FAST FOURIER TRANSFORM, EXTREME VALUE THEORY E SIMULAZIONE NELL'ANALISI DELLA DIPENDENZA TRA NORMAL E LARGE CLAIMS DI UNA COMPAGNIA DI ASSICURAZIONI DANNI

Rocco Roberto Cerchiara[1]

[1] Università della Calabria
(rocco.cerchiara@unical.it)

ABSTRACT: L'obiettivo del lavoro è dimostrare la possibilità di un utilizzo integrato tra Extreme Value Theory (EVT), Fast Fourier Transform (FFT) e simulazione di Monte Carlo nella modellizzazione del costo sinistri aggregato di una compagnia di assicurazione danni, anche nell'ottica dello sviluppo di Internal Risk Models in relazione al progetto comunitario Solvency II. L'EVT viene sfruttato per individuare il punto di demarcazione tra i sinistri di entità ordinaria ed i cosiddetti large claims. La FFT a due dimensioni consente di modellizzare non solo la distribuzione del costo sinistri aggregato, ma anche eventuali dipendenze tra le sue singole componenti. Viene infine sfruttata la simulazione di Monte Carlo per descrivere l'evoluzione dei large claims. Tale modello segue l'impostazione dell'approccio collettivo della Teoria del Rischio.

KEYWORDS: Assicurazioni Danni, Fast Fourier Transform, Extreme Value Theory, Simulazione di Monte Carlo, approccio collettivo della Teoria del Rischio.

1 Introduzione

Un'ipotetica compagnia di assicurazioni danni ha sviluppato un modello, basato su un approccio simulativo, con l'obiettivo di individuare la distribuzione di probabilità del danno aggregato. Tale modello però è basato su una ipotesi semplificatrice, in quanto piccoli e grandi sinistri (o detti anche small e large claims) vengono simulati separatamente e quindi i relativi costi sinistri aggregati sono ipotizzati reciprocamente indipendenti. In particolare la procedura simula un valore aggregato per tutti i sinistri ordinari e simula individualmente i grandi sinistri. L'obiettivo del presente lavoro è quello di dimostrare la possibilità di un utilizzo integrato tra Extreme Value Theory (EVT), Fast Fourier Transform (FFT) e simulazione di Monte Carlo, al fine di, rispettivamente, individuare una soglia come discriminante tra piccoli e grandi sinistri (c.d. truncation point), modellizzare le eventuali dipendenze tra small e large claims e simulare i grandi sinistri. In tal modo il modello opera in maniera più efficiente preservando la struttura di dipendenza. Nelle pagine seguenti viene riportato, a titolo di esempio, il calcolo del valore atteso del danno aggregato derivante sia dai normal claims che dai sinistri di entità elevata ($E[X^{Small,\ Large}]$).

2 Un esempio di applicazione dell'EVT, FFT e Simulazione

Nell'ambito dell'approccio collettivo della Teoria del Rischio (si veda Daykin et al. [1994]) la variabile aleatoria danno aggregato, ad esempio in un anno, è data dalla relazione

$$X = \sum_{i=1}^{k} Z_i \tag{1}$$

dove k è la variabile aleatoria numero sinistri e le Z_i sono i costi aleatori dei singoli sinistri (i.i.d), supposti reciprocamente indipendenti.
Si ipotizzi che la distribuzione del costo del singolo sinistro Z sia caratterizzata da 6 punti (0, $2{*}10^6$, $4{*}10^6$, $6{*}10^6$, $8{*}10^6$, $10{*}10^6$) a cui corrispondono le seguenti probabilità: 0,00%, 43,80%, 24,60%, 13,80%, 7,80%, 10,00%. Tale distribuzione di probabilità può essere rappresentata attraverso la matrice $\mathbf{M_Z}$ (si veda Homer e Clark [2004] e Cerchiara [2006]), che consente di evidenziare eventuali dipendenze tra sinistri ordinari e grandi sinistri. Nella prima colonna é riportata la distribuzione di probabilità della $\mathbf{Z}$ fino al valore della soglia pari a $10{*}10^6$ individuata attraverso l'EVT (si veda Embrechts et al. [1997] ed in particolare Cerchiara [2006] per l'applicazione dell'EVT al Database dei Dati Danesi Incendio nel periodo 1990-2002 ai fini dell'individuazione di tale soglia) e nella cella (0,1) viene riportata la probabilità di avere un sinistro di entità maggiore della soglia (10%).

Importo Sinistri ***Small*** *(Z^{Small})*	Numero sinistri ***Large*** (k^{Large})					
	0	**1**	**2**	**3**	**4**	**5**
-	0,000	**0,100**	0,000	0,000	0,000	0,000
2.000.000	**0,438**	0,000	0,000	0,000	0,000	0,000
4.000.000	**0,246**	0,000	0,000	0,000	0,000	0,000
6.000.000	**0,138**	0,000	0,000	0,000	0,000	0,000
8.000.000	**0,078**	0,000	0,000	0,000	0,000	0,000
10.000.000	0,000	0,000	0,000	0,000	0,000	0,000

Tabella 2.1: Matrice congiunta dell'importo sinistri Small e numero sinistri Large

Tale rappresentazione fornisce un significativo esempio della flessibilità del metodo FFT a 2 dimensioni, che può operare su grandezze con unità di misura diverse (numeri ed importi).
In presenza di dipendenze tra piccoli e grandi sinistri e se viene utilizzata una singola frequenza per generare il numero complessivo di sinistri, allora la covarianza tra il costo sinistri aggregato relativo ai due gruppi, rispettivamente, X^{Small} e X^{Large} (si veda Homer e Clark [2004]) può essere scritta:

$$Cov(X^{Small}, X^{Large}) = p \cdot E\left[Z^{Small}\right] * (1-p) \cdot E\left[Z^{Large}\right] \cdot \left(\sigma_k^2 - E[k]\right) \quad (2)$$

dove $\boldsymbol{p}$ è la probabilità che un sinistro sia "piccolo". Nel caso in cui la variabile aleatoria numero sinistri **k** si distribuisca secondo la Binomiale Negativa il segno della (2) è positivo; qualora segua la distribuzione Poisson Semplice la covarianza è nulla e si può ricorrere al caso di indipendenza su indicato. Nell'ipotesi di dipendenza, per il numero sinistri viene quindi utilizzata una distribuzione Binomiale Negativa con media = 10 e varianza = 20 (per ipotesi).
La distribuzione aggregata sarà data dall'espressione:

$$M_X = IFFT(PGF(FFT(M_Z))) \qquad e \qquad PGF = (2-t)^{-10} \quad (3)$$

dove *PGF* è la Funzione Generatrice delle Probabilità della Binomiale Negativa, FFT è l'algoritmo di calcolo della Fast Fourier Transform a due dimensioni e IFFT è la Inverse FFT a due dimensioni (si veda Klugman et al. [1998], Wang [1998] e Homer e Clark [2004]). Una volta eseguita la procedura si ottiene la seguente matrice (M_X) dove per brevità non sono riportate tutte le righe:

X^{Small}	Numero sinistri **Large** (k_i)							
	0	**1**	**2**	**3**	**4**	**5**	**6**	$pr\left(X^{Small}\right)$
-	0,37%	0,15%	0,03%	0,01%	0,00%	0,00%	0,00%	0,6%
2.000.000	0,57%	0,25%	0,07%	0,02%	0,00%	0,00%	0,00%	0,9%
4.000.000	0,74%	0,36%	0,09%	0,03%	0,00%	0,00%	0,00%	1,2%
…………	……	……	……	……	……	……	……	…….
48.000.000	1,16%	1,04%	0,56%	0,23%	0,07%	0,02%	0,00%	3,1%
50.000.000	1,03%	0,93%	0,50%	0,21%	0,07%	0,02%	0,00%	2,8%
							Totale	**100,0%**

Tabella 2.2:Distribuzione bivariata numero large claims e importo normal claims

Figura 2.1: *Funzione di Ripartizione di* X^{Small}

La distribuzione marginale della X^{Small} si ottiene sommando le probabilità in ciascuna riga (si veda anche la Figura 2.1). In tal caso $E[X^{Small}] = 28{,}0*10^6$ Euro.

Successivamente viene ricavata la distribuzione di frequenza condizionata per i grandi sinistri, attraverso il "rescaling" della M_X, rapportando ogni elemento al totale della rispettiva riga:

X^{Small}	Numero grandi **Large** (k)							
	0	**1**	**2**	**3**	**4**	**5**	**6**	**Totale**
-	64,7%	26,2%	5,2%	1,6%	0,8%	0,8%	0,8%	100.0%
2.000.000	62,4%	27,4%	7,2%	1,6%	0,4%	0,4%	0,4%	100.0%
4.000.000	60,3%	29,3%	7,3%	2,0%	0,3%	0,3%	0,3%	100.0%
………	…..	…..	…..	…..	…..	…..	…..	…..
48.000.000	37,5%	33,7%	18,2%	7,4%	2,4%	0,6%	0,2%	100.0%
50.000.000	37,3%	33,8%	18,1%	7,4%	2,5%	0,7%	0,2%	100.0%

Tabella 2.3: Rescaling della Tabella 2.2

Quest'ultima tabella consente di evidenziare ancora meglio le dipendenze tra i sinistri ordinari e large claims, osservando che in ogni riga sono presenti le probabilità di avere 0, 1, 2, …, 6 sinistri Large. A questo punto è possibile ricavare il valore atteso del numero di grandi sinistri (variabile aleatoria condizionata all'importo dei sinistri ordinari): $E[k^{Large}] = 0.9$. Effettuando ad esempio 200.000 simulazioni per la Z^{Large} in base alla Generalized Pareto Distribution, con parametri $\xi = 0{,}5$ e $\beta = 8{,}8$ (si veda Cerchiara [2006]), si ottiene un valore medio $E[Z^{Large}] =$ 27,6*10^7. Quindi il danno aggregato atteso sarà dato da: $E[X^{Small,Large}] = E[X^{Small}] + E[k^{Large}] \cdot E[Z^{Large}]$ =(28,0+240,8)*10^6=268,8*10^6.

Bibliografia

Cerchiara R. R. [2006], Metodo Simulativo, Fast Fourier Transform ed Extreme Value Theory per l'analisi del costo sinistri aggregato nelle assicurazioni danni, *Tesi di dottorato in Scienze Statistiche Attuariali*, Università La Sapienza, Roma.

Daykin C. D., Pentikainen T., Pesonen M. [1994], Practical risk theory for actuaries, *Chapman and Hall*, London.

Embrechts P., Kluppelbrg C., Mikosch T. [1997], Modelling extremal events, *Springer Verlag*, Berlin

Homer D. L., Clark D. R. [2004], Insurance Applications of Bivariate Distributions, *CAS Forum*.

Klugman S., Panjer H., Willmot G. [1998], Loss Models - From Data to Decisions, *John Wiley & Sons*, New York, First Edition.

Wang S.S. [1998], Aggregation of correlated risk portfolios: models and algorithms, *PCAS LXXXV*.