<sup>1</sup>.

---

<sup>1</sup>Dati trovati nel seguente sito: <https://data.world/rflpr/iphone-sales>.

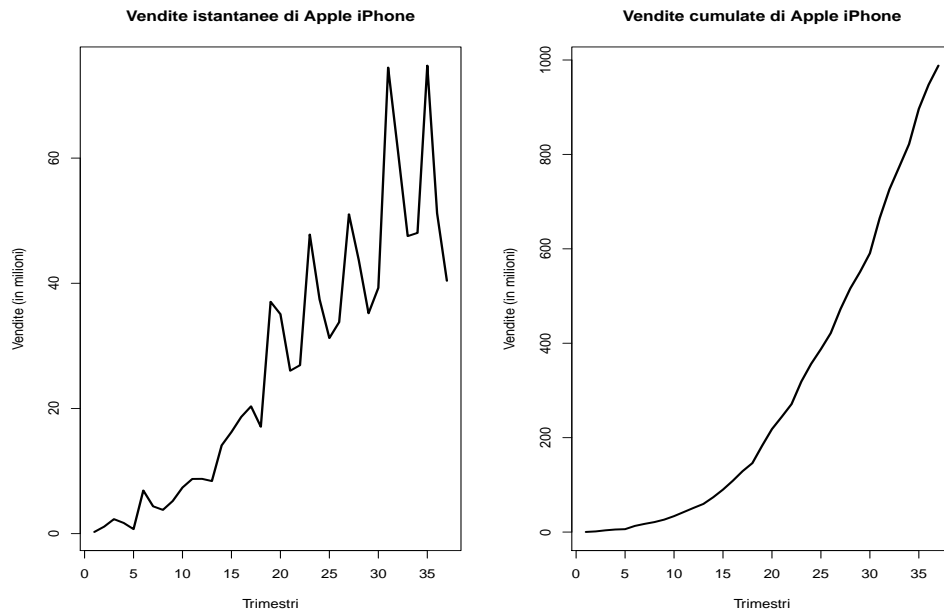


Figura 3.1: Vendite istantanee e vendite cumulate di Apple iPhone.

Osservando il grafico rappresentante le vendite istantanee è possibile intuire che il prodotto non ha ancora raggiunto la fase di maturità, dato che la serie storica non presenta ancora segnali di declino.

Inoltre, si può notare che essa è caratterizzata da un andamento molto perturbato, con picchi corrispondenti al rilascio di nuove versioni o a occasioni di "consumo di massa", come le feste natalizie: questo può portare a pensare che un modello di Bass standard non sia in grado di modellare completamente bene i dati e che sia quindi necessaria una sua modifica, come il modello di Bass generalizzato.

Comunque, attraverso l'utilizzo del modello standard si proverà ad individuare il trend del fenomeno e, successivamente, come verrà presentato nel Capitolo 4, si valuterà l'utilità di densità *a priori* ricche di informazioni, per vedere se il modello cattura l'andamento effettivo dei dati.

## 3.2 Il modello

Il modello che viene preso in considerazione presuppone una distribuzione Gaussiana per le vendite cumulate, dove la media è costituita dal Modello di Bass e la varianza  $\sigma^2$  è supposta costante:

$$Y(\mathbf{t}; M, p, q, \sigma) \sim \mathcal{N}\left(M \cdot \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}e^{-(p+q)t}}, \sigma^2\right) \quad (3.1)$$

con  $\mathbf{t} > 0$ ,  $M > 0$ ,  $p, q \in (0, 1)$  e  $\sigma > 0$ .

Visto che verrà utilizzata l'inferenza bayesiana per ottenere la stima dei quattro parametri menzionati precedentemente, è necessaria una distribuzione *a priori* per ogni parametro: in questo capitolo presupponiamo di non avere alcuna informazione da includere in esse, di conseguenza verranno adottate distribuzioni abbastanza diffuse.

### 3.2.1 Distribuzioni *a priori* dei parametri

**Distribuzione *a priori* di  $M$ .** Il parametro che indica il *mercato potenziale*,  $M$ , funge da parametro di scala della media ed è il più difficile da determinare.

Dato che assume solo valori positivi e molto grandi<sup>2</sup>, è possibile ipotizzare come sua distribuzione *a priori* una Normale<sup>3</sup>,  $M \sim \mathcal{N}(\mu_0, \sigma_0^2)$ , così definita:

$$f_M(m) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left\{-\frac{1}{2\sigma_0^2}(m - \mu_0)^2\right\} \quad (3.2)$$

con  $\mu_0 \in \mathbb{R}$  e  $\sigma_0 > 0$  iperparametri di forma e di scala della distribuzione.

---

<sup>2</sup>Ci si aspetta un mercato potenziale molto ampio data la fama di Apple e dato che ai giorni nostri ogni individuo possiede almeno uno smartphone.

<sup>3</sup>Di media molto alta.

**Distribuzione *a priori* di  $p$  e  $q$ .** Sia il parametro di *innovazione*  $p$  che quello di *imitazione*  $q$  assumono valori compresi tra zero e uno, per questo motivo una distribuzione *a priori* idonea per descriverli è una Beta,  $\mathcal{B}e(\alpha, \beta)$ , definita come segue:

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad (3.3)$$

con  $\alpha, \beta > 0$  iperparametri di forma e scala rispettivamente e  $B(\alpha, \beta) = \int_0^1 u^{\alpha-1}(1-u)^{\beta-1} du$  costante di normalizzazione.

Ovviamente le due distribuzioni *a priori* differiranno negli iperparametri, dato che solitamente  $q$  assume valori più alti di  $p$ <sup>4</sup>.

**Distribuzione *a priori* di  $\sigma^2$ .** L'ultimo parametro di cui dobbiamo stabilire la densità *a priori* è  $\sigma^2$ , il parametro di diffusione della distribuzione dei dati.

Per ottenere stime più precise di quelle che si otterrebbero con la distribuzione *a priori* coniugata al modello, cioè una Gamma-Inversa, si ipotizza come distribuzione *a priori* del parametro una Beta del secondo tipo scalata (Pérez *et al.*, 2017. Si veda Appendice B),  $\sigma^2 \sim SBeta2(\delta, \kappa, b)$ , così definita:

$$f_{\sigma^2}(\sigma^2) = \frac{\Gamma(\delta + \kappa)}{\Gamma(\delta)\Gamma(\kappa) \cdot b} \frac{(\frac{\sigma^2}{b})^{\delta-1}}{(\frac{\sigma^2}{b} + 1)^{\delta+\kappa}} \quad (3.4)$$

con  $b$  e  $\delta > 0$  iperparametri di forma e  $\kappa > 0$  iperparametro di scala.

Questa particolare distribuzione permette di modellare con più precisione il comportamento di  $\sigma^2$ , includendo i vantaggi di *proposals* come la Gamma o Gamma-Inversa. Inoltre è stato dimostrato che, dato un modello Normale con media nota, la distribuzione  $SBeta2(\delta, \kappa, b)$  è la distribuzione *a priori* intrinseca per il parametro di scala (Pérez *et al.*, 2016).

---

<sup>4</sup>Questa congettura deriva dal fatto che le persone che imitano il comportamento di altre sono in numero maggiore rispetto a quelle che "rischiano" comprando un nuovo prodotto.



### 3.2.2 Determinazione delle distribuzioni *a posteriori*

Qui verranno illustrate le distribuzioni *a posteriori* di ciascun parametro, vale a dire le distribuzioni *a posteriori* di ogni parametro condizionate a tutti gli altri parametri (o *full conditional*) e verrà spiegato da un punto di vista teorico il procedimento utilizzato per ottenerle. Nel §3.3 verranno invece presentati i risultati ottenuti.

Prima di procedere è utile ricordare che si suppone una distribuzione Gaussiana per i dati, di conseguenza la funzione di verosimiglianza associata è la seguente:

$$\begin{aligned} \mathcal{L}(\mathbf{y}; \mathbf{t}, M, p, q, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_i - \mu_i)^2 \right\} \propto \\ &\propto (\sigma^2)^{-n/2} \exp \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 \right\} \end{aligned} \quad (3.5)$$

$$\text{dove } \mu_i = M \cdot \frac{1 - e^{-(p+q)t_i}}{1 + \frac{q}{p} e^{-(p+q)t_i}}.$$

**Distribuzione *a posteriori* di  $M$ .** Come presentato nel paragrafo precedente, la distribuzione *a priori* di  $M$  è stata scelta in modo da semplificare il calcolo di quella *a posteriori*.

Dato che la verosimiglianza dei dati è Gaussiana possiamo utilizzare la distribuzione *a priori* coniugata per  $M$ , in modo da ottenere la densità *a posteriori*, condizionata a tutti gli altri parametri, con una forma funzionale nota<sup>5</sup>.

Avendo una Gaussiana sia come distribuzione *a priori* che come verosimiglianza dei dati, la distribuzione *a posteriori* di  $M$  derivante dal loro prodotto è sempre una Gaussiana, ma con diversi parametri,  $M_{post} \sim \mathcal{N}(\eta, \omega^2)$ , dove

$$\eta = \frac{\frac{\mu_0}{\sigma_0^2} + \frac{\sum_{i=1}^n y_i}{\sigma^2}}{\frac{1}{\sigma_0^2} + \frac{n}{\sigma^2}} \quad e \quad \omega^2 = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2}$$

sono i nuovi parametri di forma e scala della distribuzione.

---

<sup>5</sup>La distribuzione del parametro è quella specificata nel paragrafo precedente.

**Distribuzione *a posteriori* di  $p$ ,  $q$  e  $\sigma^2$ .** Il discorso cambia parlando di  $p$ ,  $q$ , parametri che descrivono il *tasso d'innovazione*, il *tasso d'imitazione* e  $\sigma^2$ : dato che le loro distribuzioni *a priori* non sono coniugate alla verosimiglianza dei dati, non è nota la loro forma.

La distribuzione *a posteriori full conditional* risultante è, nel caso di  $p$ , la seguente:

$$\begin{aligned} f_{P_{post}}(p|\mathbf{y}, \mathbf{t}, M, q, \sigma^2) &\propto \mathcal{L}(\mathbf{y}; \mathbf{t}, M, p, q, \sigma^2) \cdot f_P(p) = \\ &= (\sigma^2)^{-n/2} \exp \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 \right\} \cdot \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1} \end{aligned} \quad (3.6)$$

Per ottenere quella di  $q$ , si procede similmente, ponendo gli iperparametri della sua densità *a priori*.

Invece, possiamo ottenere la densità *a posteriori* di  $\sigma^2$  con il seguente prodotto:

$$\begin{aligned} f_{\sigma^2_{post}}(\sigma^2|\mathbf{y}, \mathbf{t}, M, p, q) &\propto \mathcal{L}(\mathbf{y}; \mathbf{t}, M, p, q, \sigma^2) \cdot f_{\sigma^2}(\sigma^2) = \\ &= (\sigma^2)^{-n/2} \exp \left\{ \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu_i)^2 \right\} \cdot \frac{\Gamma(\delta + \kappa)}{\Gamma(\delta)\Gamma(\kappa)} \cdot \frac{(\frac{\sigma^2}{b})^{\delta-1}}{(\frac{\sigma^2}{b} + 1)^{\delta+\kappa}} \end{aligned} \quad (3.7)$$

Quindi attraverso l'utilizzo dell'algoritmo di Metropolis-Hastings si cercherà di ottenere dei campioni casuali dalle distribuzioni *a posteriori* di  $p$ , di  $q$  e di  $\sigma^2$ , in modo da individuare il loro andamento.

### 3.3 Applicazione ai dati e risultati ottenuti

In questo paragrafo saranno applicate ai dati le metodologie computazionali descritte in §1.4 al fine di simulare valori dalle varie distribuzioni *a posteriori* dei parametri. In questo modo è quindi possibile stimare gli stessi e calcolare così la media della distribuzione indicata nella (3.1).

**Scelta degli iperparametri delle distribuzioni *a priori*.** Come già precedentemente menzionato, per cominciare sono state adottate delle distribuzioni abbastanza diffuse per i parametri, dato che non sono disponibili informazioni specifiche in modo da indirizzare la loro stima verso determinati valori. Per l'assegnazione di valori plausibili ai vari iperparametri, sono stati presi come valori iniziali i risultati della massimizzazione della log-verosimiglianza dei dati,  $\log(\mathcal{L}(\mathbf{y}|\mathbf{t}, M, p, q, \sigma))$ <sup>6</sup>.

I valori ottenuti dalla massimizzazione appena descritta sono i seguenti<sup>7</sup>:

$M$	$p$	$q$	$\sigma^2$
1.474361e+03	1.340844e-03	1.444915e-01	1.040365e+01

Tabella 3.1: Parametri iniziali.

Questi valori inoltre saranno utilizzati come punto di partenza dell'algoritmo costruito per stimare i parametri della distribuzione finale.

Specificatamente, le distribuzioni usate sono le seguenti:

- $M \sim \mathcal{N}(1600, 50)$ ;
- $p \sim \mathcal{Be}(2, 20.5)$ ;
- $q \sim \mathcal{Be}(2, 15.5)$ ;
- $\sigma^2 \sim \mathcal{SBeta2}(0.75, 0.75, 50)$ ;

<sup>6</sup>Le stime sono state ottenute tramite la funzione `nlminb`, che consente la minimizzazione (o massimizzazione, con opportuni accorgimenti) di funzioni non lineari con parametri soggetti a specifici vincoli.

<sup>7</sup>Tutti i calcoli e le operazioni descritte qui e in seguito sono state svolte con l'ausilio dell'ambiente statistico R.

Ora vedremo come sono stati ottenuti i vari parametri attraverso l'applicazione di algoritmi *MCMC*.

**Stima delle distribuzioni *a posteriori*, diagnostica e stime puntuali e intervallari dei parametri.** Per stimare le densità *a posteriori* dei vari parametri sono state usate queste due tipologie di algoritmi basati sulle catene di Markov: Metropolis-Hastings per trovare le distribuzioni incognite di  $p$ ,  $q$  e  $\sigma^2$  ed il campionamento di Gibbs per simulare valori da quella conosciuta di  $M$ .

Per utilizzarli è stata creata una funzione che stima  $M$ ,  $p$ ,  $q$  e  $\sigma^2$  mediante un ciclo che aggiorna le stime di ogni parametro ad ogni iterazione, dato che la stima di un parametro influenza quella di tutti gli altri<sup>8</sup>. Essa mediante 100000 simulazioni ha fornito dei risultati abbastanza buoni<sup>9</sup>, considerando il fatto che le densità *a priori* utilizzate non portano informazioni aggiuntive.

Per inizializzare l'algoritmo vengono utilizzate le stime dei parametri indicate nella tabella nella pagina precedente.

Inoltre, dalle 100000 simulazioni fatte vengono tolte le prime 25000, perchè esse costituiscono il *burn-in* delle stime, cioè il periodo in cui i parametri non hanno ancora raggiunto una sorta di "convergenza" e quindi non possono essere considerate rappresentative delle distribuzioni cercate.

Analizziamo ora i risultati ottenuti mediante l'algoritmo osservando come prima cosa i vari **trace plots**, ovvero una rappresentazione grafica del valore del parametro ottenuto verso il numero dell'iterazione corrispondente, poi verifichiamo la bontà delle simulazioni ottenute ed infine stimiamo i parametri.

---

<sup>8</sup>Si veda Appendice A.

<sup>9</sup>Riescono a catturare il *trend* di fondo delle vendite.

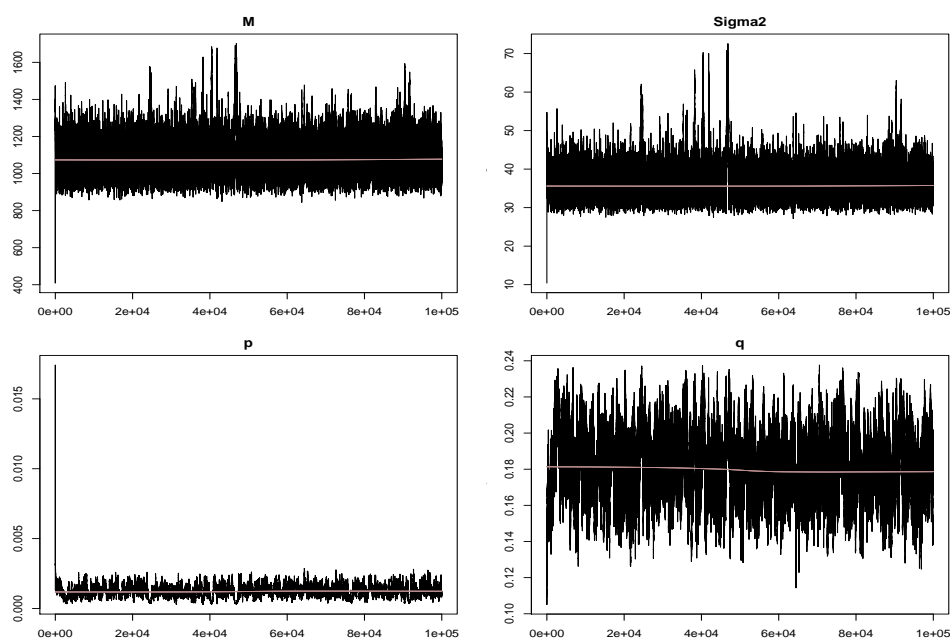


Figura 3.2: trace plots dei parametri.

Come possiamo vedere, le simulazioni hanno raggiunto un'accettabile convergenza, infatti la linea di lisciamiento che attraversa il plot riferito ad ogni parametro è stabile.

Ora, con l'aiuto dei grafici riguardanti le funzioni di autocorrelazione delle varie *catene* di valori ottenuti, è possibile stabilire se esse hanno un buon *mixing*, cioè se l'algoritmo ha esplorato gran parte dello spazio parametrico: in una buona catena markoviana non deve essere presente una fortissima correlazione a lag alti, perchè ciò indicherebbe la mancata esplorazione nella sua interezza.

Come è possibile vedere nella Figura 3.3 le ACF che otteniamo sono abbastanza buone, soprattutto per quanto riguarda  $M$  e  $\sigma^2$ ; le ACF di  $q$  e  $p$  invece sono più alte, ma tendono ad esaurirsi lentamente all'aumentare del lag.

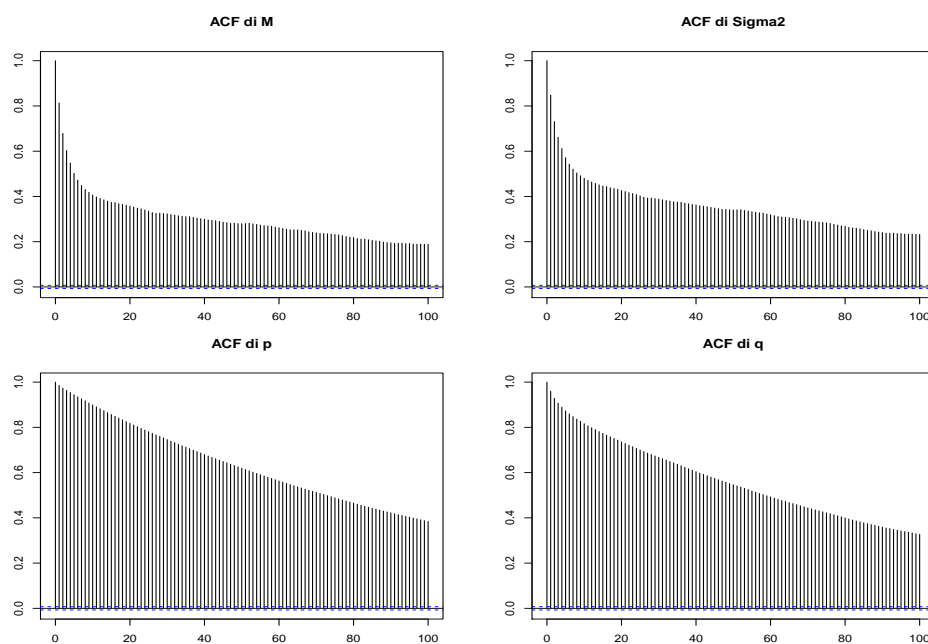


Figura 3.3: ACF delle simulazioni dei vari parametri.

Quando si applicano algoritmi di tipo MCMC per trovare distribuzioni *a posteriori* sconosciute non si può mai essere sicuri dell'eventuale convergenza e della correttezza dei valori ottenuti, infatti si dice che *"you've only seen where you've been"* (Robert e Casella, 2010), nel senso che parte del supporto della densità cercata che non è ancora stato esplorato dalla catena al tempo  $t$ , non è facile da individuare (Robert, e Casella, 2010). Per questo è meglio attuare diverse verifiche di "convergenza" delle *catene* risultanti l'algoritmo. Qui di seguito vedremo un'altra verifica grafica e due test non parametrici, il tutto attuato mediante l'ausilio del pacchetto R *coda*.

Per quanto riguarda la verifica grafica, vediamo il `cumplot` per ogni gruppo di simulazioni delle diverse distribuzioni *a posteriori*: esso consente di monitorare la convergenza attraverso il controllo della stabilità della funzione di ripartizione empirica derivata dalle catene markoviane simulate.

Esso ci fornisce l'andamento della mediana, del quantile 0.025 e il quantile 0.975 verso le iterazioni. In caso di convergenza ci si aspetta un leggero rumore all'inizio del grafico, poi una sorta di stabilità. Se per caso è visibile un trend persistente, significa che non è stata ancora raggiunta la convergenza.

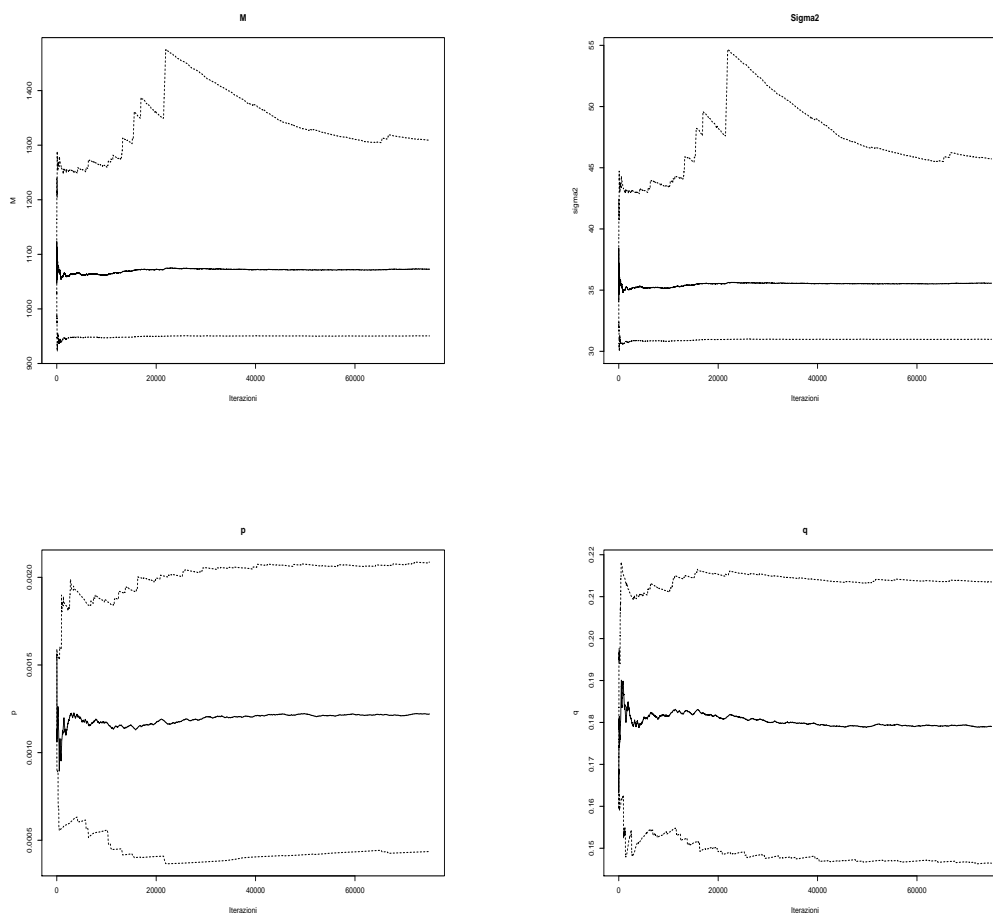


Figura 3.4: cumuplots delle catene di  $M$ ,  $\sigma^2$ ,  $p$  e  $q$ .

Come previsto, l'andamento della mediana della distribuzione di ogni parametro sembra stabile, a differenza dei quantili 0.025 e 0.975 che si stabilizzano lentamente con l'avanzare delle iterazioni.

Comunque ciò fornisce un segnale positivo riguardo la stazionarietà delle varie catene.

Per avere ulteriori conferme, non più di natura grafica, riguardo la stazionarietà delle catene generate conduciamo ora i test non parametrici di *Heidelberger-Welch* e di *Geweke* accennati nel §1.4.4.

Per utilizzare il test di *Heidelberger-Welch* è possibile utilizzare la funzione `heidel.diag` del pacchetto `coda`: qui ci concentriamo solo sulla prima parte

del suo output, perchè essa ci restituisce anche il risultato di un altro test<sup>10</sup>, qui non trattato.

I risultati ottenuti sono i seguenti:

- $M$ :

	Stationarity	start	p-value
	test	iteration	
[,1]	failed	NA	0.0275

- $\sigma^2$ :

	Stationarity	start	p-value
	test	iteration	
[,1]	failed	NA	0.0301

- $p$ :

	Stationarity	start	p-value
	test	iteration	
[,1]	passed	1	0.0929

- $q$ :

	Stationarity	start	p-value
	test	iteration	
[,1]	passed	1	0.125

Come è possibile notare solamente le *catene* di  $p$  e  $q$  hanno passato il test di stazionarietà, anche se i grafici visti precedentemente ci suggerivano la convergenza di tutte le *catene*.

---

<sup>10</sup> *Half-width test*.



Come ulteriore verifica attuiamo il test di *Geweke*.

La funzione `geweke.diag` costruisce una sorta di  $t$  test per verificare l'uguaglianza delle medie nella prima e nell'ultima metà della catena di Markov sottoposta al test: essa calcola le medie delle due parti della catena

$$\delta_A = \frac{1}{T_A} \sum_{t=1}^{T_A} h(x^{(t)}), \quad \delta_B = \frac{1}{T_B} \sum_{t=T-T_B+1}^T h(x^{(t)})$$

e le stime  $\sigma_A^2$  e  $\sigma_B^2$  di  $S_h(0)$ , densità spettrale in zero, riguardanti le due parti. La statistica test utilizzata dalla funzione

$$Z = \frac{\sqrt{T}(\delta_A - \delta_B)}{\sqrt{\frac{\sigma_A^2}{\tau_A} + \frac{\sigma_B^2}{\tau_B}}}$$

con  $T_A = \tau_A T$ ,  $T_B = \tau_B T$  e  $\tau_A + \tau_B < 1$ , si distribuisce asintoticamente come una Normale Standard ed è chiamata *Z-score* (Robert e Casella, 2010).

Dall'applicazione della funzione `geweke.diag` alle simulazioni fatte, sono stati ottenuti i seguenti risultati<sup>11</sup>:

- $M$ :  $Z=-1.811$
- $\sigma^2$ :  $Z=-1.697$
- $p$ :  $Z=-1.549$
- $q$ :  $Z=1.547$

Assumendo un'ipotesi alternativa bilaterale,  $H_1 : \delta_A \neq \delta_B$  e un  $\alpha = 0.05$ , la regione di rifiuto del test è  $R_\alpha \doteq (-\infty, -1.96) \cup (1.96, +\infty)$ : possiamo dunque notare che, per tutte e quattro le *catene*, a differenza dei risultati ottenuti con il test di *Heidelberger Welch* ma in accordo con la verifica grafica data dai `cumulplots`, è accettata l'ipotesi nulla di uguaglianza delle medie, indicando così la loro stazionarietà.

---

<sup>11</sup>Per tutte e quattro le catene testate, le due parti delle catene sono così suddivise: `Fraction in 1st window = 0.1, Fraction in 2nd window = 0.5` .

Possiamo quindi procedere con la visione delle distribuzioni *a posteriori* ottenute, con la stima di alcuni indicatori di sintesi e dei vari parametri.

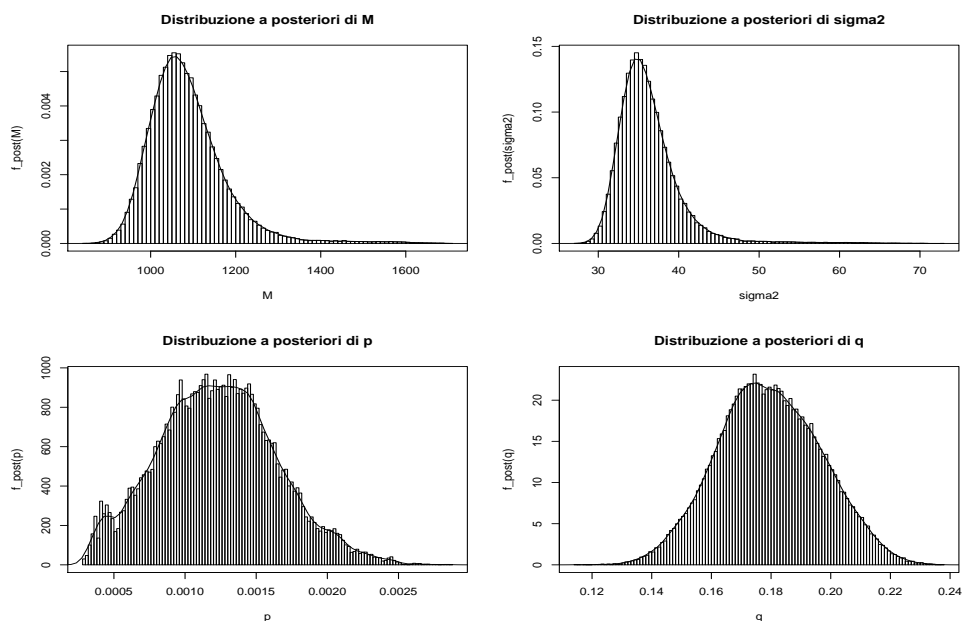


Figura 3.5: Distribuzioni *a posteriori* dei parametri.

È possibile notare che, anche se sono state utilizzate distribuzioni *a priori* abbastanza diffuse per ogni parametro, le densità *a posteriori* risultanti sono abbastanza concentrate su determinati valori, specialmente per quanto riguarda  $p$ .

Per avere un'idea più precisa sulle distribuzioni *a posteriori* stimate, sono stati calcolati alcuni indicatori di sintesi, alcuni loro quantili e le stime puntuali dei parametri con i relativi intervalli di confidenza: in seguito per ogni parametro verranno mostrate la densità e la funzione di ripartizione della distribuzione *a posteriori* corrispondente, accompagnate dai quantili 0.025, 0.25, 0.5, 0.75, 0.975, dalla stima della media, della sua deviazione standard e dell'intervallo di credibilità risultante<sup>12</sup>.

<sup>12</sup>Per il calcolo della sd è stata usata la funzione `bm` del pacchetto `batchmeans`, che la calcola con il metodo delle *medie batch*, ottenendo così una stima consistente.

Distribuzioni *a posteriori* di  $M$  e  $\sigma^2$  e indicatori di sintesi.

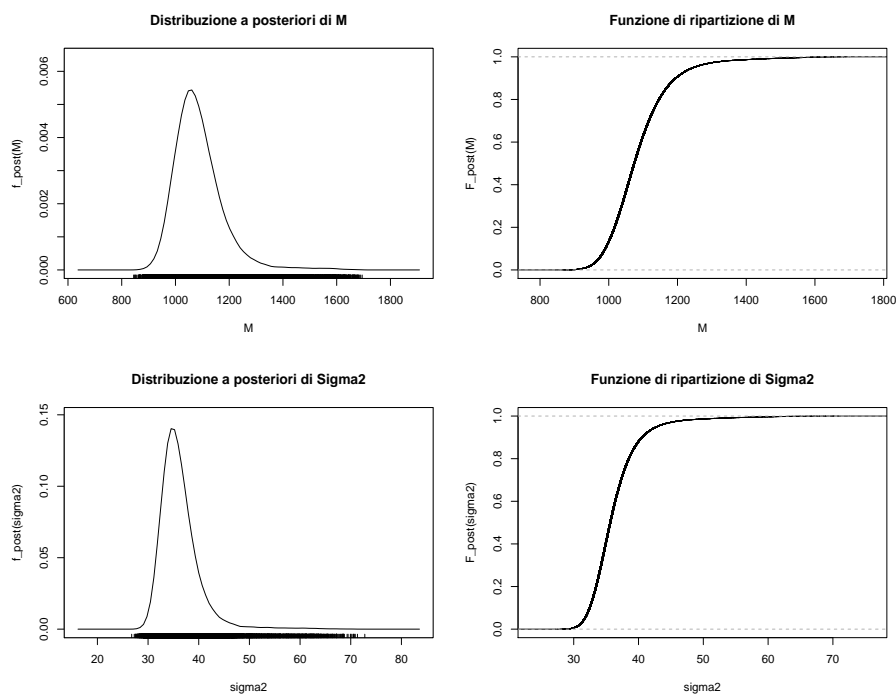


Figura 3.6: Distribuzioni *a posteriori* e funzioni di ripartizione di  $M$  e  $\sigma^2$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
950.53	1026.10	1072.73	1128.75	1309.04	1086.42	2.753	927.71	1260.71

Tabella 3.2: Informazioni sulla distribuzione *a posteriori* di  $M$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
30.9817	33.7763	35.561	37.7759	45.7239	36.246	0.12982	29.978	43.231

Tabella 3.3: Informazioni sulla distribuzione *a posteriori* di  $\sigma^2$ .

Distribuzioni *a posteriori* di  $p$  e  $q$  e indicatori di sintesi.

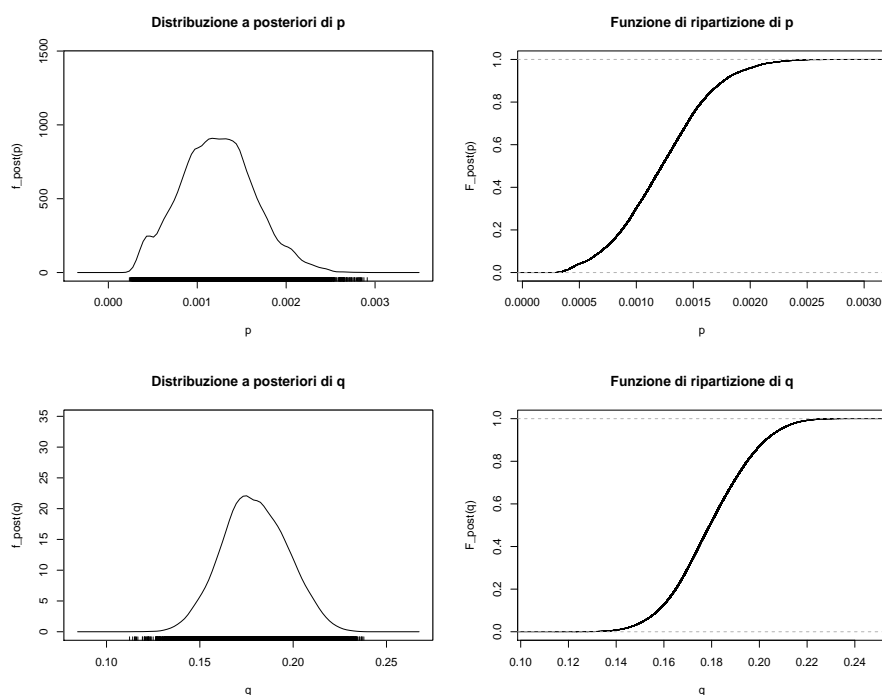


Figura 3.7: Distribuzioni *a posteriori* e funzioni di ripartizione di  $p$  e  $q$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
0.00044	0.00094	0.0012	0.0015	0.0021	0.0012	1.8e-05	0.00035	0.002

Tabella 3.4: Informazioni sulla distribuzione *a posteriori* di  $p$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
0.1464	0.1675	0.1791	0.1916	0.2135	0.1795	0.0006942	0.1459	0.2130

Tabella 3.5: Informazioni sulla distribuzione *a posteriori* di  $q$ .

**Applicazione ai dati delle stime ottenute.** Dopo aver trovato le distribuzioni *a posteriori* dei vari parametri è possibile sfruttare le stime puntuali derivanti da esse, per poter costruire il modello (3.1) proposto in §3.2.

Le stime ottenute sono le seguenti:

Media				Moda			
$M$	$p$	$q$	$\sigma^2$	$M$	$p$	$q$	$\sigma^2$
1086.421	0.00123	0.1795	36.246	1068.482	0.001558	0.1687	34.184

Tabella 3.6: Stime dei parametri ottenute tramite l'inferenza bayesiana.

Possiamo notare che la stima di  $M$  è più bassa di quella ottenuta tramite la massimizzazione della log-verosimiglianza dei dati, mentre la stima di  $\sigma^2$  è lievemente più alta. Questo può essere dovuto alle densità *a priori* adottate, molto disperse e quindi non indirizzate verso valori precisi dei parametri. Per quanto riguarda  $p$  e  $q$ , le stime trovate tramite l'inferenza bayesiana non si discostano molto con quelle precedentemente calcolate.

Il modello stimato quindi è il seguente:

$$Y(\mathbf{t}; M, p, q, \sigma) \sim \mathcal{N}\left(1086.421 \cdot \frac{1 - e^{-0.18074 \cdot t}}{1 + 146.3022 \cdot e^{-0.18074 \cdot t}}, 36.246\right) \quad (3.8)$$

Dalla Figura 3.8 possiamo constatare che le stime delle vendite calcolate con i parametri ottenuti mediante la massimizzazione della log-verosimiglianza, rispecchiano quasi perfettamente le vendite effettive, mentre quelle ottenute utilizzando il nuovo approccio bayesiano tendono a sovrastimarle leggermente dal 24-esimo trimestre e a sottostimare poi dal 34-esimo in poi. Nella Figura 3.9 invece, è possibile percepire la dispersione delle distribuzioni *a priori* adottate dalla visione dell'intervallo di credibilità, infatti esso è molto ampio.

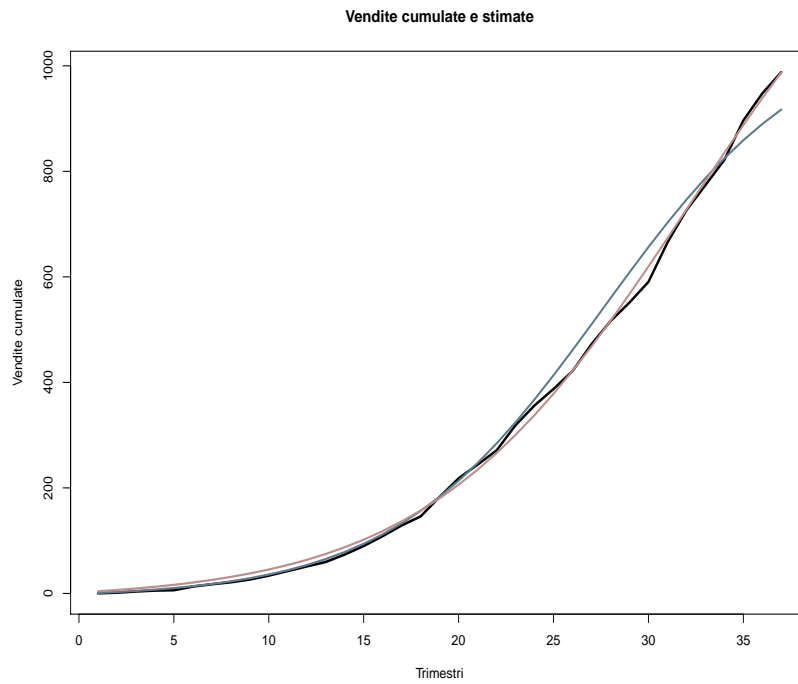


Figura 3.8: Vendite cumulate effettive (in nero), previste dal modello con i parametri stimati inizialmente (in rosa) e previste con le stime bayesiane ottenute (in blu).

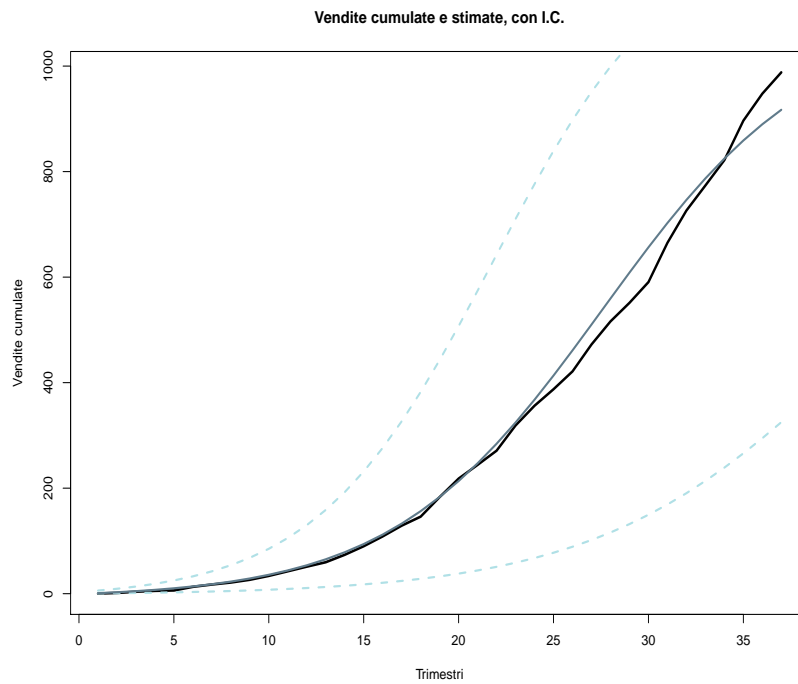


Figura 3.9: Vendite cumulate effettive (in nero), previste dal modello (in blu) e relativo intervallo di credibilità (linee tratteggiate azzurre).

## Capitolo 4

# Diverse distribuzioni *a priori* portano a diverse previsioni?

Le aspettative dei consumatori hanno grandi ripercussioni nell'andamento del mercato e quindi nell'economia, specialmente se si considerano prodotti venduti su scala mondiale.

Molte decisioni economiche quindi non dipendono solamente dalla situazione attuale (Blanchard, Amighini e Giavazzi, 2010), addirittura le decisioni prese all'interno di un sistema economico potrebbero non essere per niente influenzate dagli eventi presenti. Per questo motivo in questo capitolo saranno valutati diversi scenari, in modo da vedere se col cambiamento delle aspettative sul futuro cambiano anche, nel nostro caso, le previsioni di vendita.

Le diverse aspettative e conoscenze dell'andamento economico verranno incluse nelle diverse distribuzioni *a priori* dei parametri del modello (3.1): nel §4.1, da un punto di vista retrospettivo, verranno utilizzate densità *a priori* adattate al periodo storico in cui sono state rilevate le prime vendite<sup>1</sup>, mentre nel §4.2 verrà preso in esame l'andamento del mercato, ipotizzando un suo possibile ampliamento.

---

<sup>1</sup>Quindi con un'aspettativa pessimistica del futuro.

## 4.1 Densità *a priori* basate sul periodo storico

*"Quando gli Stati Uniti starnutiscono, il resto del mondo prende il raffreddore<sup>2</sup>."*

Apple Inc. ha commercializzato i primi esemplari di iPhone il 29 giugno 2007 negli USA, ma esso è diventato un bene disponibile globalmente<sup>3</sup> (in 70 paesi, tra cui l'Italia) solamente l'11 luglio 2008.

Si nota che il suo debutto nel mercato internazionale avviene proprio in concomitanza con l'avanzare della crisi che è esplosa nel 2007 negli Stati Uniti, in seguito allo scoppio di una bolla immobiliare e al conseguente fallimento della società d'investimenti Lehman Brothers.

Questa enorme crisi finanziaria si è poi rapidamente estesa a tutte le principali economie avanzate ed emergenti del mondo, cambiando così la percezione del mercato da parte degli investitori e dei consumatori, alterando le loro aspettative.

Solamente nel biennio 2011-2012 ci sono stati alcuni segnali di ripresa da questa "Grande Recessione", che ha visto l'avanzare della crisi di alcuni debiti sovrani e l'intervento della Banca Centrale Europea e del Fondo Monetario Internazionale per "salvare" la situazione.

In questo periodo, dati i segnali di ripresa, anche le opinioni dei consumatori riguardo il futuro si sono modificate, prevedendo periodi più rosei: questo è possibile notarlo anche nella Figura 3.1, dove la crescita delle vendite a partire dal I-II trimestre del 2012 è aumentata quasi ad un tasso esponenziale rispetto a quella osservata nei primi periodi di vendita di iPhone.

In questo paragrafo prenderemo in considerazione le aspettative dei consumatori maturate nel primo periodo delle vendite di iPhone, quindi quelle create durante la "Grande Recessione", ipotizzando la presenza di un mercato potenziale contenuto ed un tasso d'*innovazione* basso.

---

<sup>2</sup>Fonte: Harold Callender, "Europe Wonders How Long U.S. Vote Will Hamstring It", *New York Times*, 9 Novembre 1952.

<sup>3</sup>Fonte: <https://it.wikipedia.org/wiki/IPhone>.



**Distribuzioni *a priori* utilizzate.** Per raggiungere le ipotesi appena accennate, sono state utilizzate delle densità *a priori* poco diffuse e con media più bassa di quelle assunte nel capitolo precedente, lasciando invariata la famiglia distributiva di ogni parametro.

Infatti, dato il periodo di credenze pessimistiche sulla condizione del mercato, si suppone che il mercato raggiungibile sia contenuto e da ciò deriva la media e varianza relativamente basse della distribuzione *a priori* di  $M$ . Inoltre, per enfatizzare la congettura appena menzionata, sono state scelte come densità *a priori* di  $p$  e  $q$  densità concentrate attorno valori più bassi di quelli abitualmente osservati nel mercato per questi tipi di prodotti.

In dettaglio, le distribuzioni *a priori* assunte per i vari parametri sono:

- $M \sim \mathcal{N}(390, 50)$ ;
- $p \sim \mathcal{Be}(1.1, 950)$ ;
- $q \sim \mathcal{Be}(5, 50)$ ;
- $\sigma^2 \sim \mathcal{SBeta2}(0.75, 0.5, 25)$ .

I valori di riferimento per i parametri di ogni densità *a priori* sono stati ottenuti con lo stesso procedimento usato in §3.3<sup>4</sup>. Con essi inoltre sarà inizializzato l'algoritmo di stima delle distribuzioni *a posteriori*.

**Diagnostica, distribuzioni *a posteriori* e stime puntuali e intervallari dei parametri.** Dopo aver ricavato 100000 simulazioni dalla funzione `stima.parametri` utilizzata anche nel Capitolo 3, dopo un periodo di *burn-in* di 25000 iterazioni, sono state ottenute delle *catene* molto buone per ciascun parametro, caratterizzate da stazionarietà, come si può notare dai rispettivi `trace plots` e `cumuplots`.

---

<sup>4</sup>Sono state utilizzate solamente le prime 15 osservazioni, per cercare di cogliere il trend caratterizzante il periodo di crisi.

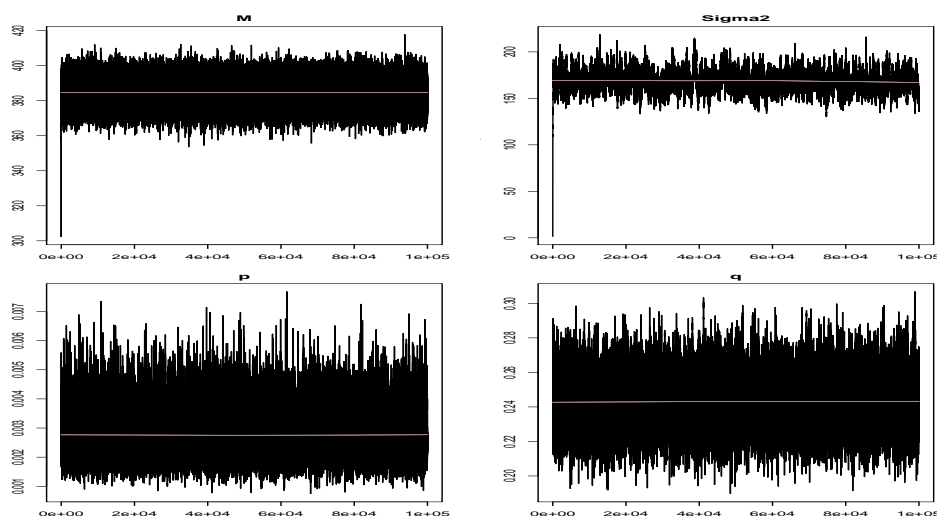


Figura 4.1: trace plots dei parametri.

Per quanto riguarda le ACF in figura 4.2, esse riflettono un'autocorrelazione che diminuisce molto velocemente all'aumentare del lag (nel caso di  $p$  e  $q$ ) o addirittura assente per quanto riguarda  $M$ . Nel caso di  $\sigma^2$  invece l'autocorrelazione dei valori della catena è più alta, ma anch'essa decresce, anche se più lentamente, all'aumentare del lag.

Questo è un segnale positivo riguardo l'esplorazione completa dello spazio parametrico e per la valutazione della convergenza delle stime ottenute.

Inoltre, i cumuplots in figura 4.3 ci mostrano ciò che avevamo già intuito dai trace plots: le catene sono stazionarie, infatti l'andamento della mediana di ciascuna di esse è stabile, così come lo è l'andamento dei quantili 0.025 e 0.975.

Tutto ciò infine è confermato dai test non parametrici di *Heidelberger-Welch* e di *Geweke*, che restituiscono risultati in linea con la stazionarietà delle catene.

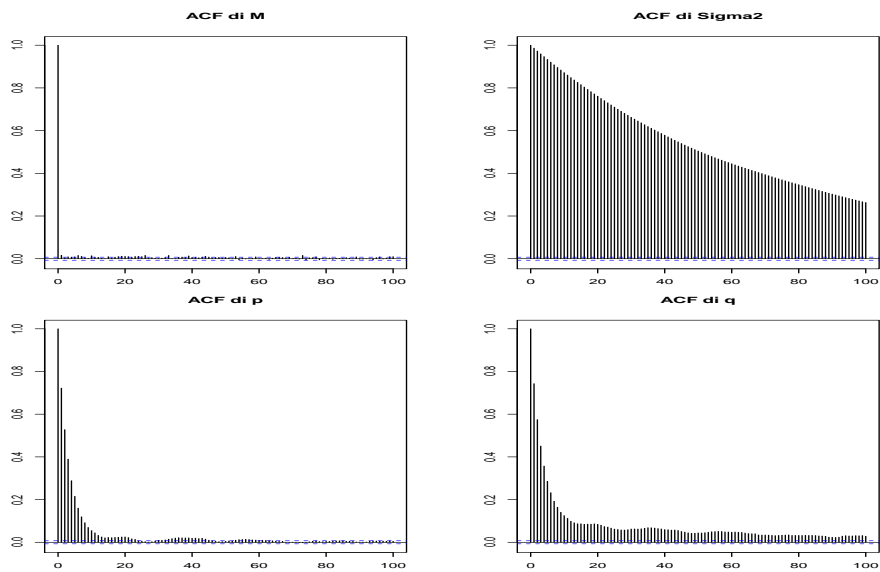


Figura 4.2: ACF dei parametri.

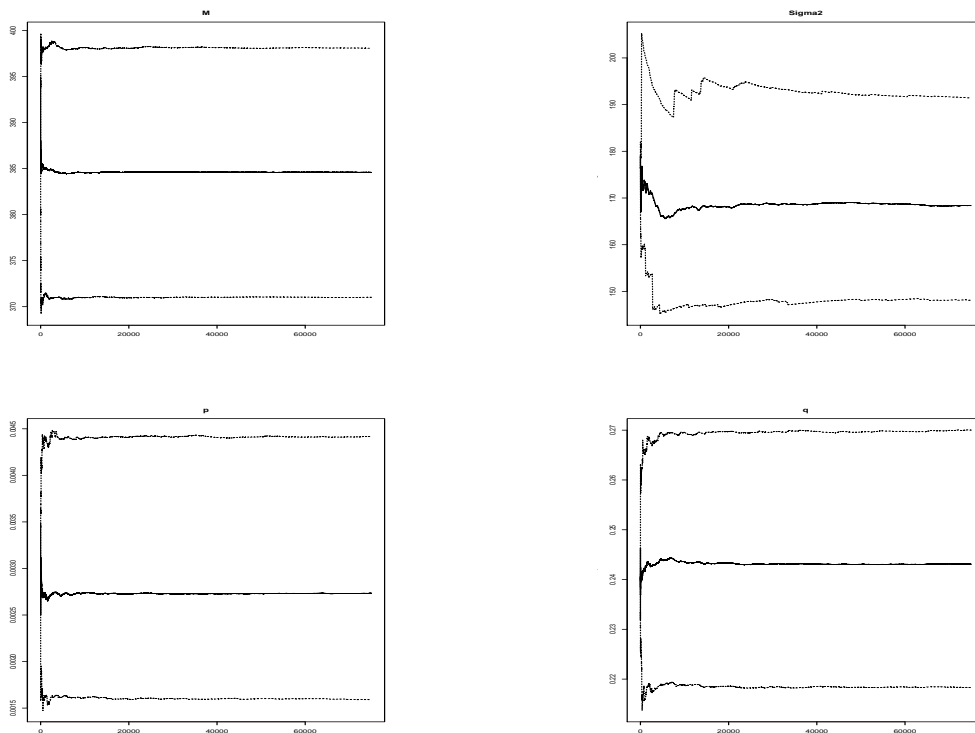


Figura 4.3: cumplots delle catene di  $M$ ,  $\sigma^2$ ,  $p$  e  $q$ .

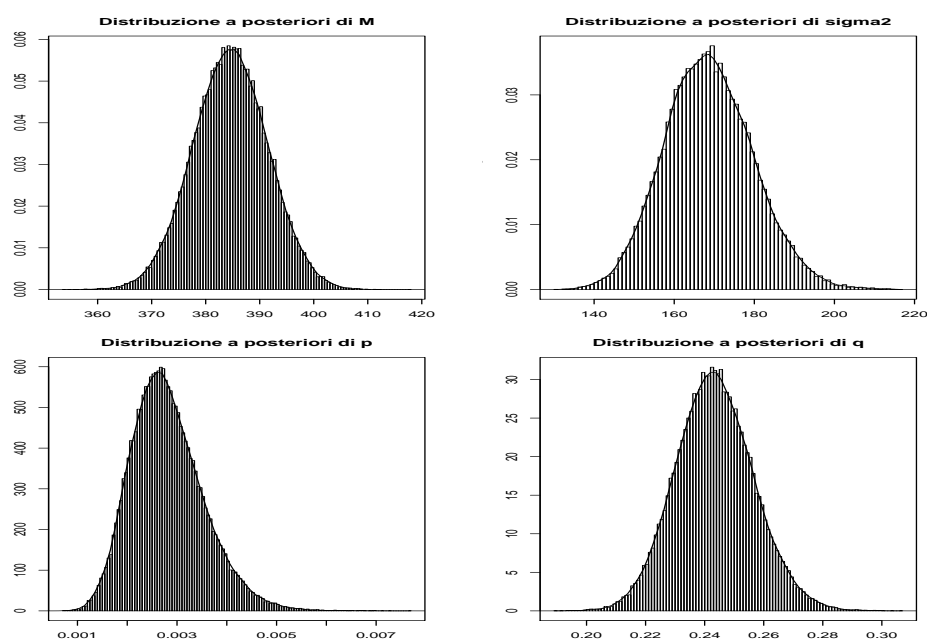


Figura 4.4: Distribuzioni *a posteriori* dei parametri.

Come possiamo vedere, sono state ottenute delle distribuzioni molto concentrate attorno a certo valori, proprio per la scelta di densità *a priori* poco diffuse.

In questo caso, dato che si è presupposto uno scenario "pessimistico" riguardo il sistema economico e il suo sviluppo futuro, si ha la distribuzione indicante il mercato  $M$  che è concentrata attorno ad un valore abbastanza basso<sup>5</sup> così come avviene per quella indicante il tasso d'innovazione  $p$ .

Tutto ciò si riflette in una previsione del ciclo di vita del prodotto più breve, come si vede nella figura 4.7, in cui già prima del 37-esimo trimestre sembra che il prodotto abbia raggiunto il picco di vendite.

Vediamo ora i dettagli delle distribuzioni *a posteriori* trovate.

<sup>5</sup>Se si confronta con quello ottenuto nel capitolo precedente.

Distribuzioni *a posteriori* di  $M$  e  $\sigma^2$  e indicatori di sintesi.

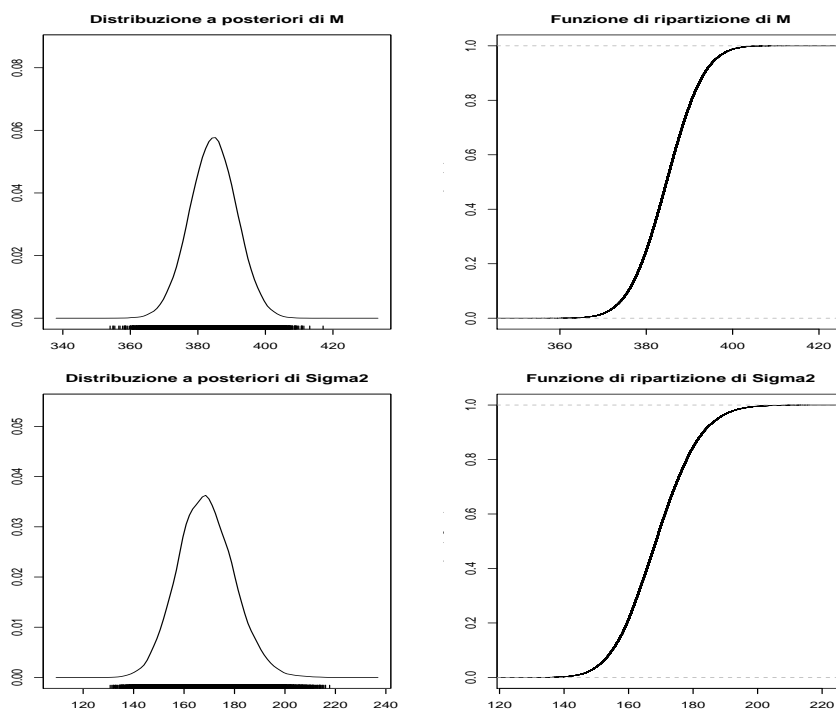


Figura 4.5: Distribuzioni *a posteriori* e funzioni di ripartizione di  $M$  e  $\sigma^2$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
371.02	379.91	384.59	389.21	398.06	384.56	0.0347	371.021	398.064

Tabella 4.1: Informazioni sulla distribuzione *a posteriori* di  $M$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
147.988	161.05	168.35	175.96	191.40	168.689	0.412	146.88	190.07

Tabella 4.2: Informazioni sulla distribuzione *a posteriori* di  $\sigma^2$ .

Distribuzioni *a posteriori* di  $p$  e  $q$  e indicatori di sintesi.

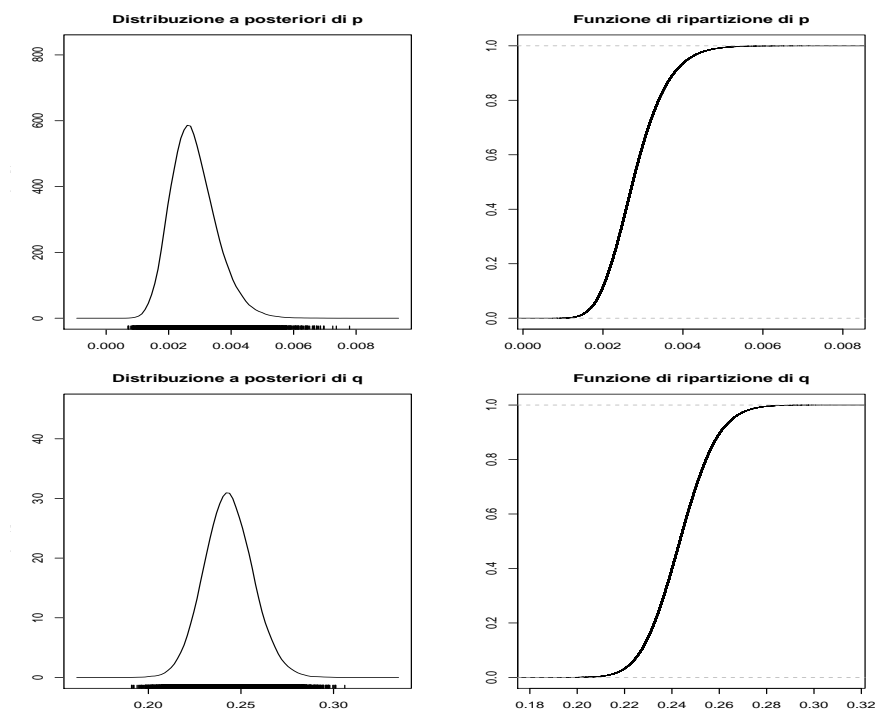


Figura 4.6: Distribuzioni *a posteriori* e funzioni di ripartizione di  $p$  e  $q$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
0.0016	0.0022	0.0027	0.0032	0.0044	0.00281	7.4e-06	0.00147	0.0043

Tabella 4.3: Informazioni sulla distribuzione *a posteriori* di  $p$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
0.2184	0.2345	0.2431	0.2520	0.2700	0.2434	0.00019	0.2180	0.2696

Tabella 4.4: Informazioni sulla distribuzione *a posteriori* di  $q$ .

**Applicazione ai dati e stime ottenute.** Le stime ottenute dalle distribuzioni *a posteriori* sono le seguenti:

Media				Moda			
M	p	q	$\sigma^2$	M	p	q	$\sigma^2$
384.564	0.00281	0.243	168.689	394.854	0.00344	0.233	176.387

Tabella 4.5: Stime dei parametri ottenute.

E di conseguenza, il modello atto a descrivere i dati che deriva da esse è il seguente:

$$Y(\mathbf{t}; M, p, q, \sigma) \sim \mathcal{N}\left(384.564 \cdot \frac{1 - e^{-0.246 \cdot t}}{1 + 86.659 \cdot e^{-0.246 \cdot t}}, 168.689\right) \quad (4.1)$$

Come possiamo vedere dalla figura 4.7, se le aspettative "pessimistiche" caratterizzanti il primo periodo di vendita di iPhone, dovute alla crisi appena iniziata, fossero rimaste fino ad oggi, le sue vendite sarebbero state molto più basse di quelle attuali.

Ciò è dato dal fatto che la gente, aspettandosi una recessione dell'economia, non compra prodotti che necessitano di una sorta di investimento (sappiamo tutti che il prezzo di iPhone è tutt'altro che conveniente): questo si riflette in un mercato totale esiguo, con pochi innovatori e quindi in volumi di vendita molto bassi. Infatti, la curva che stima le vendite non arriva neanche a metà della curva che rappresenta le vendite reali<sup>6</sup>.

---

<sup>6</sup>Esse infatti sono aumentate esponenzialmente dopo gli accenni di ripresa economica avvenuta nel biennio 2011-2012.

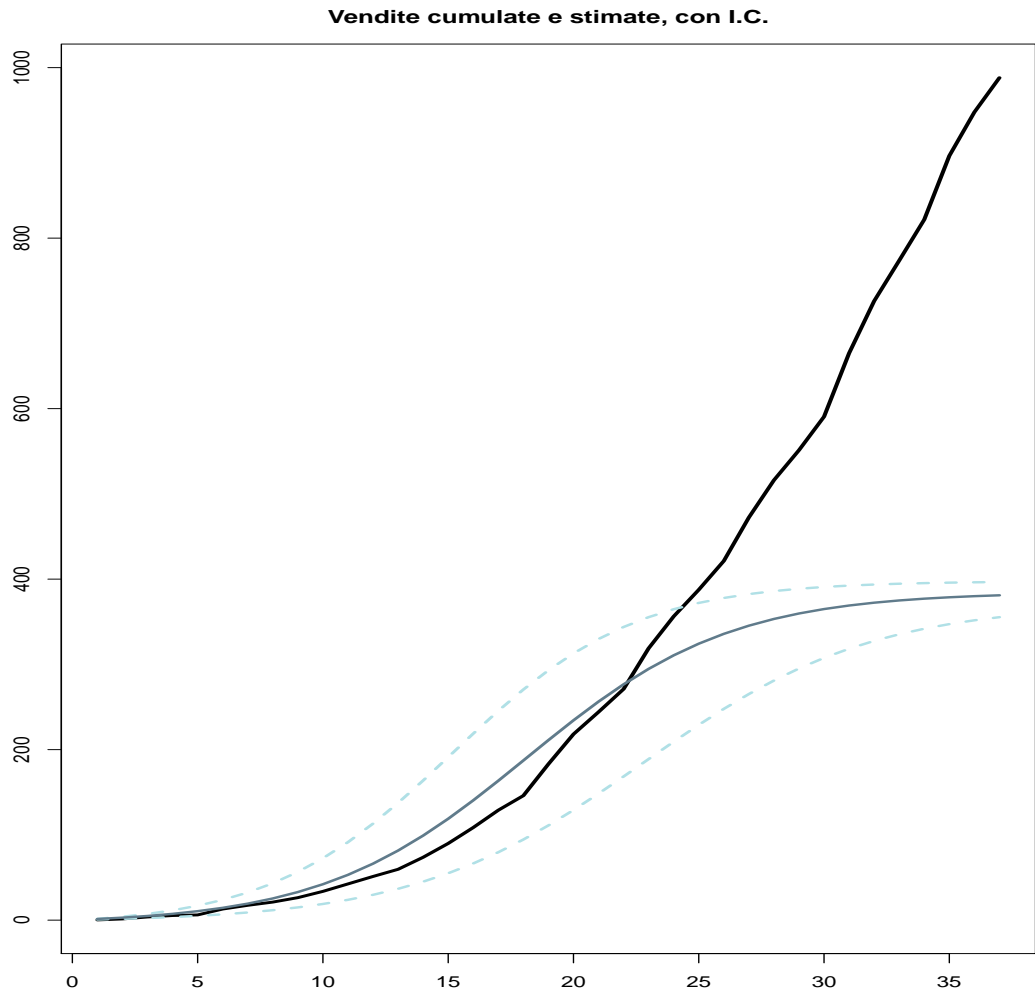


Figura 4.7: Vendite cumulate effettive (in nero) e previste dal modello ottenuto (in blu), con relative bande di confidenza.



## 4.2 Densità *a priori* basate sulla valutazione del mercato

Grazie alle strategie di marketing adottate per promuovere Iphone, Apple Inc. è riuscita a costruire un *brand* molto potente, riuscendo, sin dall'inizio, a vendere i suoi nuovi prodotti globalmente.

É noto che il prezzo di questo smartphone non è accessibile a tutti, perciò, in questo paragrafo, ipotizzando un mercato più ampio dovuto all'aumento della disponibilità di Iphone, si modellerà la densità *a priori* di  $M$  utilizzando valori più alti per i parametri che la caratterizzano, in modo da vedere l'effetto di una credenza più ottimistica, di quanto già lo è, riguardo la dimensione del mercato totale raggiungibile, sulle previsioni di vendita.

Oltre ad assumere un mercato potenziale più ampio, si parte già col presupposto di un'*innovazione* di successo aspettandosi più *imitatori* che *innovatori*: l'influenza del cosiddetto *passaparola* (*word-of-mouth*) è molto più forte delle influenze esterne, facendo intuire la soddisfazione dei clienti dovuta alla bontà del prodotto.

**Distribuzioni *a priori* utilizzate.** Per implementare nel nostro modello le ipotesi appena indicate, sono state utilizzate le seguenti distribuzioni *a priori* per i parametri:

- $M \sim \mathcal{N}(2500, 50)$ ;
- $p \sim \mathcal{Be}(4, 10)$ ;
- $q \sim \mathcal{Be}(6, 3)$ ;
- $\sigma^2 \sim \mathcal{SBeta2}(0.5, 0.5, 625)$ .

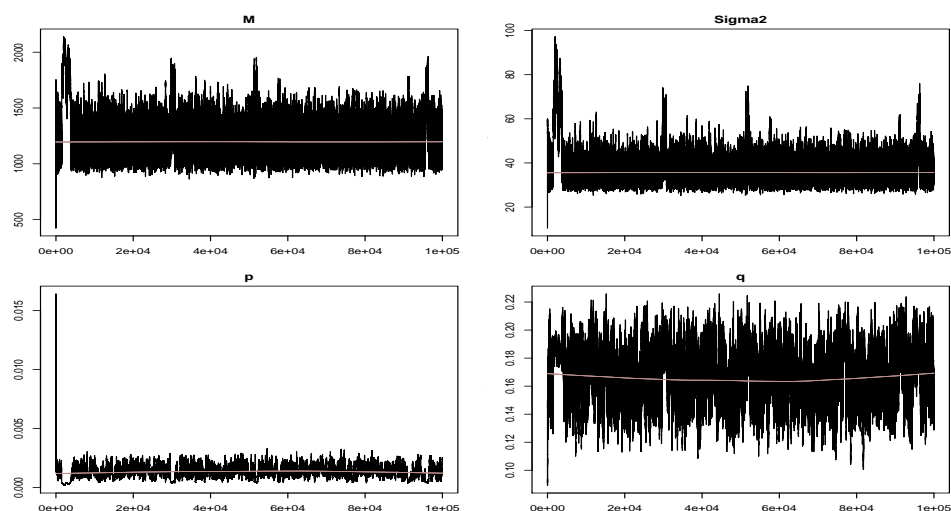


Figura 4.8: trace plots dei parametri.

**Diagnostica, distribuzioni *a posteriori* e stime puntuali e intervallari dei parametri.** Dopo le consuete 100000 iterazioni ed un periodo di *burn-in* di 25000 iterazioni sono stati ottenuti buoni risultati: dai trace plots in Figura 4.8 possiamo vedere che è stato esplorato adeguatamente lo spazio parametrico e che le catene risultanti sembrano stazionarie.

Possiamo notare che è avvenuta una sufficiente esplorazione dello spazio parametrico anche dalle ACF in Figura 4.9: l'autocorrelazione dei valori di ogni catena si abbassa molto all'aumentare del lag.

Per quanto riguarda la verifica della stazionarietà delle *catene*, i *cumplots* forniscono una risposta positiva: anche se, per esempio notiamo una instabilità nei quantili 0.975% di  $M$  e  $\sigma^2$ , essi rispecchiano solo un' "escursione esterna" (Robert e Casella, 2010) della catena di Markov, dato che la mediana resta stabile.

I test non parametrici di *Heidelberger-Welch* e di *Geweke*, qui non riportati, comunque forniscono risultati in linea con quelli ricavati dall'osservazione dei *cumplots*.

4.2. "A PRIORI" BASATE SULLA VALUTAZIONE DEL MERCATO 65

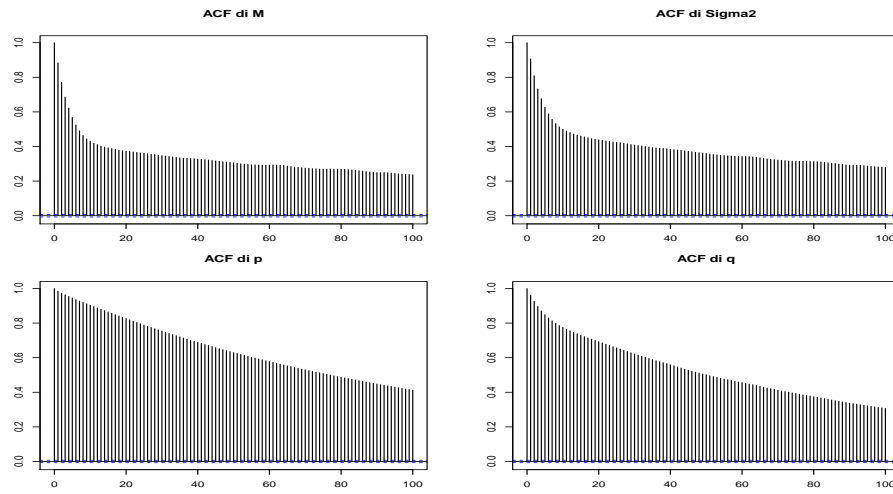


Figura 4.9: ACF dei parametri.

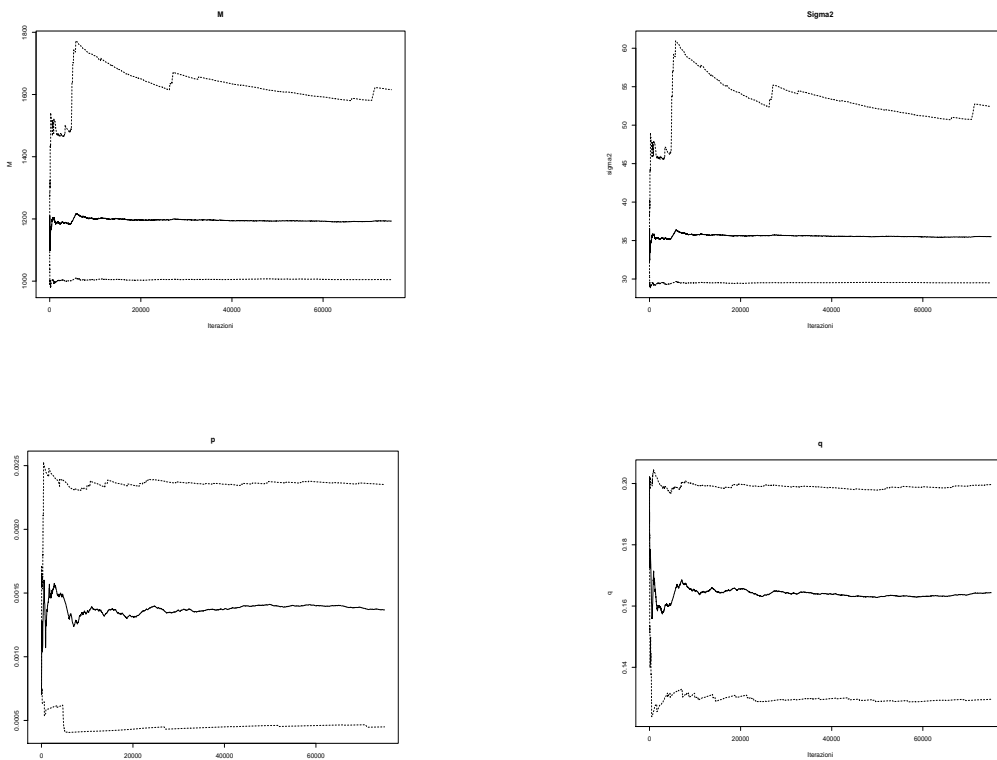


Figura 4.10: cumuplots delle catene di  $M$ ,  $\sigma^2$ ,  $p$  e  $q$ .

Le densità *a posteriori* risultanti sono le seguenti:

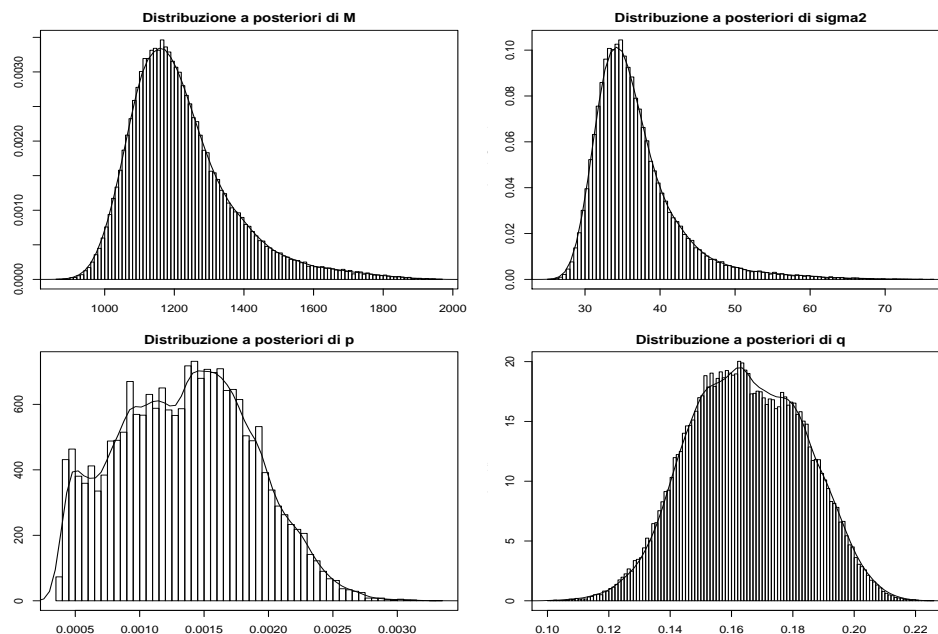


Figura 4.11: Distribuzioni *a posteriori* dei parametri.

In questo caso abbiamo ottenuto delle distribuzioni *a posteriori* asimmetriche per quanto riguarda i parametri  $M$  e  $\sigma^2$  ed una senza particolare forma per  $p$ .

Tutto ciò è dovuto alla forzatura che è stata operata durante la stima<sup>7</sup>, infatti i dati tendono a portarla "nella retta via". L'asimmetria della densità *a posteriori* di  $M$  è dovuta alla sua *prior* debole, nel senso che essa non è adatta a descrivere i dati, ma appunto forzata a portare a stime del relativo parametro più alte, così come la strana forma della *posterior* di  $p$  è dovuta alla stessa cosa, la debolezza della sua densità *a priori*.

<sup>7</sup>In modo da ottenere il desiderato effetto di "ampliamento" del mercato.

4.2. "A PRIORI" BASATE SULLA VALUTAZIONE DEL MERCATO 67

Distribuzioni *a posteriori* di  $M$  e  $\sigma^2$  e indicatori di sintesi.

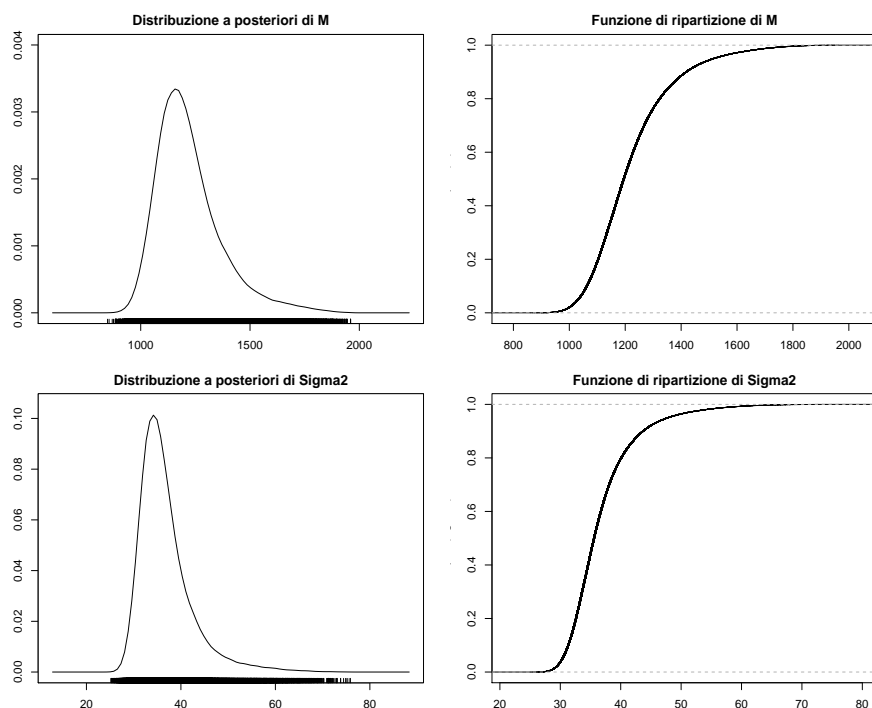


Figura 4.12: Distribuzioni *a posteriori* e funzioni di ripartizione di  $M$  e  $\sigma^2$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
1004.78	1117.66	1193.12	1290.87	1615.45	1220.69	4.874	968.79	1536.46

Tabella 4.6: Informazioni sulla distribuzione *a posteriori* di  $M$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
29.5013	33.0342	35.5102	38.8876	52.4032	36.7495	0.19908	28.3517	48.5803

Tabella 4.7: Informazioni sulla distribuzione *a posteriori* di  $\sigma^2$ .

Distribuzioni *a posteriori* di  $p$  e  $q$  e indicatori di sintesi.

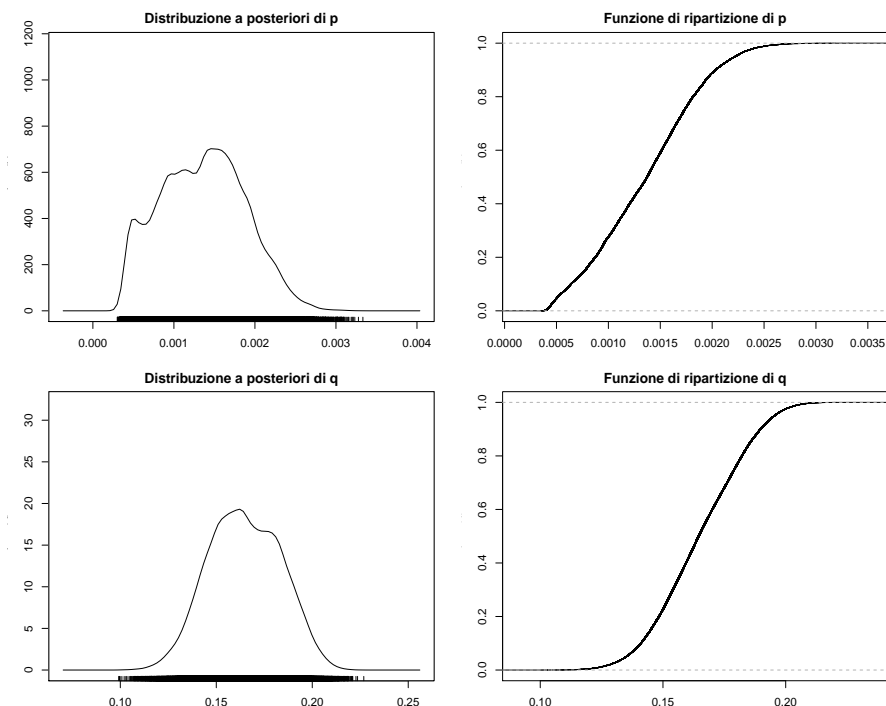


Figura 4.13: Distribuzioni *a posteriori* e funzioni di ripartizione di  $p$  e  $q$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
0.00045	0.00095	0.0014	0.0017	0.0024	0.0014	2.3e-05	0.00040	0.0022

Tabella 4.8: Informazioni sulla distribuzione *a posteriori* di  $p$ .

Quantili					Indicatori di sintesi			
2.5%	25%	50%	75%	97.5%	Media	Sd	I.C.	
0.1296	0.1511	0.16435	0.1789	0.1996	0.1648	0.00074	0.1306	0.2005

Tabella 4.9: Informazioni sulla distribuzione *a posteriori* di  $q$ .

**Applicazione ai dati e stime ottenute.** Le stime ottenute dalle densità *a posteriori* sono le seguenti:

Media				Moda			
$M$	$p$	$q$	$\sigma^2$	$M$	$p$	$q$	$\sigma^2$
1220.693	0.00135	0.16481	36.74952	1116.435	0.000707	0.20218	32.128

Tabella 4.10: Stime dei parametri ottenute.

Ed il modello atto a descrivere i dati è quindi il seguente:

$$Y(\mathbf{t}; M, p, q, \sigma) \sim \mathcal{N}\left(1220.693 \cdot \frac{1 - e^{-0.1661 \cdot t}}{1 + 121.649 \cdot e^{-0.1661 \cdot t}}, 36.749\right) \quad (4.2)$$

Dalla Figura 4.14 possiamo vedere che le previsioni ottenute ipotizzando un mercato più ampio non si discostano molto dalle vendite cumulate reali: ciò può significare che in realtà Apple, con il suo *smartphone* ha già raggiunto una quota molto ampia del mercato<sup>8</sup>.

Questa congettura è molto evidente nei trimestri successivi al 32-esimo, dove le due curve delle vendite previste e delle vendite reali combaciano: qui il mercato raggiunto è aumentato ulteriormente, dato che le vendite reali incontrano le aspettative di un mercato più ampio.

Anche l'intervallo di credibilità è simile a quello trovato precedentemente visibile in figura 3.9: ciò è dovuto al fatto che le stime della varianza ottenute nei due casi sono molto simili.

Quindi, anche se il prezzo di iPhone fosse ridotto o se la vendita di tale prodotto fosse estesa a qualche altro paese non ancora raggiunto, le vendite cumulate non aumenterebbero di molto, perché Apple è già stata capace di saturare il mercato con il suo prodotto.

<sup>8</sup>Ipotesi molto veritiera se pensiamo che iPhone è venduto, come già detto in §4.1, in 70 paesi.

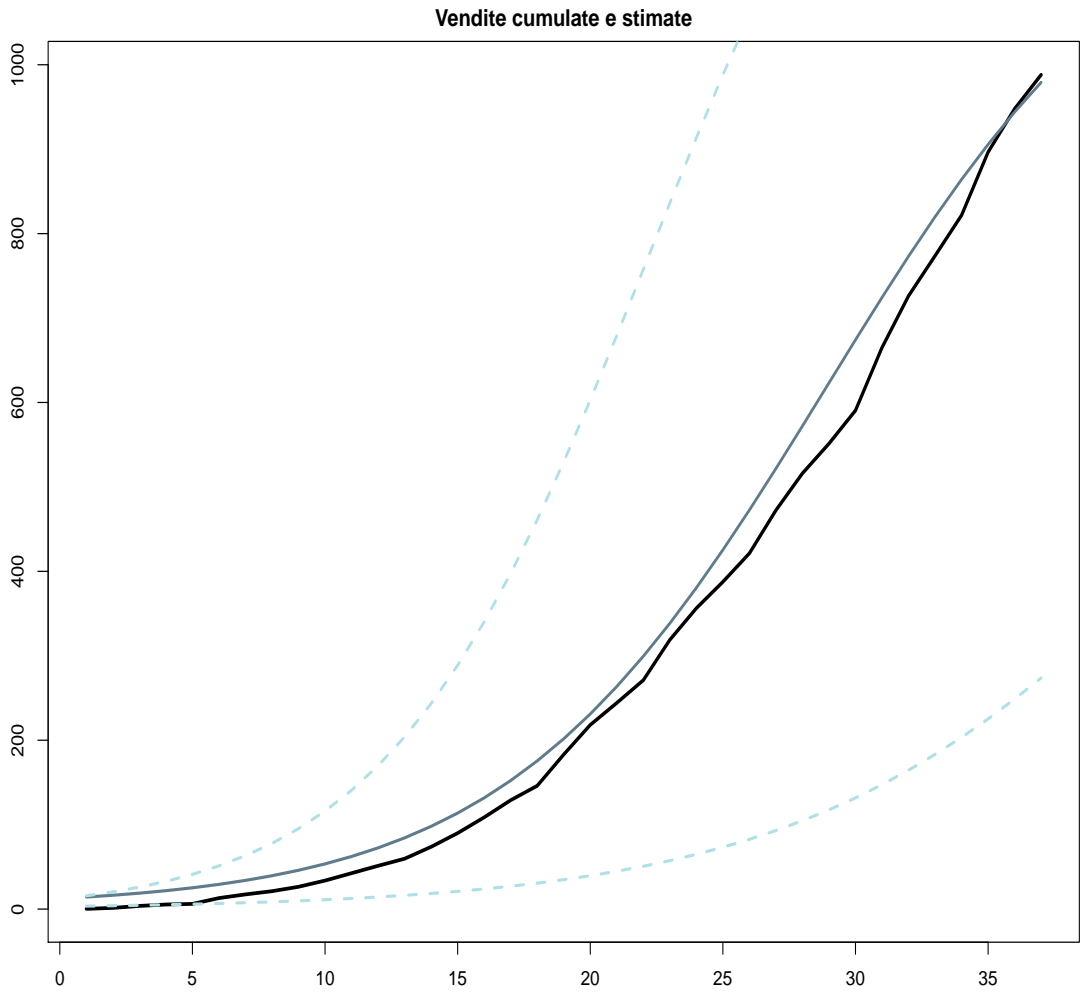


Figura 4.14: Vendite cumulate effettive (in nero) e previste dal modello ottenuto (in blu), con relativo intervallo di credibilit .



# Capitolo 5

## Conclusioni

L'obiettivo principale di questa relazione era studiare il comportamento delle previsioni ottenute tramite il Modello di Bass, con i parametri stimati mediante procedure di stima bayesiane, in modo da incorporare, eventualmente, informazioni utili al perfezionamento delle previsioni.

Nel Capitolo 3 si è visto che anche adottando distribuzioni *a priori* abbastanza diffuse, e quindi carenti di informazione aggiuntive a quelle che portano i dati, sono state ottenute previsioni molto vicine alle vendite effettive.

Se invece si inseriscono determinate *credenze* nel modello tramite densità *a priori* molto concentrate verso particolari valori, come è stato fatto nel Capitolo 4, si nota un cambiamento nelle stime delle vendite ottenute.

In conclusione, l'inferenza bayesiana applicata al Modello di Bass è molto utile, perché consente di inserire in esso ulteriori informazioni e credenze che si hanno sul mercato, sul tasso d'*innovazione* e su quello d'*imitazione*, in modo da indirizzare le previsioni di vendita verso scenari più adatti alla realtà.

Comunque, per ottenere previsioni ancora più precise, tenendo conto anche di altri fattori sia esterni che interni che possono entrare in gioco, si suggerisce di utilizzare altri modelli, come per esempio il Modello di Bass Generalizzato od altre sue modifiche come il Modello Guseo-Guidolin.



# Appendice A

## Codice R

### A.1 Codice utilizzato nel capitolo 3

#### A.1.1 Stima dei parametri iniziali

```
1 dati <- read.csv("apple--iphone-unit-sales.csv", sep="," , header=T, row.
    names = 1) #il dataset è in milioni
sales <- dati[,1]
3 cumsales <- cumsum(sales)

5 #log verosimiglianza dei dati:
llik.dati <- function(par,t, dati){
7   m <- par[1]
   p <- par[2]
9   q <- par[3]
   sigma <- par[4]
11  -length(dati)/2*log(sigma^2) - (1/(2*sigma^2))*sum((dati-m*((1-exp(-(p+q)*t)
    )/(1+(q/p)*exp(-(p+q)*t))))^2)
}
13

#stima dei parametri di inizializzazione delle catene
15 nllik.dati <- function(par, t, dati) -llik.dati(par, t, dati)
par.stimati <- nlminb(start = c(500,0.001,0.4,20),nllik.dati,lower=c
    (100,0.0001,0.01,1), upper=c(10^4,1 , 1, 10^2), dati=cumsales, t=1:37)$
par
```

## A.1.2 Stima dei parametri con Metropolis-Hastings e Gibbs Sampling

```

1 stima.parametri <- function(n.iter,par, t, dati,verbose=T){
  m0 <- par[1]
3  p0 <- par[2]
  q0 <- par[3]
5  sigma0 <- par[4]
  vari.p <- vari.q <- vari.m <- vari.sigma <- numeric(n.iter)
7  n1 <- 501 #iterazioni di ogni sottociclo
  vett.p <- vett.q <- vett.sigma <- numeric(n1)
9  vari.p[1] <- p0
  vari.q[1] <- q0
11 vari.m[1] <- m0
  vari.sigma[1] <- sigma0
13 #Distribuzioni a priori dei parametri da stimare con Metropolis-Hastings
  llik.data <- function(m,p,q,sigma, t, dati) -length(dati)/2*log(sigma^2)
    -(1/(2*sigma^2))*sum((dati-m*((1-exp(-(p+q)*t))/(1+(q/p)*exp(-(p+q)*t)
    )))^2)
15 llik.prior.p <- function(p) dbeta(p, 2, 20.5, log=T)
  llik.prior.q <- function(q) dbeta(q, 2, 15.5, log=T)
17 llik.prior.sigma <- function(sigma) dbeta2(sigma,0.75,0.75,50, log=T)
  theta <- 0.25
19 for (i in 1:(n.iter)){
  if(i%100==0 && verbose)
21   print(i)

23   #Metropolis-Hastings per p
  llik.posterior.p <- function(dati,t, m, p, q, sigma) llik.prior.p(p) +
    llik.data(m, p, q, sigma, t, dati)
25  p<- vari.p[i]
  for(j in 2:(n1)) {
27    ps <- p+runif(1,-theta,theta)
    al <-min(1, exp(llik.posterior.p(dati,t, vari.m[i], ps, vari.q[i],
      vari.sigma[i])-llik.posterior.p(dati,t, vari.m[i], p, vari.q[i],
      vari.sigma[i])) )
29    u <- runif(1)
    if(u<al) p <- ps
31    vett.p[j] <- p
  }
33  vari.p[i+1] <- mean(vett.p)

35  #Metropolis-Hastings per q
  llik.posterior.q <- function(dati,t, m, p, q, sigma) llik.prior.q(q) +
    llik.data(m, p, q, sigma, t, dati)
37  q<- vari.q[i]
  for(j in 2:n1) {
39    qs <- q+runif(1,-theta,theta)
    al <-min(1, exp (llik.posterior.q(dati,t, vari.m[i], vari.p[i+1], qs,

```

```

        vari.sigma[i])-llik.posterior.q(dati, t, vari.m[i], vari.p[i+1], q
        , vari.sigma[i])) )
41   u <- runif(1)
        if(u<al) q <- qs
43   vett.q[j] <- q
    }
45   vari.q[i+1]<-mean(vett.q)

47   #Gibbs Sampling per M
        #Distribuzione a priori di M: N(2200,50)
49   m<- vari.m[i]
        mu0 <- 2200
51   sigma20 <- 50 #sigmaquadro
        tmp <- vari.sigma[i]
53   sigma2.primo <- (sigma20*tmp^2)/(length(dati)*sigma20+tmp^2)
        mu.primo <- (mu0/sigma20+sum(dati)/tmp^2)/(1/sigma20+length(dati)/tmp^2)
55   vari.m[i+1] <- rnorm(1, mu.primo, sd=sqrt(sigma2.primo))

57   #Metropolis-Hastings per sigma^2
        llik.posterior.sigma<- function(dati,t, m, p, q, sigma) llik.prior.sigma
            (sigma) + llik.data(m, p, q, sigma, t, dati)
59   sigma <- vari.sigma[i]
        for(j in 2:n1) {
61     sigmas <- sigma+runif(1,-theta,theta)
            al <-min(1, exp (llik.posterior.sigma(dati,t, vari.m[i+1], vari.p[i
                +1], vari.q[i+1], sigmas)-llik.posterior.sigma(dati, t, vari.m[i
                +1], vari.p[i+1], vari.q[i+1], sigma)) )
63     u <- runif(1)
            if(u<al) sigma <- sigmas
65     vett.sigma[j] <- sigma
        }
67     vari.sigma[i+1]<-mean(vett.sigma)
    }
69   return(list(p=vari.p, q=vari.q, m=vari.m, sigma=vari.sigma))
}
71
pa <- stima.parametri(100000,par.stimati,1:37, cumsales)

```

### A.1.3 Diagnostica e rappresentazioni grafiche

```

#tracce plots con linea di lisciamento
2 par(mfrow=c(2,2), mar=c(3,3,2,2))

4 plot(pa$m, type="l", main="M", xlab="Iterazioni", ylab="M")
  lines(loess.smooth(1:length(pa$m), pa$m), col="rosybrown", cex=10)
6
  plot(pa$sigma, type="l", main="Sigma2", xlab="Iterazioni", ylab="sigma2")
8 lines(loess.smooth(1:length(pa$sigma), pa$sigma), col="rosybrown", cex=10)

```

```
10 plot(pa$p, type="l", main="p", xlab="Iterazioni", ylab="p")
   lines(loess.smooth(1:length(pa$p), pa$p), col="rosybrown", cex=10)
12
   plot(pa$q, main="q", type="l", xlab="Iterazioni", ylab="q")
14 lines(loess.smooth(1:length(pa$q), pa$q), col="rosybrown", cex=10)

16
   par(mfrow=c(2,2), mar=c(3,3,3,3))
18
   #ACF
20 acf(pa$m[25000:100000], lag.max=70, main="ACF di M")
   acf(pa$sigma[25000:100000], lag.max=70, main="ACF di Sigma2")
22 acf(pa$p[25000:100000], lag.max=70, main="ACF di p")
   acf(pa$q[25000:100000], lag.max=70, main="ACF di q")
24
   s1 <- as.mcmc(pa$sigma)
26 p1 <- as.mcmc(pa$p)
   q1 <- as.mcmc(pa$q)
28 m1 <- as.mcmc(pa$m)

30
   #diagnostica
32
   #cumuplots
34 cumuplot(as.mcmc(p1[25000:100000]), main="p", xlab="Iterazioni", ylab="p")
   cumuplot(as.mcmc(q1[25000:100000]), main="q", xlab="Iterazioni", ylab="q")
36 cumuplot(as.mcmc(m1[25000:100000]), main="M", xlab="Iterazioni", ylab="M")
   cumuplot(as.mcmc(s1[25000:100000]), main="Sigma2", xlab="Iterazioni", ylab="
       sigma2")
38
   #test non parametrici
40
   #Heidelberger-Welch
42 heidel.diag(as.mcmc(m1[25000:100000]))
   heidel.diag(as.mcmc(s1[25000:100000]))
44 heidel.diag(as.mcmc(p1[25000:100000]))
   heidel.diag(as.mcmc(q1[25000:100000]))
46
   #Geweke
48 geweke.diag(as.mcmc(m1[25000:100000]))
   geweke.diag(as.mcmc(s1[25000:100000]))
50 geweke.diag(as.mcmc(p1[25000:100000]))
   geweke.diag(as.mcmc(q1[25000:100000]))
52
   #Grafici delle varie distribuzioni a posteriori
54 #stimo la densità
   hp <- h.select(pa$p[25000:100000])
56 hq <- h.select(pa$q[25000:100000])
   hm <- h.select(pa$m[25000:100000])
```

```

58 hsigma <- h.select(pa$sigma[25000:100000])

60 par(mfrow=c(2,2))

62 hist(pa$m[25000:100000], nclass=100, prob=T, main="Distribuzione a
    posteriori di M", xlab="M", ylab="f_post(M)")
    sm.density(pa$m[25000:100000],h=hm, add=T, cex=10)
64
    hist(pa$sigma[25000:100000], nclass=100, prob=T, main="Distribuzione a
    posteriori di sigma2", xlab="sigma2", ylab="f_post(sigma2)")
66 sm.density(pa$sigma[25000:100000],h=hsigma, add=T, cex=10)

68 hist(pa$p[25000:100000], nclass=100, prob=T, main="Distribuzione a
    posteriori di p", xlab="p", ylab="f_post(p)")
    sm.density(pa$p[25000:100000],h=hp, add=T, cex=10)
70
    hist(pa$q[25000:100000], nclass=100, prob=T, main="Distribuzione a
    posteriori di q", xlab="q", ylab="f_post(q)")
72 sm.density(pa$q[25000:100000],h=hq,add=T, cex=10)

74
    #Informazioni di sintesi sulle distribuzioni
76
    #grafici densità e funzione di ripartizione
78 par(mfrow=c(2,2))

80 sm.density(pa$m[25000:100000],h=hm, cex=10, xlab="M", ylab="f_post(M)")
    title(main="Distribuzione a posteriori di M")
82 plot(ecdf(pa$m[25000:100000]), do.points=F, main="Funzione di ripartizione
    di M", xlab="M", ylab="F_post(M)")

84 sm.density(pa$sigma[25000:100000],h=hsigma, cex=10, xlab="sigma2", ylab="f_
    post(sigma2)")
    title(main="Distribuzione a posteriori di Sigma2")
86 plot(ecdf(pa$sigma[25000:100000]), do.points=F, main="Funzione di
    ripartizione di Sigma2", xlab="sigma2", ylab="F_post(sigma2)")

88 sm.density(pa$p[25000:100000],h=hp, cex=10, xlab="p", ylab="f_post(p)")
    title(main="Distribuzione a posteriori di p")
90 plot(ecdf(pa$p[25000:100000]), do.points=F, main="Funzione di ripartizione
    di p", xlab="p", ylab="F_post(p)")

92 sm.density(pa$q[25000:100000],h=hq, cex=10, xlab="q", ylab="f_post(q)")
    title(main="Distribuzione a posteriori di q")
94 plot(ecdf(pa$q[25000:100000]), do.points=F, main="Funzione di ripartizione
    di q", xlab="q", ylab="F_post(q)")

96 #informazioni sulle distribuzioni
    quantile(pa$m[25000:100000], c(0.025,0.25,0.5,0.75,0.975))
98 quantile(pa$sigma[25000:100000], c(0.025,0.25,0.5,0.75,0.975))

```

```

quantile(pa$p[25000:100000], c(0.025,0.25,0.5,0.75,0.975))
100 quantile(pa$q[25000:100000], c(0.025,0.25,0.5,0.75,0.975))

102 m2 <- mean(pa$m[25000:100000])
s2 <- mean(pa$sigma[25000:100000])
104 p2 <- mean(pa$p[25000:100000])
q2 <- mean(pa$q[25000:100000])
106
#calcolo della moda
108
mode <- function(x){
110   ux <- unique(x)
   ux[which.max(tabulate(match(x,ux)))]
112 }

114 mode(a$p[25000:100000])
mode(a$q[25000:100000])
116 mode(a$m[25000:100000])
mode(a$sigma[25000:100000])
118
#sd calcolati con il metodo delle medie batch
120 sd.m <- bm(m1[25000:100000])$se
sd.sigma <- bm(s1[25000:100000])$se
122 sd.p <- bm(p1[25000:100000])$se
sd.q <- bm(q1[25000:100000])$se
124

126 #I.C.
ic.m <- hdi(as.mcmc(m1[25000:100000]),0.95)
128 ic.sigma <- hdi(as.mcmc(s1[25000:100000]),0.95)
ic.p <- hdi(as.mcmc(p1[25000:100000]),0.95)
130 ic.q <- hdi(as.mcmc(q1[25000:100000]),0.95)

132
#Applicazione ai dati
134
bass.model <- function(par,dati,t){
136   m <- par[1]
   p <- par[2]
138   q <- par[3]
   sigma2 <- par[4]
140   m*((1-exp(-(p+q)*t))/(1+(q/p)*exp(-(p+q)*t)))+rnorm(dati, sd=sqrt(sigma2))
}
142
par(mfrow=c(1,1))
144 test <- bass.model(c(m2,p2, q2, s2),cumsales,1:37)
test1 <- bass.model(par.stimati,cumsales,1:37)
146 plot(cumsales,type="l", xlab="Trimestri", ylab="Vendite cumulate", main="
Vendite cumulate e stimate", lwd=3)
lines(test, col="lightskyblue4", lwd=2.5)

```



```
148 lines(test1, col="rosybrown", lwd=2.5)

150
#Grafico con I.C.
152 test.lw <- bass.model(c(ic.m[1],ic.p[1],ic.q[1],ic.sigma[1]),cumsales,1:37)
test.up <- bass.model(c(ic.m[2],ic.p[2],ic.q[2],ic.sigma[2]),cumsales,1:37)
154 plot(cumsales,type="l", xlab="Trimestri", ylab="Vendite cumulate", main="
      Vendite cumulate e stimate, con I.C.", lwd=3)
lines(test.lw, col="powderblue", lwd=2.5, lty=2)
156 lines(test.up, col="powderblue", lwd=2.5, lty=2)
lines(test, col="lightskyblue4", lwd=2.5)
```



# Appendice B

## Distribuzione Beta del Secondo tipo Scalata<sup>1</sup>

La famiglia delle distribuzioni Beta del Secondo tipo Scalate è molto utile per ottenere densità a priori robuste, sia informative che non informative, per il parametro di scala, alternative alla classica Gamma-Inversa.

Essa è abbastanza flessibile da permettere la modellazione di una varietà di comportamenti all'origine e sulle code, e inoltre può essere generata in vari modi, come per esempio da una mistura Gamma scalata di densità Gamma o da una quota scalata di una distribuzione Beta.

Una variabile casuale  $Y$  ha distribuzione Beta del Secondo tipo Scalata, con parametri di forma  $\alpha$  e  $\gamma > 0$  e parametro di scala  $\beta > 0$ , e si scrive sinteticamente  $Y \sim SBeta2(\alpha, \beta, \gamma)$ , se  $Y$  ha supporto  $S_Y = [0, 1]$  e funzione di densità di probabilità

$$f_Y(y) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \cdot \gamma \frac{\left(\frac{y}{\gamma}\right)^{\alpha-1}}{\left(\frac{y}{\gamma} + 1\right)^{\alpha+\beta}} \quad (\text{B.1})$$

per  $y \in S_Y$  e  $f_Y(y; \alpha, \beta, \gamma) = 0$  altrove, dove  $\Gamma(\cdot)$  indica la funzione Gamma. Nelle figure B.1, B.2 e B.3 è possibile osservare il variare della sua funzione di densità, con rispettiva funzione di ripartizione, al variare dei tre parametri.

---

<sup>1</sup>M. Pérez, L. R. Pericchi, I. C. Ramirez, *The Scaled Beta2 distribution as a robust prior for scales*, 2017.

## B.1 Proprietà della SBeta2

Per la distribuzione SBeta2 definita in (B.1) si ha che:

$$E[Y] = \frac{\alpha}{\beta - 1} \gamma \quad \text{con } \beta > 1$$

$$\text{Var}[Y] = \frac{\alpha(\alpha + \beta - 1)}{(\beta - 1)^2(\beta - 2)} \gamma^2 \quad \text{con } \beta > 2$$

Una proprietà interessante di questa distribuzione è che se  $Y \sim SBeta2(\alpha, \beta, \gamma)$  allora  $X = \frac{1}{Y} \sim SBeta2(\beta, \alpha, \frac{1}{\gamma})$ : se  $f_Y(y)$  appartiene alla famiglia delle distribuzioni SBeta2 allora anche  $f_X(x)$  appartiene ad essa (proprietà di reciprocità)<sup>2</sup>.

Altre caratteristiche della distribuzione sono:

1. comportamento allo zero:

$$f_Y(0) = \begin{cases} \infty & \alpha < 1 \\ \frac{\beta}{\gamma} & \alpha = 1 \\ 0 & \alpha > 1 \end{cases}$$

2. quando  $\beta \leq 1$  si ha che  $E[Y] = \infty$ ;
3. posizionamento della moda:

$$\text{moda}(Y) = \begin{cases} \frac{\alpha-1}{\beta+1} \gamma & \alpha \geq 1 \\ 0 & \text{altrimenti} \end{cases}$$

4. quando  $\alpha = \beta$ , la mediana della SBeta2 coincide con il parametro di scala  $\gamma$ . Invece, con  $\alpha = 1$ , la mediana si può così calcolare:  $\gamma \cdot (2^{1/\beta} - 1)$ .

---

<sup>2</sup>Di conseguenza, se  $\sigma^2$  è distribuito come una SBeta2, anche il parametro di precisione  $\tau$  lo è.

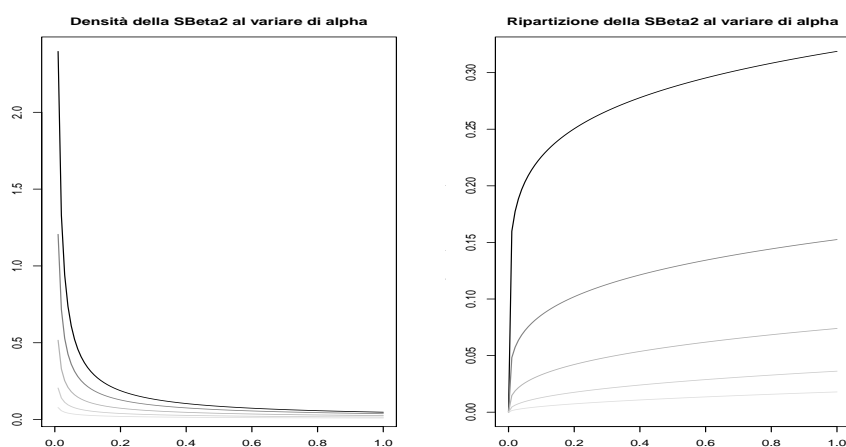


Figura B.1: Funzione di densità e di ripartizione della  $SBeta2(\alpha, \beta, \gamma)$  al variare di  $\alpha$  ( $\alpha=0.15, 0.25, 0.35, 0.45, 0.55$ , in ordine di gradazione:  $\alpha=0.15$  nero,  $\alpha=0.55$  grigio chiarissimo),  $\beta=0.5$ ,  $\gamma=625$ .

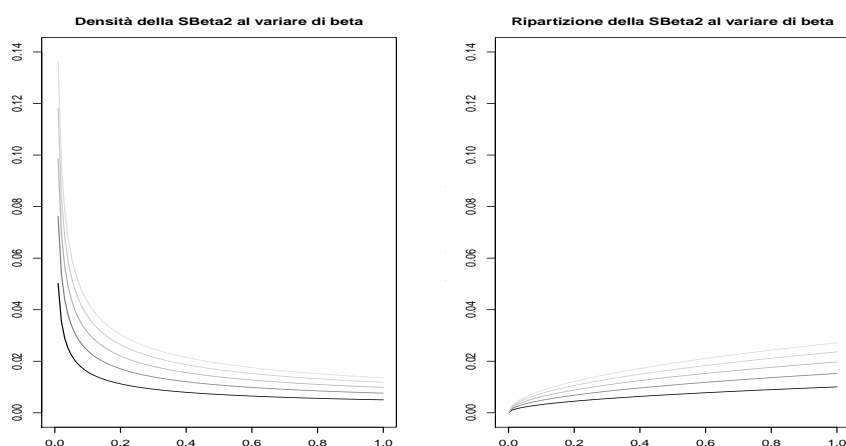


Figura B.2: Funzione di densità e di ripartizione della  $SBeta2(\alpha, \beta, \gamma)$  al variare di  $\beta$  ( $\beta=0.15, 0.25, 0.35, 0.45, 0.55$ , in ordine di gradazione:  $\beta=0.15$  nero,  $\beta=0.55$  grigio chiarissimo),  $\alpha=0.5$ ,  $\gamma=625$ .

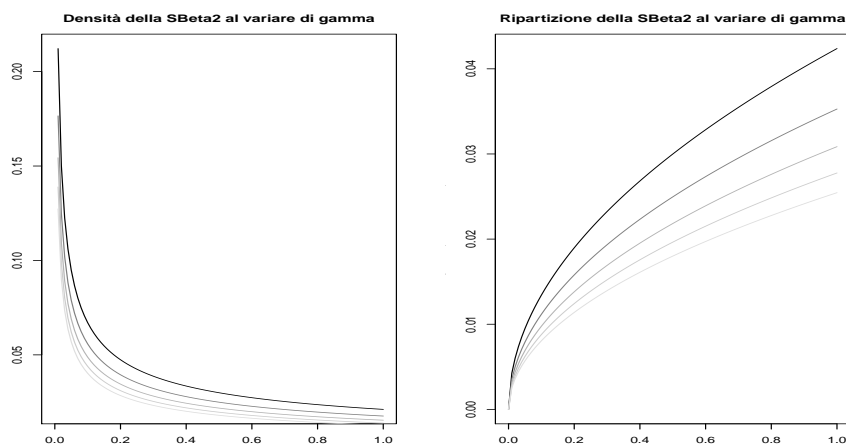


Figura B.3: Funzione di densità e di ripartizione della  $SBeta2(\alpha, \beta, \gamma)$  al variare di  $\gamma$  ( $\gamma=225, 325, 425, 525, 625$ , in ordine di gradazione:  $\gamma=225$  nero,  $\gamma=625$  grigio chiarissimo),  $\alpha=0.5$ ,  $\beta=0.5$ .



# Bibliografia

- [1] Albert, J.: *Bayesian Computation with R*. Springer, 2007.
- [2] Bass, F. M.: *A new product growth for model consumer durables*. Management Science, 15(5):215–227, Gennaio 1969.
- [3] Bertsch, M. *et al.*: *Analisi Matematica*. McGraw Hill, 2009. Seconda edizione.
- [4] Blanchard, O., Amighini A. e Giavazzi F.: *Macroeconomia. Una prospettiva europea*. Il Mulino, 2010.
- [5] Bunea, A. M.: *Modelli alla Bass: stima ed inferenza*. tesi di laurea, Università degli Studi di Pisa, 2014.
- [6] Chiandotto, B.: *Inferenza Statistica*. 2015.
- [7] Fourt, L.A. e Woodlock J.W.: *Early prediction of market success for new grocery products*. Journal of Marketing, 25:31–38, Ottobre 1960.
- [8] Geweke, J.: *Bayesian Statistics 4*. Oxford University Press, 1992. Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments.
- [9] Heidelberger, P. e Welch P.D.: *Simulation Run Length Control in the Presence of an Initial Transient*. Operations Research, (31), 1983.
- [10] Kotler *et al.*: *Marketing Management*. Pearson, 2012.
- [11] Lesa, G.: *Concorrenza e pirateria: Modello di Bass esteso e serie latenti*. tesi di laurea, Università degli Studi di Padova, 2000.

- [12] Liseo, B.: *Introduzione alla statistica bayesiana. Appunti delle lezioni*. 2008.
- [13] Mahajan, V. e Peterson R. A.: *Models for innovation diffusion*. SAGE Publications Inc, 1985.
- [14] Mansfield, E.: *Technical change and the rate of imitation*. *Econometrica*, 29:741–766, Ottobre 1961.
- [15] Peschechera, L. F.: *Previsione delle vendite di prodotti con applicazioni nel settore delle marmitte per auto*. tesi di laurea, Università degli Studi di Padova, 2004.
- [16] Pérez, M., Pericchi L. R. e Ramírez I. C.: *The Scaled Beta2 distribution as a Robust Prior for scales*. *Bayesian Analysis*, 12(3):615–637, 2017.
- [17] Robert, C.P. e Casella G.: *Introducing Monte Carlo methods with R*. Springer, 2010.
- [18] Valdani, E. e Busacca B.: *Previsione delle vendite e ciclo di vita del prodotto*. Etas, 1987.



