

# Indice

<b>Introduzione</b>	<b>3</b>
<b>1 Analisi dei valori estremi: impostazione classica</b>	<b>5</b>
1.1 Introduzione . . . . .	5
1.2 Modelli asintotici . . . . .	6
1.2.1 Formulazione del modello . . . . .	6
1.2.2 La distribuzione dei valori estremi generalizzata . . . . .	8
1.2.3 Il dominio di attrazione . . . . .	10
1.3 Inferenza per la distribuzione GEV . . . . .	11
1.3.1 Considerazioni generali . . . . .	11
1.3.2 Massima Verosimiglianza . . . . .	12
1.3.3 Verosimiglianza profilo . . . . .	13
1.4 Il modello basato sulle statistiche d'ordine . . . . .	13
1.4.1 Formulazione del modello . . . . .	13
1.4.2 Modellare le $r$ più grandi statistiche d'ordine . . . . .	16
1.5 Una possibile caratterizzazione alternativa: valori che superano una soglia	17
1.5.1 Distribuzione Generalizzata di Pareto (GPD) . . . . .	17
1.5.2 Modello Poisson-GPD e stime dei parametri . . . . .	19
<b>2 Caratterizzazione dei valori estremi mediante il processo di Poisson</b>	<b>21</b>
2.1 Introduzione . . . . .	21
2.2 Convergenza in legge . . . . .	22
2.3 Il modello statistico . . . . .	23
2.4 Connessioni con gli altri modelli . . . . .	25
2.5 Processi aleatori a tempo continuo . . . . .	26
2.6 Processi di Poisson . . . . .	28
2.6.1 Processo di Poisson omogeneo . . . . .	28
2.6.2 Processo di Poisson non-omogeneo, da finire . . . . .	29

<b>3</b>	<b>Analisi dei valori estremi: impostazione bayesiana parametrica</b>	<b>31</b>
3.1	Introduzione . . . . .	31
3.2	Scelta della distribuzione iniziale . . . . .	32
3.3	Modelli gerarchici . . . . .	34
3.3.1	Raggruppamento di popolazioni eterogenee tramite distribuzioni mistura . . . . .	35
3.3.2	Un modello mistura gerarchico . . . . .	36
<b>4</b>	<b>Alcuni elementi dell'impostazione bayesiana non-parametrica</b>	<b>39</b>
4.1	Introduzione . . . . .	39
4.2	Alcune priors utili nell'impostazione bayesiana non-parametrica . . . . .	39
4.2.1	Processo di Dirichlet . . . . .	40
4.2.2	Processi di Lévy . . . . .	41
4.2.3	Processi neutrali a destra . . . . .	42
4.2.4	Il processo Gamma Esteso . . . . .	47
4.3	Derivazione della misura di Lévy per alcuni processi . . . . .	51
<b>5</b>	<b>Inferenza bayesiana non-parametrica per un processo di Poisson non- omogeneo</b>	<b>53</b>
5.1	Introduzione . . . . .	53
5.2	Verosimiglianza e distribuzioni a priori non-parametriche . . . . .	54
5.3	Distribuzioni a posteriori . . . . .	55
<b>6</b>	<b>Valori estremi e processi d'urna</b>	<b>57</b>
6.1	Introduzione . . . . .	57
6.2	Concetti preliminari . . . . .	57
6.3	Processi di urna rinforzati (RUP) . . . . .	59
6.4	Processi di urna rinforzati per il processo beta-Stacy . . . . .	61
<b>7</b>	<b>Una applicazione finanziaria</b>	<b>65</b>
7.1	Introduzione . . . . .	65
7.2	Utilizzo dei modelli dell'analisi classica sui valori estremi . . . . .	66
7.3	Utilizzo dell'impostazione bayesiana non-parametrica sui valori estremi . . . . .	70
7.4	Applicazione dei processi di urna rinforzati ai valori estremi . . . . .	74
	<b>Bibliografia</b>	<b>77</b>

# Introduzione

La teoria dei valori estremi è l'unica disciplina statistica in cui si sviluppano tecniche e modelli per descrivere fenomeni inusuali. E' circa il 1950 quando si cominciano a proporre metodologie ad hoc per lo studio di fenomeni fisici.

Per definizione, i valori estremi sono scarsi, questo significa che le stime sono spesso richieste per livelli di un processo che sono più grandi di quelli fino a quel momento osservati. E' il caso, ad esempio, dello studio del livello dei fiumi, dove ciò che interessa realmente non è saper prevedere il generico andamento del fiume, ma la probabilità con cui si possono verificare alluvioni, considerati appunto *valori estremi*. Chiaramente le applicazioni di un approccio di questo genere sono infinite: record sportivi, inquinamento atmosferico, rimborsi assicurativi e titoli azionari.

Nel presente lavoro sono state prese in esame le serie giornaliere di cinque dei trenta titoli azionari del comparto industriale che compongono l'indice sintetico del Dow-Jones. I titoli considerati sono: IBM, General Motors, American Express, Coca Cola, Johnson&Johnson. Il periodo di osservazione a cui si fa riferimento per tutta l'analisi è il decennio 1990-2000 e le osservazioni sono date dai relativi prezzi di chiusura giornalieri. Lo scopo di questa analisi è di analizzare i valori estremi (superamenti rispetto alla soglia) di queste serie. L'obiettivo di questa tesi è quello di proporre un "percorso metodologico" in cui si parte dall'analisi dei valori estremi in termini classici, per individuare la soglia ottimale al di sopra della quale si parla di valore estremo, fino ad arrivare all'utilizzo dell'impostazione bayesiana non-parametrica in cui tali valori vengono considerati come realizzazioni di un processo di salto su cui fare previsione. Successivamente, proprio sulla base delle proprietà del processo di salto, tra cui la ricorrenza, queste osservazioni sono state pensate come possibili processi di urna rinforzati (RUP).

Nella prima parte del lavoro tutto questo è stato descritto in termini teorici, mentre nella seconda parte se ne è data una applicazione



# Capitolo 1

## Analisi dei valori estremi: impostazione classica

### 1.1 Introduzione

La teoria dei valori estremi è una disciplina statistica in cui si sviluppano tecniche e modelli che descrivono comportamenti inconsueti. Intesa come studio astratto di fenomeni aleatori, questa materia affonda le sue radici in alcuni lavori del 1920 sviluppati da Fischer e Tippett. Successivamente, nel 1940 Gnedenko propose l'utilizzo di questa teoria nelle applicazioni statistiche. L'applicazione statistica dei modelli probabilistici per gli estremi fu studiata e formalizzata da Gumbel nel 1950. In quegli stessi anni, Jenkinson applicò questi modelli ai valori estremi rilevati nelle velocità del vento e sviluppò un modello di parametrizzazione che unificò diversi modelli, così da aggiungere chiarezza alle procedure di modellazione. Nel 1970 le leggi limite classiche vennero generalizzate da Pickands, portando sostanzialmente ad implementare tali procedure che si svilupparono poi negli anni '80 e '90. Durante questi anni Leadbetter et al. investigarono sul comportamento estremo di classi di processi che ammettevano vari tipi di non-stazionarietà e dipendenza. Infine, la caratterizzazione e l'inferenza statistica di estremi multivariati si svilupparono dalla metà del 1980 e sono tuttora materia di ricerca.

Storicamente, applicazioni della teoria degli eventi estremi si ritrovano nei campi più diversi:

- meteorologia (piogge, venti) - conseguenze su bacini idrici, canali, relazioni con l'aumento della temperatura del globo terrestre;
- livelli di inquinanti (superamento del dato medio orario di qualche livello);
- assicurazioni - previsioni di eventi catastrofici (uragani, piene di fiumi);
- finanza;

- telecomunicazioni;
- ...

I riferimenti bibliografici principali sono Embrechts et al. (1997), Leadbetter et al. (1983) e Resnick (1987).

## 1.2 Modelli asintotici

### 1.2.1 Formulazione del modello

In generale, nello studio di tutti i fenomeni che si ripetono nel tempo, si può essere interessati a conoscere quando (o meglio, con che probabilità) avrà luogo un evento che superi in intensità tutti i precedenti. E' il caso ad esempio dello studio del livello dei fiumi, dove ciò che interessa realmente, non è saper prevedere il generico andamento del fiume, ma la "probabilità" con cui si possono verificare alluvioni, che sono appunto *eventi estremi*. In questo primo capitolo verrà sviluppato il modello che rappresenta il punto di svolta della teoria degli eventi estremi. Tale modello si focalizza sul comportamento di

$$M_n = \max\{X_1, \dots, X_n\},$$

dove  $X_1, \dots, X_n$ , è una sequenza di variabili aleatorie i.i.d. che hanno una funzione di distribuzione  $F$ . Tutti i risultati che verranno presentati si riferiscono solo alla distribuzione dei massimi  $M_n$ , ricordando che è sempre possibile ottenere risultati analoghi per i minimi attraverso la trasformazione

$$\min\{X_1, \dots, X_n\} = -\max\{-X_1, \dots, -X_n\}.$$

Nelle applicazioni, le  $X_i$  usualmente rappresentano i valori di un processo misurato su una scala di tempo tale che  $M_n$  rappresenti il massimo di un processo su  $n$  tempi. Se  $n$  è il numero di osservazioni in un anno, allora  $M_n$  corrisponde al massimo annuale.

Essendo le variabili i.i.d. è immediato che la distribuzione di  $M_n$  per tutti i valori di  $n$  sia:

$$\begin{aligned} Pr\{M_n \leq x\} &= Pr\{X_1 \leq x, \dots, X_n \leq x\} \\ &= Pr\{X_1 \leq x\} \times \dots \times Pr\{X_n \leq x\} \\ &= \{F(x)\}^n. \end{aligned} \tag{1.1}$$

Tuttavia ciò, non è immediatamente utile in pratica, dato che la funzione di distribuzione  $F$  non è nota. Una possibilità può essere quella di usare una tecnica statistica standard: calcolare la stima di  $F$  dai dati osservati e poi sostituire questa stima nella (1.1). Sfortunatamente, discrepanze molto piccole nella stima di  $F$  possono portare a grosse discrepanze in  $\{F(x)\}^n$ . Un approccio alternativo è quello di accettare che  $F$  sia

non nota, e cercare famiglie di modelli approssimate per  $\{F(x)\}^n$ , che possano essere stimate soltanto sulla base di dati estremi. Si procede osservando il comportamento di  $\{F(x)\}^n$  con  $n \rightarrow \infty$ . Più specificatamente, si tratta di individuare quali distribuzioni limite sono possibili per  $M_n$  con  $n \rightarrow \infty$ .

La questione così posta è banale: la distribuzione di  $M_n$  converge, con probabilità 1, al punto estremo destro di  $F$ . Questo stesso problema si incontra nel Teorema del Limite Centrale: con probabilità 1,  $\bar{X}_n$  converge alla media della popolazione,  $\mu$ . Tale problema viene superato, ammettendo un riscaldamento lineare, tale che:

$$\frac{\bar{X}_n - \mu_n}{\sigma_n} \rightarrow N(0, 1)$$

dove  $\mu_n = \mu$  e  $\sigma_n = \sigma/\sqrt{n}$ .

Si adotta lo stesso approccio per ottenere il limite della distribuzione di  $M_n$ , cercando distribuzioni limite di  $M_n^* = (M_n - b_n)/a_n$ , dove  $\{a_n > 0\}$  e  $\{b_n\}$  sono sequenze di coefficienti normalizzanti. Una scelta appropriata di queste due successioni determina la posizione e la scala di  $M_n$  al crescere di  $n$ .

L'intera classe delle possibili distribuzioni limite per  $M_n$  è dato dal Teorema seguente:

**Teorema 1.2.1.** *Se esistono sequenze di costanti  $\{a_n > 0\}$  e  $\{b_n\}$  tali che*

$$Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G(x) \quad \text{con } n \rightarrow \infty,$$

dove  $G$  è una funzione di distribuzione non-degenere, allora  $G(x)$  appartiene ad una delle seguenti famiglie:

$$I : G(x) = \exp\{-\exp(-x)\} \quad -\infty < x < \infty;$$

$$II : G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\alpha}) & x > 0, \alpha > 0; \end{cases}$$

$$III : G(x) = \begin{cases} \exp\{-(-x)^\alpha\} & x < 0, \alpha > 0 \\ 1 & x \geq 0. \end{cases}$$

Viceversa, ognuna di queste distribuzioni,  $G$ , può apparire come limite per la distribuzione di  $(M_n - b_n)/a_n$  e, infatti, succede così quando la  $G$  è essa stessa distribuzione di  $X$ .

In tal caso  $G$  prende il nome di **distribuzione dei valori estremi**.

In altre parole, il Teorema dice che il massimo di  $\{(M_n - b_n)/a_n\}$  converge in distribuzione ad una variabile che ha come funzione di distribuzione una delle tre famiglie individuate con I, II, III. Collettivamente, queste tre classi di distribuzioni sono chiamate **extreme**

**value distributions**, con i tipi I, II e III rispettivamente noti come famiglie **Gumbel**, **Fréchet**, **Weibull**. Ogni famiglia ha rispettivamente un parametro di posizione ed uno di scala,  $b$  e  $a$ ; inoltre, le famiglie Fréchet e Weibull hanno anche un parametro di forma  $\alpha$ .

### 1.2.2 La distribuzione dei valori estremi generalizzata

Von Mises (1954) e Jenkinson (1955) indipendentemente, hanno ottenuto una singola famiglia che combina queste tre distribuzioni limite, la **generalized extreme value distribution** (GEV), indicata con  $\mathcal{G}(\mu, \sigma, \xi)$ , con funzione di distribuzione

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad (1.2)$$

definita sull'insieme  $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$ , dove i parametri soddisfano  $-\infty < \mu < \infty$ ,  $\sigma > 0$  e  $-\infty < \xi < \infty$ . Il modello ha tre parametri: un parametro di posizione  $\mu$ , un parametro di scala  $\sigma$  ed un parametro di forma  $\xi$  ( $= 1/\alpha$ ). Le classi di tipo II e III della distribuzione dei valori estremi corrispondono in questa parametrizzazione rispettivamente ai casi  $\xi > 0$  e  $\xi < 0$ . Il sottoinsieme della famiglia GEV con  $\xi = 0$  è da interpretarsi come limite della (1.2) con  $\xi \rightarrow 0$ , ottenendo così la **Gumbel family** con funzione di distribuzione

$$G(x) = \exp\left[-\exp\left\{-\left(\frac{x - \mu}{\sigma}\right)\right\}\right], -\infty < x < \infty.$$

Convienne quindi, riformulare il Teorema 1.2.1 in modo diverso.

**Teorema 1.2.2.** *Se esistono sequenze di costanti  $\{a_n > 0\}$  e  $\{b_n\}$  tali che*

$$\Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G(x) \quad \text{con } n \rightarrow \infty, \quad (1.3)$$

*per una funzione di distribuzione  $G$  non-degenere, allora  $G$  è un membro della famiglia GEV*

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

*definita su  $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$ , dove  $-\infty < \mu < \infty$ ,  $\sigma > 0$  e  $-\infty < \xi < \infty$ .*

*Dimostrazione:* Si veda DI CHI E' IL TEOREMA.

Interpretare il limite di questo teorema come una approssimazione per valori grandi di  $n$ , vuol dire utilizzare la famiglia GEV per modellare la distribuzione dei massimi di sequenze lunghe. La difficoltà apparente riguardante il fatto che le costanti normalizzatrici siano sconosciute, in pratica, si risolve facilmente. Partendo dalla (1.3),

$$\Pr\{(M_n - b_n)/a_n \leq x\} \approx G(x)$$



per valori abbastanza grandi di  $n$ . Equivalentemente,

$$\begin{aligned} Pr\{(M_n \leq x)\} &\approx G\{(x - b_n)/a_n\} \\ &= G^*(x), \end{aligned}$$

dove  $G^*$  è un altro membro della famiglia GEV. In altre parole, se il Teorema 1.2.2 consente l'approssimazione della distribuzione di  $M_n^*$  con un membro della famiglia GEV per grandi valori di  $n$ , la distribuzione di  $M_n$  può essere approssimata con un diverso membro della stessa famiglia. Dato che i parametri della distribuzione devono essere comunque stimati, è irrilevante che i parametri della distribuzione  $G$  siano differenti da quelli della  $G^*$ .

Un'applicazione tipica di questo modello consiste nell'adattare la distribuzione (1.2) ad una serie di dati di massimi annuali (per esempio, prendendo  $n$  come il numero di osservazioni in un anno). Le stime dei quantili estremi dei massimi annuali si ottengono invertendo l'equazione (1.2):

$$x_p = \mu - \frac{\sigma}{\xi} \left[ 1 - \{-\log(1-p)\}^{-\xi} \right], \quad (1.4)$$

dove  $G(x_p) = 1 - p$ . Nella terminologia dei valori estremi,  $x_p$ , è il livello di ritorno associato al periodo di ritorno  $1/p$  (ossia  $x_p$  è quel valore che viene superato una volta ogni  $p$  anni). Nella relazione (1.4) assume particolare importanza il parametro di forma  $\xi$  infatti, se  $\xi$  è negativo, i quantili sono limitati superiormente, così che l'estrapolazione ad un qualche livello porterà ad un limite finito. Mentre se  $\xi$  è pari a zero o è positivo, allora la distribuzione limite non è limitata, e l'estrapolazione porta ad un limite infinito. La giustificazione formale del teorema dei modelli estremi è tecnica, sebbene non particolarmente complicata - vedi Leadbetter et al. (1983). Per prima cosa conviene dare la seguente definizione.

**Definizione 1.2.1.** *Una distribuzione  $G$  è detta **max-stabile** se, per ogni  $n = 2, 3, \dots$ , esistono delle costanti  $\alpha_n > 0$  e  $\beta_n$  tali che*

$$G^n(\alpha_n x + \beta_n) = G(x).$$

Dato che  $G^n$  è la funzione di distribuzione di  $M_n = \max\{X_1, \dots, X_n\}$ , dove le  $X_i$  sono variabili indipendenti ognuna con funzione di distribuzione  $G$ , la proprietà di max-stabilità è soddisfatta da quelle distribuzioni per cui l'operazione di prendere i valori massimi del campione porta ad una distribuzione identica tranne per un cambiamento dei parametri di scala e posizione. Il collegamento con le leggi limite dei valori estremi è dato dal seguente risultato.

**Teorema 1.2.3. DI CHI E' IL TEOREMA**

*Una distribuzione è max-stabile se, e solo se, è una distribuzione generalizzata dei valori estremi.*

Questo teorema viene usato direttamente nella dimostrazione del teorema 1.2.2. L'idea è di considerare  $M_{nk}$  come la variabile aleatoria massimo in una sequenza di  $n \times k$  variabili per un valore grande di  $n$ . Questo può essere considerato come il massimo di una singola sequenza  $n \times k$ , oppure come il massimo di  $k$  massimi, ognuno dei quali è il massimo di  $n$  osservazioni. Più precisamente, si suppone che la distribuzione limite di  $(M_n - b_n)/a_n$  sia  $G$ . Così, per valori di  $n$  abbastanza grandi,

$$Pr\{(M_n - b_n)/a_n \leq x\} \approx G(x)$$

dal Teorema 1.2.2. Quindi, per qualche intero  $k$ , dato che  $nk$  è grande,

$$Pr\{(M_{nk} - b_{nk})/a_{nk} \leq x\} \approx G(x). \quad (1.5)$$

Ma, dato che  $M_{nk}$  è il massimo di  $k$  variabili che hanno la stessa distribuzione di  $M_n$ ,

$$Pr\{(M_{nk} - b_{nk})/a_{nk} \leq x\} = [Pr\{(M_n - b_n)/a_n \leq x\}]^k. \quad (1.6)$$

Quindi, dalla (1.5) e dalla (1.6) rispettivamente si ottiene,

$$Pr\{M_{nk} \leq x\} \approx G\left(\frac{x - b_{nk}}{a_{nk}}\right)$$

e

$$Pr\{M_n \leq z\} \approx G^k\left(\frac{z - b_n}{a_n}\right).$$

Quindi,  $G$  e  $G^k$  sono identiche a parte per i coefficienti di posizione e di scala. Ne consegue che  $G$  è max-stabile e, per il Teorema 1.2.3, è anche un membro della famiglia GEV.

### 1.2.3 Il dominio di attrazione

Data una delle distribuzioni limite  $G(x)$ , la *teoria dei valori estremi* determina le condizioni su  $F$  che implicano la convergenza di  $M_n$  ad essa. Tali condizioni si riferiscono alla natura delle due successioni  $a_n$  e  $b_n$  e definiscono il *massimo dominio di attrazione di  $G$*  (indicato con  $MDA(G)$ ).

Ad un livello alto di generalità questa diventa una questione alquanto complicata. Si può comunque dare una caratterizzazione che lavori con funzioni di distribuzione  $F$  assolutamente continue (con funzione di densità  $f$ ). In questo caso, la *funzione  $h$  di rischio reciproco* è definita come

$$h(x) = \frac{1 - F(x)}{f(x)},$$

e ponendo

$$b_n = F^{-1}\left(1 - \frac{1}{n}\right), \quad a_n = h(b_n), \quad \xi = \lim_{x \rightarrow \infty} h'(x)$$

la distribuzione limite di  $(M_n - b_n)/a_n$  è

$$\exp\{-[1 + \xi x]_+^{-1/\xi}\} \quad \text{se } \xi \neq 0$$

e

$$\exp\{-e^{-x}\} \quad \text{se } \xi = 0$$

In particolare, questo risultato implica che il parametro di forma della GEV distribuzione limite dei massimi, è determinato solamente dal limite della derivata prima della funzione di rischio reciproco.

## 1.3 Inferenza per la distribuzione GEV

### 1.3.1 Considerazioni generali

Dato un insieme di valori massimi (ad esempio massimi annuali)  $X_1, X_2, \dots, X_k$ , lo scopo è di fare inferenza sui parametri della  $\mathcal{G}(\mu, \sigma, \xi)$ . A tale proposito, numerose tecniche sono state proposte:

- metodo grafico basato su varianti dei grafici di probabilità;
- metodo dei momenti in cui le funzioni di questo modello sono rese uguali alle funzioni empiriche equivalenti;
- metodo della verosimiglianza.

Ognuna di queste tecniche ha i suoi pro e i suoi contro, ma la versatilità e l'applicabilità del metodo della verosimiglianza a modelli complessi rende questo approccio particolarmente attraente. Una potenziale difficoltà derivante dall'uso di questo metodo per la GEV, riguarda la validità delle condizioni di regolarità richieste per le usuali proprietà asintotiche, associate agli stimatori di massima verosimiglianza. Queste condizioni non sono soddisfatte nel modello GEV perchè i punti estremi della distribuzione sono funzione dei valori assunti dal parametro:  $\mu - \sigma/\xi$  che è l'estremo superiore della distribuzione quando  $\xi < 0$ , e l'estremo inferiore quando  $\xi > 0$ . Smith (1985) studiò questo problema nel dettaglio ed ottenne i seguenti risultati:

- quando  $\xi > -0.5$ , gli stimatori di massima verosimiglianza sono regolari, nel senso che godono delle usuali proprietà asintotiche;
- quando  $-1 < \xi < -0.5$ , gli stimatori di massima verosimiglianza sono generalmente ottenibili, ma non sono regolari;
- quando  $\xi < -1$ , gli stimatori di massima verosimiglianza non esistono.

Il caso in cui  $\xi \leq -0.5$  corrisponde a distribuzioni con coda “molto corta”. Queste situazioni raramente si incontrano nelle applicazioni di modelli dei valori estremi.

### 1.3.2 Massima Verosimiglianza

Sotto l'assunzione che  $X_1, \dots, X_m$  siano delle variabili indipendenti con distribuzione GEV, la log-verosimiglianza per i parametri della GEV, quando  $\xi \neq 0$  è

$$\begin{aligned} \ell(\mu, \sigma, \xi) = & -m \log \sigma - (1 + 1/\xi) \sum_{i=1}^m \log \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right] \\ & - \sum_{i=1}^m \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-1/\xi}, \end{aligned} \quad (1.7)$$

posto che

$$1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) > 0, \quad \text{per } i = 1, \dots, m. \quad (1.8)$$

Per alcuni valori del parametro  $\xi$  per cui viene violata la (1.8), si ottiene una verosimiglianza pari a zero ed una log-verosimiglianza pari a  $-\infty$ .

Il caso in cui  $\xi = 0$  richiede un trattamento separato che usa la distribuzione Gumbel come limite della distribuzione GEV. Questo porta ad una log-verosimiglianza di questo tipo:

$$\ell(\mu, \sigma) = -m \log \sigma - \sum_{i=1}^m \left( \frac{x_i - \mu}{\sigma} \right) - \sum_{i=1}^m \exp \left\{ - \left( \frac{x_i - \mu}{\sigma} \right) \right\}. \quad (1.9)$$

La massimizzazione delle due equazioni (1.7) e (1.9) rispetto al vettore di parametri  $(\mu, \sigma, \xi)$  porta alla stima di massima verosimiglianza rispetto all'intera famiglia GEV. Non esiste però una soluzione analitica ma, per qualche data-set la massimizzazione si può ottenere utilizzando algoritmi di ottimizzazione numerica standard (Hosking et al. (1985)), quando il supporto della distribuzione non dipende dai parametri. E' necessario introdurre dei fattori di correzione per assicurare che questi algoritmi non modifichino le combinazioni dei parametri della (1.8), in modo tale da evitare le difficoltà numeriche che si potrebbero incontrare nella valutazione dell'equazione (1.7) in prossimità di  $\xi = 0$ . Quest'ultimo problema si può risolvere facilmente utilizzando la (1.9) al posto della (1.7), per valori di  $\xi$  che cadono nell'intorno di zero oppure cercando degli stimatori appropriati per il parametro  $\xi$ . Il più noto tra questi stimatori è quello di *Pickands* (VEDERE ANNO).

**Definizione 1.3.1.** *Sia dato un campione  $X_1, \dots, X_n$  estratto da una distribuzione  $F$  appartenente al massimo bacino d'attrazione di una delle distribuzioni del Teorema 1.2.1 e siano  $X_{k,n}$  le sue statistiche d'ordine  $k$ . Allora, fissato  $k = k(n)$  tale che  $k(n) \rightarrow \infty$  e  $k(n)/n \rightarrow 0$ , la funzione*

$$\hat{\xi}_{k,n} = \frac{1}{\log 2} \log \frac{X_{k,n} - X_{2k,n}}{X_{2k,n} - X_{4k,n}}$$

è detta **stimatore di Pickands** per  $\xi$ .

**Proposizione 1.3.1.** *Lo stimatore di Pickands presentato nella definizione 1.3.1 è consistente per  $\xi$ , ovvero*

$$\hat{\xi}_{k,n} \rightarrow \xi, \quad n \rightarrow \infty.$$

DIMOSTRAZIONE - Si veda Dekkers e de Haan (1989).

Oltre a quello di Pickands sono stati proposti altri stimatori per  $\xi$  tra cui lo **stimatore di Hill** per  $\xi = \alpha^{-1} > 0$  (si veda Embrechts et al. (1997) per una breve rassegna).

### 1.3.3 Verosimiglianza profilo

Un metodo per fare inferenza sulla massima verosimiglianza è quello che si basa sulla verosimiglianza profilo. La valutazione numerica della verosimiglianza profilo per uno qualsiasi dei parametri  $\mu, \sigma$  o  $\xi$  è immediata. Per esempio, per ottenere il profilo della verosimiglianza per  $\xi$ , si fissa  $\xi = \xi_0$  e si massimizza la log-verosimiglianza (1.7) rispetto ai parametri rimanenti,  $\mu$  e  $\sigma$ . Questa procedura viene ripetuta per un range di valori di  $\xi_0$ . I corrispondenti valori massimizzati della log-verosimiglianza costituiscono il profilo della log-verosimiglianza per  $\xi$ , dalla quale è possibile ottenere intervalli di confidenza approssimati.

Questa metodologia si può applicare anche quando viene richiesta l'inferenza su alcune combinazioni di parametri. In particolare, si possono ottenere intervalli di confidenza per livelli di ritorno,  $x_p$ , specificati. Ciò richiede una riparametrizzazione del modello GEV, tale che  $x_p$  diventi uno dei parametri, dopo di che si ottiene il profilo della log-verosimiglianza dalla massimizzazione rispetto ai rimanenti parametri nel modo usuale. La riparametrizzazione è:

$$\mu = x_p + \frac{\sigma}{\xi} [1 - \{-\log(1 - p)\}^{-\xi}], \quad (1.10)$$

a questo punto si inserisce la (1.10) nella (1.7) ottenendo il modello GEV in termini dei parametri  $(x_p, \sigma, \xi)$ .

## 1.4 Il modello basato sulle statistiche d'ordine

### 1.4.1 Formulazione del modello

Una difficoltà implicita nell'analisi dei valori estremi è data dal limitato ammontare dei dati per la stima del modello. Per definizione, gli estremi sono scarsi, questo comporta

una grande varianza nelle stime del modello. L'emergere di questo problema, ha motivato la ricerca di caratterizzazioni alternative del comportamento dei valori estremi, che siano in grado di modellare i dati in modo diverso.

Esistono due ben note caratterizzazioni generali. Una è basata sulle eccedenze rispetto ad una soglia alta, l'altra si basa sul comportamento delle  $r$  più grandi statistiche d'ordine. Entrambe le caratterizzazioni possono essere unificate usando la rappresentazione di un processo di punto. In questa sezione si focalizzerà l'attenzione sulla seconda caratterizzazione.

Si suppone che  $X_1, X_2, \dots$  sia una sequenza di variabili aleatorie i.i.d., e lo scopo è di caratterizzare il comportamento estremo delle  $X_i$ . Nell'equazione (1.3) si è ottenuta la distribuzione limite, con  $n \rightarrow \infty$ , di  $M_n$  che opportunamente normalizzata è una GEV. Quindi come prima cosa si estende questo risultato alle statistiche d'ordine estreme, definendo

$$M_n^{(k)} = k - \text{esima statistica d'ordine di } \{X_1, \dots, X_n\},$$

e identificando il comportamento limite di questa variabile, per  $k$  fissato, con  $n \rightarrow \infty$ . Il risultato che segue è una generalizzazione del teorema 1.2.1.

**Teorema 1.4.1.** *Se esistono due sequenze di costanti  $\{a_n > 0\}$  e  $\{b_n\}$  tali che*

$$Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G(x) \quad \text{con } n \rightarrow \infty,$$

dove  $G$  è una funzione di distribuzione non-degenere, tale che  $G$  sia una funzione di distribuzione data dall'equazione (1.2), allora, per valori fissati di  $k$ ,

$$Pr\{(M_n^{(k)} - b_n)/a_n \leq x\} \rightarrow G_k(x)$$

su  $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$ , dove

$$G_k(x) = \exp\{-\tau(x)\} \sum_{s=0}^{k-1} \frac{\tau(x)^s}{s!} \tag{1.11}$$

$$\tau(x) = \left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}.$$

*Dimostrazione:* Si veda AUTORE TEOREMA.

Questo teorema implica che, se la  $k$ -esima più grande statistica d'ordine in un blocco è normalizzata esattamente allo stesso modo del massimo, allora la sua distribuzione limite è della forma data dalla (1.11), dove i parametri corrispondono ai parametri della distribuzione limite GEV del massimo di un blocco.

Il problema che si verifica più di frequente, utilizzando la (1.11), è che ognuna delle

$r$  più grandi statistiche d'ordine si trova in blocchi diversi, per alcuni valori di  $r$ . Cioè, si ha il vettore completo

$$M_n^{(r)} = (M_n^{(1)}, \dots, M_n^{(r)})$$

per ogni blocco. Però, mentre il Teorema 1.4.1 dà una famiglia per la distribuzione approssimata di ognuna delle componenti di  $M_n^{(r)}$ , non dà la distribuzione congiunta di  $M_n^{(r)}$ . Inoltre, le componenti non possono essere indipendenti:  $M_n^{(2)}$  non può essere più grande di  $M_n^{(1)}$ , per esempio, quindi il risultato di ognuna delle componenti influenza la distribuzione delle altre. Quello che si sta cercando è la distribuzione limite congiunta dell'intero vettore  $M_n^{(r)}$ . Il prossimo teorema dà la funzione di densità congiunta della distribuzione limite.

**Teorema 1.4.2.** (METTERE AUTORE)

Se esistono due sequenze di costanti  $\{a_n > 0\}$  e  $\{b_n\}$  tali che

$$Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G(x) \quad \text{con } n \rightarrow \infty,$$

per qualche funzione di distribuzione  $G$  non-degenere, allora, per valori di  $r$  fissati, la distribuzione limite per  $n \rightarrow \infty$  di

$$\tilde{M}_n^{(r)} = \left( \frac{M_n^{(1)} - b_n}{a_n}, \dots, \frac{M_n^{(r)} - b_n}{a_n} \right)$$

appartiene alla famiglia che ha come funzione di densità di probabilità congiunta

$$\begin{aligned} f(x^{(1)}, \dots, x^{(r)}) &= \exp\left\{-\left[1 + \xi\left(\frac{x^{(r)} - \mu}{\sigma}\right)\right]^{-1/\xi}\right\} \\ &\times \prod_{k=1}^r \sigma^{-1} \left[1 + \xi\left(\frac{x^{(k)} - \mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1}, \end{aligned} \quad (1.12)$$

dove  $-\infty < \mu < \infty$ ,  $\sigma > 0$  e  $-\infty < \xi < \infty$ ;  $x^{(r)} \leq x^{(r-1)} \leq \dots \leq x^{(1)}$ ; e  $\{x^{(k)} : 1 + \xi(x^{(k)} - \mu)/\sigma > 0\}$  per  $k = 1, \dots, r$ .

Nel caso in cui  $r = 1$ , la (1.12) si riduce alla funzione di densità della famiglia GEV. Il caso  $\xi = 0$  nella (1.12) è da interpretarsi come forma limite con  $\xi \rightarrow 0$ , portando alla famiglia di funzioni di densità

$$\begin{aligned} f(x^{(1)}, \dots, x^{(r)}) &= \exp\left\{-\exp\left[-\left(\frac{x^{(r)} - \mu}{\sigma}\right)\right]\right\} \\ &\times \prod_{k=1}^r \sigma^{-1} \exp\left[-\left(\frac{x^{(k)} - \mu}{\sigma}\right)\right], \end{aligned} \quad (1.13)$$

per cui il caso  $r = 1$  riduce la (1.13) alla densità della famiglia Gumbel.

### 1.4.2 Modellare le $r$ più grandi statistiche d'ordine

Si consideri una serie di variabili i.i.d. e dati raggruppati in  $m$  blocchi. Nel blocco  $i$  le più grandi  $r_i$  osservazioni sono registrate, ottenendo così una serie  $\mathbf{M}_i^{r_i} = (x_i^{(1)}, \dots, x_i^{(r_i)})$ , per  $i = 1, \dots, m$ .

Come con il modello GEV, il problema della dimensione del blocco corrisponde ad un trade-off tra distorsione e varianza, che normalmente viene risolto facendo una scelta pragmatica, per cui la dimensione del blocco è di lunghezza pari ad un anno. Il numero delle statistiche d'ordine usate in ogni blocco include anche un trade-off varianza-distorsione: valori piccoli di  $r$  generano pochi dati portando ad una varianza alta; al contrario, valori grandi di  $r$  portano ad una distorsione. Nella pratica quindi, conviene scegliere valori di  $r_i$  più grandi possibili, sottoposti ad adeguati modelli diagnostici.

La verosimiglianza per questo modello è ottenuta dalla (1.12) e dalla (1.13), inserendo i coefficienti di scala non noti al posto degli usuali parametri di posizione e di scala, e facendo i prodotti tra blocchi. Così, quando  $\xi \neq 0$ ,

$$L(\mu, \sigma, \xi) = \prod_{i=1}^m \left( \exp \left\{ - \left[ 1 + \xi \left( \frac{x_i^{(r_i)} - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \right. \\ \left. \times \prod_{k=1}^{r_i} \sigma^{-1} \left[ 1 + \xi \left( \frac{x_i^{(k)} - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi} - 1} \right), \quad (1.14)$$

posto che  $\{1 + \xi(x^{(k)} - \mu)/\sigma > 0\}$ , con  $k = 1, \dots, r_i$ ,  $i = 1, \dots, m$ ; altrimenti la verosimiglianza è zero. Quando  $\xi = 0$ ,

$$L(\mu, \sigma, \xi) = \prod_{i=1}^m \left( \exp \left\{ - \exp \left[ - \left( \frac{x_i^{(r_i)} - \mu}{\sigma} \right) \right] \right\} \right. \\ \left. \times \prod_{k=1}^{r_i} \sigma^{-1} \exp \left[ - \left( \frac{x_i^{(k)} - \mu}{\sigma} \right) \right] \right). \quad (1.15)$$

Le verosimiglianze (1.14) e (1.15) possono essere massimizzate per ottenere la stima della massima verosimiglianza.

In conclusione si può affermare che, il modello della  $r$  più grande statistica d'ordine dà una verosimiglianza i cui parametri corrispondono a quelli della distribuzione GEV per i massimi di un blocco, ma incorpora molti dei dati estremi osservati. Così, con riferimento all'analisi standard dei massimi dei blocchi, l'interpretazione dei parametri è inalterata, ma la precisione dovrebbe aumentare sulla base del fatto che si considerano delle informazioni extra.



## 1.5 Una possibile caratterizzazione alternativa: valori che superano una soglia

Come è stato ampiamente dimostrato nelle sezioni precedenti, la tecnica di modellare i massimi di ogni blocco con una distribuzione GEV è inefficiente: anche disponendo di rilevazioni giornaliere o orarie, si dovrebbe comunque considerare solamente una osservazione per ogni anno, con una evidente perdita di informazioni. Sebbene il modello della  $r$  più grande statistica d'ordine rappresenta una alternativa migliore, è assolutamente impossibile nella realtà disporre di dati in questa forma.

In anni recenti, l'attenzione si è rivolta verso metodi basati sullo studio delle *eccedenze rispetto ad una soglia*. Questi metodi sono più flessibili rispetto ai metodi del massimo annuale per una serie di ragioni. Primo, prendendo in considerazione tutte le eccedenze rispetto ad una soglia alta, i dati vengono utilizzati più efficientemente. Secondo, sono metodi che si possono estendere facilmente a situazioni in cui si voglia studiare, come livelli estremi di una variabile  $Y$  dipendano da una variabile  $X$  - per esempio,  $Y$  potrebbe essere il livello di ozono troposferico in un particolare giorno e  $X$  un vettore di variabili meteorologiche per quello stesso giorno (Smith e Shively, 1995). Questo tipo di problemi sarebbe impossibile da trattare con il metodo del massimo annuale.

### 1.5.1 Distribuzione Generalizzata di Pareto (GPD)

Sia  $X_1, X_2, \dots$  una sequenza di variabili aleatorie i.i.d. con funzione di distribuzione  $F$ . L'idea di base è di fissare una soglia alta  $u$  e studiare tutti gli eventi che superano questa soglia. Si indichi con  $X$  il generico termine della sequenza  $X_i$ , e si ponga  $Y = X - u$  condizionato ad  $X > u$ . Allora

$$\Pr\{X > u + y | X > u\} = \frac{1 - F(u + y)}{1 - F(u)}, \quad y > 0 \quad (1.16)$$

Per valori grandi di  $u$  si ottengono dei risultati asintotici strettamente collegati a quelli sulla distribuzione dei valori estremi.

#### **Teorema 1.5.1. METTERE AUTORE**

*Sia  $X_1, X_2, \dots$  una sequenza di variabili aleatorie i.i.d. con funzione di ripartizione comune  $F$ . Allora la funzione di distribuzione di  $(X - u)$ , condizionata a  $X > u$ , è approssimativamente*

$$H(y; \psi, \xi) = 1 - \left(1 + \frac{\xi y}{\psi}\right)^{-1/\xi} \quad (1.17)$$

*definita su  $\{y : y > 0 \text{ e } (1 + \xi y/\psi) > 0\}$ , dove*

$$\psi = \sigma + \xi(u - \mu). \quad (1.18)$$

La famiglia di distribuzioni definita dall'equazione (1.17) è chiamata **famiglia generalizzata di Pareto**,  $GPD(\psi, \xi)$ .

Il parametro  $\xi$  è dominante nel determinare il comportamento qualitativo della GPD.

- Se  $\xi < 0$ , allora la GPD ha un limite superiore pari a  $u - \psi/\xi$ ;
- Se  $\xi > 0$ , la distribuzione non ha limite superiore;
- Se  $\xi \rightarrow 0$ , allora si ha una distribuzione esponenziale con parametro  $1/\psi$

$$H(y; \psi, 0) = 1 - \exp\left(-\frac{y}{\psi}\right). \quad (1.19)$$

Il Teorema 1.5.1 suggerisce il contesto per modellare i valori estremi. Si dispone di una sequenza di misurazioni  $x_1, \dots, x_n$ , gli eventi estremi sono identificati definendo una soglia alta  $u$ , per cui le eccedenze sono  $\{x_i : x_i > u\}$ . Il problema della scelta della soglia è analogo a quello della scelta della dimensione del blocco nell'approccio dei massimi, si deve cercare un giusto equilibrio tra distorsione e varianza. Da un lato la soglia deve essere sufficientemente elevata affinché si possa ritenere valida l'approssimazione asintotica, anche se una soglia alta genererà pochi eccessi portando ad una varianza molto grande; dall'altro il valore di  $u$  dovrebbe essere sufficientemente piccolo da permettere una stima accurata. La soluzione migliore è di adottare una soglia più bassa possibile in modo tale che il modello riesca a garantire una approssimazione ragionevole. Il metodo utilizzato per questo scopo è basato sulla media della GPD: se  $Y$  ha una distribuzione generalizzata di Pareto con parametri  $\sigma$  e  $\xi$  (da questo punto in poi, per convenienza nella notazione, non si distinguerà più tra il parametro di scala  $\sigma$  della GEV e il parametro  $\psi$  della GPD, si userà  $\sigma$  per indicare il parametro di scala in entrambe le famiglie), allora

$$E(Y) = \frac{\sigma}{1 - \xi} \quad (1.20)$$

posto che  $\xi < 1$ . Quando  $\xi \geq 1$  il valore della media è infinito. Ora, si supponga che la GPD sia valida come modello per gli eccessi, rispetto ad una soglia  $u_0$ , generati da una serie  $X_1, \dots, X_n$  e si consideri  $X$  come un termine generico. Allora,

$$E(X - u_0 | X > u_0) = \frac{\sigma_{u_0}}{1 - \xi},$$

posto che  $\xi < 1$ , dove per convenzione si indica con  $\sigma_{u_0}$  il parametro di scala corrispondente agli eccessi rispetto alla soglia  $u_0$ . Ma se la GPD è valida per gli eccessi rispetto ad  $u_0$ , sarà ugualmente valida per tutte le soglie  $u > u_0$ . Quindi, per  $u > u_0$ ,

$$E(X - u | X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi u}{1 - \xi} \quad (1.21)$$

in virtù della (1.18). Così, per  $u > u_0$ ,  $E(X - u|X > u)$  è una funzione lineare di  $u$ . Inoltre,  $E(X - u|X > u)$  è semplicemente la media dei superamenti rispetto alla soglia  $u$ , per cui la media campionaria di tali superamenti rispetto ad  $u$  fornisce una stima empirica. Seguendo la (1.21), ci si aspetta che queste stime cambino linearmente con  $u$ , a livelli di  $u$  tali che il modello generalizzato di Pareto risulti appropriato.

Dopo aver determinato la soglia, i parametri della GPD possono essere stimati tramite la verosimiglianza massimizzata. Si supponga che i valori  $y_1, \dots, y_k$  siano i  $k$  superamenti rispetto alla soglia  $u$ . Per  $\xi \neq 0$  la log-verosimiglianza si ottiene dalla (1.17) come

$$\ell(\sigma, \xi) = -k \log \sigma - (1 + 1/\xi) \sum_{i=1}^k \log(1 + \xi y_i / \sigma), \tag{1.22}$$

posto che  $(1 + \sigma^{-1} \xi y_i) > 0$  per  $i = 1, \dots, k$ ; altrimenti,  $\ell(\sigma, \xi) = -\infty$ . Nel caso in cui  $\xi = 0$  la log-verosimiglianza è ottenuta dalla (1.19) come

$$\ell(\sigma) = -k \log \sigma - \sigma^{-1} \sum_{i=1}^k y_i.$$

La massimizzazione analitica della log-verosimiglianza non è possibile, così si ricorre nuovamente a tecniche numeriche, facendo attenzione ad evitare instabilità numeriche quando  $\xi \approx 0$  nella (1.22), e assicurandosi che l'agoritmo non fallisca quando la valutazione viene fatta fuori dallo spazio dei parametri ammissibile. Gli errori standard e gli intervalli di confidenza per la distribuzione generalizzata di Pareto si ottengono nell'usuale modo dalla teoria standard della verosimiglianza.

### 1.5.2 Modello Poisson-GPD e stime dei parametri

Si supponga di osservare le v.a.  $X_1, \dots, X_n$  i.i.d., e di osservare gli indici  $i$  per i quali  $X_i > u$ . Se si riscalassero gli indici  $i$  in questo modo  $i/n$ , allora tali indici potrebbero essere visti come un processo di punto dei tempi di superamento riscalati su  $[0, 1]$ . Se  $n \rightarrow \infty$  e  $1 - F(u) \rightarrow 0$  tale che  $n(1 - F(u)) \rightarrow \lambda$  ( $0 < \lambda < \infty$ ), allora potrebbe dirsi che il processo converge velocemente ad un processo di Poisson omogeneo su  $[0, 1]$ , con intensità  $\lambda$ .

Partendo da ciò, si può immaginare una forma limite del processo di punto congiunto dei tempi di superamento e dei superamenti rispetto alla soglia, che abbia la seguente forma:

- (a) Il numero,  $N$ , di superamenti del livello  $u$  in qualsiasi anno ha una distribuzione di Poisson con media  $\lambda$ ;
- (b) Condizionatamente a  $N \geq 1$ , i valori  $Y_1, \dots, Y_n$  che superano la soglia sono i.i.d. da una GPD.

Questo viene chiamato *modello Poisson-GPD*.

Naturalmente, non c'è nessuna motivazione particolare per cui si è preso come riferimento l'anno - potrebbe essere utilizzata una qualsiasi altra unità di tempo - ma in particolare nei processi ambientali e nell'analisi di serie finanziarie, l'anno è il periodo di tempo più conveniente a cui fare riferimento.

Il processo Poisson-GPD è strettamente collegato alla distribuzione GEV per i massimi annuali. Si supponga che  $x > u$ . La probabilità che il massimo annuale del processo Poisson-GPD sia minore o uguale ad  $x$  è

$$\begin{aligned} Pr\left\{\max_{1 \leq i \leq N} Y_i \leq x\right\} &= Pr\{N = 0\} + \sum_{n=1}^{\infty} Pr\{N = n, Y_1 \leq x, \dots, Y_n \leq x\} \\ &= e^{-\lambda} + \sum_{n=1}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \cdot \left\{1 - \left(1 + \xi \frac{x-u}{\sigma}\right)_+^{-1/\xi}\right\}^n \\ &= \exp\left\{-\lambda \left(1 + \xi \frac{x-u}{\sigma}\right)_+^{-1/\xi}\right\}. \end{aligned} \quad (1.23)$$

Se si sostituisce

$$\sigma = \psi - \xi(u - \mu) \quad \lambda = \left(1 + \xi \frac{u - \mu}{\psi}\right)^{-1/\xi}, \quad (1.24)$$

la (1.23) si ridurrebbe alla forma della distribuzione GEV. Così i modelli GEV e GPD sono completamente conformi con un altro modello in cui si supera la soglia  $u$ , e la (1.24) dà una esplicita relazione tra i due insiemi di parametri.

E' possibile, anche in questo caso, effettuare delle stime di massima verosimiglianza. Si supponga di avere un totale di  $N$  osservazioni al di sopra di una soglia al tempo  $T$ . Si supponga inoltre che il numero atteso di superamenti di Poisson sia  $\lambda T$ , e i parametri GPD siano  $\sigma$  e  $\xi$  come nella (1.17). Allora la log-verosimiglianza sarà

$$\ell_{N,Y}(\lambda, \sigma, \xi) = N \log \lambda - \lambda T - N \log \sigma - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^N \log \left(1 + \xi \frac{Y_i}{\sigma}\right) \quad (1.25)$$

posto che  $1 + \xi Y_i/\sigma > 0$  per tutti i valori di  $i$ .

Gli stimatori di massima verosimiglianza sono i valori dei parametri non noti che massimizzano la log-verosimiglianza. In pratica, questi sono massimi locali trovati mediante ottimizzazione non-lineare. I risultati asintotici di consistenza, l'efficienza asintotica e la normalità asintotica valgono per queste distribuzioni se  $\xi > -\frac{1}{2}$  (Smith, 1985).

Il modello Poisson-GPD è strettamente collegato al modello *Peaks Over Threshold* (POT) originariamente sviluppato dagli idrologi. Nei casi con alta correlazione periodica i superamenti rispetto alla soglia non si verificano singolarmente ma in gruppi, e in questo caso, il metodo viene direttamente applicato ai valori dei picchi all'interno di ogni cluster. Per ulteriori approfondimenti si veda Davison e Smith (1990).

# Capitolo 2

## Caratterizzazione dei valori estremi mediante il processo di Poisson

### 2.1 Introduzione

Esistono modi differenti per caratterizzare il comportamento dei valori estremi di un processo, e una formulazione particolarmente elegante deriva dalla teoria dei processi di punto. Come si vedrà nella prima parte del capitolo, questa caratterizzazione non porta a nulla di nuovo in termini di modelli statistici; tutte le deduzioni fatte utilizzando la metodologia del processo di punto possono essere ugualmente ottenute utilizzando uno dei modelli descritti nelle sezioni precedenti. Comunque ci sono due buone ragioni per preferire questo approccio. Innanzitutto, contempla un'interpretazione del comportamento del valore estremo che unifica tutti i modelli finora introdotti; inoltre, questo modello conduce direttamente ad una verosimiglianza che consente di ottenere una formulazione più naturale della non-stazionarietà dei superamenti rispetto a quella ottenuta dalla distribuzione generalizzata di Pareto.

Nella seconda parte invece, si considererà il processo di Poisson come processo aleatorio a tempo continuo. La proprietà caratterizzante di questo tipo di processo è che *non conserva memoria* del passato. Questo significa che solo lo stato corrente del processo può influenzare gli stati futuri.

Verrà infine data anche una definizione del processo di Poisson inteso come processo di salto che è esattamente la caratterizzazione del processo di cui che verrà utilizzata nella parte dell'impostazione bayesiana non-parametrica.

## 2.2 Convergenza in legge

Il contesto del processo di punto fornisce un modo elegante di formulare i risultati limite dei valori estremi. Così come è stato fatto nel capitolo precedente, si assume che  $X_1, X_2, \dots$  sia una serie di v.a. i.i.d., con funzione di distribuzione comune  $F$ . Si supponga inoltre che le  $X_i$  si comportino bene nel senso di valore estremo. Cioè, sia  $M_n = \max\{X_1, \dots, X_n\}$ , allora esistono due sequenze di costanti  $\{a_n > 0\}$  e  $\{b_n\}$  tali che

$$Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G(x),$$

con

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

per alcuni parametri  $\mu, \sigma > 0$  e  $\xi$ . Successivamente viene definita una sequenza di processi di punto  $N_n$  su  $\mathbb{R}^2$  data da

$$N_n = \{(i/(n+1), (X_i - b_n)/a_n) : i = 1, \dots, n\}.$$

Il fatto di aver cambiato scala nella prima ordinata assicura che l'asse dei tempi sia sempre disegnata nell'intervallo  $(0, 1)$ , mentre il cambio di scala nella seconda ordinata stabilisce il comportamento degli estremi quando  $n \rightarrow \infty$ . Il risultato fondamentale è che nella regione  $(0, 1) \times [u, \infty)$ ,  $N_n \xrightarrow{d} N$  con  $n \rightarrow \infty$ , dove  $N$  è un processo di Poisson non-omogeneo.

Si consideri una regione della forma  $A = [0, 1] \times (u, \infty)$  per valori grandi di  $u$ , allora ognuno degli  $n$  punti di  $N_n$  ha probabilità  $p$  di cadere in  $A$ , dove

$$p = Pr\{(X_i - b_n)/a_n > u\} \approx \frac{1}{n} \left[1 + \xi\left(\frac{u - \mu}{\sigma}\right)\right]^{-1/\xi}.$$

Dato che le  $X_i$  sono mutuamente indipendenti,  $N_n(A)$  ha la distribuzione binomiale

$$N_n(A) \sim Bin(n, p).$$

Dalla convergenza standard di una distribuzione binomiale ad una distribuzione Poisson segue che, la distribuzione limite di  $N_n(A)$ , quando  $n \rightarrow \infty$ , è una distribuzione di  $Po\{\Lambda(A)\}$ , con

$$\Lambda(A) = \left[1 + \xi\left(\frac{u - \mu}{\sigma}\right)\right]^{-1/\xi}, \quad (2.1)$$

dove  $\Lambda(A)$  indica la misura di intensità della regione  $A$  (si veda figura (2.1)).

A causa dell'omogeneità del processo rispetto al tempo si ha che, per ogni regione della forma  $A = [t_1, t_2] \times (u, \infty)$ , con  $[t_1, t_2] \subset [0, 1]$ , la distribuzione limite di  $N_n(A)$  è ancora una distribuzione di  $Po\{\Lambda(A)\}$ , dove

$$\Lambda(A) = (t_2 - t_1) \left[1 + \xi\left(\frac{u - \mu}{\sigma}\right)\right]^{-1/\xi}. \quad (2.2)$$

Il risultato viene formalizzato dal seguente teorema.

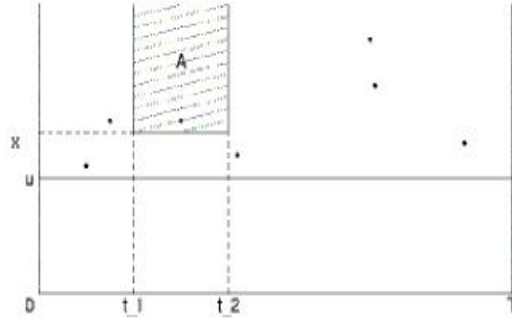


Figura 2.1: Illustrazione della rappresentazione tramite il processo di punto. I punti sul grafico rappresentano i tempi  $t$  e i valori  $x$  dei superamenti rispetto alla soglia  $u$ .

**Teorema 2.2.1.** *Sia  $X_1, X_2, \dots$  una serie di v.a. i.i.d. per le quali esistono sequenze di costanti  $\{a_n > 0\}$  e  $\{b_n\}$  tali che*

$$\Pr\{(M_n - b_n)/a_n \leq x\} \rightarrow G(x),$$

dove

$$G(x) = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}\right\},$$

e siano  $x_-$  e  $x_+$  i punti estremi, rispettivamente inferiore e superiore di  $G$ .

Allora, la sequenza di processi di punto

$$N_n = \{(i/(n+1), (X_i - b_n)/a_n) : i = 1, \dots, n\}$$

converge, su regioni di forma  $(0, 1) \times [u, \infty)$  e per ogni  $u > x_-$ , ad un processo di Poisson, con misura di intensità su  $A = [t_1, t_2] \times [x, x_+)$  data da

$$\Lambda(A) = (t_2 - t_1) \left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}. \quad (2.3)$$

## 2.3 Il modello statistico

E' stato già detto come sia più vantaggioso lavorare con il modello del processo di punto. Come con tutti gli altri modelli sui valori estremi, il modello limite - in questo caso il processo di Poisson - viene interpretato come una ragionevole approssimazione del comportamento di un campione grande ma finito. Cioè, nella convergenza del processo di punto in cui  $N_n \rightarrow N$ , si assume che la legge di probabilità associata al limite  $N$  fornisce una approssimazione ragionevole alla legge associata a  $N_n$  per valori sufficientemente grandi di  $n$ . In altre parole, il teorema 2.2.1 fornisce una rappresentazione soltanto per il comportamento del valore estremo delle  $X_i$ , ma per le applicazioni è necessario fare degli aggiustamenti. Il primo riguarda la scelta più accurata della soglia  $u$  in modo

che i valori che la superano possano fornire una ragionevole approssimazione; il secondo problema riguarda invece, la normalizzazione delle sequenze  $\{a_n\}$  e  $\{b_n\}$ . Come per gli altri modelli finora analizzati, queste sono dipendenti dalla distribuzione  $F$  delle  $X_i$  sottostante, e quindi sono in pratica sconosciute. Questa difficoltà è superata, come negli altri modelli, includendo i coefficienti nei parametri di posizione e di scala del modello stesso.

In conclusione, il teorema 2.2.1 può essere rinunciato nel modo seguente.

**Teorema 2.3.1.** *Sia  $X_1, \dots, X_n$  una serie di v.a. i.i.d. e sia*

$$N_n = \{(i/(n+1), X_i) : i = 1, \dots, n\}.$$

*Allora, per valori di  $u$  sufficientemente grandi, sulla regione di forma  $(0, 1) \times [u, \infty)$ ,  $N_n$  è approssimativamente un processo di Poisson con misura di intensità su  $A = [t_1, t_2] \times (x, \infty)$  data da*

$$\Lambda(A) = (t_2 - t_1) \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi}. \quad (2.4)$$

Questo teorema consente di eseguire la seguente procedura per modellare gli estremi utilizzando il processo di punto. Primo, si seleziona una soglia alta  $u$ , per cui l'approssimazione al processo di Poisson sembra buona, e si pone  $A = (0, 1) \times [u, \infty)$ . I punti  $N(A)$  osservati nella regione  $A$  saranno rinominati in questo modo  $\{(t_1, x_1), \dots, (t_{N(A)}, x_{N(A)})\}$ . Assumendo inoltre, che il processo di Poisson limite sia una approssimazione accettabile al processo  $N_n$  su  $A$ , può derivarsi una verosimiglianza approssimata. Massimizzando questa verosimiglianza si ottengono le stime dei parametri  $(\mu, \sigma, \xi)$  della funzione di intensità limite. La funzione di verosimiglianza per il processo di Poisson sarà quindi data da

$$\begin{aligned} L_A(\mu, \sigma, \xi; x_1, \dots, x_n) &= \exp\{\Lambda(A)\} \prod_{i=1}^{N(A)} \lambda(t_i, x_i) \\ &\propto \exp\left\{ -n_y \left[ 1 + \xi \left( \frac{u - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \prod_{i=1}^{N(A)} \sigma^{-1} \left[ 1 + \xi \left( \frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}-1} \end{aligned} \quad (2.5)$$

dove con  $n_y$  si indica il numero di anni di osservazione e con  $\lambda(t_i, x_i)$  la funzione di intensità calcolata nella regione  $(t_i, x_i)$ <sup>1</sup>.

Questa funzione può essere trattata nel modo consueto per ottenere le stime di massima verosimiglianza, gli errori standard e gli intervalli di confidenza approssimati dei parametri del modello.

Come nel modello dei superamenti rispetto alla soglia, le stime derivate dalla

---

<sup>1</sup>Assumendo la regione  $A = [a_1, x_1] \times \dots \times [a_k, x_k]$ , e posto che esista, la funzione di derivata  $\lambda(\mathbf{x}) = \frac{\partial \Lambda(A)}{\partial x_1 \dots \partial x_k}$  indica la **funzione di intensità** del processo



verosimiglianza del processo di punto sono basate su tutti quei dati considerati estremi, nel senso di essere più grandi di una soglia specificata. L'inferenza è invece più accurata rispetto alle stime basate su un diretto adattamento della distribuzione GEV ai dati massimi annuali.

Quindi, la verosimiglianza del processo di punto è ottenuta come prodotto della verosimiglianza per i superamenti rispetto alla soglia della GPD con una verosimiglianza dell'evento binario di superamento della soglia, riparametrizzato in termini dei parametri della GEV. L'equivalenza delle verosimiglianze per il processo di punto limite e del modello di superamento conferma che l'inferenza ottenuta dal modello del processo di punto potrebbe ugualmente essere ottenuta in un contesto di superamento della soglia.

## 2.4 Connessioni con gli altri modelli

E' facile dimostrare come il modello dei blocchi, il modello relativo ai superamenti di una soglia e il modello delle  $r$  più grandi statistiche d'ordine siano tutti casi speciali della rappresentazione del processo di punto data nel teorema 2.2.1.

Si consideri il risultato ottenuto per i **massimi dei blocchi**. Indicando  $M_n = \max\{X_1, \dots, X_n\}$  nel modo usuale, e ponendo

$$N_n = \{(i/(n+1), (X_i - b_n)/a_n) : i = 1, \dots, n\},$$

l'evento  $\{(M_n - b_n)/a_n \leq x\}$  è equivalente all'evento  $N_n(A_x) = 0$ , dove  $A_x = (0, 1) \times (x, \infty)$ . Quindi

$$\begin{aligned} Pr\{(M_n - b_n)/a_n \leq x\} &= Pr\{N_n(A_x) = 0\} \\ &\rightarrow Pr\{N(A_x) = 0\} \\ &= \exp\{-\Lambda(A_x)\} \\ &= \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}, \end{aligned}$$

così la distribuzione limite dei massimi normalizzati per ogni blocco è la distribuzione GEV.

Argomentazioni simili sono applicate al **modello dei superamenti rispetto alla soglia**. Come prima cosa, risulta utile fattorizzare  $\Lambda(A_x)$  nel seguente modo

$$\Lambda(A_x) = \Lambda_1([t_1, t_2]) \times \Lambda_2([x, \infty)),$$

dove

$$\Lambda_1([t_1, t_2]) = (t_2 - t_1) \quad e \quad \Lambda_2([x, \infty)) = \left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}.$$

Allora,

$$\begin{aligned}
Pr\{(X_i - b_n)/a_n > x | (X_i - b_n)/a_n > u\} &= \frac{\Lambda_2[x, \infty)}{\Lambda_2[u, \infty)} \\
&= \frac{n^{-1}[1 + \xi(x - \mu)/\sigma]^{-1/\xi}}{n^{-1}[1 + \xi(u - \mu)/\sigma]^{-1/\xi}} \\
&= \left[1 + \frac{\xi(x - \mu)/\sigma}{1 + \xi(u - \mu)/\sigma}\right]^{-1/\xi} \\
&= \left[1 + \xi\left(\frac{x - u}{\psi}\right)\right]^{-1/\xi},
\end{aligned}$$

con  $\psi = \sigma + \xi(u - \mu)$ .

Invece, per quanto riguarda il **modello delle  $r$  piú grandi statistiche d'ordine**, con  $M_n^{(k)}$  si indicava la  $k$ -esima piú grande tra le variabili i.i.d.  $X_1, \dots, X_n$ . Sia inoltre,

$$N_n = \{(i/(n+1), (X_i - b_n)/a_n) : i = 1, \dots, n\}$$

e  $A_x = (0, 1) \times [x, \infty)$ ,

$$\begin{aligned}
Pr\{(M_n^{(k)} - b_n)/a_n \leq x\} &= Pr\{N_n(A_x) \leq k - 1\} \\
&= \sum_{s=0}^{k-1} Pr\{N_n(A_x) = s\}.
\end{aligned} \tag{2.6}$$

Dal processo limite di Poisson,  $N_n(A_x)$  converge, per  $n \rightarrow \infty$ , ad una variabile di Poisson con media

$$\Lambda(A_x) = \left[1 + \xi\left(\frac{x - \mu}{\sigma}\right)\right]^{-1/\xi}.$$

Prendendo il limite della (1.25),

$$Pr\{(M_n^{(k)} - b_n)/a_n \leq x\} \rightarrow \sum_{s=0}^{k-1} e^{-\tau(x)} \frac{\tau(x)^s}{s!},$$

dove  $\tau(x) = \Lambda(A_x)$ . Questo è il teorema 1.4.1.

Il teorema 1.4.2 si ottiene dalla verosimiglianza del processo limite di Poisson. Siano  $(x^{(1)}, \dots, x^{(r)})$  i valori osservati di  $M_n$ . La sostituzione di  $u = x^r$  e delle  $x_i$  con le  $x_i$  nella (1.26), dà la verosimiglianza per le  $r$  piú grandi statistiche d'ordine.

## 2.5 Processi aleatori a tempo continuo

Si consideri  $I$  un insieme numerabile. Un *processo aleatorio a tempo continuo*

$$(X_t)_{t \geq 0} = (X_t : 0 \leq t < \infty)$$

con valori in  $I$  è una famiglia di variabili aleatorie  $X_t : \Omega \rightarrow I$ . Si andranno a considerare modi in cui si potrà specificare il comportamento probabilistico (o *in legge*) di  $(X_t)_{t \geq 0}$ . Questo potrebbe facilitare la ricerca di una qualche probabilità connessa con il processo, così come  $\mathbb{P}(X_t = i)$  o  $\mathbb{P} = (X_{t_0} = i_0, \dots, X_{t_n} = i_n)$ , o  $\mathbb{P}(X_t = i$  per alcuni valori di  $t$ ).

In questa parte del capitolo si restringerà l'attenzione a quei processi  $(X_t)_{t \geq 0}$  che sono *continui a destra*. Questo significa che per tutti i valori  $\omega \in \Omega$  e  $t \geq 0$  esiste un  $\varepsilon > 0$  tale che

$$X_s(\omega) = X_t(\omega) \quad \text{per } t \leq s \leq t + \varepsilon.$$

Da un risultato standard della teoria della misura, la probabilità di un qualsiasi evento che dipende da un processo continuo a destra può essere determinata dalla sua

Si indichino con  $J_0, J_1, \dots$  i *tempi di salto* di  $(X_t)_{t \geq 0}$  e con  $S_1, S_2, \dots$  *tempi di interarrivo*. Questi sono ottenuti dal processo nel modo seguente

$$J_0 = 0, \quad J_{n+1} = \inf\{t \geq J_n : X_t \neq X_{J_n}\}$$

per  $n = 0, 1, \dots$ , dove  $\inf \emptyset = \infty$ , e, per  $n = 1, 2, \dots$ ,

$$S_n = \begin{cases} J_n - J_{n-1} & J_{n-1} < \infty \\ \infty & \text{altrimenti.} \end{cases}$$

Si noti che la continuità a destra implica che  $S_n > 0$  per tutti i valori di  $n$ . Se  $J_{n+1} = \infty$  per alcuni  $n$ , si definisce  $X_\infty = X_{J_n}$ , come valore finale, altrimenti  $X_\infty$  è indefinito. Il (*primo*) *tempo di esplosione*  $\zeta$  è definito da

$$\zeta = \sup_n J_n = \sum_{n=1}^{\infty} S_n.$$

Il processo a tempo discreto  $(Y_n)_{n \geq 0}$  dato da  $Y_n = X_{J_n}$  è chiamato *processo di salto* di  $(X_t)_{t \geq 0}$ , o *catena di salto* se è una catena di Markov a tempo discreto. Questa è semplicemente la sequenza dei valori presi da  $(X_t)_{t \geq 0}$  fino all'esplosione.

Si potrebbe pensare di non considerare cosa accadrebbe al processo dopo l'esplosione. A tal proposito è quindi conveniente aggiungere ad  $I$  un nuovo stato, detto  $\infty$ , e richiedere che  $X_t = \infty$  se  $t \geq \zeta$ . Qualsiasi processo che soddisfi questi requisiti è detto *minimale*. Il termine 'minimale' non si riferisce allo stato del processo ma all'intervallo di tempo lungo il quale il processo è attivo. E' utile puntualizzare il fatto che un processo minimale possa essere ricostruito attraverso i suoi tempi di interarrivo e il processo di salto.

**Teorema 2.5.1.** *Sia  $S_1, S_2, \dots$  una sequenza di variabili aleatorie indipendenti con  $S_n \sim E(\lambda_n)$  e  $0 < \lambda_n < \infty$  per tutti i valori di  $n$ .*

(i) Se  $\sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty$ , allora  $P(\sum_{n=1}^{\infty} S_n < \infty) = 1$ .

(ii) Se  $\sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty$ , allora  $P(\sum_{n=1}^{\infty} S_n = \infty) = 1$ .

*Dimostrazione:* Si veda Norris [rif].

## 2.6 Processi di Poisson

I processi di Poisson costituiscono una particolare classe di processi stocastici a tempo continuo. Volendo dare a tali processi una spiegazione intuitiva, basterà considerarli come dei processi di punto che “contano” il numero di volte in cui si verifica un evento aleatorio in un dato intervallo di tempo.

### 2.6.1 Processo di Poisson omogeneo

**Definizione 2.6.1.** *Un processo continuo a destra  $(X_t)_{t \geq 0}$  con valori in  $\{0, 1, 2, \dots\}$  è un processo di Poisson omogeneo di intensità  $\lambda$  ( $0 < \lambda < \infty$ ) se i suoi tempi di interarrivo  $S_1, S_2, \dots$  sono variabili aleatorie esponenziali indipendenti di parametro  $\lambda$  e la sua catena di salto è data da  $Y_n = n$ .*

Dal teorema 2.5.1, si ha  $P(J_n \rightarrow \infty) = 1$  in modo tale che non ci sia esplosione, e la legge di  $(X_t)_{t \geq 0}$  è unicamente determinata. Un modo semplice per costruire un processo di Poisson di tasso  $\lambda$  è quello di prendere una sequenza  $S_1, S_2, \dots$  di variabili aleatorie esponenziali indipendenti di parametro  $\lambda$ , porre  $J_0 = 0$ ,  $J_n = S_1 + \dots + S_n$  e poi porre

$$X_t = n \quad \text{se} \quad J_n \leq t < J_{n+1}.$$

**Teorema 2.6.1. (Proprietà di Markov).** *Sia  $(X_t)_{t \geq 0}$  un processo di Poisson di tasso  $\lambda$ . Allora, per qualsiasi valore  $s \geq 0$ ,  $(X_{s+t} - X_s)_{t \geq 0}$  è ancora un processo di Poisson di tasso  $\lambda$ , indipendente da  $(X_r : r \leq s)$ .*

*Dimostrazione:* Si veda Norris [rif].

**Teorema 2.6.2. (Proprietà forte di Markov).** *Sia  $(X_t)_{t \geq 0}$  un processo di Poisson di tasso  $\lambda$  e sia  $T$  un tempo di arresto di  $(X_t)_{t \geq 0}$ . Allora, condizionatamente a  $T < \infty$ ,  $(X_{T+t} - X_T)_{t \geq 0}$  è ancora un processo di Poisson di intensità  $\lambda$ , indipendente da  $(X_s : s \leq T)$ .*

Se  $(X_t)_{t \geq 0}$  è un processo a valori reali, si può considerare il suo *incremento*  $(X_s - X_t)$  su un qualsiasi intervallo  $(s, t]$ . Il processo  $(X_t)_{t \geq 0}$  ha incrementi *stazionari* se la distribuzione di  $(X_{s+t} - X_s)$  dipende solamente da  $t \geq 0$ . Inoltre,  $(X_t)_{t \geq 0}$  ha incrementi *indipendenti* se i suoi incrementi su una qualsiasi collezione finita di intervalli disgiunti sono indipendenti.

**Teorema 2.6.3.** *Sia  $(X_t)_{t \geq 0}$  un processo crescente, a valori interi e continuo a destra che parta da 0. Sia inoltre,  $0 < \lambda < \infty$ . Allora le seguenti tre condizioni sono equivalenti:*

- (a) *i tempi di interarrivo  $S_1, S_2, \dots$  di  $(X_t)_{t \geq 0}$  sono v.a. esponenziali indipendenti di parametro  $\lambda$  e la catena di salto è data da  $Y_n = n$  per tutti i valori di  $n$ ;*
- (b)  *$(X_t)_{t \geq 0}$  ha incrementi indipendenti e, per  $h \downarrow 0$ , uniformemente in  $t$ ,*

$$P(X_{t+h} - X_t = 0) = 1 - \lambda h + o(h), \quad P(X_{t+h} - X_t = 1) = \lambda h + o(h);$$

- (c)  *$(X_t)_{t \geq 0}$  ha incrementi stazionari indipendenti e, per ogni  $t$ ,  $X_t$  ha una distribuzione di Poisson di parametro  $\lambda t$ .*

*Se  $(X_t)_{t \geq 0}$  soddisfa una qualsiasi di queste condizioni allora è chiamato processo di Poisson di intensità  $\lambda$ .*

**Osservazione 2.6.1.** *I processi di Poisson omogenei sono processi stocastici a incrementi stazionari, cioè*

$$(X_{t+h} - X_t) \stackrel{d}{=} X_h, \quad \forall t, h \geq 0.$$

## 2.6.2 Processo di Poisson non-omogeneo, da finire

Partendo dal processo di Poisson omogeneo, si può caratterizzare un'altra categoria di processi di punto: quella dei processi di Poisson non-omogenei. Questi processi si ottengono attraverso l'indebolimento di alcune assunzioni fatte sulla funzione di intensità.



# Capitolo 3

## Analisi dei valori estremi: impostazione bayesiana parametrica

### 3.1 Introduzione

Finora sono state analizzate differenti tecniche per la stima dei parametri nei modelli dei valori estremi e si è argomentato su come siano migliori i metodi basati sulla verosimiglianza. Ma questo non è però l'unico modo per fare inferenza partendo dalla funzione di verosimiglianza; a questo proposito le tecniche bayesiane offrono una valida alternativa che è spesso da preferirsi (si vedano Coles e Powell (1996), Coles e Tawn (1996), Smith (1999), Smith e Goodman (2000), Coles e Pericchi (2001)):

- gli *eventi estremi* sono, per definizione, *rari* e incorporare una struttura di prior all'analisi è spesso essenziale per ottenere risposte significative;
- la *previsione delle perdite* è una fase cruciale nello studio degli eventi estremi e la statistica bayesiana fornisce un metodo che tiene conto dell'incertezza della stima;
- l'*eterogeneità dei dati* assume un ruolo molto importante nelle applicazioni; anche questa volta l'approccio bayesiano fornisce una serie di tecniche di clustering che consentono di modellare tali eventi.

Proprio sulla base di quest'ultima motivazione risulta evidente come nel lavoro da noi proposto sia impossibile, in termini di applicazione, utilizzare questo tipo di approccio. Comunque per completezza espositiva, verranno brevemente esposte le diverse metodologie proposte in letteratura in ambito bayesiano parametrico.

## 3.2 Scelta della distribuzione iniziale

I metodi bayesiani si basano sulla specificazione di una funzione di densità a priori per i parametri non noti, e successivamente si calcola la densità a posteriori per questi parametri date le osservazioni. I calcoli di questa distribuzione a posteriori sono quasi sempre ottenuti dall'utilizzo di forme di campionamento MCMC.

Purtroppo, per esempio, il modello GEV non ammette distribuzioni a priori coniugate. A tale proposito si conoscono due approcci basati sulla ricerca di famiglie coniugate per sottoclassi di modelli. Engelund e Rackwitz (1991) ottengono priors coniugate con un solo parametro nel caso di sottoinsiemi della distribuzione GEV: un parametro di posizione nel caso in cui  $\xi = 0$  e un parametro di scala altrimenti. Nelle applicazioni, questi modelli sono ovviamente abbastanza restrittivi, specialmente perchè il parametro  $\xi$  è quello che presenta maggiore incertezza.

Pickands (1994), invece ha proposto una specificazione non-coniugata delle distribuzioni a priori entro la classe delle GPD

$$H(y; \psi, \xi) = 1 - \left(1 + \frac{\xi y}{\psi}\right)^{-1/\xi} \quad (3.1)$$

definita su  $\{y : y > 0 \text{ e } (1 + \xi y/\psi) > 0\}$ , dove

$$\psi = \sigma + \xi(u - \mu)$$

e  $\mu$  e  $\sigma$  sono i parametri della distribuzione GEV. Tale specificazione è data da

$$f(\psi, \xi) \propto \frac{1}{\psi} I_{\{\psi > 0\}}$$

che equivale a specificare le priors per  $\log(\psi)$  e  $\xi$  che sono indipendenti e uniformi su  $(-\infty, \infty)$ , sebbene una limitazione di questa proposta sia che il parametro  $\psi$  sia dipendente dalla soglia  $u$ .

L'approccio predittivo bayesiano è stato esaminato in alcuni recenti articoli, Coles e Powell (1996), Coles e Tawn (1996a), Smith (1997a). In particolare, Coles e Tawn (1996a) hanno argomentato come una formulazione a priori in termini di parametri della GEV su un comportamento estremo, non sia un approccio sensato. Risulta invece preferibile specificare la prior in termini di quantili estremi invece che in termini di specifici parametri della distribuzione, questo perchè con questa scala di misura un esperto è in grado di quantificare le aspettative del comportamento estremo del fenomeno studiato. Partendo dal modello (1.2), il quantile  $(1 - p)$  assume la forma

$$x_p = \begin{cases} \mu - \frac{\sigma}{\xi} \left[1 - \{-\log(1 - p)\}^{-\xi}\right], & \text{per } \xi \neq 0, \\ \mu - \sigma \log\{-\log(1 - p)\}, & \text{per } \xi = 0. \end{cases} \quad (3.2)$$



Così, specificare una prior per  $x_p$  per tre valori di  $p$  equivale a specificare una prior per  $(\mu, \sigma, \xi)$  ma con una scala di misura che l'esperto è in grado interpretare ( $x_p$  è il livello di ritorno del periodo  $1/p$  anni). Si formuli una prior informativa in termini di  $(x_{p_1}, x_{p_2}, x_{p_3})$  per valori specifici di  $p_1 > p_2 > p_3$ . E' necessaria a questo punto una precisazione: i parametri devono essere ordinati  $x_{p_1} < x_{p_2} < x_{p_3}$ . Quindi, si lavorerà con le differenze

$$\begin{aligned}\tilde{x}_1 &= x_{p_1} - e_1, \\ \tilde{x}_2 &= x_{p_2} - x_{p_1}, \\ \tilde{x}_3 &= x_{p_3} - x_{p_2},\end{aligned}\tag{3.3}$$

dove  $e_1$  indica il limite inferiore dei valori possibili del processo. Inoltre si assume che le priors su queste quantità siano indipendenti, ognuna con supporto corrispondente all'asse dei reali positivi. In particolare, sono state prese delle priors marginali della forma

$$\tilde{x}_i \sim \Gamma(\alpha_i, \beta_i), \quad i = 1, 2, 3.\tag{3.4}$$

L'assunzione di indipendenza delle distribuzioni a priori per  $\tilde{x}_i$  impone alcuni vincoli sulla prior congiunta di  $x_{p_i}$ , ma a priori tali vincoli sembrano ragionevoli.

Considerate le specificazioni sulla distribuzione a priori date dalle equazioni (3.3) e (3.4), la prior congiunta per  $x_{p_i}$  è ottenuta come

$$\begin{aligned}f(x_{p_1}, x_{p_2}, x_{p_3}) &\propto x_{p_1}^{\alpha_1-1} \exp(-\beta_1 x_{p_1}) \prod_{i=2}^3 (x_{p_i} - x_{p_{i-1}})^{\alpha_i-1} \\ &\quad \times \exp\{-\beta_i(x_{p_i} - x_{p_{i-1}})\},\end{aligned}\tag{3.5}$$

su  $0 \leq x_{p_1} \leq x_{p_2} \leq x_{p_3}$ . Sostituendo l'espressione del quantile (3.2) nella (3.5) e moltiplicando per un appropriato Jacobiano della trasformazione

$$(x_{p_1}, x_{p_2}, x_{p_3}) \rightarrow (\mu, \sigma, \xi)$$

si ottiene un'espressione per le priors in termini dei parametri della GEV. Moltiplicando poi, per un'appropriata funzione di verosimiglianza si ottengono le distribuzioni a posteriori,  $\pi(\mu, \sigma, \xi | \mathbf{x})$ , per una costante di proporzionalità. Non è possibile fare dei calcoli analitici delle distribuzioni marginali di  $\pi$ . Comunque, le tecniche MCMC suggeriscono come sia semplice la diretta simulazione da una catena di Markov la cui distribuzione di equilibrio è proprio  $\pi$ . L'analisi si basa su un campionamento di Gibbs, per cui viene aggiornato ognuno dei tre parametri, da  $\pi$ , condizionatamente ai valori correnti degli altri due. Dato che le distribuzioni condizionali non sono di una forma standard, è necessario includere un passo del Metropolis ad ogni iterazione del campionamento di

Gibbs. Questo equivale a simulare da densità arbitrarie, e applicare un criterio del rifiuto al valore simulato. Con un appropriato criterio del rifiuto, si ottiene una catena con la distribuzione di equilibrio marginale corretta. Ulteriori dettagli su queste tecniche sono stati proposti da Smith e Roberts (1993). La procedura è ottimizzata quando la probabilità di rifiuto è bassa.

### 3.3 Modelli gerarchici

I risultati finora ottenuti hanno implicitamente assunto un singolo campione omogeneo. Comunque, questi risultati sono applicabili anche nei problemi di più campioni, in cui i parametri del modello sono differenti da un campione all'altro. Tali problemi portano a quelli che vengono usualmente chiamati metodi di analisi empirica di Bayes. Negli anni recenti è diventato piuttosto comune risolvere tali problemi da un punto di vista completamente bayesiano, utilizzando un modello gerarchico per collegare i parametri di diversi sottocampioni. Questo è il punto di vista adottato nella monografia di Carlin e Louis (1996).

Solitamente, i dati di cui si dispone quando si pensa ad un modello gerarchico sono dati di natura eterogenea divisi in gruppi e i dati di ogni singolo gruppo si distribuiscono secondo una certa distribuzione di probabilità.

Smith e Goodman (2000) hanno analizzato dati che si riferiscono alle richieste di risarcimento dei danni di una famosa compagnia di assicurazione per un periodo di 15 anni. Le richieste di risarcimento sono state classificate in tipi di richieste (incendio, inondazioni, ecc.). Gli autori per analizzare questi gruppi utilizzano un modello gerarchico assumendo la scambiabilità dei parametri. Il modello gerarchico ha la seguente struttura:

Livello I. I parametri  $m_\mu$ ,  $m_\sigma$ ,  $m_\xi$ ,  $s_\mu^2$ ,  $s_\sigma^2$ ,  $s_\xi^2$  sono generati da una distribuzione a priori che verrà specificata successivamente.

Livello II. Condizionatamente ai parametri del 1° Livello, i parametri  $\mu_1, \dots, \mu_J$  (dove  $J$  è il numero di tipi) sono estratti indipendentemente da  $N(m_\mu, s_\mu^2)$ , la distribuzione normale con media  $m_\mu$  e varianza  $s_\mu^2$ . Allo stesso modo,  $\log \psi_1, \dots, \log \psi_J$  sono estratti indipendentemente da  $N(m_\psi, s_\psi^2)$ ,  $\xi_1, \dots, \xi_J$  sono estratti indipendentemente da  $N(m_\xi, s_\xi^2)$ .

Livello III. Condizionatamente al Livello II, per ogni  $j \in \{1, \dots, J\}$ , il processo di punto dei superamenti di tipo  $j$  è generato dal processo di Poisson definito dalla (2.1) con parametri  $\mu_j$ ,  $\psi_j$  (dove  $\psi = \sigma + \xi(u - \mu)$ ),  $\xi_j$ .

La distribuzione a priori per  $(m_\mu, s_\mu^2)$  è del tipo “Normal-Gamma”: sia  $\tau_\mu$  estratto da una distribuzione Gamma con parametro di forma  $\alpha$  e parametro di scala  $\beta$  e, dato  $\tau_\mu$ , si definisce  $s_\mu^2 = 1/\tau_\mu$  e sia  $m_\mu \sim N(\nu, \frac{1}{\kappa\tau_\mu})$ . Smith e Goodman indicano questa distribuzione come  $(m_\mu, s_\mu^2) \sim GN(\alpha, \beta, \nu, \kappa)$ . In maniera simile,  $(m_\psi, s_\psi^2)$  e  $(m_\xi, s_\xi^2)$  saranno indipendentemente estratte dalla stessa distribuzione. Fissando  $\alpha = \beta = \kappa = 0.001$  e  $\nu = 0$  si ottiene una distribuzione a priori molto diffusa.

Si può facilmente dimostrare che la famiglia GN è coniugata: dato  $\mu_1, \dots, \mu_J$  e definendo  $\bar{\mu} = \frac{1}{J}(\mu_1 + \dots + \mu_J)$ , la distribuzione a posteriori di  $(m_\mu, s_\mu^2)$  è ancora una  $GN(\alpha', \beta', \nu', \kappa')$  dove

$$\begin{aligned}\alpha' &= \alpha + \frac{J}{2}, \\ \beta' &= \beta + \frac{1}{2} \frac{J\kappa}{J+\kappa} (\bar{\mu} - \nu)^2 + \frac{1}{2} \sum_j (\mu_j - \bar{\mu})^2, \\ \nu' &= \frac{\kappa\nu + J\bar{\mu}}{\kappa + J}, \\ \kappa' &= \kappa + J.\end{aligned}$$

Formule simili si applicano, naturalmente, per i parametri  $\psi$  e  $\xi$ .

La procedura di stima utilizzata è un algoritmo di simulazione MCMC in cui i valori delle quantità non note nei Livelli I e II si aggiornano nel modo seguente:

- 1: (Gibbs step) Dati i valori correnti di  $\{\mu_j, \psi_j, \xi_j, j = 1, \dots, J\}$ , i nuovi valori dei parametri del Livello I  $m_\mu, m_\psi, m_\xi, s_\mu^2, s_\psi^2, s_\xi^2$  sono generati da appropriate distribuzioni a posteriori.
- 2: (Hastings-Metropolis step) Dati i valori osservati e i valori correnti di  $m_\mu, m_\psi, m_\xi, s_\mu^2, s_\psi^2, s_\xi^2$ , ognuno dei parametri  $\{\mu_j, \psi_j, \xi_j, j = 1, \dots, J\}$  viene aggiornato generando una piccola perturbazione e accettando o rifiutando il nuovo valore secondo il criterio Hastings-Metropolis.

Questa procedura è stata ripetuta fino a  $10^6$  iterazioni. Inoltre, ogni 100 cicli, è stato generato un nuovo valore per la “perdita futura” della compagnia assicuratrice, condizionatamente ai valori correnti di  $\mu_j, \psi_j, \xi_j$  per ogni  $j$ , e i risultati sono stati aggregati su tutte le classi. Così si è ottenuto un campione MCMC dalla distribuzione predittiva.

### 3.3.1 Raggiungimento di popolazioni eterogenee tramite distribuzioni mistura

Si supponga esistano  $I$  gruppi, in modo che i dati possano essere ordinati come  $y_i = (y_{il}, l = 1, \dots, n_i), i=1, \dots, I$ , dove  $y_{il}$  rappresenta l' $l$ -esima eccedenza rispetto alla soglia

$u_i$  per il gruppo  $i$ . Il classico approccio bayesiano per modellare l'eterogeneità dei gruppi è quello di assumere uno specifico parametro  $\eta_i$  per la distribuzione delle  $y_i$  e modellare gerarchicamente  $\eta = (\eta_1, \dots, \eta_I)$ .

Si utilizzi a questo punto un approccio basato sulle distribuzioni mistura. Condizionatamente al numero delle componenti-mistura  $k$ , i pesi  $w_j$  e gli iperparametri  $\delta_j$ , gli  $\eta_i$  sono assunti i.i.d. secondo la distribuzione mistura-finita

$$\eta_i | k, w, \delta \stackrel{iid}{\sim} \sum_{j=1}^k w_j p_j(\cdot | \delta_j), \quad (3.6)$$

dove  $p_j(\cdot | \cdot)$  indica la  $j$ -esima componente della distribuzione, e sia  $w = (w_1, \dots, w_k)$  un vettore con  $\sum_{j=1}^k w_j = 1$ , e sia  $\delta = (\delta_1, \dots, \delta_k)$ . Si possono ora assegnare le distribuzioni a priori, possibilmente utilizzando una ulteriore gerarchia, per  $k$ ,  $w$  e  $\delta$ .

Si può riscrivere la (3.6) utilizzando, indipendentemente per ogni  $\eta_i$ , una variabile latente  $z_i$  tale che  $Pr = (z_i = j | k, w) = w_j$ . La variabile  $z_i$  identifica la componente della mistura da cui  $\eta_i$  è estratta. La formula (3.6) può essere sostituita con

$$\begin{aligned} \eta_i | \delta, z_i &\stackrel{ind}{\sim} p_{z_i}(\cdot | \delta_{z_i}) \\ z_i | k, w &\stackrel{iid}{\sim} Pr(z_i = j | k, w) = w_j. \end{aligned}$$

L'impiego della variabile  $z_i$  è uno strumento utile dal punto di vista computazionale.

I modelli mistura bayesiani con un numero sconosciuto di parametri sono diventati un potente strumento per l'analisi statistica, come si vede nel lavoro di Richardson e Green (1997), i quali utilizzano l'algoritmo RJMCMC proposto da Green (1995).

### 3.3.2 Un modello mistura gerarchico

Seguendo il lavoro di Smith (1989), risulta ormai noto che è possibile vedere i superamenti rispetto ad una soglia come generati da un processo di Poisson di intensità

$$\Lambda_y = \left(1 + \xi \frac{y - \mu}{\psi}\right)_+^{-1/\xi}, \quad (3.7)$$

dove  $x_+ = \max(0, x)$ . Uno dei vantaggi di questo tipo di rappresentazione è che questo può essere immediatamente estendibile a situazioni più complesse, come quelle con parametri dipendenti dal tempo, si veda Smith e Goodman (2000). Condizionatamente a  $\mu_i$ ,  $\psi_i$  e  $\xi_i$  si assume che, indipendentemente, il processo di punto dei superamenti di tipo  $i$  è generato secondo il processo di Poisson (3.7) con parametri  $\mu_i$ ,  $\psi_i$  e  $\xi_i$ . Seguendo Nobile e Green (2000) si può specificare una distribuzione mistura gerarchica a priori per l'insieme dei parametri  $\mu_i$ ,  $\psi_i$  e  $\xi_i$ . Quando non vengono date specifiche indicazioni, si può assumere l'indipendenza condizionata.

Siano

$$\begin{aligned}
\mu_i | k^\mu, w^\mu, \theta^\mu, \sigma^\mu &\stackrel{iid}{\sim} \sum_{j=1}^{k^\mu} w_j^\mu N(\theta_j^\mu, (\sigma_j^\mu)^2), \\
\log \psi_i | k^\psi, w^\psi, \theta^\psi, \sigma^\psi &\stackrel{iid}{\sim} \sum_{j=1}^{k^\psi} w_j^\psi N(\theta_j^\psi, (\sigma_j^\psi)^2), \\
\xi_i | k^\xi, w^\xi, \theta^\xi, \sigma^\xi &\stackrel{iid}{\sim} \sum_{j=1}^{k^\xi} w_j^\xi N(\theta_j^\xi, (\sigma_j^\xi)^2).
\end{aligned} \tag{3.8}$$

Si noti che i tre parametri non sono trattati simmetricamente, dato che viene specificata una distribuzione mistura differente, con numero specifico di componenti, di pesi, di medie e di varianze per ognuno dei tre parametri. Questo tipo di distribuzione a priori ammette un elevato grado di flessibilità, e probabilmente comprende molte più situazioni di quante se ne possano incontrare in pratica.

Per completare la specificazione del modello, le distribuzioni degli iperparametri che compaiono nella (3.8) devono essere spiegate più nel dettaglio. Per semplicità verrà presentata solo quella riferita a  $\mu$ . Espressioni simili valgono anche per  $\log \psi$  e  $\xi$ .

Si assuma che

$$\begin{aligned}
k^\mu &\sim Pr(k^\mu = k), \quad k = 1, \dots, I, \\
w^k | k^\mu &\sim Dir(d_1^\mu, \dots, d_{k^\mu}^\mu), \\
\theta_j^\mu | \tau^\mu &\stackrel{ind}{\sim} N(m_j^\mu, \frac{1}{\tau^\mu}), \\
(\sigma_j^\mu)^{-2} &\stackrel{ind}{\sim} Ga(a_j^\mu, b_j^\mu), \\
\tau^\mu &\sim Ga(a^{\tau^\mu}, b^{\tau^\mu}),
\end{aligned} \tag{3.9}$$

dove  $Dir$  indica la distribuzione di Dirichlet,  $Ga(a, b)$  la distribuzione gamma con valore atteso  $a/b$ .

Successivamente, vengono assegnate le distribuzioni a priori sugli iperparametri, in termini di distribuzioni di probabilità indotte sulle *partizioni* degli  $I$  tipi per ogni fissato  $k^\mu$ . Questo è l'approccio proposto da Nobile e Green (2000) ormai ampiamente utilizzato. Facilmente questa procedura può essere implementata utilizzando il RJMCMC e dato il risultato delle simulazioni si può fare inferenza sulla struttura dei raggruppamenti, attraverso la distribuzione delle frequenze del vettore delle partizioni.



# Capitolo 4

## Alcuni elementi dell'impostazione bayesiana non-parametrica

### 4.1 Introduzione

Nel 1972 D. V. Lindely scrisse: “*Nonparametric Statistics. This is a subject about which the Bayesian method is embarrassingly silent*”. Da allora la statistica bayesiana non-parametrica si è sviluppata molto sia da un punto di vista teorico, sia da un punto di vista computazionale (Walker *et al.*, 1999). Però, la prima impostazione bayesiana di un problema non-parametrico fu proposta da Bruno de Finetti (1935) in un lavoro che riguardava il problema della sostituzione di una “distribuzione teorica” a dati empirici. Egli propose due metodi alternativi per affrontare la questione, il secondo dei quali coincide con ciò che oggi è noto come impostazione bayesiana non-parametrica.

Diversi sono i problemi che si pongono nell'inferenza bayesiana non-parametrica: la scelta fra impostazione parametrica o non-parametrica; l'assegnazione di una distribuzione iniziale su una misura di probabilità aleatoria; la determinazione della distribuzione finale e della distribuzione di funzionali; l'analisi non-parametrica di dati con strutture di dipendenza complesse (in particolare la scambiabilità parziale). In questo capitolo verranno evidenziati soprattutto gli aspetti relativi all'assegnazione delle priors e quindi i relativi aggiornamenti e si daranno le definizioni dei due tipi di scambiabilità parziale che verranno di frequente richiamate in questo lavoro.

### 4.2 Alcune priors utili nell'impostazione bayesiana non-parametrica

Le prime distribuzioni a priori si devono a Freedman (1963) e Fabius (1964), che introdussero i *processi tailfree*. Furono però i contributi di Ferguson (1973; 1974), sulla base

dei lavori di Fabius, Dubins e Freedman (1965) e Freedman (1965), che diedero la spinta decisiva alla applicazione delle metodologie bayesiane per la soluzione di svariati problemi di inferenza non-parametrica. Ferguson (1973) introdusse il *processo di Dirichlet* come misura di probabilità aleatoria su spazi misurabili astratti.

Diverse distribuzioni a priori sono state successivamente proposte in letteratura, alcune ristrette alla stima puntuale, coinvolgendo il processo di Dirichlet, come in Susarla & Van Ryzin (1976), e altre che coinvolgono la classe dei *processi neutrali a destra* (Doksum, 1974), come in Ferguson & Phadia (1979).

Molti sono i processi che appartengono a questa classe: il processo omogeneo semplice (Ferguson e Phadia, 1979), il processo gamma (Ferguson, 1974) e il processo beta-Stacy (Walker e Muliere, 1997a), che è un processo neutrale a destra che generalizza i precedenti. I processi neutrali a destra sono caratterizzati mediante i processi di Levy.

Di seguito, verranno descritti alcuni tra i principali processi che possono essere utilizzati come distribuzioni a priori.

### 4.2.1 Processo di Dirichlet

Nell'articolo del 1973 Ferguson oltre a stabilire la possibilità teorica di definire una misura di probabilità aleatoria introduce la prima classe di distribuzioni iniziali, il processo di Dirichlet.

Questo processo rappresentò una “spinta” per la letteratura successiva rompendo definitivamente quell'*imbarazzante silenzio* (Lindley, 1972) che per decenni aveva caratterizzato la statistica bayesiana non-parametrica.

Sia  $\mathcal{X}$  uno spazio e  $\mathcal{A}$  una  $\sigma$ -algebra di sottoinsiemi e si definisca una probabilità aleatoria,  $P$ , su  $(\mathcal{X}, \mathcal{A})$  definendo la distribuzione congiunta delle v.a.  $(P(B_1), \dots, P(B_k))$  per tutti i  $k$  e tutte le partizioni misurabili  $(B_1, \dots, B_k)$  di  $\mathcal{X}$ . (Si dice che  $(B_1, \dots, B_k)$  è una partizione misurabile di  $\mathcal{X}$  se  $B_i \in \mathcal{A}$  per tutti i valori di  $i$ ,  $B_i \cap B_j = \emptyset$  per  $i \neq j$ , e  $\bigcup_{j=1}^k B_j = \mathcal{X}$ ).

**Definizione 4.2.1.** *Sia  $\alpha$  una misura finita non-nulla (non-negativa e finitamente additiva) su  $(\mathcal{X}, \mathcal{A})$ . Si dice che  $P$  è un processo di Dirichlet su  $(\mathcal{X}, \mathcal{A})$  con parametro  $\alpha$  se per ogni  $k = 1, 2, \dots$ , e per ogni partizione misurabile  $(B_1, \dots, B_k)$  di  $\mathcal{X}$ , la distribuzione di  $(P(B_1), \dots, P(B_k))$  è Dirichlet,  $\mathcal{D}(\alpha(B_1), \dots, \alpha(B_k))$ .*

Una volta assegnata la distribuzione iniziale sul parametro infinito-dimensionale, si dovrà calcolare la distribuzione finale. Purtroppo, non sempre tale calcolo è agevole. Un motivo per cui il processo di Dirichlet è così ampiamente utilizzato come iniziale non-parametrica nell'inferenza bayesiana, è dato dalla semplicità della regola di aggiornamento dei parametri della distribuzione a posteriori.



Per  $x \in \mathcal{X}$ ,  $\delta_x$  indichi la misura su  $(\mathcal{X}, \mathcal{A})$  che dà massa uno al punto  $x$ :

$$\delta_x(A) = \begin{cases} 1 & \text{se } x \in A \\ 0 & \text{se } x \notin A \end{cases}$$

dove  $A \in \mathcal{A}$ .

**Teorema 4.2.1.** *Sia  $P$  un processo di Dirichlet su  $(\mathcal{X}, \mathcal{A})$  con parametro  $\alpha$ , e sia  $X_1, \dots, X_n$  un campione di dimensione  $n$  estratto da  $P$ . Allora la distribuzione condizionata di  $P$  dato  $X_1, \dots, X_n$  è ancora un processo di Dirichlet con parametro  $\alpha + \sum_{i=1}^n \delta_{X_i}$ .*

Dimostrazione. Si veda Ferguson [rif] Teorema 1.

Una caratteristica dei processi di Dirichlet è che la distribuzione a posteriori della v.a.  $P(A)$  con  $A \in \mathcal{A}$  dato il campione  $(X_1, \dots, X_n)$ , dipende solo dal numero di osservazioni che cadono in  $A$ , e non di quelle che cadono fuori. Se questo da una parte rende facile determinare la distribuzione a posteriori del processo, dall'altra impedisce di tener conto di dati censurati del tipo  $X \geq x$ .

## 4.2.2 Processi di Lévy

I processi di Lévy sono dei processi stocastici i cui incrementi, in intervalli di tempo che non si sovrappongono, sono indipendenti e stazionari nel tempo. Inoltre si assume una continuità debole chiamata *continuità stocastica*.

**Definizione 4.2.2.** *Un processo stocastico  $\{L_t : t \geq 0\}$  su  $\mathbb{R}$  è un processo di Lévy, definito su uno spazio di probabilità  $(\Omega, \mathcal{F}, P)$ , se sono soddisfatte le seguenti condizioni:*

- (1) *per una qualsiasi scelta di  $n \geq 1$  e  $0 \leq t_0 < t_1 < \dots < t_n$ , le v.a.  $L_{t_0}, L_{t_1} - L_{t_0}, L_{t_2} - L_{t_1}, \dots, L_{t_n} - L_{t_{n-1}}$  sono indipendenti (proprietà degli incrementi indipendenti);*
- (2)  *$L_0 = 0$  q.c.;*
- (3) *la distribuzione di  $L_{s+t} - L_s$  non dipende da  $s$  (proprietà degli incrementi stazionari o dell'omogeneità temporale);*
- (4) *è continuo stocasticamente;*
- (5) *esiste un  $\Omega_0 \in \mathcal{F}$  con  $P[\Omega_0] = 1$  tale che, per ogni  $\omega \in \Omega_0$ ,  $L_t(\omega)$  è continuo a destra in  $t \geq 0$  ed ha limite sinistro in  $t > 0$ .*

Qualora venisse a mancare la condizione di stazionarietà per gli incrementi del processo di Lévy, allora  $L(t)$  verrà chiamato *processo additivo*. Le distribuzioni di Lévy e i processi additivi in qualsiasi tempo sono *infinitamente divisibili*, cioè posseggono le  $n$ -esime radici nel senso della convoluzione per ogni  $n$ . Quando viene fissato un tempo, la classe dei processi di Lévy è in corrispondenza uno-a-uno con la classe delle distribuzioni infinitamente divisibili.

Seguendo Gikhman e Skorokhod (1965), ogni processo di Lévy  $L(\cdot)$  può essere scritto come somma di una componente di salto fissata ed una componente 'continua': se  $t_1, t_2, \dots$  corrispondono ai punti fissati di discontinuità con salti non-negativi indipendenti  $L\{t_1\}, L\{t_2\}, \dots$  (indipendenti anche dal resto del processo), allora

$$L(t) = L_c(t) + \sum_j L\{t_j\}I(t_j \leq t) \quad (4.1)$$

dove  $L_c(\cdot)$  è un processo non-decrescente senza punti fissati di discontinuità e una rappresentazione di Lévy per la trasformata di Laplace

$$E\left\{e^{-\psi L_c(t)}\right\} = \exp\left\{-\int_0^\infty (1 - e^{-\psi\nu}) dN_t(\nu)\right\} \quad (4.2)$$

con  $N_t(\nu)$  misura di Lévy che soddisfi le seguenti proprietà:

1. per ogni insieme di Borel  $B$ ,  $N_t(B)$  è continuo e non-decrescente;
2. per ogni insieme reale  $t > 0$ ,  $N_t(\cdot)$  è una misura sugli insiemi di Borel di  $(0, \infty)$ ;
3.  $\int_0^1 \nu dN_t(\nu) < \infty$ ;
4.  $\int_0^1 dN_t(\nu) < \infty$ .

### 4.2.3 Processi neutrali a destra

#### Concetti preliminari

Sia  $\mathcal{B}(\alpha, \beta)$  per  $\alpha, \beta > 0$  una distribuzione beta. Per comodità, conviene definire  $\mathcal{G}(\alpha_1, \beta_1, \dots, \alpha_m, \beta_m)$  per  $\alpha_j, \beta_j > 0$  come la distribuzione di Dirichlet generalizzata, introdotta da Connor e Mosimann (1969). La funzione di densità, a meno di una costante, è data da

$$\begin{aligned} & y_1^{\alpha_1-1}(1-y_1)^{\beta_1-1} \\ & \times \frac{y_2^{\alpha_2-1}(1-y_1-y_2)^{\beta_2-1}}{(1-y_1)^{\alpha_2+\beta_2-1}} \\ & \dots \\ & \times \frac{y_m^{\alpha_m-1}(1-y_1-\dots-y_{m-1}-y_m)^{\beta_m-1}}{(1-y_1-\dots-y_{m-1})^{\alpha_m+\beta_m-1}} I\{(y_1, \dots, y_m) : y_j \geq 0, \sum_{j=1}^m y_j \leq 1\}, \end{aligned} \quad (4.3)$$

dove  $I$  indica la funzione indicatrice.

**Definizione 4.2.3.**  $\mathcal{C}(\alpha, \beta, \xi)$  con  $\alpha, \beta > 0$  e  $0 < \xi \leq 1$  è detta *distribuzione beta-Stacy* se la funzione di densità è data da

$$\frac{1}{B(\alpha, \beta)} y^{\alpha-1} \frac{(\xi - y)^{\beta-1}}{\xi^{\alpha+\beta-1}} I_{(0, \xi)}(y),$$

dove  $B(\alpha, \beta)$  è la usuale funzione beta.

Si noti che se  $Y \sim \mathcal{C}(\alpha, \beta, \xi)$  allora  $Y/\xi \sim (\alpha, \beta)$  e la distribuzione beta si ottiene se  $\xi = 1$ .

Nel 1974 Doksum ha dimostrato che, risultati simili a quelli di Ferguson possono essere ottenuti per una classe più ampia di processi. Definisce quindi, i *processi neutrali a destra* (NTR) a partire da una misura di probabilità aleatoria neutrale rispetto a partizioni dei reali secondo gli intervalli  $(-\infty, t_1), (t_1, t_2), \dots, (t_k, +\infty)$  con  $t_1 < \dots < t_k$ .

**Definizione 4.2.4.** (Doksum, 1974). Una funzione di distribuzione aleatoria  $F$  su  $(0, \infty)$  è detta *neutrale a destra* se per ogni  $k > 1$  e  $t_1 < \dots < t_k$  esistono delle v.a. indipendenti non-negative  $V_1, \dots, V_k$  tali che

$$(1 - F(t_1), 1 - F(t_2), \dots, 1 - F(t_k)) \stackrel{\mathcal{L}}{=} (V_1, V_1 V_2, \dots, \prod_{j=1}^k V_j).$$

Se  $F$  è NTR, allora una sua trasformazione

$$L(t) = -\log[1 - F(t)]$$

ha incrementi indipendenti. E' vero anche l'inverso: se  $L(t)$  è

- un processo neutrale a destra,
- non decrescente q.c.,
- continuo a destra q.c.,
- $L(0) = 0$  q.c.,
- $\lim_{t \rightarrow \infty} L(t) = +\infty$  q.c.,

allora

$$F(t) = 1 - \exp[-L(t)]$$

è NTR, con  $L(\cdot)$  processo di Levy ha incrementi indipendenti.

Il risultato fondamentale per i processi neutrali a destra è:

**Teorema 4.2.2.** (Doksum, 1974; Ferguson, 1974). *Se  $F$  è NTR e  $X_1, \dots, X_n$  è un campione estratto da  $F$ , che include la possibilità di campioni censurati a destra (dove  $X_i$  rappresenta il tempo di censura se applicabile), allora la posterior di  $F$  dato  $X_1, \dots, X_n$  è ancora NTR.*

La distribuzione a priori per un processo sarà caratterizzata da

$$M = \{t_1, t_2, \dots\}, \{f_{t_1}, f_{t_2}, \dots\},$$

rispettivamente i punti fissati di discontinuità e le corrispondenti funzioni di densità per i salti. La distribuzione a posteriori verrà ora data per una singola osservazione  $X$ .

**Teorema 4.2.3.** (Ferguson, 1974; Ferguson e Phadia, 1979). *Sia  $F$  NTR e sia  $X$  un campione aleatorio estratto da  $F$ .*

(i) *Dato  $X > x$  i parametri a posteriori (indicati con  $*$ ) sono  $M^* = M$ ,*

$$f_{t_j}^*(\nu) = \begin{cases} c \cdot e^{-\nu} f_{t_j}(\nu) & \text{se } t_j \leq x \\ f_{t_j}(\nu) & \text{se } t_j > x \end{cases}$$

*e  $K^*(\nu, s) = \exp[-\nu I(x \geq s)]K(\nu, s)$  (dove  $c$  è la costante di normalizzazione).*

(ii) *Dato  $X = x \in M$  i parametri a posteriori sono  $M^* = M$ ,*

$$f_{t_j}^*(\nu) = \begin{cases} c \cdot e^{-\nu} f_{t_j}(\nu) & \text{se } t_j < x \\ c \cdot (1 - e^{-\nu}) f_{t_j}(\nu) & \text{se } t_j = x \\ f_{t_j}(\nu) & \text{se } t_j > x \end{cases}$$

*e, di nuovo,  $K^*(\nu, s) = \exp[-\nu I(x \geq s)]K(\nu, s)$ .*

(iii) *Dato  $X = x \notin M$  i parametri a posteriori sono  $M^* = M \cup \{x\}$ , con  $f_x(\nu) = c \cdot (1 - e^{-\nu})K(\nu, x)$ ,*

$$f_{t_j}^*(\nu) = \begin{cases} c \cdot e^{-\nu} f_{t_j}(\nu) & \text{se } t_j < x \\ f_{t_j}(\nu) & \text{se } t_j > x \end{cases}$$

*e, di nuovo,  $K^*(\nu, s) = \exp[-\nu I(x \geq s)]K(\nu, s)$ .*

## Dati censurati

Un numero di individui viene osservato dal tempo di entrata fino al verificarsi di un particolare evento (ad esempio, la morte). Spesso, il tempo esatto di morte non è noto per tutti gli individui; per alcuni si sa solamente che quell'evento in un tempo specifico non si è ancora verificato, ed in questo caso l'osservazione è *censurata a destra*. Si veda Andersen et al. (1993) per una serie di esempi. Formalmente, il modello considerato è

il seguente: si considerino  $n$  individui con tempi di sopravvivenza  $X_1, X_2, \dots, X_n$ . Ogni  $X_i$  corrisponde al tempo di morte oppure si può solo dire che il tempo di morte è più grande di  $X_i$ . I dati quindi saranno rappresentati così

$$(X_1, \delta_1), \dots, (X_n, \delta_n)$$

dove  $\delta_i = 1$  se si verifica la morte e  $\delta_i = 0$  se si verifica la censura.

Si consideri il caso discreto, in cui ogni  $X_i \in \Omega = \{1, 2, \dots\}$ . E' necessario a questo punto studiare le conseguenze delle assunzioni che riguardano la predittiva:

$$P(X_{n+1} = k | X_1, \dots, X_n) = f_k(n_1, \dots, n_k, m_k), \quad (4.4)$$

per certi valori di  $f_k$ , dove  $n_k = \sum_{1 \leq i \leq n} I(X_i = k)$  e  $m_k = \sum_{1 \leq i \leq n} I(X_i > k)$ .

E' possibile dimostrare che l'assunzione di scambiabilità, combinata con la (6.2), implica un processo a priori neutrale a destra per la sequenza.

**Teorema 4.2.4.** (Walker e Muliere, 1997b). *Una sequenza scambiabile  $X_1, X_2, \dots$  in cui ogni  $X_i$  è definita su  $\Omega = \{1, 2, \dots\}$  ha come prior un processo neutrale a destra se, e solo se, per ogni  $n = 1, 2, \dots$  e  $k \in \Omega$ ,*

$$P(X_{n+1} = k | X_1, \dots, X_n) = f_k(n_1, \dots, n_k, m_k), \quad (4.5)$$

dove  $n_k = \sum_{1 \leq i \leq n} I(X_i = k)$  e  $m_k = \sum_{1 \leq i \leq n} I(X_i > k)$ .

Per rendere questo teorema utile nelle applicazioni, occorre specificare la forma della funzione  $f_k$ . Di seguito, verrà presentata la costruzione generale di un processo NTR come prior.

**Corollario 1.** *Una sequenza scambiabile  $X_1, X_2, \dots$ , in cui ogni  $X_i$  è definita sullo spazio  $\Omega = \{1, 2, \dots\}$ , ha come prior un processo NTR se, e solo se,*

$$P(X_{n+1} = k | X_1, \dots, X_n) = \frac{E\{V_k^{n_k+1}(1-V_k)^{m_k} \prod_{j < k} V_j^{n_j}(1-V_j)^{m_j+1}\}}{E\{V_k^{n_k}(1-V_k)^{m_k} \prod_{j < k} V_j^{n_j}(1-V_j)^{m_j}\}} \quad (4.6)$$

dove  $n_k = \sum_{1 \leq i \leq n} I(X_i = k)$  e  $m_k = \sum_{1 \leq i \leq n} I(X_i > k)$ .

Per poter trattare la distribuzione predittiva, le variabili  $V_k$  avranno distribuzione beta,  $V_k \sim \mathcal{B}(\alpha_k, \beta_k)$ .

## Il processo beta-Stacy

Si consideri un processo NTR quando le  $V_k$  sono delle variabili  $\mathcal{B}(\alpha_k, \beta_k)$  indipendenti, che verrà chiamato processo beta-Stacy. Questo processo comprende virtualmente tutti i processi NTR menzionati in letteratura.

*Caso discreto.* Dalla (4.4) si noti che  $P(X_1 = k) = E(V_k)$  ed è facile da dimostrare che  $P(X_{n+1} = k | X_1, \dots, X_n) = E(V_k^*)$ , dove  $V_k^* \sim \mathcal{B}(\alpha_k + n_k, \beta_k + m_k)$ , dando

$$P(X_{n+1} = k | X_1, \dots, X_n) = \frac{\alpha_k + n_k}{\beta_k + \alpha_k + n_k + m_k} \prod_{j=1}^{k-1} \frac{\beta_j + m_j}{\beta_j + \alpha_j + n_j + m_j}. \quad (4.7)$$

E' inoltre interessante notare che se

$$Y_1 \sim \mathcal{C}(\alpha_1, \beta_1, 1),$$

$$Y_2 | Y_1 \sim \mathcal{C}(\alpha_2, \beta_2, 1 - Y_1),$$

...

$$Y_k | Y_{k-1}, \dots, Y_1 \sim \mathcal{C}(\alpha_k, \beta_k, 1 - F_{k-1}), \quad (4.8)$$

dove  $F_k = \sum_{j=1}^k Y_j$ , allora  $F$ , definita da una sequenza numerabile di v.a.  $Y_k$ , è estratta da un processo beta-Stacy e, per ogni  $m > 1$ ,

$$\mathcal{L}(Y_1, \dots, Y_m) = \mathcal{G}(\alpha_1, \beta_1, \dots, \alpha_m, \beta_m).$$

Se si impongono dei vincoli sui parametri della distribuzione beta è possibile ottenere differenti processi che appartengono alla classe dei processi NTR. In particolare, si osservi che il processo di Dirichlet si ha quando

$$\beta_j = \sum_{k>j} \alpha_k \quad \left( \sum_{j=1}^{\infty} \alpha_j < \infty \right).$$

Per quanto riguarda l'aggiornamento dei parametri di un processo beta-Stacy vale il seguente teorema:

**Teorema 4.2.5.** (Walker e Muliere, 1997b). *Sia  $X_1, \dots, X_n$  con ogni  $X_i \in \{t_k : k \geq 1\}$ , un campione iid, possibilmente con censura a destra, con una distribuzione  $F$  non nota. Se  $F$  è estratta da un processo beta-Stacy a tempo discreto con parametri  $\{\alpha_k, \beta_k\}$  e salti in  $\{t_k\}$ , allora, dato  $X_1, \dots, X_n$ , la distribuzione a posteriori per  $F$  è ancora un processo beta-Stacy a tempo discreto con salti in  $\{t_k\}$  e parametri  $\{\alpha_k^*, \beta_k^*\}$  dove*

$$\alpha_k^* = \alpha_k + n_k \quad e \quad \beta_k^* = \beta_k + m_k \quad (4.9)$$

e  $m_k$  è la somma del numero di osservazioni esatte in  $\{t_j : j > k\}$  e di osservazioni censurate in  $\{t_j : j \geq k\}$ ; cioè,  $m_k = \sum_{j>k} n_j + \sum_{j \geq k} r_j$ , con  $n_k$  numero di osservazioni esatte in  $t_k$  e  $r_k$  numero dei tempi di censura in  $t_k$  ( $X > t_k$ ).

*Caso continuo.* Sia  $c(\cdot)$  una funzione positiva, sia  $G \in \mathcal{F}$  continua a destra e sia  $\{t_k\}$  l'insieme numerabile di punti di discontinuità di  $G$ , cioè  $G\{t_k\} = G(t_k) - G(t_{k-}) > 0$  per tutti i  $k$ . Ora si ponga  $G_c(t) = G(t) - \sum_{t_k < t} G\{t_k\}$  dove  $G_c(\cdot)$  sia la parte continua.

**Definizione 4.2.5.** (Walker e Muliere, 1997b).  $F$  sia un processo beta-Stacy su  $([0, \infty), \mathcal{A})$  con parametri  $c(\cdot)$  e  $G$ ,  $F \sim \mathcal{S}(c(\cdot), G)$ , se per tutti i  $t \geq 0$ ,  $F(t) = 1 - \exp(-L(t))$ , dove  $L$  è il processo di Lévy con misura di Lévy per  $L(t)$  data, per  $\nu > 0$ , da

$$dN_t(\nu) = \frac{d\nu}{(1 - \exp(-\nu))} \int_0^t \exp(-\nu c(s)G(s, \infty))c(s)dG_c(s) \quad (4.10)$$

e la funzione generatrice data da

$$\log E \exp\{-\phi L(t)\} = \sum_{t_k \geq t} \log E \exp(-\phi S_k) + \int_0^\infty (\exp(-\nu\phi) - 1)dN_t(\nu), \quad (4.11)$$

dove  $1 - \exp(-S_k) \sim \mathcal{B}(c(t_k)G\{t_k\}, c(t_k)G\{t_k, \infty\})$ .

LEMMA 1. Se  $U_k$  è una sequenza di v.a. non-negative tali che  $U_k \leq 1$  a.s.,  $\{U_k\}$  è non-decrescente a.s. e  $EU_k \rightarrow 1$ , allora  $U_k \rightarrow 1$  a.s.

Dato che  $G(t) \rightarrow 1$  con  $t \rightarrow \infty$ ,  $F(t) \leq 1$  a.s. e  $F(t)$  è non-decrescente a.s., allora dal lemma segue che  $F(t) \rightarrow 1$  a.s. Ne segue che a.s.  $F \in \mathcal{F}$  e che  $F$  è un processo neutrale a destra (Doksum, 1974). Inoltre, come  $L$  ha componenti aleatorie,  $L$  aumenta a.s. soltanto in corrispondenza dei salti e  $F$  è con probabilità 1 un numero discreto di  $\mathcal{F}$ .

**Osservazione 4.2.1.** Il processo beta-Stacy generalizza il processo di Dirichlet, questo risulta più evidente se  $G$  è presa continua, dato che, se  $c(s) = c > 0$  per tutte le  $s \geq 0$ , allora la (4.10) diventa

$$\begin{aligned} dN_t(\nu) &= \frac{d\nu}{(1-e^{-\nu})} [\nu^{-1} \exp(-\nu c\{1 - G(s)\})]_0^t \\ &= \frac{d\nu}{(1-e^{-\nu})} (\exp -\nu c\{1 - G(t)\}) - e^{-\nu c}, \end{aligned}$$

ossia la misura di Lévy data in Ferguson (1974) che rappresenta il processo di Lévy corrispondente al processo di Dirichlet quando viene visto come un processo NTR.

#### 4.2.4 Il processo Gamma Esteso

Sia  $F(\cdot)$  una funzione di distribuzione cumulata continua a sinistra tale che  $F(x) = 0$  per  $x \leq 0$ . La funzione di sopravvivenza e la funzione di rischio cumulato corrispondenti ad  $F$  sono definite da  $\bar{F}(x) = 1 - F(x)$  e  $H(x) = -\ln \bar{F}(x)$ . Se per tutte le  $x$ ,

$$H(x) = \int_{[0,x)} r(t)dt$$

dove  $r(x)$  è chiamato tasso di rischio della distribuzione e in un contesto di affidabilità sta ad indicare la propensione al fallimento di un certo articolo in un futuro prossimo dato che l'articolo stesso è sopravvissuto fino al tempo  $x$ .

Si indichi con  $G(\alpha, \beta)$  la distribuzione gamma con densità

$$g(x|\alpha, \beta) = x^{\alpha-1} \exp(-x/\beta) I_{(0, \infty)}(x) / \Gamma(\alpha) \beta^\alpha$$

per  $\alpha, \beta > 0$ ;  $G(0, \beta)$  indica una distribuzione degenera in 0.

Sia  $\alpha(t)$ , con  $t \geq 0$ , una funzione a valori reali non-decrescente continua da sinistra tale che  $\alpha(0) = 0$ , e sia  $\beta(t)$ , con  $t \geq 0$ , una funzione a valori reali positivi e continua da destra, limitata in un intorno di  $t$  sufficientemente grande, con limite sinistro esistente.

Sia inoltre,  $Z(t)$  con  $t \geq 0$ , definito su un appropriato spazio di probabilità  $(\Omega, \mathcal{F}, P)$ , un processo gamma con incrementi indipendenti che corrispondono ad  $\alpha(t)$ . Cioè,  $Z(0) \equiv 0$ ,  $Z(t)$  ha incrementi indipendenti, e per  $t > s$ ,  $Z(t) - Z(s) \sim G(\alpha(t) - \alpha(s), 1)$ .

Si definisca un nuovo processo stocastico da

$$r(t) = \int_{[0, t)} \beta(s) dZ(s), \quad (4.12)$$

dove l'integrazione è fatta rispetto alle traiettorie campionarie del processo  $Z(t)$ . Un processo definito in questa maniera ha una *distribuzione gamma estesa*, e si indica con  $r(t) \sim \Gamma(\alpha(\cdot), \beta(\cdot))$ .

Naturalmente se  $r(t)$  è considerato un tasso di rischio aleatorio, ci sarà una corrispondente funzione di distribuzione cumulata aleatoria data da

$$F(x) = 1 - \exp\left[-\int_{[0, x)} r(t) dt\right].$$

Dal lavoro di Doksum (1974), si desume che  $F(x)$  sarà neutrale a destra solo se  $H(x) = \int_{[0, x)} r(t) dt$  ha incrementi indipendenti. E' facile da vedere che sebbene  $r(t)$  abbia incrementi indipendenti,  $H(x)$  non li avrà, quindi il risultato della distribuzione di Doksum non si potrà applicare.

**Teorema 4.2.6.** *Se  $r(t)$  si distribuisce come una  $\Gamma(\alpha(\cdot), \beta(\cdot))$ , allora  $r(t)$  ha incrementi indipendenti e per valori di  $t$  fissati, la funzione caratteristica di  $r(t)$  in un qualche intorno di 0 è data da*

$$\psi_{r_t}(\theta) = \exp\left[-\int_{[0, t)} \ln(1 - \beta(s)\theta) d\alpha(s)\right],$$

quindi,

$$E[r(t)] = \int_{[0, t)} \beta(s) d\alpha(s), \quad (4.13)$$

e

$$Var[r(t)] = \int_{[0, t)} \beta^2(s) d\alpha(s). \quad (4.14)$$



Posto che  $\alpha(t)$  non sia identicamente pari a zero, si può assumere che le traiettorie campionarie di un processo gamma esteso  $r(t)$  siano tassi di rischio non-decrescenti ben definiti che corrispondono a distribuzioni assolutamente continue. Così la distribuzione condizionata delle osservazioni  $X_1, \dots, X_n$  dato  $r(t)$  sarà definita da

$$P(X_1 \geq x_1, \dots, X_n \geq x_n | r(t)) = \prod_{i=1}^n \exp \left[ - \int_{[0, x_i]} r(t) dt \right]. \quad (4.15)$$

Ovviamente, la (4.15) insieme alla distribuzione di  $r(t)$  determineranno la distribuzione congiunta di  $X_1, \dots, X_n$  e di  $r(t)$ , la quale sarà utilizzata per derivare la distribuzione marginale di  $X_1, \dots, X_n$  e la posterior di  $r(t)$  dati i valori osservati di  $X_1, \dots, X_n$ . Dato che le traiettorie campionarie del processo  $r(t)$  sono funzioni non-decrescenti q.c., si può scegliere la prior all'interno della classe di distribuzioni che abbiano tassi di rischio non-decrescenti.

Assegnando la misura di probabilità con questo metodo, occorre inserire le funzioni  $\alpha(t)$  e  $\beta(t)$ . Un approccio consiste nel definire le funzioni media  $\mu(t)$  e varianza  $\sigma^2(t)$  non-decrescenti. Sembrerebbe quindi ragionevole assegnare a  $\mu(t)$  la migliore "guess" del tasso di rischio ed utilizzare  $\sigma^2(t)$  per misurare l'ammontare dell'incertezza o della variazione nel tasso di rischio nel punto  $t$ . Assumendo che  $\mu(t)$ ,  $\sigma^2(t)$  e  $\alpha(t)$  siano tutte differenziabili, si può utilizzare la (4.13) e la (4.14) e scrivere

$$\mu(t) = \int_{[0, t]} \beta(s) \alpha'(s) ds,$$

e

$$\sigma^2(t) = \int_{[0, t]} \beta^2(s) \alpha'(s) ds.$$

Risolvendo rispetto ad  $\alpha(t)$  e  $\beta(t)$  si ottiene

$$\beta(t) = \frac{d\sigma^2(t)}{dt} \bigg/ \frac{d\mu(t)}{dt}, \quad (4.16)$$

e

$$\frac{d\alpha(t)}{dt} = \left[ \frac{d\mu(t)}{dt} \right]^2 \bigg/ \frac{d\sigma^2(t)}{dt}, \quad (4.17)$$

che verranno poi utilizzate per determinare la distribuzione a priori. La forma della posterior dà informazioni sull'effetto della prior e può essere utile nella scelta di  $\alpha(\cdot)$  e  $\beta(\cdot)$ .

**Teorema 4.2.7.** *Se la prior sui tassi di rischio è una  $\Gamma(\alpha(\cdot), \beta(\cdot))$  allora la funzione marginale di sopravvivenza di una osservazione  $X$  è data da:*

$$\bar{F}(t) = P(X \geq t) = \exp \left[ - \int_{[0, t]} \ln(1 + \beta(s)(t - s)) d\alpha(s) \right]. \quad (4.18)$$

**Corollario 2.** *Se la prior sui tassi di rischio è  $\Gamma(\alpha(\cdot), \beta(\cdot))$ , allora la funzione congiunta marginale di sopravvivenza di  $n$  osservazioni  $X_1, \dots, X_n$  è*

$$\begin{aligned} (\bar{F}(t_1, \dots, t_n)) &= P(X_1 \geq t_1, \dots, X_n \geq t_n) \\ &= \exp \left[ - \int_{[0, \infty)} \ln(1 + \beta(s)) \sum_{i=1}^n (s - t_i)^+ d\alpha(s) \right] \end{aligned} \quad (4.19)$$

dove  $a^+ = \max\{a, 0\}$

Il problema chiave in un contesto bayesiano è quello di derivare la distribuzione a posteriori. E' inoltre importante saper manipolare le osservazioni censurate dato che i dati sull'affidabilità sono di questo tipo. Se viene utilizzata come distribuzioni a priori la gamma estesa, la a posteriori sarà ancora un processo gamma esteso.

**Teorema 4.2.8.** *Se la prior sui tassi di rischio è una  $\Gamma(\alpha(\cdot), \beta(\cdot))$ , allora la posterior sui tassi di rischio, date le  $m$  osservazioni censurate della forma  $X_1 \geq x_1, X_2 \geq x_2, \dots, X_m \geq x_m$ , è una  $\Gamma(\alpha(\cdot), \hat{\beta}(\cdot))$  dove*

$$\hat{\beta}(t) = \frac{\beta(t)}{1 + \beta(t) \sum_{i=1}^m (x_i - t)^+}. \quad (4.20)$$

L'effetto delle osservazioni censurate è di diminuire la pendenza delle traiettorie campionarie a sinistra dei punti di censura mentre di lasciare immutata la pendenza dei valori a destra.

Di seguito si porrà l'attenzione sulla distribuzione a posteriori di  $r(t)$  date le osservazioni esatte. Sivedrà come la posterior possa essere espressa come una mistura continua di distribuzioni gamma estese. La dimensione della "mixing measure" aumenta con la dimensione campionaria.

**Teorema 4.2.9.** *Se la a priori sui tassi di rischio è una  $\Gamma(\alpha(\cdot), \beta(\cdot))$ , allora la distribuzione a posteriori, date le  $m$  osservazioni esatte della forma  $(X_1 = x_1, X_2 = x_2, \dots, X_m = x_m)$ , sarà una mistura di processi gamma estesi. La distribuzione della mistura è data da*

$$\begin{aligned} &P(r(t) \in B | X_1 = x_1, X_2 = x_2, \dots, X_m = x_m) \quad (4.21) \\ &= \frac{\int_{[0, x_m)} \cdots \int_{[0, x_1)} \prod_{i=1}^m \hat{\beta}(z_i) F(B; \Gamma(\alpha + \sum_{i=1}^m I_{(z_i, \infty)}, \hat{\beta})) \prod_{i=1}^m d[\alpha + \sum_{j=i+1}^m I_{(z_j, \infty)}](z_i)}{\int_{[0, x_m)} \cdots \int_{[0, x_1)} \prod_{i=1}^m \hat{\beta}(z_i) d[\alpha + \sum_{j=i+1}^m I_{(z_j, \infty)}](z_i)} \end{aligned}$$

Qui  $F(B; Q)$  indica la probabilità assegnata a  $B \in \mathcal{B}_{\mathbb{R}}$  da un processo stocastico che è distribuito come un processo  $Q$ ,  $\hat{\beta}(\cdot)$  è definito come nella (4.20), e le integrazioni sono fatte rispetto a  $z_1, \dots, z_m$ . Naturalmente  $\sum_{j=m+1}^m I_{(z_j, \infty)} \equiv 0$ .

*Dimostrazione:* Si veda VEDERE RIFERIMENTO

La complessità di questa distribuzione rende difficile vedere come una particolare osservazione  $X_1 = x_1$  possa influenzare la posterior. Un fallimento al tempo  $x_1$  aumenta il tasso di rischio a priori in  $x_1$ . Questo aumento è tanto più piccolo per valori piccoli di  $t$ . Questo viene evidenziato dalla funzione peso  $\hat{\beta}(t) = \beta(t)[1 + \beta(t)(x_1 - t)^+]^{-1}$  che si trova nell'integrale misto. Tale effetto viene limitato dal tasso al quale  $\alpha(t)$  aumenta in modo tale che  $\hat{\beta}(t)$  e  $\alpha(t)$  insieme determinino dove e come si verifica l'aumento nel rischio (il salto unitario nella funzione  $\alpha$ ).

### 4.3 Derivazione della misura di Lévy per alcuni processi

In questa sezione si presenterà l'elenco delle misure di Lévy, calcolate sui processi finora descritti, per la costruzione delle priors bayesiane non-parametriche.

La forma generale per la misura di Lévy che caratterizza il processo di Levy ha la seguente forma:

$$dN_t(\nu) = d\nu \int_0^t K(\nu, s) ds. \quad (4.22)$$

dove  $N_t(\nu)$  soddisfa le condizioni già descritte nel paragrafo 3.2.2.

Forme particolari di questa misura di Levy per alcuni processi sono:

(\*) Processo di Dirichlet:

$$K(\nu, s) ds = \frac{1}{1 - \exp(-\nu)} \exp\{-\nu\alpha[s, \infty)\} d\alpha(s) \quad (4.23)$$

(\*) Processo log-beta:

$$K(\nu, s) ds = \frac{1}{1 - \exp(-\nu)} \exp\{-\nu\beta(s)\} d\alpha(s) \quad (4.24)$$

(\*) Processo beta-Stacy:

$$K(\nu, s) ds = \frac{1}{1 - \exp(-\nu)} \exp\{-\nu c(s)G(s, \infty)\} c(s) dG_c(s) \quad (4.25)$$

(\*) Processo Gamma Estesio:

$$K(\nu, s) ds = \nu^{-1} \exp\{-\nu\beta(s)\} d\alpha(s) \quad (4.26)$$

In tutti i casi,  $\beta(\cdot)$  è una funzione non-negativa continua a tratti e  $\alpha(\cdot)$  è una misura continua sull'intervallo  $[0, \infty)$ . Nel caso del processo beta-Stacy,  $c(\cdot)$  e  $G$  rappresentano i parametri del processo, in particolare  $c(\cdot)$  è una funzione positiva e  $G$  è una funzione di distribuzione cumulata su  $[0, \infty)$ . Inoltre,  $G_c(t) = G(t) - \sum_{t_k \leq t} G\{t_k\}$  dove  $G_c(\cdot)$  rappresenta la parte continua,  $\{t_k\}$  è l'insieme numerabile dei punti di discontinuità di  $G$ .

# Capitolo 5

## Inferenza bayesiana non-parametrica per un processo di Poisson non-omogeneo

### 5.1 Introduzione

Il problema dell'inferenza statistica nei processi di Poisson ha una lunga storia. Nell'impostazione classica, con l'utilizzo di un processo di punto di Poisson, si è in grado di modellare i valori estremi rispetto ad una soglia che si trovano in una regione  $A = [t_1, t_2] \times (u, \infty)$ , caratterizzandoli con la misura di intensità,  $\Lambda(A)$  (si veda figura 2.1).

Praticamente, i problemi di inferenza statistica sui processi di Poisson si concentrano sui metodi di stima e test di questa misura di intensità, cioè si è in grado di calcolare la probabilità che tali valori si trovino all'interno della regione  $A$  e massimizzando tale funzione si riescono ad ottenere le stime dei parametri della funzione di intensità limite. I metodi impiegati sono parametrici in natura, nel senso che la misura di intensità viene assunta di una qualche forma lineare o parametrica (Brown, 1972; Lewis, 1972).

La situazione è cambiata considerevolmente con l'arrivo di Aalen (1978). Egli nel suo articolo propose una classe di stimatori non-parametrici per le misure di intensità dei processi di punto.

Questo significa che, introducendo una prior sulla misura di intensità, si è in grado di stimare la funzione di verosimiglianza puntuale. Diverse sono le classi di priors non-parametriche che vengono impiegate per modellare l'intensità di un processo di Poisson non-omogeneo. Lo (1982), per poter fare inferenza bayesiana non-parametrica sul processo di Poisson non-omogeneo, considera l'ipotesi di poter stimare la funzione di intensità da campioni multipli indipendenti e identicamente distribuiti. Successiva-

mente, Lo e Weng (1989) sviluppano risultati analitici per una famiglia di processi di punto moltiplicativi prendendo come prior il processo gamma esteso.

In particolare in questa parte del lavoro, sono state utilizzate come distribuzioni a priori due processi, il processo gamma esteso (Dykstra e Laud, 1981) e il processo log-beta (Walker e Muliere, 1997); quindi sono stati confrontati i risultati ottenuti inserendo tali processi nello schema di campionamento di Gibbs.

## 5.2 Verosimiglianza e distribuzioni a priori non-parametriche

Seguendo Kuo e Ghosh (1997), si indichi con  $N(t)$  il numero di guasti in  $(0, t]$ . Si assuma che  $N(t)$  sia un processo di Poisson non-omogeneo (NHPP) con funzione media  $\Lambda(t) = EN(t)$ . Cioè, il numero di guasti in intervalli disgiunti sono indipendenti con distribuzioni di Poisson di medie  $\Lambda(t_i) - \Lambda(t_{i-1})$  per l'intervallo  $(t_{i-1}, t_i]$  ( $t_{i-1} < t_i$ ). In questa definizione di NHPP, sono inclusi anche gli intervalli disgiunti della forma  $(t_{i-1}, t_i^-]$ ,  $(t_i^-, t_i]$ , ecc., dove  $t^-$  indica il limite sinistro di  $t$ . Sia  $\lambda(t) = \frac{d}{dt}\Lambda(t)$ . Se  $\Lambda(t)$  ha un salto in  $t_i$ , allora si definisce  $\lambda(t_i) = [\Lambda(t_i) - \Lambda(t_i^-)]\delta(t_i)$ , dove  $\delta(t_i)$  è la funzione delta di Dirac.

Si consideri un modello di tempo troncato in cui il processo viene osservato fino ad un tempo fissato  $\tau$ . Siano  $0 = x_0 < x_1 < \dots < x_n \leq \tau$  le epoche ordinate degli  $n$  salti osservati. Se si ipotizzano vincoli tra i tempi di salto osservati allora la probabilità di non osservare salti nell'intervallo  $(0, x_1)$ ,  $d_1$  salti in  $x_1$ , zero salti in  $(x_1, x_2)$ , e così via fino a non osservare salti in  $(x_n, \tau)$ , è data da

$$L(\Lambda|D, \tau) = \left\{ \prod_{i=1}^n [\Lambda(x_i) - \Lambda(x_{i-1})]^{d_i} e^{-[\Lambda(x_i) - \Lambda(x_{i-1})]} e^{-[\Lambda(x_{i-1}) - \Lambda(x_{i-2})]} \right\} e^{-[\Lambda(\tau) - \Lambda(x_n)]},$$

dove  $D$  indica il data-set e  $d_i$  il numero dei salti multipli osservati al tempo  $x_i$ . Il numero totale dei salti osservati nell'intervallo  $(0, \tau]$  è dato da  $d = \sum_{i=1}^n d_i$ .

Questa funzione di verosimiglianza suggerisce che se la distribuzioni a priori su  $\Lambda(t)$  ha incrementi indipendenti allora la distribuzione a posteriori per  $\Lambda(t)$  ha ancora incrementi indipendenti. In particolare se gli incrementi della prior hanno distribuzioni gamma, allora gli incrementi della a posteriori hanno ancora distribuzioni gamma.

Si assegnerà un processo di Lévy a priori per  $\Lambda(\cdot)$ . Come si è evidenziato nella Sezione 4.2.3, la distribuzione a priori può essere caratterizzata da

$$M = \{t_1, t_2, \dots\}, \{f_{t_1}, f_{t_2}, \dots\},$$

rispettivamente i punti fissati di discontinuità e le corrispondenti funzioni di densità per i salti, e da  $N_t(\cdot)$ , che rappresenta la misura di Lévy per quella parte del processo

senza punti fissati di discontinuità (si veda paragrafo 4.3 per tutte le derivazioni). In particolare in questo lavoro si farà riferimento solo a due di queste derivazioni.

Se

$$K(\nu, s)ds = \frac{1}{1 - \exp(-\nu)} \exp\{-\nu\beta(s)\}d\alpha(s) \quad (5.1)$$

allora  $L(\cdot)$  è un processo log-beta (si veda Walker e Muliere, 1997).

Dall'altro lato, se

$$K(\nu, s)ds = \nu^{-1} \exp\{-\nu\beta(s)\}d\alpha(s) \quad (5.2)$$

allora  $L(\cdot)$  è un processo gamma esteso (si veda Dyktra e Laud, 1981).

### 5.3 Distribuzioni a posteriori

Le corrispondenti distribuzioni a posteriori sono date dal seguente teorema.

**Teorema 5.3.1.** *Sia  $\Lambda(\cdot)$  un processo di Lévy e siano  $x_1, \dots, x_n$  le epoche ordinate dei tempi di salto osservati derivanti da un processo di Poisson non-omogeneo,  $N(\cdot)$ . Siano inoltre  $x_1^*, \dots, x_m^*$  le epoche di tempo non uguali a nessuno dei salti fissati a priori. Allora la distribuzione a posteriori di  $\Lambda(\cdot)$  dati i dati è di nuovo un processo di Lévy con le seguenti caratteristiche:*

$$M^* = M \cup \{x_1^*, \dots, x_m^*\}, \quad \text{con}$$

$$f_{t_j}^*(\nu|D) \propto \begin{cases} \nu^{d_i} e^{-\nu} f_{t_j}(\nu) & \text{se } t_j = x_i, \\ e^{-\nu} f_{t_j}(\nu) & \text{se } t_j \neq x_i, \end{cases}$$

$$f_{x_j^*}(\nu|D) \propto \nu^{d_i} e^{-\nu} K(\nu, x_i),$$

per  $i = 1, \dots, m$  e

$$K^*(\nu, s) = e^{-\nu} K(\nu, s).$$

*Dimostrazione:* Si veda Gutiérrez-Peña e Nieto-Barajas (2002) Teorema 1.

In particolare, se  $\Lambda(\cdot) \sim \log BeP\{\alpha(\cdot), \beta(\cdot)\}$ , cioè se  $\Lambda(\cdot)$  è un processo log-beta senza punti di discontinuità fissati a priori, allora la distribuzione a posteriori di  $\Lambda(\cdot)$  è di nuovo un processo di Lévy con una misura log-beta per la parte continua, con parametri  $\alpha(\cdot)^* = \alpha(\cdot)$  e  $\beta(\cdot)^* = \beta(\cdot) + 1$ , e con punti fissati di discontinuità in  $M^* = \{x_1, \dots, x_n\}$  con distribuzioni a posteriori per i salti  $f_{x_i}^*(\nu) \propto \nu^{d_i} e^{-\nu} K(\nu, x_i)$ ,  $i = 1, \dots, n$ , dove  $k(\nu, s)$  è dato dalla (4.24).

Kuo e Ghosh (1997) derivano le distribuzioni a posteriori per particolari distribuzioni a priori, sia nel caso di intensità cumulata  $\Lambda(t)$  sia nel caso di funzione di intensità  $\lambda(t) = d\Lambda(t)/dt$ . Specificatamente, questi autori assumono processi gamma e beta

(Hjort, 1990) come prior per  $\Lambda(\cdot)$  ed un processo gamma esteso come prior su  $\lambda(\cdot)$ .

Inoltre, se  $\Lambda(\cdot) \sim EGaP\{\alpha(\cdot), \beta(\cdot)\}$ , per esempio,  $\Lambda(\cdot)$  è un processo gamma esteso senza punti fissati di discontinuità a priori, allora la distribuzione a posteriori di  $\Lambda(\cdot)$  è di nuovo un processo gamma esteso con parametri  $\alpha^*(\cdot) = \alpha(\cdot)$  e  $\beta^*(\cdot) = \beta(\cdot) + 1$ , e con punti fissati di discontinuità in  $M^* = \{x_1, \dots, x_n\}$  con distribuzioni a posteriori per i salti  $f_{x_i}^*(\nu) = Ga(\nu \mid d_i, \beta(x_i) + 1)$ ,  $i = 1, \dots, n$ .



# Capitolo 6

## Valori estremi e processi d'urna

### 6.1 Introduzione

I modelli d'urna hanno una lunga tradizione nella probabilità e nella statistica. Vengono utilizzati allo scopo di rendere concrete le idee che generano un modello probabilistico, riuscendo comunque a mantenere un livello di astrazione tale, per cui possano considerarsi svincolati dalla particolare applicazione per cui erano stati concepiti. Lo schema d'urna, introdotto da Polya nel 1923, è un tipico esempio di questa caratteristica: fu utilizzato per descrivere malattie contagiose e divenne presto il prototipo per modelli probabilistici costruiti ad hoc per i trattamenti adattivi (Feller, 1957).

Questo approccio venne seguito da Blackwell e MacQueen (1973) per la costruzione del processo di Dirichlet, da Mauldin, Sudderth e Williams (1992) per la costruzione dei Polya trees e da Walker e Muliere (1997) per quella del processo beta-Stacy.

Prima di introdurre i concetti legati ai modelli di urna è necessario ricordare alcune nozioni fondamentali sulla scambiabilità e sulle strutture di dipendenza più complesse.

### 6.2 Concetti preliminari

#### Scambiabilità

Lo schema di base nell'impostazione bayesiana è la scambiabilità di osservazioni  $X_1, X_2, \dots$  che esprime la situazione di esperimenti ripetuti in condizioni omogenee.

**Definizione 6.2.1.** *Siano  $X_1, \dots, X_n$   $n$  variabili casuali e sia  $P(X_1, \dots, X_n)$  la loro distribuzione di probabilità congiunta. Le variabili  $X_1, \dots, X_n$  sono dette scambiabili (scambiabilità finita) per la misura di probabilità  $P$  se la loro distribuzione di probabilità congiunta soddisfa*

$$P(X_1, \dots, X_n) = P(X_{i_1}, \dots, X_{i_n})$$

in cui  $(i_1, \dots, i_n)$  è una qualunque permutazione di  $(1, \dots, n)$ .

In termini di corrispondente densità o funzione di massa si può scrivere

$$p(X_1, \dots, X_n) = p(X_{i_1}, \dots, X_{i_n})$$

**Definizione 6.2.2.** Una infinita sequenza di v.a.  $X_1, X_2, \dots$  si dice *infinitamente scambiabile* se ogni sottosequenza finita è scambiabile nel senso della definizione precedente.

Se  $X = \{X_1, X_2, \dots\}$  è una successione infinita di variabili aleatorie a valori in  $\mathcal{X} \subseteq \mathbb{R}^d$  e supponiamo che la legge di probabilità  $P$  di  $X$  sia scambiabile, allora per il teorema di rappresentazione di de Finetti, per ogni  $n = 1, 2, \dots$  e per ogni  $(x_1, \dots, x_n)$  si ha

$$P(X_1 \leq x_1, \dots, X_n \leq x_n) = \int_{\mathcal{F}} \prod_{i=1}^n F(x_i) d\pi(F). \quad (6.1)$$

dove  $\mathcal{F}$  rappresenta la classe delle f.d.r. limite (nel senso della convergenza quasi certa) delle f.d.r. empiriche

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \delta_{(-\infty, x)}(x_i) \quad (6.2)$$

al divergere di  $n$ , e  $\pi$  è la misura di probabilità che ha per supporto  $\mathcal{F}$  ed è univocamente determinata.

### Parziale scambiabilità alla de Finetti

Sia  $X = \{X_{n,m}, n = 1, 2, \dots; m = 1, 2, \dots, k\}$  una successione di variabili aleatorie a valori in  $\mathcal{X} \subseteq \mathbb{R}^d$ . La successione  $X$  si dice *parzialmente scambiabile* se le successioni  $\{X_{n,1}\}, \dots, \{X_{n,k}\}$  sono ognuna una successione scambiabile. In base al teorema di rappresentazione (de Finetti, 1937; Regazzini, 1991)  $X$  è parzialmente scambiabile se e solo se la sua legge di probabilità può esprimersi come

$$P(X_{1,1} \leq x_{1,1}, \dots, X_{n_1,1} \leq x_{n_1,1}, \dots, X_{1,k} \leq x_{1,k}, \dots, X_{n_k,k} \leq x_{n_k,k}) = \int_{\mathcal{F}^k} \prod_{i=1}^{n_1} F_1(x_{i,1}) \cdots \prod_{i=1}^{n_k} F_k(x_{i,k}) \pi(dF_1, \dots, dF_k) \quad (6.3)$$

per ogni  $(x_{1,1}, \dots, x_{n_k,k})$  e  $n_1, \dots, n_k \geq 1$ . Inoltre, la successione dei vettori delle funzioni di ripartizione empiriche  $(F_{n,1}, \dots, F_{n,k})$  (dove  $F_{n,j}(x) = \frac{\sum_{i=1}^n \delta_{(-\infty, x)}(x_{i,j})}{n}$ ), converge debolmente a  $(\tilde{F}_1, \dots, \tilde{F}_k)$ . La legge di probabilità  $\pi$  è la legge di probabilità di  $(\tilde{F}_1, \dots, \tilde{F}_k)$ , ed è detta misura di de Finetti della successione  $X$ , o distribuzione iniziale.

### Parziale scambiabilità alla Diaconis e Freedman

Sia  $X = \{X_1, X_2, \dots\}$  una successione di numeri aleatori a valori in un insieme numerabile  $I$  che si chiamerà *spazio degli stati*. Sia  $\sigma = (x_0, x_1, \dots, x_n)$  una stringa finita di elementi di  $I$  e  $t_{i,j}(\sigma)$  il numero di transizioni dallo stato  $i$  allo stato  $j$  in  $\sigma$  per ogni  $i, j \in I$ .

A questo punto si rende necessaria la seguente relazione di equivalenza. Se  $\sigma = (x_0, \dots, x_n)$  e  $\phi = (y_0, \dots, y_n)$  sono due stringhe finite, si dirà che  $\sigma$  è equivalente a  $\phi$  se hanno lo stesso elemento iniziale e  $t_{i,j}(\sigma) = t_{i,j}(\phi)$ , per ogni  $i, j \in I$ .

La successione  $X$  si dirà *parzialmente scambiabile secondo Diaconis e Freedman* se

$$P(X_0 = x_0, \dots, X_n = x_n) = P(X_0 = y_0, \dots, X_n = y_n)$$

per ogni coppia di stringhe equivalenti  $\sigma$  e  $\phi$ . Diaconis e Freedman (1980) forniscono il seguente teorema di rappresentazione. Si supponga che la successione  $X$  sia ricorrente, ossia che  $P(X_n = X_0 \text{ per infiniti } n) = 1$ . Allora, la successione ricorrente  $X$  è parzialmente scambiabile alla Diaconis e Freedman se e solo se è una mistura di catene di Markov; cioè, esiste una misura  $\pi$  sull'insieme  $\mathcal{P}$  di matrici stocastiche su  $I \times I$  tale che per ogni  $n \geq 1$  e per ogni  $(x_0, x_1, \dots, x_n) \in I^{n+1}$  vale

$$P(X_0 = x_0, \dots, X_n = x_n) = \int_{\mathcal{P}} \prod_{j=0}^{n-1} s(x_j, x_{j+1}) \pi(ds). \quad (6.4)$$

Questo vuol dire che, condizionatamente alla matrice di transizione  $s$ ,  $X$  è una catena di Markov con matrice di transizione  $s$ .

## 6.3 Processi di urna rinforzati (RUP)

In questa sezione verrà introdotta una classe di processi stocastici definiti su uno spazio numerabile di urne di Polya che sarà utile per costruire classi più generali di priors comunemente utilizzate nell'inferenza bayesiana non-parametrica come i Polya trees e i processi beta-Stacy.

Sia  $S$  uno spazio numerabile ed  $E = \{c_1, \dots, c_k\}$  un insieme finito di colori con cardinalità  $k \geq 1$ . Si immagini che una funzione  $U$  associ ad ogni  $x \in S$  un'urna di Polya la cui composizione iniziale sia  $U(x) = (n_x(c_1), \dots, n_x(c_k))$ , dove  $n_x(c)$  rappresenta il numero di palline di colore  $c$  contenute nell'urna allo stato  $x$ , per  $c \in E$  e  $x \in S$ . Infatti  $U$  è chiamata *funzione di composizione dell'urna* se manda  $S$  in un insieme di  $k$ -uple di numeri reali non-negativi la cui somma è strettamente positiva.

Sia  $q : S \times E \rightarrow S$  la *legge di moto* che gode della proprietà per cui, per tutti i valori  $x, y \in S$ , esiste al più un colore  $c(x, y) \in E$  tale che  $q(x, c(x, y)) = y$ . Adesso

si può procedere alla costruzione di un processo stocastico  $\{X_n\}$  su  $S$  la cui definizione ricorsiva è la seguente: si fissa uno stato iniziale  $x_0 \in S$  e si pone  $X_0 = x_0$ . Per tutti i valori di  $n \geq 1$ , se  $X_{n-1} = x \in S$ , una pallina viene estratta a caso dall'urna associata ad  $x$  e viene reinserita nell'urna aggiungendone un'altra dello stesso colore. Se  $c \in E$  è il colore della pallina estratta, allora  $X_n = q(x, c)$ .

In particolare, si comincia estraendo una pallina dall'urna associata allo stato  $x_0$ , e si osserva che è di colore  $c_0$ . Si ricolloca la pallina nell'urna con un'altra dello stesso colore e ci si sposta in  $X_1 = q(x_0, c_0)$ . Successivamente si estrae a caso una pallina dall'urna associata ad  $X_1$ : e si osserva che è di colore  $c_1$ . La si ricolloca nell'urna con un'altra dello stesso colore e ci si sposta allo stato  $X_2 = q(X_1, c_1)$ , e così via.

**Definizione 6.3.1.** Si dirà che  $\{X_n\} \in RUP(S, E, U, q)$  con stato iniziale  $x_0$ .

Se  $\sigma = (s_0, \dots, s_n)$  è una sequenza finita di elementi di  $S$ , si dirà che  $\sigma$  è *ammissibile* se  $s_0 = x_0$  e se esiste una sequenza finita  $(c_0, \dots, c_{n-1})$  di elementi di  $E$  tali che  $q(s_i, c_i) = s_{i+1}$  per  $i = 0, \dots, n-1$ . Data una sequenza ammissibile  $\sigma$ , per ogni  $x, y \in S$ , si definisce  $t(x, y)$  come il numero delle transizioni in  $\sigma$  dallo stato  $x$  allo stato  $y$ . Si indichi inoltre, per tutti gli stati  $x \in S$  e per tutti i colori  $c \in E$

$$l_x(c) = t(x, q(x, c))$$

e

$$t(x) = \sum_{y \in S} t(x, y).$$

**Teorema 6.3.1.** Per tutti i valori di  $n \geq 0$  e tutte le sequenze finite  $(s_0, \dots, s_n)$  di elementi di  $S$ ,  $P[X_0 = s_0, \dots, X_n = s_n] = 0$  se  $(s_0, \dots, s_n)$  non è ammissibile; altrimenti

$$P[X_0 = s_0, \dots, X_n = s_n] = \prod_{x \in S} \left[ \frac{\prod_{c \in E} \prod_{i=0}^{l_x(c)-1} (n_x(c) + i)}{\prod_{j=0}^{t(x)-1} (j + \sum_{c \in E} n_x(c))} \right] \quad (6.5)$$

dove per convenzione  $\prod_0^{-1} = 1$ .

Una conseguenza di questo risultato è che un processo di urna rinforzato possa considerarsi come *parzialmente scambiabile* nel senso di Diaconis e Freedman (1980). Due sequenze finite  $\sigma$  e  $\phi$  sono *equivalenti* se cominciano con lo stesso elemento e, per tutti gli stati  $x, y \in S$ , il numero delle transizioni nelle due sequenze è lo stesso. Inoltre, i due autori definiscono come parzialmente scambiabile qualsiasi processo  $\{Y_n\}$  definito su  $S$  tale che, per tutti gli  $n \geq 0$  e per tutte le sequenze equivalenti  $\sigma = (s_0, \dots, s_n)$  e  $\phi = (y_0, \dots, y_n)$  di elementi di  $S$ , si ha

$$P[Y_0 = s_0, \dots, Y_n = s_n] = P[Y_0 = y_0, \dots, Y_n = y_n].$$

**Corollario 3.** *Un processo di urna rinforzato  $\{X_n\}$  è parzialmente scambiabile.*

Per un processo  $\{Y_n\}$  definito su  $S$  con stato iniziale  $Y_0 = y_0$ , si pone  $\tau_0 = 0$  e, per tutti i valori di  $n \geq 1$ , si definisce

$$\tau_n = \inf\{j > \tau_{n-1} : Y_j = y_0\}. \quad (6.6)$$

Allora  $\{Y_n\}$  si dice *ricorrente* se

$$P[Y_n = Y_0 \text{ per infiniti } n] = 1$$

o, equivalentemente, se

$$P\left[\bigcap_{n=0}^{\infty} \{\tau_n < \infty\}\right] = 1. \quad (6.7)$$

Sia  $(s_0, \dots, s_n)$  una sequenza ammissibile. Dopo aver osservato  $X_0 = s_0, \dots, X_n = s_n$ , per tutti gli stati  $x \in S$  l'urna  $U(x)$  conterrà

$$n_x(c) + l_x(c) \quad (6.8)$$

palline di colore  $c$ , per ogni  $c \in E$ ; sia  $\tilde{U}(x)$  un'urna la cui composizione è quella descritta nella (6.8). Il prossimo risultato è abbastanza intuitivo ed ha una importante implicazione quando si utilizza il processo  $\{X_n\}$  a scopi predittivi.

**Teorema 6.3.2.** *Si assuma che  $\{X_n\}$  sia ricorrente. Sia  $(s_0, \dots, s_n)$  una sequenza ammissibile tale che  $P[X_0 = s_0, \dots, X_n = s_n] > 0$ . Allora, dati  $X_0 = s_0, \dots, X_n = s_n$ , il processo  $(X_n, X_{n+1}, \dots) \in RUP(S, E, \tilde{U}, q)$  con stato iniziale  $s_n$  ed è ricorrente.*

Seguendo Diaconis e Freedman (1980), si definisce un  $y_0$ -blocco per il processo  $\{Y_n\}$  definito su  $S$ , come una sequenza finita di stati che inizia da  $y_0$  e che non contiene altri  $y_0$ . Si consideri  $S^*$  come lo spazio numerabile di tutte le sequenze finite di elementi di  $S$ . Quando  $\{Y_n\}$  è ricorrente, sia  $B_1 \in S^*, B_2 \in S^*, \dots$ , la sequenza dei successivi  $y_0$ -blocchi in  $\{Y_n\}$ .

**Teorema 6.3.3.** (Diaconis e Freedman, 1980). *Se  $\{Y_n\}$  è ricorrente e parzialmente scambiabile con stato iniziale  $y_0$ , la sequenza  $B_n$  dei suoi  $y_0$ -blocchi è scambiabile.*

*Dimostrazione:* Si veda Diaconis e Freedman [rif].

## 6.4 Processi di urna rinforzati per il processo beta-Stacy

Un processo beta-Stacy (Walker e Muliere, 1997) su  $S = \{0, 1, 2, \dots\}$  è una funzione di distribuzione aleatoria costruita su  $S$  tale che, per tutti i valori di  $K \geq 0$ , la massa

aleatoria assegnata a  $\{0, \dots, k\}$  è data da

$$1 - \prod_{j=0}^k (1 - W_j)$$

dove le variabili aleatorie della sequenza  $\{W_j\}$  sono indipendenti con distribuzione beta. Per costruire un processo beta-Stacy utilizzando un RUP si consideri l'insieme dei colori  $E = \{w, b\}$  dove con  $w$  sta per 'white' e  $b$  sta per 'black'. Sia  $x_0 = 0$  e si assuma che  $n_0(w) = 0$ , per tutti i valori  $j \geq 1$ ,  $n_j(w) > 0$ . Infine si definisce la legge di moto  $q$  per  $\{X_n\}$  ponendo

$$q(x, b) = x + 1 \text{ e } q(x, w) = 0 \quad (6.9)$$

per tutti gli stati  $x \in S$ . Quindi,  $\{X_n\} \in RUP(S, E, U, q)$  con stato iniziale  $x_0 = 0$ . Il processo  $\{X_n\}$  inizia in 0, al passo  $n \geq 0$  si trova nello stato  $x \in S$ , e si sposta nello stato  $x + 1$  se la pallina estratta dall'urna  $U(x)$  è nera mentre torna in 0 se la pallina estratta è bianca. Se  $n_j(b) = 0$  per  $j \geq 1$ , sia  $N = \min\{j \geq 1 : n_j(b) = 0\}$  e si noti che, con probabilità uno, il processo  $\{X_n\}$  visita solamente gli stati  $\{0, \dots, N\}$ .

Applicando la (6.5), per sequenze finite ammissibili  $(x_0, \dots, x_n)$  di elementi di  $S$  si ha

$$P[X_0 = x_0, \dots, X_n = x_n] = \prod_{j=1}^{\infty} \frac{B(n_j(w) + t(j, 0), n_j(b) + t(j, j + 1))}{B(n_j(w), n_j(b))}$$

dove, per valori di  $a, b > 0$ ,  $B(a, b)$  è l'integrale beta.

Per  $n \geq 1$ , sia  $T_n = X_{\tau_{n-1}}$  dove la sequenza dei tempi di arresto  $\{\tau_n\}$  viene definita come nella (6.6); cioè  $\tau_0 = 0$  e

$$\tau_n = \inf\{j > \tau_{n-1} : X_j = 0\}$$

per  $n \geq 1$ . In pratica,  $T_n$  misura la lunghezza della sequenza degli stati compresi strettamente tra l' $n$ -esimo zero e l' $n + 1$ -esimo zero nella sequenza  $\{X_n\}$ .

**Lemma 1.** *Il processo  $\{X_n\}$  è ricorrente se e solo se*

$$\lim_{n \rightarrow \infty} \prod_{i=0}^n \frac{n_i(b)}{n_i(b) + n_i(w)} = 0. \quad (6.10)$$

Quando vale la (6.10), la sequenza  $\{T_n\}$  è scambiabile e, per il teorema di rappresentazione di de Finetti, esiste una funzione di distribuzione aleatoria  $F$  tale che, dato  $F$ , le v.a. della sequenza  $\{T_n\}$  sono iid con distribuzione  $F$ . Con il prossimo teorema si dimostra che  $F$  è un processo beta-Stacy su  $S$  con parametri  $\{(n_j(w), n_j(b))\}$  (Walker e Muliere, 1997); ciò significa che  $F$  è un processo neutrale a destra tale che  $F(0) = 0$  con probabilità uno e, per ogni  $j \geq 1$ ,  $[F(j) - F(j - 1)]$  ha la stessa distribuzione di

$$P_j = W_j \prod_{i=1}^{j-1} (1 - W_i) \quad (6.11)$$

dove  $\{W_j\}$  è una sequenza di v.a. indipendenti tali che, per ogni  $j \geq 1$ ,  $\{W_j\}$  ha una distribuzione Beta  $\{(n_j(w), n_j(b))\}$ . Si noti che nella (6.11) si assume  $\prod_{i=1}^0 = 1$  mentre Beta  $(a, 0)$  indica la distribuzione che concentra tutta la massa di probabilità in 1 qualunque sia  $a > 0$ .

**Teorema 6.4.1.** *F è un processo beta-Stacy su S con parametri*

$$\{(n_j(w), n_j(b))\}.$$

**Osservazione 6.4.1.** *Si osservi che, per ogni  $k \geq 1$ ,*

$$P[T_1 = k] = \frac{n_k(w)}{n_k(b) + n_k(w)} \prod_{j=0}^{k-1} \frac{n_j(b)}{n_j(b) + n_j(w)} \quad (6.12)$$

dove  $\prod_0^{-1} = 1$ .

Inoltre, per ogni  $n \geq 1$  e  $k \geq 1$ ,

$$P[T_{n+1} = k | T_1, T_2, \dots, T_n] = \frac{n_k(w) + n_k}{n_k(b) + n_k(w) + n_k + m_k} \prod_{j=1}^{k-1} \frac{n_j(b) + m_j}{n_j(b) + n_j(w) + n_j + m_j} \quad (6.13)$$

sull'insieme  $\bigcap_{i=1}^{\infty} \{\tau_i < \infty\}$  se, per ogni  $0 \geq j \geq k-1$ , si definisce  $n_j = \sum_{i=1}^{k-1} I[T_i = j]$  e  $m_j = \sum_{i=1}^n I[T_i > j]$ . Pertanto, la (6.10) è verificata, la (6.13) è vera con probabilità uno.





# Capitolo 7

## Una applicazione finanziaria

### 7.1 Introduzione

In questa parte del lavoro si prenderanno in esame le serie giornaliere di cinque dei trenta titoli azionari del comparto industriale che compongono l'indice sintetico del Dow-Jones. I titoli considerati sono: IBM, General Motors, American Express, Coca Cola, Johnson&Johnson. Il periodo di osservazione a cui si farà riferimento per tutta l'analisi è di dieci anni (1990-2000) e le rilevazioni sono effettuate quotidianamente. Lo scopo di questa analisi è di analizzare i valori estremi (superamenti rispetto alla soglia) di queste serie. L'utilità di considerare più titoli azionari è stata dettata da motivazioni sia di carattere pratico, in modo da consentire un confronto sulla rischiosità di un titolo rispetto ad un altro, sia di carattere prettamente metodologico dato che, nella caratterizzazione dei tempi di superamento tramite processi di urna si è assunta la scambiabilità sia dei titoli che dei 0-blocchi di uno stesso titolo.

A causa della forte non-stazionarietà di questo tipo di serie storiche, utilizzando una metodologia ben nota si è resa stazionaria la serie effettuando la seguente trasformazione

$$\tilde{X}_t = \log X_t - \log X_{t-1}.$$

Per convenienza di presentazione, i dati sono ora riscaldati in questo modo  $\tilde{X}_t \rightarrow 100\tilde{X}_t$ . La figura (7.1) mostra come la serie trasformata del titolo Coca Cola sia “ragionevolmente” vicina alla stazionarietà.

#### OBIETTIVO CAPITOLO

Il capitolo è così suddiviso: nella prima parte, si farà un'analisi dei cinque data-sets utilizzando queste due tecniche di impostazione classica che consentono di modellare i valori estremi secondo distribuzioni note in letteratura, permettendo quindi una scelta ottimale della soglia di riferimento al di sopra della quale un titolo verrà ritenuto rischioso; nella seconda parte invece, il data-set sarà rappresentato dai tempi (giorni)

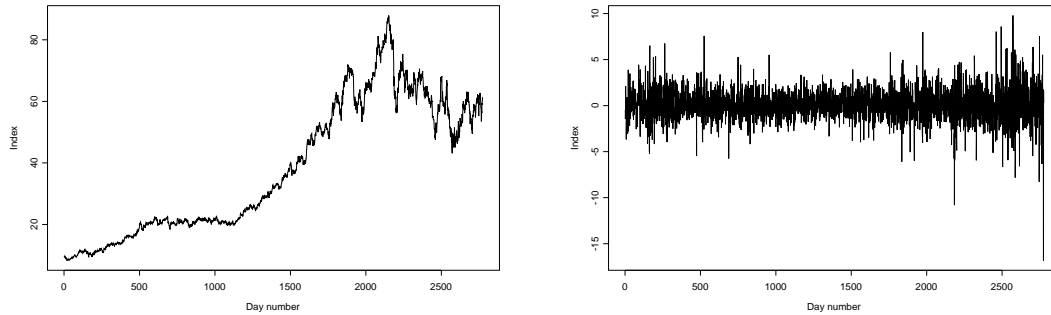


Figura 7.1: Grafico a sinistra: prezzi di chiusura giornalieri del titolo Coca Cola. Grafico a destra: log-daily returns del titolo Coca Cola.

corrispondenti ai superamenti di queste soglie, analizzati utilizzando una impostazione bayesiana non-parametrica.

## 7.2 Utilizzo dei modelli dell'analisi classica sui valori estremi

Le tecniche sui valori estremi sono diventate sempre più popolari nelle applicazioni finanziarie. Questo non deve sorprendere: la solubilità finanziaria viene più verosimilmente determinata da cambiamenti estremi delle condizioni del mercato piuttosto che da cambiamenti tipici. Data la complessa struttura stocastica dei mercati finanziari, potrebbe accadere che una semplice applicazione delle tecniche dei valori estremi possa essere fuorviante.

Infatti, come già è stato detto nella sezione 1.5, modellare soltanto i massimi di ogni blocco è un approccio valido nel caso in cui si disponga di poche osservazioni; nei casi di serie storiche invece, è un approccio che comporta uno spreco di enormi quantità di dati. D'altro canto, lavorare con le statistiche d'ordine è una alternativa teoricamente migliore ma nella pratica è inusuale, in quanto risulta difficile disporre di dati in questa forma. Inoltre anche questo metodo, nei casi in cui un blocco contenga più valori estremi rispetto ad un altro, non considera gran parte delle osservazioni. Quindi se è disponibile, come nel caso dei titoli azionari, l'intera serie storica giornaliera (o addirittura oraria) allora la cosa migliore è quella di evitare del tutto procedure che utilizzino blocchi di dati. A seguito di ciò, si è ritenuto che la migliore analisi da fare, per il tipo di dati di cui si dispone, fosse un confronto tra tecniche alternative a queste e cioè, il modello che analizza i superamenti di una soglia e l'approccio che caratterizza questi valori estremi tramite un processo di punto.

Da questo momento in poi si utilizzerà sempre la serie modificata nel modo descritto sopra e si analizzerà, per comodità di esposizione, soltanto uno dei cinque titoli e cioè quello della Coca Cola. Alla fine dell'analisi verranno presentati i grafici e le tabelle riassuntive di tutti i titoli presi in considerazione.

Il grafico sulla vita media residua della serie trasformata viene mostrato nella figura (7.2) in funzione delle soglie, con un intervallo di confidenza approssimato del 95 % per la serie giornaliera del titolo Coca Cola.

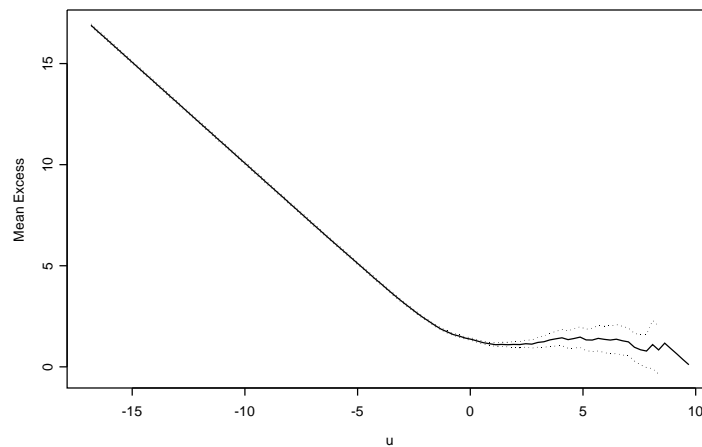


Figura 7.2: Grafico della vita media residua per il titolo Coca Cola.

Questo grafico è inizialmente lineare ma mostra una sostanziale curvatura da  $u = -2.5$  a  $u \approx 3.8$ , oltre questo valore risulta approssimativamente lineare fino a  $u \approx 7.5$ , dopodichè decresce. Quindi, si può concludere che  $u = 3.8$  sia il punto di inizio della linearità e che si possa scegliere  $u_0 = 3.8$  come soglia per individuare i superamenti. La validità della scelta di questa soglia viene supportata dall'idea di stimare il modello GPD in un intervallo di soglie, andando a cercare una stabilità rispetto alle stime di massima verosimiglianza per il modello riparametrizzato (si veda figura (7.3)).

Successivamente si è adattato il modello alla soglia specificata ottenendo così 54 superamenti rispetto alla serie completa di lunghezza pari a 2774 osservazioni, cosicchè il rapporto tra le eccedenze e il totale delle osservazioni è pari a 0.019.

Le stime di massima verosimiglianza dei parametri della distribuzione generalizzata di Pareto sono  $(\hat{\sigma}, \hat{\xi}) = (1.44, -0.04)$ , con errori standard di 0.30 e 0.16 rispettivamente e log-verosimiglianza negativa pari a 71.51. Errori standard e intervalli di confidenza

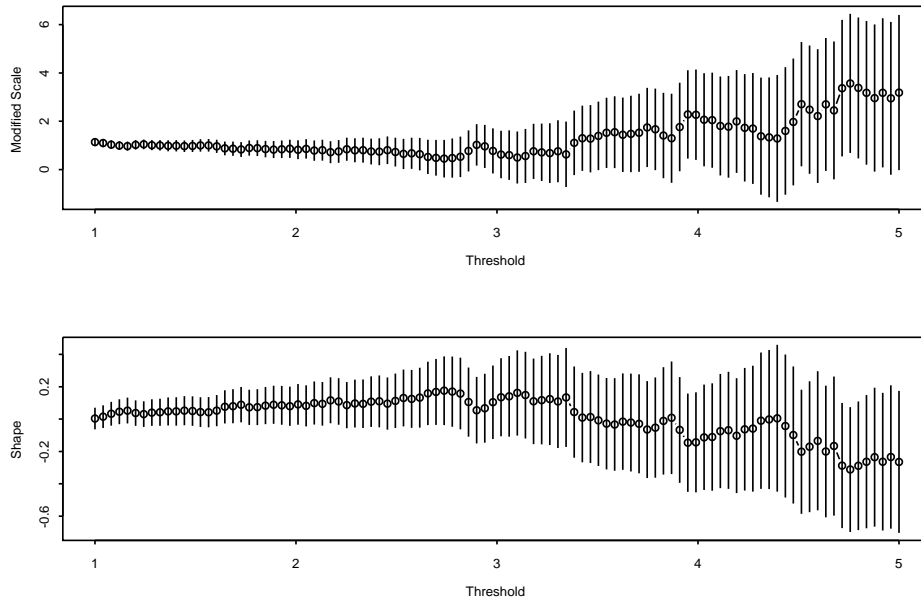


Figura 7.3: Stime di massima verosimiglianza e intervalli di confidenza al 95% per il parametro ( $\xi$ ) trasformato nel modello GPD applicato al data-set giornaliero della Coca Cola.

per la GPD sono ottenuti, nel consueto modo, dalla teoria della verosimiglianza.

I grafici diagnostici per la distribuzione di Pareto generalizzata sono mostrati nella figura (7.4).

Questi grafici evidenziano l'accuratezza o meno con cui il modello GPD utilizzato si adatta ai dati della Coca Cola. Nè il *probability plot* nè il *quantile plot* danno modo di dubitare della validità di questo modello, in quanto ogni insieme dei punti è vicino alla linearità. La bontà di adattamento dei dati al modello è riscontrabile anche dal fatto che gli intervalli di confidenza, sul grafico del livello di ritorno, suggeriscono come gli allontanamenti dal modello non siano affatto grandi anche dopo aver considerato la variabilità campionaria. Infine c'è da notare che la stima della densità corrispondente non sembra molto consistente con l'istogramma dei dati. In conclusione, i grafici diagnostici confermano che è possibile utilizzare il modello GPD.

Come è stato già evidenziato nel capitolo 1, un modello più flessibile per lo studio dei valori estremi è quello che si basa sulla rappresentazione di questi superamenti tramite un processo di punto. Si osservi come la distribuzione dei superamenti sia dipendente dalla storia del processo (figura 7.5) e come invece, di considerare i tempi in cui si verificano i superamenti rispetto ad una soglia e i valori che superano la soglia come

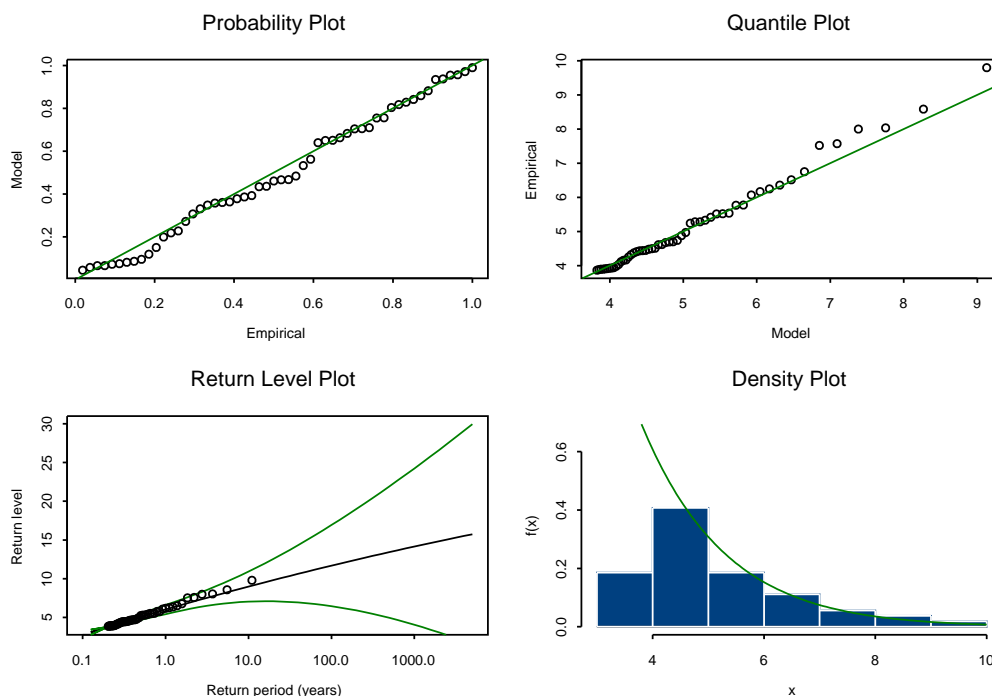


Figura 7.4: Grafici diagnostici per il modello dei superamenti rispetto alla soglia adattato ai dati trasformati del titolo Coca Cola.

due processi separati, questi siano combinati in un unico processo basato su un grafico bidimensionale in cui la prima dimensione è data dai tempi di superamento e la seconda dai corrispondenti valori di superamento.

I vantaggi immediati che ne derivano sono dovuti al fatto che, la naturale parametrizzazione del modello avviene in termini di parametri della distribuzione GEV e che il tasso di superamento della soglia viene preso in considerazione quando si fa inferenza. L'output ha un formato simile a quello ottenuto con la GPD. Si osservi come la stima del parametro di forma ( $\hat{\xi}$ ) sia identica per i due modelli (si veda tabella 1.1); il parametro di scala del modello generalizzato di Pareto è una funzione deterministica di  $\mu$  e  $\sigma$ . Si noti come la log-verosimiglianza negativa calcolata con questo modello è sempre inferiore rispetto a quella ottenuta con il modello GPD. L'output che si ottiene con il processo di punto può essere utilizzato direttamente per avere i grafici diagnostici che sono identici a quelli ottenuti con il modello GPD (si veda figura 7.6). Questi grafici sono costruiti mediante trasformazione dei valori del parametro del processo di punto nei corrispondenti parametri della distribuzione generalizzata di Pareto.

Di seguito verranno riportate le tabelle sintetiche delle stime dei parametri delle due distribuzioni considerate relativamente ai cinque titoli azionari che compongono il data-set.

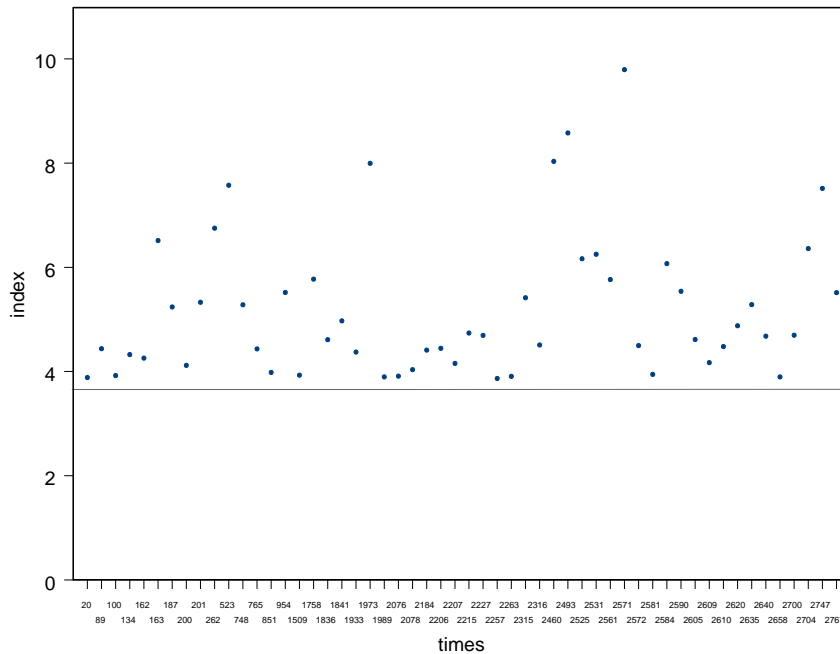


Figura 7.5: Superamenti rispetto alla soglia ( $u=3.8$ ) della serie storica Coca Cola trasformata.

Si può notare come le considerazioni finora fatte sulle stime dei parametri delle distribuzioni GPD e PP sull'adattamento di questi modelli ai dati del titolo Coca Cola valgono sostanzialmente anche per gli altri titoli.

### 7.3 Utilizzo dell'impostazione bayesiana non-parametrica sui valori estremi

In questa parte del lavoro, come è stato già accennato all'inizio del capitolo, si discuterà un approccio bayesiano non-parametrico applicato al data-set composto dai tempi di superamento della soglia, scelta con l'approccio classico, oltre la quale un titolo è considerato rischioso. I tempi di superamento verranno presentati come un processo di Poisson non-omogeneo (NHPP), in cui la funzione media non nota verrà modellizzata *a priori* come un processo a incrementi indipendenti. Successivamente si ricaveranno le corrispondenti distribuzioni a posteriori. Le inferenze a posteriori sono ottenute utilizzando uno schema di campionamento di Gibbs.

Nel seguito,  $\mathcal{A} = \{A(t) : t \geq 0\}$  indicherà l'insieme di tutte le funzioni  $A(\cdot)$  tale che

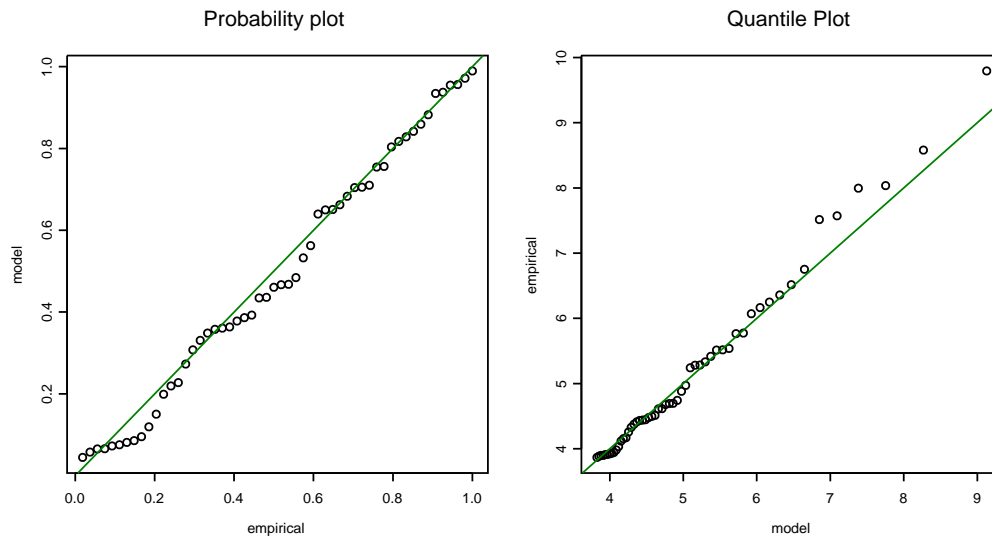


Figura 7.6: Grafici diagnostici del modello del processo di punto adattato ai dati trasformati del titolo Coca Cola.

- (i)  $A(0) = 0$ ;
- (ii)  $A(t) < \infty$  per  $0 \leq t < \infty$  e
- (iii)  $A(\cdot)$  è una funzione non-decrescente e continua a destra.

Sia  $\Lambda(\cdot) \in \mathcal{A}$  la funzione media (misura di intensità) di un NHPP, indicato con  $N(\cdot)$ , quindi

$$[N(\cdot)|\Lambda(\cdot)] \sim Po\{\Lambda(\cdot)\}.$$

Questo implica che  $N(\cdot)$  abbia una funzione media  $\mathcal{M} = \Lambda(\cdot)$ . Il processo di punto allora ha queste proprietà:

- (i)  $N(t)$  ha incrementi indipendenti;
- (ii)  $[N(t) - N(s)] \sim Po\{\Lambda(t) - \Lambda(s)\}$  per  $0 \leq t < \infty$ .

In questo caso gli incrementi non sono più stazionari a meno che  $\Lambda(t)$  non sia lineare in  $t$ .

Si assegnerà a  $\Lambda(\cdot)$  un processo di Lévy a priori in cui alla parte continua del processo viene assegnato un processo log-beta di parametri  $\{\alpha(\cdot), \beta(\cdot)\}$  (si veda Osservazione

	GPD	PP
$\hat{\mu}$	-	6.01 (0.35)
$\hat{\sigma}$	1.44 (0.30)	1.35 (0.21)
$\hat{\xi}$	-0.04 (0.16)	-0.04 (0.16)
nllh	71.5	39.7

Tabella 7.1: Stime di massima verosimiglianza, errori standard (nelle parentesi) e log-verosimiglianza negativa (nllh) dei modelli GPD e PP adattati alla serie storica della Coca Cola.

	GPD	PP		GPD	PP
$\hat{\mu}$	-	7.56 (0.56)	$\hat{\mu}$	-	5.11 (0.25)
$\hat{\sigma}$	1.68 (0.41)	2.15 (0.44)	$\hat{\sigma}$	0.92 (0.20)	0.96 (0.16)
$\hat{\xi}$	0.16 (0.20)	0.16 (0.20)	$\hat{\xi}$	0.02 (0.17)	0.02 (0.17)
nllh	87.3	58.6	nllh	49.1	20.4

Tabelle 1.2-1.3: Stime di massima verosimiglianza, errori standard (nelle parentesi) e log-verosimiglianza negativa (nllh) dei modelli GPD e PP adattati alle serie storiche della IBM (tabella a sinistra) e della Johnson& Johnson (tabella a destra).

5.1.1).

Di seguito verrà descritto il modo in cui è stato simulato il processo di Lévy, così come è stato descritto nel paragrafo 3.2.2, dato che verrà utilizzato per campionare dalla distribuzione a posteriori della media del . Il metodo si basa su una rappresentazione del processo di Lévy derivata da Ferguson e Klass (1972) e successivamente discussa da Walker e Damien (2000).

Si consideri un processo di Lévy e si ricordi che tale processo viene osservato fino ad un tempo  $\tau$ . Sia

$$M(x) = -N_\tau[x, \infty) = - \int_x^\infty dN_\tau(\nu),$$

dove  $N_\tau(\nu)$  indica la misura di Lévy del processo  $L_c(t)$ .

Si definiscano inoltre le v.a. positive  $J_1 \geq J_2 \geq \dots$  tramite

$$Pr(J_1 \leq x_1) = \exp\{M(x_1)\}$$



	GPD	PP		GPD	PP
$\hat{\mu}$	-	5.98 (0.21)	$\hat{\mu}$	-	6.98 (0.37)
$\hat{\sigma}$	1.05 (0.19)	0.78 (0.09)	$\hat{\sigma}$	1.28 (0.27)	1.42 (0.25)
$\hat{\xi}$	-0.19 (0.12)	-0.19 (0.12)	$\hat{\xi}$	0.07 (0.17)	0.07 (0.17)
nllh	42.7	17.3	nllh	64.5	40.4

Tabelle 1.4-1.5: Stime di massima verosimiglianza, errori standard (nelle parentesi) e log-verosimiglianza negativa (nllh) dei modelli GPD e PP adattati alle serie storiche della General Motors (tabella a sinistra) e dell'American Express (tabella a destra).

e

$$Pr(J_i \leq x_i | J_{i-1} \leq X_{i-1}) = \exp\{M(x_i) - M(x_{i-1})\}$$

per  $x_i < x_{i-1}$ .

Si otterranno i  $J_i$  via  $\omega_i - M(J_i)$ , dove  $\omega_1, \omega_2, \omega_3, \dots$  sono i tempi di salto di un processo di Poisson standard (per esempio, con intensità 1). Cioè,  $\omega_1, \omega_2 - \omega_1, \omega_3 - \omega_2, \dots \stackrel{iid}{\sim} Ga(1, 1)$ . La rappresentazione Ferguson-Klass del processo  $L_c(t)$  è data da

$$L_c(t) = \sum_i J_i I\{U_i \leq n_t(J_i)\},$$

dove  $U_1, U_2, \dots \stackrel{iid}{\sim} Un(0, 1)$ , e

$$n_t(\nu) = \frac{dN_t}{dN_\tau}(\nu)$$

per  $t \in [0, \tau]$ .

La simulazione a posteriori di  $\mathcal{M}(\cdot)$  può essere ottenuta dalle medie risultanti da uno schema di campionamento di Gibbs (per esempio, Smith e Roberts, 1993) utilizzando le full conditionals date nel Teorema 3.4.1. Allo scopo di campionare dalla full conditional di  $\Lambda(\cdot)$ , viene applicato l'algoritmo appena descritto basato sulla rappresentazione di Ferguson-Klass di un processo di Lévy.

Sono state fatte 4000 iterazioni con un periodo di burn-in di 1000 iterazioni. La convergenza è stata valutata basandosi sui grafici delle medie ergodiche per  $\Lambda(t)$  per alcuni valori selezionati di  $t$ . Il grafico della convergenza evidenzia una elevata variabilità iniziale seguita da uno smorzamento ed un andamento monotono verso il limite raggiunto attorno alla iterazione numero 1700-1800 e poi mantenuto fedelmente. Nel grafico sulla funzione di auto-correlazione (ACF) si nota come, dopo  $t = 0$  si abbiano dei valori assolutamente compresi nell'intervallo di confidenza al 95%, il che risulta coerente con

le ipotesi di base del Gibbs sampler. La figura successiva rappresenta una traiettoria campionaria del processo di Lévy “completo” ( $L(t)$ ), inteso come somma della parte con punti fissati di discontinuità ( $L_j(t)$ ) e la parte senza punti fissati di discontinuità ( $L_c(t)$ ), che mostra discontinuità più pronunciate in corrispondenza di tempi di interarrivo più lunghi ed una marcata tendenza (lineare) a crescere al crescere del tempo. (MOTIVI STORICI DEI DUE CLUSTERS) Lo stesso tipo di andamento è visibile nel grafico relativo al valore atteso del processo di Lévy condizionatamente ai dati ( $E(L(t)|D)$ ) che evidenzia come la media a posteriori dell’intensità del processo sia crescente (in modo poco più che lineare) con il crescere del tempo, questo vuol dire che andando avanti con i giorni gli eventi tendono ad addensarsi. L’ultimo grafico presenta la distribuzione del numero di eventi predetti per il periodo (inteso in termini di giorni) 2767-2850, si noti come sia fortemente asimmetrica con classe modale 0-50. Sull’asse delle ascisse è indicato il numero dei superamenti che si verificheranno nel periodo in cui si fa la predizione, sull’asse delle ordinate invece c’è la corrispondente probabilità con cui si verificheranno tali eventi.

## 7.4 Applicazione dei processi di urna rinforzati ai valori estremi

Come è stato già detto nella sezione riguardante l’applicazione della teoria dei valori estremi, sono stati presi in considerazione cinque titoli azionari che compongono l’indice del Dow-Jones. In questa parte del lavoro però, sarà fondamentale considerare la scambiabilità di questi titoli e degli 0-blocchi per ogni singolo titolo.

Anche in questo caso, il data-set sarà composto soltanto dalle epoche corrispondenti ai superamenti della soglia stabilita per ciascun titolo. Si è pensato di costruire un programma che consentisse di calcolare le probabilità iniziali e le probabilità a posteriori, così come sono state definite nella parte teorica.

Si consideri  $E = \{w, b\}$  l’insieme dei colori ai quali sono stati associati due eventi: con  $w$ , pallina bianca, si indichi l’evento “il titolo supera la soglia” e con  $b$ , pallina nera, l’evento “il titolo non supera la soglia”; sia inoltre,  $S$  lo spazio numerabile degli stati intesi come i giorni nei quali si osservano tali eventi. Si definisca come legge di moto per  $\{X_n\}$  quella data dalla (6.9) che implica che, quando il titolo non supera la soglia ci si sposta allo stato successivo e si aggiunge una pallina nera mentre quando tale superamento non si verifica si torna allo stato 0 aggiungendo una pallina bianca. Quindi,  $\{X_n\} \in RUP(S, E, U, q)$  con stato iniziale  $x_0 = 0$ . Inoltre, così come è stato detto nella

teoria sulla costruzione del processo beta-Stacy, in  $x_0 = 0$  l'urna sarà così composta

$$n_0(w) = 0 \quad n_0(b) = 50$$

invece, per tutti i  $j \geq 1$ ,  $n_j(w) > 0$ , in particolare per  $j = 1$  si è assunto  $n_j(w) = 50$ . Per  $n > 1$  è stata calcolata la sequenza dei tempi di arresto  $\{\tau_n\}$  definendo come tempo di arresto il tempo in cui la quotazione del titolo ha superato la soglia stabilita. Una volta definita tale sequenza è risultato immediato ottenere la sequenza dei  $\{T_n\}$  dove con  $T_n$  si indica il numero di giorni consecutivi in cui il titolo non ha superato la soglia, compresi tra l' $n$ -esimo e l' $n + 1$ -esimo superamento. Data la natura del data-set considerato, si può assumere che il processo  $\{X_n\}$  sia ricorrente e quindi la sequenza dei  $\{T_n\}$  sarà scambiabile.

Si assumerà che la funzione di distribuzione  $F$  dei tempi di interarrivo tra un superamento e l'altro sia un processo beta-Stacy.

A questo punto si è pensato di costruire un programma che fosse in grado di calcolare le probabilità date dalla (6.12) e dalla (6.13) per ciascuno dei cinque titoli. Facendo una assunzione un pò forte, si è imposto che i tempi di sopravvivenza di ciascuna azione lungo tutto il periodo di osservazione, si possano considerare scambiabili, è come se si potesse pensare che tutte le volte in cui il titolo supera la soglia, per cui torna allo stato 0, si prenda un altro titolo perchè quello precedente è stato ritenuto rischioso.

C'è comunque da dire che questa assunzione può essere indebolita qualora si decida di voler considerare una dipendenza tra le urne assegnate ad ogni singolo stato. A tal proposito un

In particolare, osservando le probabilità condizionate dei  $T_n$  di ciascuna azione si nota come, per periodi di uguale lunghezza di sopravvivenza verificatisi in tempi diversi, la probabilità diminuisce man mano che aumentano i condizionamenti.



# Bibliografia



# Bibliografia

- [1] Aalen O. O. (1975) Statistical inference for a family of a counting processes. Ph.D. thesis, University of California, Berkeley
- [2] Barbieri M. M. (1996). *Metodi MCMC nell'inferenza statistica*. Casa editrice: CISU
- [3] Blackwell D. e Mac Queen J. B. (1973). Ferguson distributions via Pòlya urn schemes. *Ann. Statist.*, **1**, 2, 353-355
- [4] Bottolo L., Consonni G. e Dellaportas P. (19??). Bayesian analysis of extreme values by mixture modelling. *Technical report, Università ???*
- [5] Brunner L. J. e Lo A. Y. (1994). Nonparametric Bayes methods for directional data. *The Canadian Journal of Statistics*, **22**, 3, 401-412
- [6] Cifarelli, D., Muliere, P. e Secchi, P. (2002). Exchangeable sequences of random variables. *Technical report*. Dipartimento di Matematica, Politecnico di Milano
- [7] Cifarelli, D., Muliere, P. e Secchi, P. (2000). Urn schemes for constructing priors. *Technical report*. Dipartimento di Matematica, Politecnico di Milano
- [8] Coles S. G. e Tawn J. A. (1991) Modelling extreme multivariate extreme. *J. R. Statis. Soc. B* **53**, 377-392
- [9] Coles S. G. e Tawn J. A. (1994) Statistical methods for multivariate extreme: an application to structural design (with discussion). *Appl. Stat.* **43**, 1-48
- [10] Coles S. G. e Powell E. A. (1996a). Bayesian methods in extreme value modelling: a review and new developments. *International Statistical Review*, **64**, 1, 119-136
- [11] Coles S. G. e Tawn J. A. (1996) A Bayesian analysis of extreme rainfall data. *Appl. Statist.*, **45**, 4, 463-478
- [12] Coles S. G. (2001) *An introduction to Statistical Modelling of extreme Values*. London: Springer

- 
- [13] Davison A. C. e Smith R. L. (1990) Models for exceedances over high threshold (with discussion). *J. R. Statis. Soc. B* **52** 393-442
- [14] De Finetti B. (1929). Sulla possibilità di valori eccezionali per una legge di incrementi aleatori. *Rendiconti della R. Accademia nazionale dei Lincei*, **X**, serie 6, fasc. 7-8
- [15] Diaconis P. e Freedman D. (1980). De Finetti's theorem for Markov chains. *Ann. Probability*, **8**, 1, 115-130
- [16] Dykstra, R.L. e Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Ann. Statist.*, **9**, 2, 356-367
- [17] Doksum, K.A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Ann. Prob.*, **2**, 183-201
- [18] Embrechts P., Kluppelberg C. e Mikosh T. (1997). *Modelling extremal events*. New York: Springer
- [19] Ferguson T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, **1**, 2, 209-230
- [20] Ferguson, T. S. e Klass, M. J. (1972). A representation of independent increment processes without gaussian components. *Ann. Math. Statist.*, **43**, 1634-1643
- [21] Feller W. (1968). *An introduction to probability theory and its applications, Vol. I.*(3rd ed.) Wiley, New York
- [22] Gilks, W. R., Richardson, S. e Spiegelhalter, D. J. (eds.) (1996), *Markov Chain Monte Carlo in practice*. Chapman and Hill, London
- [23] Gikhman, I.I. e Skorokhod, A.V. (1965). *An introduction to the theory of random processes*. Dover Publications, New York
- [24] Grandell, J. (1976), *Doubly stochastic processes*. Lecture notes in Mathematics 529. Springer- Verlag, Berlin
- [25] Gutiérrez-Peña E. e Nieto-Barajas L. E. (2002). Bayesian nonparametric inference for mixed Poisson processes. *Bayesian Statistics*
- [26] Kuo L. e Ghosh S. K. (1997). Bayesian nonparametric inference for nonhomogeneous Poisson Processes. *Technical report* 9718. Department of Statistics. University of Connecticut, Storrs, USA



- [27] Lo A. Y. (1981). Bayesian nonparametric statistical inference for Shock Models and Wear Processes. *Scand. J. Statist.*, **8**, 237-242
- [28] Lo A. Y. (1982). A note on the minimax estimation of the Poisson intensity function. *Ann. Statist.*, **10**, 3, 1009-1011
- [29] Lo A. Y. (1982). Bayesian nonparametric statistical inference for Poisson point processes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **59**, 55-66
- [30] Lo A. Y. (1984). On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann. Statist.*, **12**, 1, 351-357
- [31] Lo A. Y. e Weng C-S (1989). On a class of Bayesian nonparametric estimates: II. Hazard rate estimates. *Ann. Inst. Statist. Math.*, **41**, 2, 227-245
- [32] Longin F. (1994). Optimal marginal level in futures markets: a parametric extreme-based method.
- [33] Mauldin, R.D., Sudderth, W.D. e Williams, S.C. (1992). Pólya trees and random distributions. *Ann. Statist.*, **20**, 1203-1221
- [34] Mc Donald D. (1989). On nonhomogeneous, spatial Poisson processes. *The Canadian Journal of Statistics*, **17**, 2, 183-195
- [35] Muliere P. e Walker S. (1997). A Bayesian non-parametric approach to survival analysis using Polya trees. *Scand. Journal of Statistics*, **24**, 331-340
- [36] Muliere P. e Walker S. (2000). Neutral to the right processes from a predictive perspective: a review and new developments. *Metron*, Vol. LXVIII, n. 3-4, 13-30
- [37] Muliere P., Secchi P. e Walker S. G. (2000). Urn schemes and reinforced random walks. *Stochastic Processes and their Applications*, **88**, 59-78
- [38] Muliere P., Secchi P. e Walker S. G. Reinforced random process in continuous time.
- [39] Nobile, A. e Green, P.J. (2000). Bayesian analysis of factorial experiments by mixture modelling. *Biometrika*. **87**, 15-35
- [40] Pickands J.(1975), Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119-131
- [41] Pickands J.(1971). The two-dimensional Poisson process and extremal processes. *J. Appl. Prob.*, **8**, 745-756

- [42] Revfeim K. J. A. (1984). The cumulants of an extended family type I extreme value distributions. *The Indian Journal of Statistics B*, **46**, Series B, 281-284
- [43] Richardson, S. e Green, P. J. (1997). On bayesian analysis of mixtures with an unknown number of components. *J. R. Statist. Soc.*, **B**, 4, 731-792
- [44] Sato (1999). *Lévy processes and infinitely divisible distributions*. Cambridge studies in advanced Mathematics 68. Unversity Press, Cambridge
- [45] Shively, T. S. (1991), An analisys of the trend in ground-level ozone using non-homogeneous Poisson processes. *Atmospheric Environment*, **25B**, 387-396
- [46] Smith R. L. (2001). *Environmental Statistics*.  
Web reference: <http://www.stat.unc.edu/postscript/rs/envonotes.ps>
- [47] Smith R. L. (1989). Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, **4**, 4, 367-377
- [48] Smith R. L. (1997c), Predictive inference, rare events and hierarchical models. Preprint, University of North Carolina
- [49] Smith R. L. e Goodman D. (2000). Bayesian risk analysis. Chapter 17 of *Extremes and Integrated Rosk Management*. Risk Books, London,235-251
- [50] Smith R. L., Tawn J. A. e Coles S. G. (1997). Markov chain models for threshold exceedances. *Biometrika*, **84**, 2, 249-268
- [51] Smith R. L. (1998). Measuring risk with extreme value theory.
- [52] Tancredi A., Anderson C. e O'Hagan A. (19??). A Bayesian model for threshold uncertainty in extreme value theory *Working paper 2002.12, Dipartimento di Scienze Statistiche, Università di Padova*. Submitted
- [53] Tawn, J. A. (1998b), Bivariate extreme value theory - models estimation. *Biometrika*, **75**, 397-415
- [54] Walker, S. e Damien, P. (1998). A full bayesian non-prametric analysis involving neutral to the right processes. *Scand. J. Statist.*, **25**, 669-680
- [55] Walker, S. e Damien, P. (19??). Sampling methods for bayesian nonparametric inference involving stochastic processes. *Lecture notes in Statistics.*, 243-254
- [56] Walker, S. e Muliere, P. (1997). Beta-Stacy processes and a generalization of the Pólya-urn scheme. *Ann. Statist.*, **25**, 1762-1780

- 
- [57] Wang Y. H. (1989). A multivariate extension of Poisson's theorem. *The Canadian Journal of Statistics*, **17**, 2, 241-245
- [58] Wolpert R. e Ickstadt K. (1998a). Poisson/gamma random field models for spatial statistics. *Biometrika*, **85**, 251-267
- [59] Wolpert R. and Ickstadt K. (1998b). Simulation of Lévy random fields. In *practical nonparametric and semiparametric bayesian statistics*. (D. Dey, P. Müller and D. Sinha, Eds.) Springer- Verlag, New York, pp. 227-242