

**MODELLI MATEMATICI PER I
PROCESSI DI SINISTRI**

CORSO F.A.C - S.I.F.A. 2012

DOCENTE: GIOVANNA FERRARA

INDICE

	Pagina
INTRODUZIONE	1
PARTE I: COSTO COMPLESSIVO DEL PORTAFOGLIO (O DANNO GLOBALE)	
1.1. Formalizzazione della nozione di danno globale	2
1.2. Modelli parametrici dei processi	4
PARTE II: PROCESSO DEL NUMERO DEI SINISTRI	
2.1. Premessa	5
2.2. Formalizzazione del processo del numero di sinistri	5
2.3. Processo di Poisson	7
2.4. Distribuzione binomiale negativa	10
2.5. Distribuzione Binomiale	13
2.6. La classe (a,b,θ)	15
PARTE III: PROCESSO DEL NUMERO DEI SINISTRI – PROBLEMI DI BONUS MALUS	
3.1. Costruzione del modello	18
3.2. Indicazione sull'uso del modello	22
PARTE IV: PROCESSO DEL COSTO DEI SINISTRI	
4.1. Premessa	25
4.2. Una introduzione alla distribuzione lognormale	25
4.3. La distribuzione di Pareto	28
4.4. La distribuzione Gamma	32
4.5. La scelta di una conveniente funzione di distribuzione	33
4.6. Criteri di scelta	34
4.7. Considerazioni sul VaR	38
PARTE V: ALCUNE INDICAZIONI PER LA RAPPRESENTAZIONE GRAFICA	41
RIFLESSIONI FINALI	43

INTRODUZIONE

Nel proporre per un corso SIFA l'argomento che tratterò oggi sono stata indotta da alcune considerazioni che vorrei condividere con voi.

Innanzitutto mi duole constatare che non sempre gli attuari sono perfettamente coscienti delle ipotesi celate nelle loro formulazioni. Secondo il mio modo di pensare, invece sono proprio le ipotesi che rendono la teoria allettante

Inoltre noto che il linguaggio attuariale è a volte indebitamente ermetico e può risultare incomprensibile anche ai non profani. A mio parere occorrerebbe cercare di trattare i nostri argomenti con una certa leggerezza che, tuttavia, non significa né superficialità né mancanza di rigore, ma esprime solo il desiderio di voler carpire e fare propri gli elementi essenziali della teoria per poi essere in grado di affrontare agevolmente le sue applicazioni. Cercherò, pertanto, di trattare i processi con una certa "leggerezza" evitando di addentrarmi in minuziose dimostrazioni.

Un'altra considerazione riguarda la memoria storica. Per chi come me ha percorso il lungo cammino da un tempo pionieristico a un'epoca di complessità della materia attuariale, risulta a volte incomprensibile la mancanza di memoria che i giovani studiosi riservano agli studi dei loro "antenati". Va detto che spesso la mancanza di memoria storica porta a creare delle nuove definizioni e a attribuire nuovi nomi anche a problemi che sono stati proposti e risolti in epoche passate. In questa mia esposizione cercherò di ricostruire per quanto possibile anche un certo percorso storico dei principali concetti.

Il corpo di questo corso è strutturato da una parte teorica che, come detto, cercherò di trattare con una certa leggerezza e da applicazioni concrete. Le applicazioni concrete saranno sviluppate in ambiente Excel in maniera alquanto artigianale. I colleghi vorranno scusare la forma di presentazione tenendo conto che nella mia vita d'attuario ho dovuto percorrere anche il cammino delle nuove tecnologie. Tutti gli esercizi sono stati da me trattati varie volte aggiornando ovviamente i miei programmi per tener conto dei "progressi" tecnologici.

PARTE I

COSTO COMPLESSIVO DEL PORTAFOGLIO (O DANNO GLOBALE)

1.1. Formalizzazione della nozione di danno globale

Il costo complessivo dei sinistri relativo ad un certo portafoglio, in un determinato periodo di tempo (ad esempio un anno), può essere posto come:

$$X = Z_1 + Z_2 + \dots + Z_k \quad [1.1]$$

in cui:

- k il numero aleatorio dei sinistri con funzione di probabilità
 $p_k = \text{Prob} \{ N = k \} \quad k = 0, 1, 2, \dots$
- Z_i la variabile casuale indicante il costo dei sinistri relativamente al rischio i con
 funzione di ripartizione $S(z) = \text{Pr ob} \{ Z \leq z \}$
- X l'importo aleatorio del danno.

Supponendo noti il numero dei sinistri p_k e la funzione di ripartizione, il problema consiste nel determinare la funzione di ripartizione del danno globale [1.1] che si verifichi l'evento $X \leq X$. Questo evento può verificarsi in molti modi, cioè

- 1) Nell'intervallo non avviene alcun sinistro.
- 2) Il numero dei sinistri =1 e l'ammontare è $\leq X$
- 3) Il numero dei sinistri =2 e la somma degli importi è $\leq X$
- 4) Il numero dei sinistri =3 e la somma degli importi è, eccetera

La probabilità condizionata che il numero dei sinistri sia esattamente eguale a k e che la somma degli esborsi sia $\leq X$ può essere indicata con $p_k S_k(X)$. Applicando le regole additive e moltiplicative delle probabilità, la funzione di ripartizione del danno globale può essere posta come:

$$F(X) = \sum_{k=0}^{\infty} p_k S_k(X) \quad [1.2]$$

Se si ipotizza che gli importi dei sinistri siano mutualmente indipendenti, la funzione $S_k(X)$ è ben nota nel calcolo delle probabilità come la k -esima convoluzione della funzione di distribuzione $S(X)$, che può essere calcolata in base alla formula ricorrente:

$$S_k(X) = \int_0^X S_{k-1}(X-Y) dS(Y) = S^{k-1} * S(X) = S^{k*}(X)$$

Pertanto sarà:

$$F(X) = \sum_{K=0}^{\infty} p_k S^{k*}(X) \quad [1.3]$$

Si nota a questo punto un'evidente contraddizione nella simbologia: nella prima parte quando si è definito il danno [1.1] è stata introdotta una funzione di ripartizione $S(z)$ mentre successivamente [1.2] si parla di $S_k(X)$. In effetti si tratta di un unico fenomeno visto da due diverse angolazioni.

Nell'ultimo modello (detto collettivo) il sinistro aggregato è la somma di un numero k di pagamenti individuali che si suppone siano stati registrati o pagati nella sequenza indicata. Il

sinistro X_1 rappresenta il primo pagamento del periodo, X_2 il secondo e via di seguito. In questo caso non è necessario conoscere il numero di polizze che producono questi pagamenti. Ragionevolmente si suppone che le variabili termini della somma siano indipendenti e indipendenti dal numero dei sinistri.

Nel primo caso, denominato modello di rischio individuale, le variabili che rappresentano i termini della somma, descrivono gli importi pagati relativamente ad un rischio di un portafoglio. In questo modello si presume che le variabili causali Z_r siano mutuamente indipendenti (senza supporre che abbiano un'identica distribuzione). Ciascun contatto infatti produce sinistri conformemente alle caratteristiche specifiche dell'assicurato. In sostanza il modello individuale viene impiegato per sommare i sinistri o i pagamenti derivanti da un insieme di contratti assicurativi.

Va detto che in entrambi i casi si distinguono i due processi del numero di sinistri e della loro dimensione.

Se il numero dei sinistri risulta molto grande, i calcoli che si dovrebbero effettuare per il modello individuale risulta considerevole. Per semplificare il lavoro allora si fa ricorso ad un modello collettivo, la cui costruzione, in termini di distribuzioni di probabilità, risulta simile al modello individuale.

In una forma semplificata¹ del modello individuale si usa porre come espressione della media e della varianza del danno globale le relazioni seguenti:

$$E[X] = E[N] \times E[Z] \quad [1.4]$$

$$Var[X] = E[N] \times E[Z^2] + (Var[N] - E[N]) \times E[Z]^2 \quad [1.5]$$

Come abbiamo già osservato, il costo complessivo o danno globale dipende da due componenti fondamentali, cioè

- dal processo del numero di sinistri;
- dal processo del costo dei sinistri.

Nel primo processo si tratta di distribuzioni discrete con probabilità definite solo per numeri interi non negativi, cioè solo nei punti 0, 1, 2, 3,.... Queste distribuzioni descrivono il numero di eventi, come i danni subiti dall'assicurato o il numero di sinistri per la compagnia di assicurazione. Nel secondo processo invece la variabile è di tipo continuo e descrive la dimensione del danno.

In entrambi i processi considereremo modelli parametrici, che incorporano le informazioni relative alla distribuzione in termini di forma (per esempio funzione di Poisson, binomiale, ecc) e di valori dei parametri. Inoltre i modelli parametrici necessitano di informazioni nettamente inferiori a quelli richiesti per un'analisi non parametrica e permettono di integrare o di estrapolare le osservazioni di cui si dispone.

¹ Si veda ad esempio l'appendice G del testo di Gary S. Patrik "Chapter 7 – Reinsurance" Casualty Actuarial Society 2001.

1.2. Modelli parametrici dei processi

Per il numero aleatorio N si impiegano le seguenti distribuzioni di probabilità :

Poisson	Funzione di probabilità	$p_k = \frac{e^{-\lambda} \lambda^k}{k !}$
	Media	λ
	Varianza	λ
	Osservazione	media = varianza
Binomiale negativa	Funzione di probabilità	$p_k = \left(\frac{\gamma}{1+\gamma} \right)^\alpha \left(\frac{1}{1+\gamma} \right)^k \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)k!}$ $k = 1, 2, 3, \quad \alpha > 0, \quad \gamma > 0$
	Media	$E(N) = \frac{\alpha}{\gamma}$
	Varianza	$\frac{\alpha}{\gamma} \frac{1+\gamma}{\gamma}$
	Osservazione	$\gamma > 0$ quindi varianza > media
Binomiale	Funzione di probabilità	$p_k = \binom{m}{k} q^k (1-q)^{m-k}, \quad k = 0, 1, \dots, m$
	Media	$E(N) = m q$
	Varianza	$Var(N) = m q (1-q)$
	Osservazione	Varianza < media

La variabile entità del sinistro può essere descritta da varie funzioni di probabilità, quali

Lognormale	Funzione di densità	$f_z(x) = \frac{1}{x\sqrt{2\pi}\sigma} e^{-\frac{(\ln x - m)^2}{2\sigma^2}}$
	Media	con $0 < x < +\infty$, $\sigma > 0$ $E(Z) = e^{m + \frac{\sigma^2}{2}}$
	Varianza	$\text{var}(Z) = e^{2m + \sigma^2} [e^{\sigma^2} - 1]$
Gamma	Funzione di densità	$f_z(x) = \frac{1}{\beta \Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right)$
	Media	con $0 < x < +\infty$, $\alpha > 0$, $\beta > 0$ $E(Z) = \beta \alpha$
	Varianza	$\text{Var}(Z) = \beta^2 \alpha$
Pareto	Funzione di densità	$f_z(x) = \alpha \frac{c_0^\alpha}{x^{\alpha+1}}$ con $c_0 \leq x < +\infty$, $c_0 > 0$, $\alpha > 0$
	Media	$E(Z) = c_0 \frac{\alpha}{\alpha - 1}$ (se $\alpha > 1$)
	Varianza	$\text{var}(Z) = c_0^2 \frac{\alpha}{(\alpha - 1)^2 (\alpha - 2)}$ (se $\alpha > 2$)

Osserviamo che, per il numero aleatorio dei sinistri, la distribuzione da adottare è determinabile in base alla considerazione dei valori della media e della varianza, per la distribuzione dell'entità del costo dei sinistri occorre far riferimento non solo all'accostamento dei dati effettivi alla distribuzione prescelta, ma anche, nei problemi di solvibilità o di mitigazione del rischio, alla velocità con la quale la funzione di densità si riduce a zero quando il danno diventa infinitamente grande. Nel seguito forniremo alcuni criteri che possono guidare nella scelta delle distribuzioni.

Inoltre se si ipotizza che il numero dei sinistri sia distribuito secondo una legge di Poisson anche le somme (convoluzioni) sono dei processi di Poisson. I momenti centrali della variabile danno globale sono:

$$\text{Media } E[X] = E[N] \times E[Z] \quad [1.6]$$

$$\text{Varianza } \text{Var}[X] = E[N] \times E[Z^2] \quad [1.7]$$

Questa semplificazione delle formule è ampiamente usata a volte in maniera inconsapevole dagli attuari..

PARTE II

PROCESSO DEL NUMERO DEI SINISTRI

2.1. Premessa

Gli sviluppi di questa seconda parte sono tratti principalmente dall'eccellente testo di S.A. Klugman, H.H. Panjer, G.E. Wilmot "Loss Models – From Data to Decisions".

2.2. Formalizzazione del processo del numero di sinistri

Per meglio illustrare il processo partiremo da un esempio concreto: si tratta di una polizza di responsabilità civile per cure ospedaliere, che nel corso degli ultimi anni ha subito un certo numero di sinistri, come mostrato nella tavola che segue.

Tavola 2.1. Numero di sinistri di una polizza negli ultimi anni

Anno	Numero sinistri
2003	6
2004	2
2005	3
2006	0
2007	2
2008	1
2009	2
2010	5
2011	1
2012	3

In alternativa si possono rappresentare questi dati secondo il numero di anni nei quali ci sono stati 0 sinistri, 1 sinistro, ecc.

Tavola 2.2. Numero dei sinistri della polizza di tavola 2.1 in base alla frequenza

Frequenza	Numero di osservazioni
0	1
1	2
2	3
3	2
4	0
5	1
6	1
7+	0

Questa ultima distribuzione empirica fornisce informazioni sulla variabilità del numero di sinistri per anno e dovrebbe permettere di prevedere sia il numero di sinistri negli anni futuri sia la variabilità attesa futura.

La tavola però non fornisce direttamente le probabilità del verificarsi di un sinistro. Ad esempio nel corso degli ultimi anni non si sono verificati più di 7 sinistri. Occorre, quindi, adottare un modello parametrico che consenta di esprimere probabilità positive per tutti gli interi non-negativi e consenta un'estrapolazione automatica delle osservazioni. Va rilevato che estendendo la distribuzione oltre il valore massimo fornito dalle osservazioni si può studiare la forma della coda destra della distribuzione.

I modelli parametrici ci permettono anche di sviluppare distribuzioni di frequenza in situazioni assicurative in cui non si abbiano dati disponibili, come ad esempio in una polizza con livelli di franchigia prefissati.

2.3. Modello di Poisson

Il processo del numero dei sinistri è stato sempre descritto come connesso al verificarsi nel tempo di un evento aleatorio in base alle seguenti condizioni:

- la probabilità del verificarsi dell'evento durante un periodo molto piccolo di tempo dt è proporzionale alla durata del periodo, cioè a $p dt$;
- questa probabilità è indipendente dal numero di eventi che sono prodotti precedentemente e resta costante nel corso del periodo d'osservazione;
- la probabilità del verificarsi di due eventi successivi dello stesso avvenimento è trascurabile.

In base alle ipotesi esposte² il numero x dei sinistri avvenuti nel corso di un intervallo di tempo T è una variabile casuale di Poisson di parametro $\lambda = pT$ con funzione di probabilità:

$$P_k = \frac{e^{-\lambda} \lambda^k}{k!} \quad [2.1]$$

Ricordiamo, inoltre che la funzione di Poisson è un caso particolare della legge binomiale: cioè rappresenta la funzione verso la quale tende la binomiale quando il numero delle prove diviene grande e la probabilità di realizzazione dell'evento è piccola.

La funzione generatrice di probabilità è:

$$P(z) = e^{\lambda(z-1)}, \quad \lambda > 0 \quad [2.2]$$

da questa funzione si calcolano la media e la varianza:

$$\begin{aligned} E(N) &= P'(1) = \lambda \\ E[N(N-1)] &= P''(1) = \lambda^2 \\ Var(N) &= E[N(N-1)] + E(N) - [E(N)]^2 = \\ &= \lambda^2 + \lambda - \lambda^2 = \lambda \end{aligned} \quad [2.3]$$

La distribuzione di Poisson ha l'importante proprietà di avere la media eguale alla varianza. Conoscendo quindi il solo valor medio è possibile stimare i valori della distribuzione. Sia ad esempio $E(x) = \lambda = 0.2$

è possibile costruire i valori di $f(x)$ e di $F(x)$

X	$f(x)$	$F(x)$
0	0.8187	0.8187
1	0.1637	0.9825
2	0.0164	0.9989
3	0.0011	0.9999

² Una dimostrazione della formula del processo di Poisson viene fornita da Luciano Daboni in "Calcolo delle probabilità" – Boringhieri 1967 pag. 115-117

La funzione di Poisson possiede proprietà che risultano particolarmente utili per configurare i processi assicurativi.

La più importante delle proprietà è la seguente:

. Se N_1, \dots, N_n sono variabili di Poisson indipendenti con parametri $\lambda_1, \dots, \lambda_n$, la loro somma $N = N_1 + \dots + N_n$ ha distribuzione di Poisson con parametro $\lambda = \lambda_1 + \dots + \lambda_n$.

Infatti la funzione generatrice dei momenti totale è data dal prodotto delle funzioni delle v.a. indipendenti. Analiticamente:

$$P_N(z) = \prod_{j=1}^n P_{N_j}(z) = \prod_{j=1}^n \exp[\lambda_j(z-1)]$$

$$= \exp\left[\sum_{j=1}^n \lambda_j(z-1)\right] = e^{\lambda(z-1)}$$

Si dimostra quindi che la somma N è una distribuzione di Poisson con parametro λ .

Stima dei parametri

La stima del parametro del modello di Poisson può essere effettuata mediante il metodo dei momenti, il cui risultato, come vedremo, coincide con quello della massima verosimiglianza. Consideriamo a questo proposito la precedente distribuzione dei sinistri di responsabilità nelle cure ospedaliere (illustrato nella tavola 2.1). Il numero totale dei sinistri tra il 2003 e il 2012 è 25, quindi il numero medio di sinistri per anno è 2.5. Indicando con n_k il numero di anni nei quali si verificano esattamente k sinistri. La frequenza attesa (media del campione) è:

$$\hat{\lambda} = \frac{\sum_{k=0}^{\infty} k n_k}{\sum_{k=0}^{\infty} n_k}$$

dove n_k rappresenta il numero dei valori osservati in corrispondenza della frequenza k . Quindi la stima del metodo dei momenti del parametro di Poisson è 2.5.

Per quanto riguarda la stima di massima verosimiglianza, la probabilità di un'osservazione k è p_k . Quindi la verosimiglianza per l'intera serie di osservazioni è

$$L = \prod_{k=0}^{\infty} p_k^{n_k} \quad [2.4]$$

e il logaritmo della verosimiglianza è

$$l = \sum_{k=0}^{\infty} n_k \ln p_k \quad [2.5].$$

Nelle precedenti funzioni occorre determinare il valore dei parametri. Per la distribuzione di Poisson, che ha un solo parametro, si otterrà:

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!}$$

e $\ln p_k = -\lambda + k \ln \lambda - \ln k!$ quindi:

$$l = -\lambda n + \sum_{k=0}^{\infty} k n_k \ln \lambda - \sum_{k=0}^{\infty} n_k \ln k!$$

e differenziando rispetto a λ , si ottiene

$$\frac{dl}{d\lambda} = -n + \sum_{k=0}^{\infty} k n_k \frac{1}{\lambda}$$

La stima di massima verosimiglianza è, quindi, ottenuta ponendo eguale a 0 la derivata espressa dall'equazione precedente. Lo stimatore è pertanto

$$\hat{\lambda} = \frac{\sum_{k=0}^{\infty} k n_k}{n} \quad [2.6].$$

Quindi nel caso di una distribuzione di Poisson, il metodo dei momenti e della massima verosimiglianza danno risultati identici.

Un'ulteriore illustrazione del processo di conto del numero dei sinistri può essere fornita da un esempio numerico sempre tratto da S. A. Klugman – H.H. Panjer- G.E. Willmot – Loss Models. Si tratta di una statistica relativa al comportamento di 23.589 automobilisti classificati secondo il numero di incidenti avuti in un periodo annuale.

Tavola 2.3. Esempio del processo del numero di sinistri Assicurazione Auto

Numero di sinistri k	Numero di conducenti n_k
0	20.592
1	2 651
2	297
3	41
4	7
5	0
6	1
7	0
Totale	23.589

I valori della media e della varianza sono

$$media = 0.1442$$

$$varianza = 0.1639$$

Per il processo di Poisson le probabilità sono determinate dalla formula

$$p_k = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{con } \lambda = 0.1442$$

Tavola 2.4. Processo di Poisson nell'esempio numerico della tavola 2.3

Numero di sinistri k	Numero di Conducenti n_k	Probabilità effettiva	Probabilità teorica	Numero teorico	Valori di χ^2
0	20.592	0,87295	0,86570	20.420,9	1,4
1	2.651	0,11238	0,12485	2.945,1	29,4
2	297	0,01259	0,00900	212,4	33,7
3	41	0,00174	0,00043	10,2	92,9
4	7	0,00030	0,00002	0,4	
5	0	0,00000	0,00000		
6	1	0,00004	0,00000		
7	0	0,00000	0,00000		
Totale	23.589				157,4

2.4. Distribuzione binomiale negativa

Nei primi anni dell'ASTIN (cioè negli anni '70), la distribuzione di Poisson ha presto ceduto il passo alla distribuzione binomiale negativa. Le prime elaborazioni delle tariffe bonus-malus portavano infatti ad ipotizzare che ciascun rischio avesse una propria propensione (a provocare sinistri). Pertanto, sotto vari modi, si pensò di utilizzare la distribuzione binomiale negativa. Va detto che la distribuzione e le sue particolarità erano già state trattate in maniera esauriente da O. Lundberg nel suo testo relativo agli infortuni sul lavoro. Il testo³ redatto già nel 1940 fu ripubblicato nel 1964.

La distribuzione binomiale negativa, come avviene per la distribuzione di Poisson, ha probabilità positive per interi non-negativi. La sua forma è più flessibile rispetto alla funzione di Poisson, essendo determinata da due parametri.

Esistono vari modi per presentare la distribuzione binomiale negativa. Qui sarà presentata come mistura della distribuzione di Poisson con una funzione Gamma.

Supponiamo che ciascun rischio sia caratterizzato da una propensione agli incidenti invariabile nel tempo ma diversa da rischio a rischio. Tale diversa propensione sarà rappresentata dal parametro λ .

Indichiamo con $p(x, \lambda; t)$ la probabilità per un rischio, caratterizzato da una propensione λ , di provocare x incidenti durante il periodo $0 - t$. Se $f(\lambda)$ è la legge secondo cui la variabile λ si distribuisce fra i rischi, la probabilità che si verificano x sinistri nel tempo $0 - t$ sarà data da:

$$P(x; t) = \int_0^{\infty} p(x, \lambda; t) f(\lambda) d\lambda \quad [2.7]$$

Ove Lundberg "On Random Processes and their Application to Sickness and Accident Statistics"
 Uppsala, 1964

La relazione [2.7] può essere facilmente interpretata facendo riferimento ad un'estrazione di questo tipo (modello di Lexis): siano diverse urne contenenti ciascuna palline bianche e nere di diverse proporzioni e sia $f(\lambda) d\lambda$ la probabilità di sorteggiare l'urna che ha un valore ξ compreso fra λ e $\lambda + d\lambda$. La probabilità composta di scegliere l'urna caratterizzata da ξ e di ottenere nell'estrazione successiva x palline bianche sarà data da

$$p(x, \lambda) f(\lambda) d\lambda$$

di cui la [2.7].

In problemi come questi la $p(x, \lambda; t)$ può essere descritta dalla v.c. di Poisson, cioè assumere la forma:

$$p(x, \lambda; t) = \frac{e^{-\lambda t} (\lambda t)^x}{x!} \quad [2.8]$$

Per la $f(\lambda)$ - che dovrebbe essere desunta dai dati statistici - numerose esperienze hanno dimostrato come si possa far ricorso alla distribuzione Γ , cioè alla

$$f(\lambda) = \frac{\gamma^\alpha}{\Gamma(\alpha)} e^{-\gamma\lambda} \lambda^{\alpha-1} \quad [2.9]$$

con α e $\gamma > 0$

Introducendo le relazioni [2.8] e [2.9] nella [2.7] si ottiene la seguente espressione della probabilità del verificarsi di x sinistri nel periodo $0 - t$:

$$P(x; t) = \left(\frac{\gamma}{t+\gamma} \right)^\alpha \left(\frac{t}{t+\gamma} \right)^x \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)x!} \quad [2.10]$$

per $t = 1$

$$P(x; 1) = \left(\frac{\gamma}{1+\gamma} \right)^\alpha \left(\frac{1}{1+\gamma} \right)^x \frac{\Gamma(x+\alpha)}{\Gamma(\alpha)x!}$$

Coerentemente a quanto indicato per la distribuzione di Poisson, potremmo scrivere

$$p_k = \left(\frac{\gamma}{1+\gamma} \right)^\alpha \left(\frac{1}{1+\gamma} \right)^k \frac{\Gamma(k+\alpha)}{\Gamma(\alpha)k!} \quad [2.11]$$

$$k = 1, 2, 3, \quad \alpha > 0, \quad \gamma > 0$$

La relazione [2.10] è l'espressione di una binomiale negativa, con funzione di probabilità

$$\frac{\Gamma(x+\alpha)}{\Gamma(\alpha)\Gamma(x+1)} p^\alpha q^x \quad [2.12]$$

$$\text{con } p = \frac{\gamma}{t+\gamma} \text{ e } q = \frac{t}{t+\gamma}$$

Stima dei parametri

Per $t = 1$ (cioè considerando un solo periodo di osservazione) il modello dipende da due soli parametri che possono essere determinati in base ai dati statistici. Infatti si ottiene:

$$\gamma = \frac{E(X;1)}{\sigma^2(X;1) - E(X;1)} \quad [2.13]$$

$$\alpha = \gamma E(X;1) \quad [2.14]$$

Quindi col metodo dei momenti, i due parametri sono dati da:

$$\gamma = \frac{\text{media}}{\text{varianza} - \text{media}} \quad [2.15]$$

$$\alpha = \gamma \times \text{media} \quad [2.16]$$

Poiché la funzione è definita per valori positivi dei parametri, occorre che sia:

$$\text{varianza} > \text{media}$$

Tavola 2.5. Processo di Binomiale Negativa nell'esempio numerico della tavola 2.3

Numero di sinistri k	Numero di Conducenti n_k	Probabilità effettiva	Probabilità teorica	Numero teorico	Valori di χ^2
0	20 592	0,87295	0,87353	20605,8	0,0
1	2 651	0,11238	0,11088	2615,5	0,5
2	297	0,01259	0,01368	322,8	2,1
3	41	0,00174	0,00167	39,5	0,1
4	7	0,00030	0,00020	4,8	
5	0	0,00000	0,00002	0,6	
6	1	0,00004	0,00000	0,1	
7	0	0,00000	0,00000	0,0	
Totale	23 589				2,6

Va rilevato che la stima mediante il metodo dei momenti non coincide con quella di massima verosimiglianza e occorrerebbe apportare delle leggere modifiche al valore dei parametri (come mostrato nel testo “Loss Models”).

Infine notiamo che la distribuzione binomiale negativa è una generalizzazione della Poisson in due modi diversi, cioè come una mistura di una distribuzione di Poisson con una Gamma e come un composto di una Poisson con una distribuzione secondaria logaritmica.

Il caso $\alpha = 1$ corrisponde alla distribuzione geometrica.

2.5. Distribuzione Binomiale

La distribuzione binomiale possiede delle proprietà che sono diverse rispetto alle distribuzioni trattate precedentemente. Infatti si ha che:

- La varianza è più piccola della media e quindi il modello è applicabile quando si verifichi nei dati concreti questa circostanza.
- Descrive una situazione pratica nella quale si abbia in un insieme di m rischi, ognuno dei quali è soggetto al danno o al sinistro totale.

Questa seconda circostanza può essere esemplificata nel caso di un'assicurazione sulla vita. Supponiamo un insieme di m rischi identici e indipendenti con probabilità q di dar luogo ad un sinistro; gli individui considerati hanno inoltre la stessa mortalità, per esempio fumatori maschili di 35 anni e polizza di durata 5 anni, in questo caso q rappresenta la probabilità di morire nell'anno considerato.

Il numero di sinistri di un singolo assicurato è rappresentabile mediante la distribuzione di Bernoulli, con probabilità $1 - q$ in 0 e q in 1. La funzione generatrice di probabilità del numero di sinistri per un individuo è:

$$P(z) = (1 - q)z^0 + qz^1 = 1 + q(z - 1).$$

La funzione generatrice di probabilità del numero totale di sinistri in un gruppo di m individui è:

$$P(z) = [1 + q(z - 1)]^m, \quad 0 < q < 1$$

quindi la probabilità del verificarsi di esattamente k sinistri nel gruppo di individui è data da:

$$p_k = \Pr(N = k) = \binom{m}{k} q^k (1 - q)^{m-k}, \quad k = 0, 1, \dots, m \quad [2.16]$$

Questa è la distribuzione di probabilità di una binomiale con parametri m e q . Al massimo m eventi possono verificarsi, quindi la distribuzione ha probabilità positive per ogni numero intero non-negativo fino a m (incluso). L'intervallo di valori per i quali esistono probabilità positive ha ampiezza finita.

La media e varianza della binomiale sono:

$$E(N) = mq, \quad \text{Var}(N) = mq(1 - q) \quad [2.17]$$

Stima dei parametri

Spesso il valore di m è noto e quindi occorre stimare solo q . In molti tipi di assicurazioni q rappresenta la probabilità del verificarsi di alcuni eventi, come morte o invalidità. In tal caso il valore di q può essere stimato dalla quantità

$$\hat{q} = \frac{\text{Numero eventi osservati}}{\text{Massimo numero possibili eventi}}$$

La precedente quantità rappresenta anche la stima mediante il metodo dei momenti, supponendo che m sia noto. Se m non è noto le equazioni [2.17] permettono di stimare entrambi i parametri tenendo presente che m è una quantità intera positiva.

Con il metodo della massima verosimiglianza si ha:

$$l = \sum_{k=0}^m n_k \ln p_k$$

$$= \sum_{k=0}^m n_k \left[\ln \binom{m}{k} + k \ln q + (m-k) \ln (1-q) \right] \quad [2.18]$$

Se m è noto, bisogna solo massimizzare rispetto a q e si ottiene:

$$\hat{q} = \frac{1}{m} \frac{\sum_{k=0}^m k n_k}{\sum_{k=0}^m n_k} \quad [2.19].$$

Quando m non è noto, la stima di q è:

$$\hat{q} = \frac{1}{\hat{m}} \frac{\sum_{k=0}^m k n_k}{\sum_{k=0}^m n_k} \quad [2.20],$$

dove \hat{m} è la stima di m . Per stimare m e q mediante il metodo della massima verosimiglianza si può considerare la seguente procedura:

- Passo 1: Si parte con \hat{m} pari al valore osservato più grande
- Passo 2: Si ottiene \hat{q} utilizzando la formula precedente [2.20]
- Passo 3: Si calcola la verosimiglianza [2.18] di questi valori

La procedura viene illustrata sulla base dell'esempio esposto nella tavola che segue.

**Tavola 2.6. Numero di sinistri per polizza nel corso di un anno
 per un gruppo di 15.160 polizze**

Numero sinistri/polizza	Numero polizze
0	5.367
1	5.893
2	2870
3	842
4	163
5	23
6	1
7	1
8+	0

La media e varianza del campione sono 0,9854 e 0,8904. la varianza è più piccola della media, quindi il modello binomiale è adattabile. Il metodo dei momenti porta a

$$m q = 0.9854$$

$$m q (1 - q) = 0.8904$$

quindi $\hat{q} = 0.0964$ e $\hat{m} = 10.2144$. Poiché m può avere solo valori interi, verrà scelto 10 e \hat{q} diventa 0,09854. La varianza è conseguentemente pari a 0,88832, che differisce da quella del campione.

Per parametri interi viene consigliato l'uso del metodo della massima verosimiglianza: dai dati si osserva che $m \geq 7$, se m è noto bisogna massimizzare solo q . Se m non è noto si utilizzerà la procedura esposta in precedenza partendo dal valore di 7 sinistri. I risultati sono esposti nella seguente tavola:

Tavola 2.7. Stima col metodo della massima verosimiglianza per la tavola 2.6

\hat{m}	\hat{q}	-Log verosimiglianza
7	0.1408	19273.56
8	0.1232	19265.37
9	0.1095	19262.02
10	0.0985	19260.98
11	0.0896	19261.11
12	0.0821	19261.84

Il minimo valore (massima verosimiglianza) si trova per $m=10$, e quindi le stime saranno $\hat{m} = 10$, $\hat{q} = 0,0985$.

Tavola 2.8. Processo di Binomiale nell'esempio numerico della tavola 2.6

Numero di sinistri k	Numero di polizze n_k	Probabilità teorica	Numero teorico	Valori di χ^2
0	5.367	0,3544	5372,2	0,0051
1	5.893	0,3874	5872,6	0,0709
2	2870	0,1906	2888,8	0,1225
3	842	0,0555	842,1	0,0000
4	163	0,0106	161,1	0,0225
5	23	0,0014	21,1	0,1651
6	1	0,0001	1,9	
7	1		0,1	
8	0			
9				
10				
Totale			15160,0	0,39

Va rilevato che, nonostante gli sforzi fatti per giungere ad un valore formalmente corretto si perviene allo stesso risultato ottenuto mediante il metodo dei momenti.

2.6. La classe $(a, b, 0)$

Un tentativo di trovare un ulteriore criterio discriminante per i processi del numero di sinistri viene effettuato da S. A. Klugman, H.H. Panjer & G.E. Willmot in "Loss Models"

Definita con p_k la funzione di probabilità della variabile aleatoria discreta, vengono determinati i valori di a e b della classe di distribuzioni $(a, b, 0)$ tali che

$$\frac{p_k}{p_{k-1}} = a + \frac{b}{k}, \quad k = 1, 2, 3, \dots$$

Questa formula ricorrente descrive il rapporto relativo di probabilità successive della distribuzione. Anche la probabilità in zero, p_0 , può essere calcolata facilmente tenendo presente che la somma delle probabilità deve essere eguale a 1.

La classe di distribuzioni $(a, b, 0)$ è una classe con due parametri. Tutte le funzioni studiate finora soddisfano la formula precedente e i valori dei parametri sono nella seguente tabella:

Distribuzione	a	b	p_0
Poisson	0	λ	$e^{-\lambda}$
Binomiale	$-\frac{q}{1-q}$	$(m+1)\frac{q}{1-q}$	$(1-q)^m$
Binomiale negativa	$\frac{1}{1+\gamma}$	$(r-1)\frac{1}{1+\gamma}$	$\left(1+\frac{1}{\gamma}\right)^{-\alpha}$
Geometrica	$\frac{1}{1+\gamma}$	0	$\left(1+\frac{1}{\gamma}\right)^{-1}$

La formula ricorrente può essere scritta come:

$$k \frac{p_k}{p_{k-1}} = ak + b, \quad k = 1, 2, 3, \dots$$

Si tratta di una funzione lineare di k . Il parametro a (che denota l'inclinazione) è eguale a 0 per la Poisson, negativo per la binomiale e positivo per la binomiale negativa e per la geometrica. Questo suggerisce per le differenti distribuzioni un metodo di scelta della distribuzione di tipo grafico. Si può dunque rappresentare graficamente la relazione:

$$k \frac{\hat{p}_k}{\hat{p}_{k-1}} = k \frac{n_k}{n_{k-1}}$$

Se si vuole applicare uno di questi modelli, i valori osservati dovrebbero disporsi approssimativamente lungo una linea retta. Il valore dell'inclinazione dovrebbe fornire delle indicazioni sulla distribuzioni da selezionare.

Notiamo tuttavia che nonostante il tentativo pregevole di cercare un criterio valido per la scelta della migliore funzione per descrivere il processo del numero dei sinistri, i risultati non sembrano soddisfacenti.

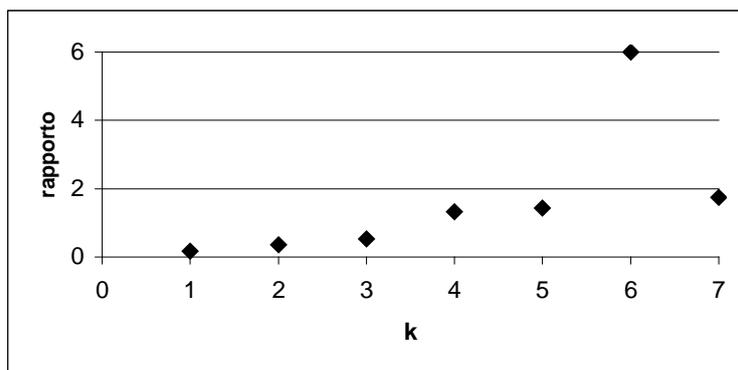
Forniamo ora un esempio che permette di comprendere quanto detto.

Tavola 2.9: Numero di incidenti riferiti a 9461 polizze auto

Numero di incidenti k	Numero di polizze n_k	$k \frac{n_k}{n_{k-1}}$
0	7.840	
1	1.317	0.17
2	239	0.36
3	42	0.53
4	14	1.33
5	4	1.43
6	4	6.00
7	1	1.75
8+	0	
Totale	9.461	

Osserviamo che l'affidabilità delle quantità calcolate nell'ultima colonna della tavola precedente diminuisce all'aumentare di k, poiché il numero delle osservazioni è ridotto.

Rappresentazione grafica del rapporto $k \frac{n_k}{n_{k-1}}$ in funzione di k



Come si può vedere dal grafico l'andamento del rapporto è fondamentalmente lineare ad esclusione del punto $k = 6$. L'inclinazione sembra positiva anche se non è evidente che sia significativamente diversa da 0. La distribuzione binomiale sicuramente è da escludere perché dovrebbe non c'è un'inclinazione negativa.

Considerando i valori della media e della varianza

Media = 0.2144; Varianza = 0.2889

La distribuzione binomiale negativa sembra essere più adatta poiché la varianza è maggiore della media. Tuttavia non esistono metodi formali che permettano di decidere se la varianza è sufficientemente maggiore della media. Il grado di accettabilità della distribuzione può essere verificato calcolando opportuni indici di accostamento.

Bibliografia essenziale:

- S. A. Klugman – H.H. Panjer- G.E. Willmot – Loss Models – Wiley – 1998

PARTE III

PROCESSO DEL NUMERO DEI SINISTRI – PROBLEMI DI BONUS - MALUS

3.1. Costruzione del modello

La distribuzione binomiale negativa è in genere impiegata anche nei problemi di Bonus – Malus. Ricordiamo che, in base alle ipotesi esposte nella parte II, si suppone che ciascun rischio (cioè ogni autovettura) sia caratterizzato da una propensione agli incidenti invariabile nel tempo ma diversa da autovettura ad autovettura. Si suppone inoltre che tale diversa propensione sia distribuita secondo la alla distribuzione Γ .

Le ipotesi sopra riportate portano a definire quale probabilità del verificarsi di x sinistri nel periodo $0 - t$:

$$P(x; t) = \left(\frac{\gamma}{t + \gamma} \right)^\alpha \left(\frac{t}{t + \gamma} \right)^x \frac{\Gamma(x + \alpha)}{\Gamma(\alpha)x!} \quad [3.1]$$

Ricordiamo inoltre che nel periodo $0 - t$ la media della variabile X è espressa da:

$$E(X; t) = \frac{\alpha}{\gamma} t \quad [3.2]$$

e la varianza di X è:

$$\sigma^2(X, t) = \frac{t\alpha}{\gamma} \frac{t + \gamma}{\gamma} \quad [3.3]$$

Per $t = 1$ (cioè considerando un solo periodo di osservazione, quale l'anno, il biennio o il semestre) il modello dipende da due soli parametri che possono essere determinati in base ai dati statistici. Infatti dalle relazioni [3.2] e [3.3] si ottiene:

$$\gamma = \frac{E(X; 1)}{\sigma^2(X; 1) - E(X; 1)} \quad [3.4]$$

$$\alpha = \gamma E(X; 1) \quad [3.5]$$

Poiché la funzione [3.1] è definita per valori positivi dei parametri, occorre che sia:

$$\sigma^2(X; 1) > E(X; 1) \quad [3.6]$$

Dalle relazioni precedenti si rileva che, a parità di varianza e al crescere della media, i due parametri aumentano. Analogamente, tenendo fissa la media ed aumentando la varianza (con il limite dato dalla [3.6]) i parametri risultano crescenti, generando distribuzioni di X meno concentrate. Questa caratteristica può risultare utile negli studi di simulazione, quando cioè si voglia costruire una funzione teorica di distribuzione dei sinistri, partendo da alcune ipotesi di frequenza media del portafoglio.

La relazione [3.1] consente di determinare le probabilità del verificarsi di 0, 1 ecc. sinistri in corrispondenza di vari intervalli temporali. In particolare per un intervallo eguale a due periodi (ad esempio 2 anni), la probabilità del verificarsi di y sinistri, può essere espressa nella forma:

$$P(y; 2) = \left(\frac{\gamma}{2+\gamma} \right)^\alpha \left(\frac{2}{2+\gamma} \right)^y \frac{\Gamma(y+\alpha)}{\Gamma(\alpha)y!} \quad [3.7]$$

Per la costruzione della matrice di transizione, occorre ora calcolare le probabilità subordinate, cioè le probabilità relative all'evento che siano y i sinistri complessivi nell'ambito di due periodi, nell'ipotesi che nel primo periodo siano avvenuti esattamente x incidenti. Schematizzando desideriamo conoscere la probabilità di quest'evento:

Periodi	1° periodo	2° periodo	Complessivo
Numero di sinistri	X	$y-x$	y

Nel modello di Poisson composto, tale probabilità assume la forma⁴:

$$\begin{aligned} P(y-x; 2/x; 1) &= \frac{y!}{(y-x)!x!} \left(\frac{1}{2} \right)^{y-x} \frac{P(y; 2)}{P(x; 1)} \left(\frac{1}{2} \right)^x = \\ &= \frac{y!}{(y-x)!x!} \left(\frac{1}{2} \right)^y \frac{P(y; 2)}{P(x; 1)} = P_{x,y} \quad [3.8] \end{aligned}$$

La relazione [3.8] consente di costruire la matrice teorica di transizione in base alla sola conoscenza dei due parametri della distribuzione, cioè in base ad ipotesi sulla media e sulla varianza della variabile numero dei sinistri. Ai fini applicativi occorre calcolare altri valori caratteristici del modello.

Un parametro interessante è la cosiddetta funzione d'intensità espressa dalla relazione:

$$\mu_n(t) = -\frac{P^{(n+1)}(t)}{P^{(n)}} \quad [3.9]$$

in cui le funzioni che figurano al numeratore ed al denominatore sono rispettivamente le derivate di ordine n e $n+1$ della funzione $P(0; t)$ espressa dalla [3.1]. Costruiamo ora analiticamente la funzione [3.9], osservando che nel nostro modello si ha:

$$P(0; t) = \frac{\gamma^\alpha}{\Gamma(\alpha)} \int_0^\infty e^{-\lambda(\gamma+t)} \lambda^{\alpha-1} d\lambda \quad [3.10]$$

mentre la derivata di ordine n della [3.10] assume la forma:

$$P^{(n)}(0; t) = \frac{\gamma^\alpha}{\Gamma(\alpha)} \int_0^\infty (-\lambda)^n e^{-\lambda(\gamma+t)} \lambda^{\alpha-1} d\lambda \quad [3.11]$$

ottenendo in definitiva:

$$P^{(n)}(0; t) = \frac{\gamma^\alpha}{\Gamma(\alpha)} (-1)^n \frac{\Gamma(\alpha+n)}{(t+\alpha)^{\alpha+n}} \quad [3.12]$$

e quindi:

⁴ Si veda Ove Lundberg "On Random Processes and their Application to Sickness and Accident Statistics" Uppsala, 1964 – pag. 85

$$\mu_n(t) = \frac{\alpha + n}{t + \gamma} \quad [3.13]$$

Le medie condizionate del modello possono essere espresse in funzione delle intensità, in base alla formula:

$$E(Z/x; t' - t) = (t' - t) \mu_x(t) \quad [3.14]$$

che ci dà la media condizionata della variabile numero dei sinistri ($z = y - x$) per il periodo $t' - t$, nell'ipotesi che nel periodo $0 - t$ (con $t < t'$) si siano avuti x sinistri. Su un periodo di due anni le medie condizionate coincidono con l'intensità del primo periodo. Infatti si ha:

$$E(Z/x; 1) = (2 - 1) \mu_x(1) = \frac{\alpha + x}{1 + \gamma} \quad [3.15]$$

La relazione [3.15] può risultare molto utile per risolvere i problemi di stima dei parametri.

Esempio numerico

Supponiamo che i dati statistici in nostro possesso ci portino ad esprimere le seguenti stime:

$$E(X) = 0,20, \quad \sigma^2 = 0,25 \quad \text{e pertanto } \alpha = 0,80 \quad \gamma = 4$$

In base a queste ipotesi otterremo

Tavola 3.1. Applicazione del modello teorico ai dati di un periodo (anno)
 $\alpha = 0,80 ; \gamma = 4$

Numero sinistri anno 0	Probabilità teoriche nell'anno
0	0,8365
1	0,1338
2	0,0241
3	0,0045
4	0,0009
5	0,0002

Calcolo delle probabilità teoriche su due periodi

Sulla base dei parametri stimati è possibile classificare le autovetture in base ai sinistri avvenuti nei due periodi presi in considerazione, calcolarne la frequenza sui due anni e applicare il nostro modello teorico in base alla formula:

$$P(y; 2) = \left(\frac{\gamma}{2 + \gamma} \right)^\alpha \left(\frac{2}{2 + \gamma} \right)^y \frac{\Gamma(y + \alpha)}{\Gamma(\alpha) y!} \quad [3.16]$$

Otterremo

Tavola 3.2. Applicazione del modello teorico al modello di parametri
 $\alpha = 0,80 ; \gamma = 4$

Numero sinistri nei due esercizi	Probabilità teoriche in base alla [3.16]
0	0,72298
1	0,19279
2	0,05784
3	0,01799
4	0,00570
5	0,00182
6	0,00059
7	0,00019
8	0,00006
9	0,00002

Siamo ora in grado di applicare la formula della probabilità condizionata, cioè

$$P_{x,y} = \frac{y!}{(y-x)! x!} \left(\frac{1}{2}\right)^y \frac{P(y;2)}{P(x;1)} \quad [3.17]$$

in cui y rappresenta il numero dei sinistri nei due periodi e x il numero dei sinistri nel primo periodo. $P(y; 2)$ e $P(x; 1)$ sono già stati calcolati rispettivamente nelle tavole precedenti. I dettagli per il calcolo delle probabilità di passaggio dallo stadio iniziale a quello del secondo periodo sulla base della formula [3.17] sono compendiate nella tavola seguente.

Tavola 3.3. Applicazione del modello teorico al modello di parametri
 $\alpha = 0,80 ; \gamma = 4$

a) Transizione dallo stadio 0; $x = 0$; $P(0; 1) = 0,83651$

y	$y - x$	$\frac{y!}{(y-x)! x!}$	$P(y; 2)$
0	0	1	0,86428
1	1	1	0,11524
2	2	1	0,01729
3	3	1	0,00269
4	4	1	0,00043
5	5	1	0,00007
6	6	1	0,00001
7	7	1	0,00000
8	8	1	0,00000
9	9	1	0,00000

b) Transizione dallo stadio 1; $x = 1$; $P(1; 1) = 0,13384$

y	$y - x$	$\frac{y!}{(y-x)! x!}$	$P(y; 2)$	$P_{x,y}$
1	0	1	0,19279	0,72023
2	1	2	0,05784	0,21607
3	2	3	0,01799	0,05042
4	3	4	0,00570	0,01064
5	4	5	0,00182	0,00213
6	5	6	0,00059	0,00041
7	6	7	0,00019	0,00008
8	7	8	0,00006	0,00001

c) Transizione dallo stadio 2; $x = 2$; $P(2; 1) = 0,02409$

y	$y - x$	$\frac{y!}{(y-x)! x!}$	$P(y; 2)$	$P_{x,y}$
2	0	1	0,05784	0,60020
3	1	3	0,01799	0,28009
4	2	6	0,00570	0,08870
5	3	10	0,00182	0,02365
6	4	15	0,00059	0,00572
7	5	21	0,00019	0,00130

Analoghi calcoli possono essere effettuati per le transizioni dallo stadio 3, 4, 5. In definitiva si ottiene la matrice di transizione così rappresentata:

Tavola 3.4. Matrice di transizione costruita in base al modello teorico con $\alpha = 0,80$; $\gamma = 4$

Stato Iniziale Sinistri	Stato finale dei sinistri					
	0 Sinistri	1 Sinistro	2 sinistri	3 sinistri	4 sinistri	5 e più sinistri
0	0,8643	0,1152	0,0173	0,0027	0,0004	0,0001
1	0,7202	0,2161	0,0504	0,0106	0,0021	0,0005
2	0,6002	0,2801	0,0887	0,0237	0,0057	0,0017
3	0,5002	0,3168	0,1267	0,0408	0,0116	0,0040
4	0,4168	0,3334	0,1612	0,0609	0,0198	0,0079
5	0,3473	0,3358	0,1903	0,0824	0,0302	0,0140

3.2. Indicazioni sull'uso del modello

Evidentemente la matrice delle frequenze, rilevata su base campionaria o sull'intero portafoglio, dovrebbe costituire la materia prima per l'impiego del modello. Tuttavia può avvenire che non si disponga di alcuna indagine in proposito o che l'indagine sia inattendibile o incompleta. Anche in questi casi è possibile impiegare il modello con opportune ipotesi ed accorgimenti: ovviamente i risultati che si otterranno dipenderanno largamente dalla

plausibilità delle ipotesi di base. Circa gli accorgimenti ci proponiamo di esaminare i seguenti casi concreti:

- a) mancanza (o inattendibilità) della matrice delle frequenze;
- b) disponibilità di una matrice incompleta;
- c) disponibilità di una matrice completa;
- d) disponibilità di un'indagine su più periodi.

a) Mancanza della matrice delle frequenze

Nel caso più disperante, in cui non si disponga di una matrice delle frequenze, occorre ipotizzare valori plausibili della media e della varianza e costruire i valori teorici. Va rilevato che in genere il valore medio (cioè la frequenza media del portafoglio) è facilmente desumibile, mentre per la varianza occorre formulare valide ipotesi tenendo presente che al crescere della varianza si ottengono distribuzioni meno concentrate.

b) Disponibilità di una matrice incompleta

A volte il fenomeno della propensione a provocare sinistri è esaminato sul solo portafoglio sinistri ottenendo una matrice di transizione incompleta, cioè mancante della probabilità di passaggio dallo stadio iniziale 0. In questo caso la stima dei parametri può essere effettuata sui dati statici in base alla [3.15], cioè considerando i valori medi subordinati. Tuttavia i valori desunti dalla [3.15] devono essere congrui rispetto ai dati desumibili dal portafoglio, quali la frequenza media.

Supponiamo ad esempio di avere a disposizione una matrice di frequenze, ma incompleta in quanto mancante della prima riga. Le medie dello stadio 2 e 3 sono rispettivamente 27,6 e 69,0 (per 100). In base alla [3.15] potremmo scrivere

$$\frac{\alpha + 1}{1 + \gamma} = 0.276 \quad \text{e} \quad \frac{\alpha + 2}{1 + \gamma} = 0.690.$$

La soluzione di queste due equazioni porta a valori di α e γ negativi e quindi non compatibili con il modello. In questo caso, ovviamente il risultato può dipendere dalla scarsa attendibilità di medie desunte da un numero ridotto di casi. Occorrerà pertanto formulare ipotesi senza fare ricorso a dati statistici.

c) Disponibilità di una matrice completa

Se si dispone di una matrice completa ed attendibile, si hanno varie possibilità di stima dei parametri. Oltre al confronto delle medie nei due periodi di osservazione, si possono prendere in considerazione le relative varianze o ricorrere ad altri metodi di interpolazione. Ad esempio i parametri potrebbero essere calcolati prendendo in considerazione la media e la varianza nei due anni di osservazione, in base alle formule:

$$E(Y; 2) = \frac{2\alpha}{\gamma}$$
$$\sigma^2(Y; 2) = \frac{2\alpha}{\gamma} \frac{2 + \gamma}{\gamma}$$

d) Disponibilità di un'indagine su più periodi

Qualora si disponga dei dati relativi ad un'indagine campionaria condotta sugli stessi rischi per più anni d'osservazione, si può provare sperimentalmente la validità delle ipotesi di base del modello e ricercare un'espressione analitica della funzione di struttura (cioè di $f(\lambda)$).

Bibliografia essenziale

- P. Thyron "Nombre de Sinistres en Assurance Automobile – ASTIN Bulletin VI – 3-1972
- O. Lundberg "On Random Processes and their Application to Sickness and Accident Statistics" Uppsala, 1964

PARTE IV

PROCESSO DEL COSTO DEI SINISTRI

4.1. Premessa

L'analisi del processo del costo dei sinistri consiste nello stimare la funzione di distribuzione di un sinistro individuale sulla base delle informazioni contenute nella polizza. La funzione è basilare per l'applicazione della teoria del rischio e per tutte le decisioni riguardanti la franchigia, il massimale, i caricamenti e le strategie riassicurative.

Per il processo del costo del sinistro sono state, a volte, formulate ipotesi elementari largamente accettabili e verificabili su dati concreti. L'ipotesi che viene qui proposta, nel caso specifico del ramo incendi, è modellata sul processo di crescita studiato nelle scienze biologiche¹ e variamente applicato nelle teorie economiche⁵. Poiché tale ipotesi non è sempre applicabile a casi concreti occorrerà valutare soluzioni alternative.

4.2. Una introduzione alla distribuzione log-normale

Nel trattare fenomeni di biologia H. Cramér⁶ afferma: supponiamo di essere interessati allo studio della variabile casuale "dimensione di un specifico organismo" raggiunta al momento attuale in un particolare individuo. Essa può essere considerata come l'effetto congiunto di un gran numero di cause mutuamente indipendenti, agenti in una sequenza ordinata durante il tempo di crescita dell'individuo. Se queste cause semplicemente aggiungono i loro effetti (effetti che presumiamo siano variabili causali), in base al teorema del limite centrale, possiamo affermare che la loro somma è asintoticamente distribuita normalmente.

In perfetta analogia, nel ramo incendi⁷ possiamo ammettere che la dimensione dei sinistri sia la risultante di un gran numero di cause ad effetti positivi. Si può facilmente immaginare qualcuna di queste cause che consentono ad esempio l'espansione di un incendio: condizioni atmosferiche particolari, distanza o inefficienza dei mezzi di prevenzione, ora di avvenimento del sinistro, ecc. Anche in questo caso possiamo ipotizzare che la somma degli effetti sia distribuita normalmente.

In generale non sembra essere plausibile ipotizzare che le cause contribuiscano in maniera additiva. Sembra più naturale supporre che ciascuna causa produca un impulso, l'effetto del quale dipende sia dalla forza dell'impulso che dalla dimensione raggiunta dall'organismo o dall'incendio al momento in cui l'impulso agisce.

Supponiamo di avere n impulsi $\xi_1, \xi_2, \dots, \xi_n$ che agiscono nell'ordine indicato dai loro indici. Questi impulsi possono essere considerati variabili causali. Se indichiamo con x la dimensione prodotta dagli impulsi $\xi_1, \xi_2, \dots, \xi_n$, possiamo ad esempio supporre che l'incremento causato dall'impulso ξ_{n+1} sia proporzionale a ξ_{n+1} a qualche funzione $g(x)$ della dimensione momentanea dell'organismo

$$x_{n+1} = x_n + \xi_{n+1} g(x_n)$$

⁵ Un testo esauriente a tale proposito è J. Aitchisson & J.C. Brown "The Lognormal Distribution" University Press Cambridge

⁶ Harald Cramér – Mathematical Methods in Statistics – Princeton 1974

⁷ Giovanna Ferrara – La distribution des sinistres Incendie selon leur coût ASTIN Bulletin – Vol. VI

Ne segue che

$$\xi_1 + \xi_2 + \dots + \xi_n = \sum_0^{n-1} \frac{x_{v+1} - x_v}{g(x_v)}$$

Se poi ciascun impulso fornisce solo un modesto contributo alla crescita dell'organismo o all'espansione dell'incendio avremmo approssimativamente:

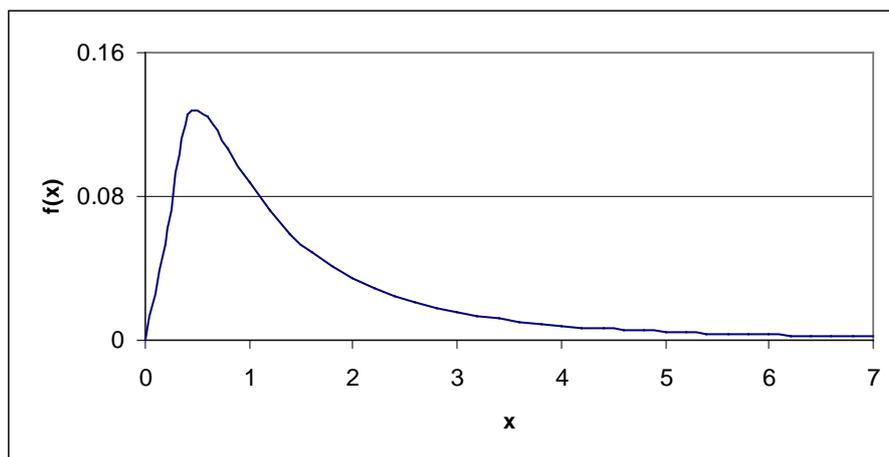
$$\xi_1 + \xi_2 + \dots + \xi_n = \int_{x_0}^x \frac{d(t)}{g(t)}$$

dove $x = x_n$ indica la dimensione finale dell'organismo o dell'incendio. Per ipotesi $\xi_1, \xi_2, \dots, \xi_n$ sono variabili indipendenti e n può essere considerato un numero grande. Sotto le condizioni di validità del teorema del limite centrale ne segue che, al limite, la funzione della variabile casuale x che figura al secondo membro dell'equazione precedente è distribuita normalmente.

Consideriamo ora il caso $g(t) = t$, cioè quando l'effetto di ciascun impulso è direttamente proporzionale alla dimensione momentanea dell'organismo. In questo caso troviamo che $\ln g(x)$ è distribuita normalmente. Se più generalmente $\ln(x - c)$ è una variabile normale (m, σ) è facilmente deducibile che la variabile x ha come funzione di frequenza:

$$f(x) = \frac{1}{\sigma(x-c)\sqrt{2\pi}} e^{-\frac{(\ln(x-c)-m)^2}{2\sigma^2}} \quad [4.1]$$

con $x > c$ mentre per $x \leq c$ la funzione di densità è zero.



Funzione di frequenza di una variabile lognormale con $m = 0$ e $\sigma = 2$ ”

Determinazione dei parametri

La determinazione dei parametri può essere effettuata considerando la relazione fra le variabili Z e X , cioè la

$$Z = a \ln (X - c) + b \quad [4.2]$$

Poiché Z è una variabile casuale normale, la funzione di distribuzione (cioè le frequenze dei valori di Z inferiori ad z_i) è espressa da

$$\frac{1}{\sqrt{\pi}} \int_{-\infty}^{z_i} \exp \{ - z^2 \} dz \quad [4.3]$$

Se $P(x_i)$ indica la funzione di distribuzione di X (cioè le frequenze dei valori di X inferiori ad x_i) si può definire :

$$P(x_i) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{z_i} \exp \{ - z^2 \} dz \quad [4.4]$$

Dopo aver determinato i valori di z_i , occorre calcolare i parametri a, c, b in base alla relazione [4.2], che può essere posta nella forma:

$$z_i = a \ln (x_i - c) + b \quad [4.5]$$

Il parametro c che si presenta in forma non lineare può essere determinato per tentativi, partendo dalla quantità c data dall'uguaglianza che riguarda le prime tre classi di sinistri:

$$\frac{z_3 - z_2}{z_2 - z_1} = \frac{\ln(x_3 - c) - \ln(x_2 - c)}{\ln(x_2 - c) - \ln(x_1 - c)} \quad [4.6].$$

Stabiliti quindi i valori di z_i tramite la relazione [4.4] e attraverso l'utilizzo della [4.3], si ottengono i valori teorici $P^*(x_i)$ e $f^*(x_i)$.

Per esprimere un giudizio sul grado di accostamento, si possono calcolare gli scostamenti assoluti tra i valori teorici e quelli empirici, cioè determinando lo scarto medio

$$I = \sum_{i=1}^n | f(x_i) - f^*(x_i) |$$

e calcolare l'indice χ^2 . Per meglio visualizzare l'intero processo d'interpolazione, forniamo ora un esempio numerico con dati tratti da una statistica del ramo incendi (Tavola 4.1).

Nella Tavola 4.1 l'interpolazione dei valori z_i è stata calcolata mediante il metodo dei minimi quadrati mediante la relazione

$$z_i = a \ln (x_i - c) + b$$

I valori dei parametri sono

$a = 0,4179$, $b = -2,4455$ e per semplicità si è posto in questo primo esempio $c = 0$.

Il coefficiente di correlazione è pari a $r = 0,9968$

Come si può notare, si è ottenuto un buon grado di accostamento e, quindi, l'ipotesi alla base del processo sembra confermata dai dati empirici.

Tavola 4.1 - Esempio di interpolazione nel ramo incendi – modello lognormale

Estremo superiore della classe	Valori Osservati			Valori Teorici		
	numero sinistri					numero sinistri
X	$n(x)$	$P(x)$	$Z(x)$	$z^*(x)$	$P^*(x)$	$n^*(x)$
250	4.346	0,4346	-0,1647	-0,1383	0,4450	4.450
500	1.231	0,5577	0,1451	0,1513	0,5601	1.151
1000	1.423	0,7000	0,5244	0,4409	0,6704	1.102
2000	846	0,7846	0,7878	0,7306	0,7675	971
4000	462	0,8308	0,9573	1,0202	0,8462	787
8000	692	0,9000	1,2816	1,3098	0,9049	587
16000	346	0,9346	1,5110	1,5995	0,9451	403
32000	346	0,9692	1,8692	1,8891	0,9706	254
64000	192	0,9884	2,2701	2,1787	0,9853	148
Oltre	116	1,0000			1,0000	147
Totale	10.000					10.000

$$a = 0,4179 \quad b = -2,4455, \quad c = 0.$$

$$r = 0.9968$$

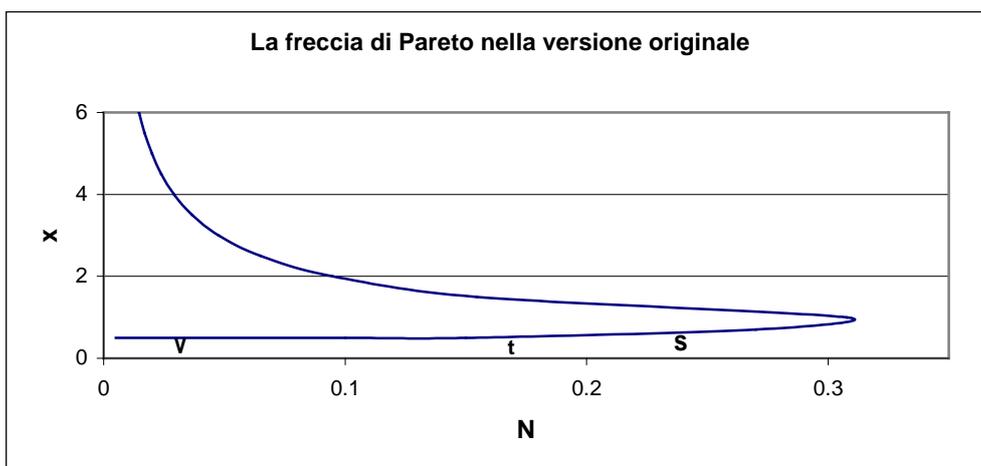
4.3. La distribuzione di Pareto

Come cercherò di mostrare in seguito, la distribuzione di Pareto ha un ruolo importante nella definizione della distribuzione della grandezza dei sinistri. Quindi è importante considerare cosa ha detto veramente il suo autore.

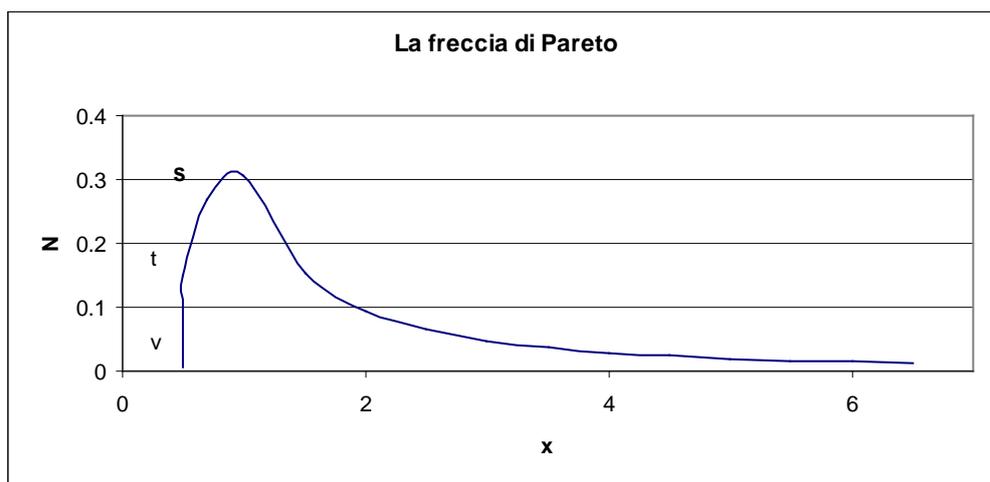
Nel suo corso di Economia Politica⁸, V. Pareto si interessa della distribuzione della ricchezza e dice: “la ripartizione della ricchezza può dipendere dalla natura degli uomini di cui la società si compone, dall’organizzazione di quest’ultima ed anche in parte dal caso, cioè da quel complesso di cause ignote, agenti ora in senso ora in altro, che, nell’ignoranza circa la loro vera natura, designiamo con l’espressione di “caso”. E’ l’osservazione che ci deve informare in merito alla parte che effettivamente hanno tali cause nella ripartizione della ricchezza”. Il suo scopo era evidentemente quello di dimostrare, sulla base di osservazioni, che la ripartizione della ricchezza dipenda dalla natura dell’uomo. Ma quale era la distribuzione che Pareto prospettava?

“Si discorre spesso della piramide sociale di cui i poveri costituiscono la base, i ricchi la cima. A dire il vero, non è già d’una piramide che si tratta, ma invece d’un corpo che ha la forma della punta d’una freccia, della punta di una trottola.” La rappresentazione è la seguente:

⁸ Vilfredo Pareto – Economia Politica – UTET 1973



in cui N il numero (relativo) dei possessori di un reddito superiore a x (cioè $1-F(x)$ con $F(x)$ funzione di ripartizione. La stessa distribuzione invertendo gli assi (nella maniera con cui in genere si rappresenta) è:



V. Pareto contesta la possibilità di interpolare la curva con una distribuzione normale e trascurando la base della sua trottola (cioè il tratto indicato con vt) fornisce come descrizione del tratto rimanente della curva l'espressione seguente:

$$\ln N = \ln A - \alpha \ln(a + x) - \beta x ;$$

cioè

$$N = \frac{A}{(x+a)^\alpha} e^{-\beta x} \quad [4.7];$$

se β è molto prossimo allo zero (come si verifica nella maggior parte dei dati da lui interpolati) si avrà:

$$N = \frac{A}{(x+a)^\alpha} \quad [4.8];$$

e per il valore di a prossimo allo zero si avrà:

$$N = \frac{A}{x^\alpha} \quad [4.9].$$

E' questa la forma che noi (da questa parte dell'Atlantico) usiamo, definendo come funzione di ripartizione di Pareto, la quantità:

$$F(x) = 1 - \left(\frac{x_0}{x}\right)^\alpha \quad [4.10]$$

con funzione di densità:

$$f(x) = \alpha \frac{x_0^\alpha}{x^{\alpha+1}} \quad [4.11];$$

e media eguale a
$$E(X/\alpha) = \frac{x_0 \alpha}{\alpha - 1} \quad (\text{con } \alpha > 1)$$

e varianza
$$\sigma^2(X/\alpha) = \frac{\alpha x_0^2}{(\alpha - 1)^2 (\alpha - 2)} \quad (\text{con } \alpha > 2).$$

Mentre dall'altra parte dell'Oceano si preferisce la forma:

$$F(x) = 1 - \left(\frac{x_0}{x + x_0}\right)^\alpha \quad [4.12];$$

con funzione di densità:

$$f(x) = \frac{\alpha x_0^\alpha}{(x + x_0)^{\alpha+1}} \quad [4.13].$$

V. Pareto, da sociologo e economista dell'Ottocento, era interessato a fornire delle spiegazioni ad un problema corroborandole con interpolazione dei dati reali. Sorge spontanea una domanda: esiste un'ipotesi probabilistica su cui fondare la funzione di Pareto?

Nel giornale dell'associazione svizzera degli attuari un collega belga, J.F. Walhin⁹, ne offre una interpretazione suggestiva. Supponiamo che ciascun fenomeno elementare sia distribuito secondo una legge esponenziale di parametro θ , cioè si abbia una funzione di densità del tipo

$$f(x/\theta) = \theta e^{-\theta x} \quad [4.14]$$

e che nell'ambito del collettivo i vari eventi elementari siano caratterizzati da propri θ distribuiti secondo una Gamma di parametri A e α , cioè

$$g(\theta) = \frac{A^\alpha \theta^{\alpha-1} e^{-A\theta}}{\Gamma(\alpha)} \quad [4.15]$$

la funzione risultante sarà:

$$h(x) = \int_0^\infty \theta e^{-\theta x} g(\theta) d\theta = \frac{\alpha A^\alpha}{(A+x)^{\alpha+1}}; \quad x > 0 \quad [4.16]$$

⁹ J.F. Walhin - Une nouvelle caractérisation de la distribution de Pareto, avec application à la cadence de paiement du réassureur en excédent de sinistre. ASA 2003/2

In sostanza la mistura di una esponenziale con una Gamma conduce proprio alla distribuzione di Pareto (nella formulazione di oltre Oceano, cioè con una funzione di distribuzione definita dalla [4.16]. E' interessante notare che la legge di Pareto è applicata non già alla dimensione dei danni, ma alla cadenza dei pagamenti di un riassicuratore per eccesso sinistri.

Determinazione dei parametri

Per la determinazione dei parametri partiamo dall'espressione:

$$F(x) = 1 - \left(\frac{c}{x}\right)^\alpha \quad [4.17]$$

con $c < x < \infty$.

L'interpolazione dei dati mediante la funzione di Pareto è agevolata dalla relazione:

$$H(x) = 1 - F(x) = \left(\frac{c}{x}\right)^\alpha \quad [4.18]$$

cioè da:

$$\ln H(x) = \alpha \ln c - \alpha \ln x \quad [4.19]$$

Mediante il metodo dei minimi quadrati è possibile determinare i parametri c e α .

Il valor medio della variabile x è dato da:

$$E(x) = \frac{c\alpha}{\alpha - 1} \quad [4.20] \quad \text{con } \alpha > 1$$

$$\text{Per la varianza si ha: } \text{var}(x) = c^2 \frac{\alpha}{(\alpha - 1)^2 (\alpha - 2)} \quad (\text{con } \alpha > 2) \quad [4.21]$$

Riportiamo ora un esempio d'interpolazione mediante la distribuzione di Pareto

Tavola 4.2. Esempio d'interpolazione con la funzione di Pareto

Classi D'importo	Valori effettivi		Valori teorici		Valori per χ^2
	Numero di sinistri	Funzione di distribuzione $F(x)$	Numero di sinistri	Funzione di distribuzione $F(x)$	d_i
da 0 a 1	300	0,300	222	0,222	27,4
da 1 a 1,5	400	0,700	433	0,655	2,6
da 1,5 a 2	100	0,800	151	0,807	17,4
da 2 a 2,5	50	0,850	70	0,876	5,6
da 2,5 a 3	50	0,900	38	0,914	3,9
da 3 a 4	50	0,950	38	0,952	4,1
da 4 a 8	40	0,990	36	0,988	0,4
Da 8 a 12	10	1,000	12	1,000	0,3
Totali	1.000		1.000		61,7

4.4. La distribuzione Gamma

Nei paragrafi precedenti abbiamo visto delle applicazioni della funzione lognormale e della Pareto ai processi del costo dei sinistri. Evidentemente esistono svariate funzioni che possono essere utilmente impiegate e la funzione Gamma offre notevoli vantaggi. La funzione di distribuzione della funzione Gamma è data da:

$$F(x) = \int_0^x \frac{1}{\beta \Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) dx \quad [4.22]$$

con $\Gamma(\alpha) = \int_0^{\infty} \exp(-u) u^{\alpha-1} du$

La funzione di densità è data da:

$$f(x) dx = \frac{1}{\beta \Gamma(\alpha)} \left(\frac{x}{\beta}\right)^{\alpha-1} \exp\left(-\frac{x}{\beta}\right) dx \quad [4.23]$$

Per la media e la varianza si ha

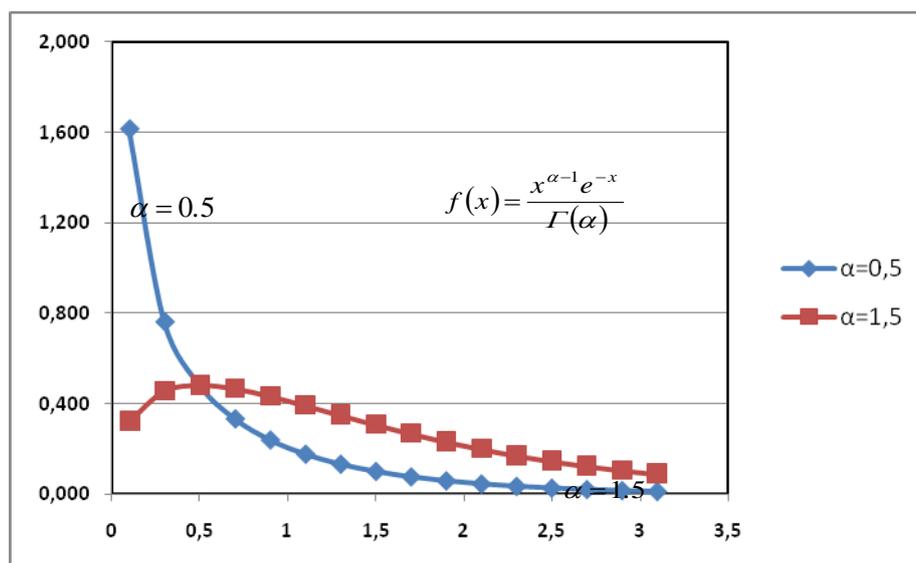
$$\mu_1 = \beta \alpha \qquad \sigma^2 = \beta^2 \alpha$$

Il coefficiente di asimmetria si pone nella forma

$$\gamma = 2 \alpha^{-1/2}$$

Il valore di α della funzione Gamma è un parametro di scala e fornisce la forma della distribuzione: se $\alpha < 1$ la funzione ha il valore modale nel punto 0 e presenta un tratto rapidamente discendente verso l'asse delle ascisse. Per $\alpha \geq 1$ la distribuzione ha un valore modale nel punto $b(\alpha - 1)$. L'andamento della funzione è mostrato nel grafico che segue.

**Rappresentazione della funzione con Gamma
 per due differenti valori di α e con $\beta=1$**



La determinazione dei parametri viene effettuata mediante il metodo dei momenti calcolati sui dati statistici. In pratica si ottiene:

$$\beta = \frac{\text{Varianza}}{\text{Media}}, \quad \alpha = \frac{\text{Media}}{\beta}$$

Un esempio di interpolazione mediante la funzione Gamma viene mostrato nella tavola che segue (Tavola 5). Nella tavola 6 i valori ottenuti sono messi a confronto con i risultati desunti dalla funzione lognormale.

Tavola 4.3. Interpolazione con una funzione Gamma

Estremo superiore della classe	Valori effettivi		Valori teorici		Valori per χ^2 d_i
	Numero sinistri $n(x)$	Distribuzione $F(x)$	Distribuzione Gamma	Numero sinistri $n^*(x)$	
0,25	4.346	0,4346	0,6195	6.195,4	552,1
0,50	1.231	0,5577	0,6655	460,0	1.292,6
1,00	1.423	0,7000	0,7145	490,0	1.776,6
2,00	846	0,7846	0,7663	517,3	208,8
4,00	462	0,8308	0,8199	536,6	10,4
8,00	692	0,9000	0,8736	537,0	44,7
16,00	346	0,9346	0,9237	500,8	47,9
32,00	346	0,9692	0,9643	406,2	8,9
64,00	192	0,9884	0,9894	251,0	13,9
oltre	116	1,0000	1,0000	105,7	1,0
Totale	10.000			10.000,0	3.956,8

4.5. La scelta di una conveniente funzione di distribuzione

Una funzione che descrive il processo di costo dei sinistri dovrebbe avere le seguenti caratteristiche:

- Definire le probabilità per ogni numero reale non negativo. Tuttavia, a fini pratici, alcune dovrebbero avere il valore modale nel punto 0
- Definire le probabilità in modo continuo.
- Alcune dovrebbero avere una “coda” moderata, altre una “coda” molto prolungata

La prima caratteristica è dovuta alla circostanza che in molti processi di costo non esiste un danno massimo possibile anche se, nella realtà, è sempre possibile fissare un massimo ammontare. Circa i valori minimi, vi sono delle distribuzioni, quale la funzione di Pareto, che non partono dal valore 0. Se gli esborsi inferiori ad un certo ammontare non sono importanti, è opportuno fare riferimento a queste funzioni .

La continuità è senza dubbio una caratteristica opportuna, in quanto permette di definire valori di franchigia o di cesura, che non sempre sono palesi nei dati concreti.

Infine la forma della “coda” destra della distribuzione è di importanza cruciale. In effetti se la possibilità di sinistri d’importo rilevante fosse remota, forse non ci sarebbe alcun interesse di ricorrere all’assicurazione o alla riassicurazione. D’altra parte, nel mondo odierno, i sinistri d’importo rilevante si producono spesso e, quindi, è opportuno studiarne gli aspetti tecnici. A questo riguardo, per tutte le funzioni definite fino all’infinito, è interessante valutare con

quale velocità la funzione di densità si riduce a 0 quando il danno diventa infinitamente grande.

4.6. Criteri di scelta

La funzione raffigurante la dimensione dei sinistri assume un'importanza basilare nel campo assicurativo. Sono evidenti i suoi usi per tutte le questioni riguardanti il calcolo dei premi (franchigia, massimali e caricamenti) e le strategie riassicurative. Tuttavia la sua importanza è fondamentale nell'ambito della teoria del rischio.

Nel semplice modello ideato da Filip Lundberg nel 1903 e sviluppato da H. Cramér e dalla scuola di Stoccolma occorre essenzialmente definire il processo del numero dei sinistri - $N(t)$ - e quello dell'ammontare dei sinistri $S(t)$ al variare del tempo t . Definire un processo significa non solo ipotizzarne la distribuzione, ma anche calcolarne le caratteristiche principali (cioè i momenti e le proprietà asintotiche al variare di t). I procedimenti di simulazione non riducono affatto la necessità di questi approfondimenti: infatti è buona norma che il procedimento di simulazione abbia buoni basi teoriche.

Sul processo caratterizzante il numero dei sinistri vi sono specifiche indicazioni sulle distribuzioni da impiegare. Ma quali sono i criteri per definire la distribuzione della dimensione? Esistono dei metodi per classificare le varie distribuzioni? Una risposta a questa domanda è stata fornita nel 1960 da un articolo di G. Benktander e C.O. Segerdhal¹⁰, che si proponevano di valutare con quale velocità la funzione di densità di una distribuzione della dimensione del danno si riduce a 0 quando il danno diventa infinitamente grande: è questo un problema tipico della valutazione degli eccessi sinistri.

Per studiare tale fenomeno gli autori hanno proposto due quantità: la prima è data dalla funzione che rappresenta il costo medio del sinistro a carico dell'eccesso; la seconda ne misura la varianza. Circa la prima quantità, il costo medio dei sinistri a carico dell'eccesso, relativamente ad una priorità x , è dato da:

$$m(x) = \frac{\int_x^{\infty} (z-x) dP(z)}{\int_x^{\infty} dP(z)} = \frac{\int_x^{\infty} (z-x)p(z) dz}{\int_x^{\infty} p(z) dz}, \quad [4.24]$$

in cui $P(z)$ è la funzione di distribuzione della variabile casuale dimensione del sinistro - z - e $p(z)$ la funzione di densità. Analogamente la varianza del costo dei sinistri a carico dell'eccesso è data dall'espressione:

$$s(x) = \frac{\int_x^{\infty} (z-x)^2 dP(z)}{\int_x^{\infty} dP(z)} = \frac{\int_x^{\infty} (z-x)^2 p(z) dz}{\int_x^{\infty} p(z) dz}$$

I valori di queste quantità sono indipendenti dai parametri caratterizzanti la funzione della dimensione dei sinistri; sono facilmente calcolabili per molte distribuzioni e

¹⁰ G. Benktander – C.O. Segerdhal - On the analytical representation of claim distribution with special reference to excess of loss reinsurance – ICA 1960

evidenziano i danni d'importo rilevante. Inoltre le quantità sono strettamente legate al premio puro dell'eccesso; sono anche adatte ad un confronto fra distribuzioni calcolate mediante modelli teorici e, quindi, utili per classificare questi modelli.

L'articolo di G. Benktander e C.O. Segerdhal è molto ben articolato e rigoroso e merita un'attenta lettura: ovviamente in questa esposizione cercherò di evidenziarne i punti che ritengo principali. Gli autori calcolano il valore dell'indice $m(x)$ per alcune distribuzioni usate per descrivere la dimensione dei sinistri (coda destra della normale, esponenziale, Gamma e Pareto) e trovano risultati interessanti che permettono di classificare i vari modelli a seconda della "pericolosità".

Se la funzione del danno è della forma esponenziale, cioè $p(x) = e^{-kx}$ si avrà:

$$m(x) = \frac{\int_x^\infty (z-x)e^{-kz} dz}{\int_x^\infty e^{-kz} dz} = \frac{\int_{kx}^\infty \frac{u}{k} \cdot e^{-u} \cdot \frac{du}{k}}{\frac{1}{k} \cdot e^{-kx}} - x =$$

$$= \frac{kx \cdot e^{-kx} + e^{-kx}}{k \cdot e^{-kx}} - x = \frac{1}{k}$$

Questo significa che al variare della priorità x il costo medio dell'eccesso è sempre costante. Il risultato non è nuovo: era già noto nella teoria della "sopravvivenza" e va sotto il nome di "mancanza di memoria" della distribuzione esponenziale.

Nella distribuzione di Pareto in cui $p(x) = c \cdot x^{-k-1}$, quindi per $k > 2$ (la condizione $k > 2$ serve a definire anche la varianza del costo dell'eccesso: ai fini della validità della formula bastava porre $k > 1$) si avrà:

$$m(x) = \frac{\int_x^\infty (z-x)z^{-k-1} dz}{\int_x^\infty z^{-k-1} dz} = \frac{\frac{1}{k-1} \cdot x^{1-k}}{\frac{1}{k} \cdot x^{-k}} - x = \frac{x}{k-1}$$

In questo caso il costo medio dell'eccesso si incrementa nella stessa misura in cui si incrementa la priorità.

Considerando per $p(x)$ altre funzioni, quali la normale o la gamma, i valori di $m(x)$ presentano andamenti diversi. Nel caso della funzione normale $m(x)$ ha un andamento prossimo a quello di un'iperbole decrescente; per la distribuzione Gamma è un'iperbole crescente, mentre come abbiamo visto nel caso dell'esponenziale la funzione è costante e per la Pareto presenta una crescita lineare.

Tutte le curve di $m(x)$ sono, al variare di x , crescenti con incrementi decrescenti cioè si ha:

$$m(x) > 0 \quad m'(x) \geq 0 \quad m''(x) \leq 0$$

In generale si constata che la funzione $m(x)$ è compresa tra quella dell'esponenziale generalizzata e quella corrispondente alla distribuzione di Pareto. E' quindi interessante esaminare il caso della funzione:

$$m(x) = \frac{1}{k} \cdot x^p \quad \text{con } 0 \leq p \leq 1$$

Tale funzione per $p = 1$ rappresenta il costo medio dell'eccesso nel caso paretiano (e quindi di massima asimmetria), mentre per $p = 0$ quello corrispondente all'esponenziale generalizzata.

Gli autori non limitano la loro attenzione solo alla funzione $m(x)$ o $s(x)$ ma danno anche l'espressione del premio di eccesso sinistri. D'altra parte, poiché nella prassi le riassicurazioni d'eccesso sinistri per un massimale illimitato sono una eccezione, viene introdotta accanto alla soglia minima la considerazione di una soglia massima (massimale). L'analisi condotta sulle distribuzioni tronche (e nella nostra esposizione limitata ad alcuni risultati del valore medio dell'eccesso) porta a definire:

$$m(x, M) = \frac{\int_x^M H(z) dz}{H(x)}.$$

In un successivo articolo¹¹ G. Benktander introduce nell'analisi un ulteriore indice: il tasso istantaneo $\mu(x) dx$ rappresenta la probabilità che una variabile casuale, pari almeno a x , non ecceda $x + dx$, ossia la probabilità che un sinistro, che "sopravvive" a x , "muoia" tra $(x, x + dx)$. Più basso è questo tasso di "mortalità" del danno, più asimmetrica e più pericolosa è la distribuzione dei sinistri.

Per definire questo tasso istantaneo indica con $H(x)$ il numero medio dei sinistri il cui importo sia superiore a x :

$$H(x) = 1 - P(x) = \text{prob} \{ X > x \}$$

Tralasciando tutti i passaggi intermedi, ottiene in definitiva:

$$\mu(x) = \frac{m'(x) + 1}{m(x)} \quad \text{e} \quad H(x) = e^{-\int_0^x \mu(z) dz}.$$

Sulla base della considerazione del tasso istantaneo di estinzione, si suppone che la variabile x non sia limitata, cioè che si abbia $H(x) > 0$ per ogni valore finito di x . Poiché è $H(\infty) = 0$ ne discende che

$$\int_x^\infty \mu(t) dt \text{ deve convergere.}$$

Se si pone $\mu(x)$ nella forma $\mu(x) = k \cdot x^p$, la condizione è soddisfatta per tutti i valori di $p \geq -1$. Se $p > -1$, le corrispondenti distribuzioni sono di tipo esponenziale e per

¹¹ G. Benktander –A note on the most "dangerous" and skewest class of distribution – ASTIN Bulletin II

$p = 0$ si ottiene la distribuzione esponenziale ordinaria.
 $p = -1$ definisce la classe di distribuzioni di Pareto.

Tra le varie classi di distribuzioni, quelle corrispondenti a $p = -1$ rappresentano la classe limite caratterizzata dalla più alta *pericolosità* possibile di sinistri. La funzione di Pareto è quella che presenta la massima asimmetria, in altri termini è la più *pericolosa* espressione analitica che possa essere usata per descrivere una distribuzione dei sinistri.

Esistono anche delle distribuzioni più asimmetriche di quella di Pareto, ma per ottenerle bisogna andare al di fuori della classe di distribuzioni definite da $\mu(x) = k \cdot x^p$.

G. Benktander definisce quindi la *classe di Pareto in senso ampio* in cui le distribuzioni sono caratterizzate da espressioni di $\mu(x)$ tali che il $\lim_{x \rightarrow \infty} x\mu(x) = k$.

Se $\mu(x) = \frac{k}{x \ln x}$, l'integrale $\int_x^\infty \mu(t) dt$ diverge e per $x > x_0$ i sinistri ovviamente hanno

una *mortalità* più bassa che nel caso di $\mu(x) = \frac{k}{x}$. Quindi si ha:

$$H(x) = C e^{-\int_x^\infty \frac{k dt}{t \ln t}} = C e^{-\int^{k d \ln t}} \quad \text{oppure} \quad H(x) = C \frac{1}{(\ln gx)^k}.$$

Questa distribuzione dei sinistri non ha una media finita ed è quindi, si spera, poco interessante per studi pratici nel campo assicurativo. Sono stati studiati altri casi con medie infinite. Tuttavia le distribuzioni definite da una mortalità più bassa della classe di Pareto non possono intuitivamente avere una media finita. La conclusione è che: la classe di Pareto in senso ampio è una classe limite che divide le funzioni con media finita e quelle con media infinita.

Questi due lavori, che ritengo basilari per la comprensione dell'asimmetria delle distribuzioni della dimensione dei sinistri, hanno avuto il grande torto di essere apparsi troppo presto nella letteratura attuariale e, conseguentemente di essere quasi sconosciuti alla maggior parte degli studiosi di oggi. Anche il linguaggio forse andrebbe adeguato alle esigenze attuali. Il valore medio $m(x)$ andrebbe chiamato secondo la terminologia di P. Embrechts¹² "eccedenza media al di sopra di una certa soglia" o meglio "shortfall", mentre la ricerca dell'asimmetria della distribuzione andrebbe definita come "ricerca del peggiore degli scenari".

Nei testi canonici della letteratura attuariale si trovano criteri che non hanno le caratteristiche dell'immediatezza della funzione $m(x)$ o del tasso istantaneo $\mu(x)$. Ad esempio nel citato testo "Loss distributions" il confronto fra le code di due distribuzioni viene effettuato esaminando il limite del rapporto fra due funzioni di densità (o delle funzioni di sopravvivenza) al tendere di x all'infinito. Viene inoltre introdotto un rapporto:

$$\lambda(x) = \frac{f(x)}{S(x)} \quad x \geq 0 ; S(x) = 1 - F(x)$$

Tale rapporto viene denominato forza di mortalità o tasso di decadimento ("failure rate") e interpretato come la probabilità di decadimento a x dopo essere sopravvissuto fino a x .

¹² P. Embrechts et al. – Quantitative Risk Management – Princeton Series in Finance - 2005

Intuitivamente se $\lambda(x)$ diviene piccolo, un immediato decadimento è meno probabile e la distribuzione è più asimmetrica e pericolosa (con una coda più lunga). Calcolando ad esempio l'indice per la distribuzione di Pareto (nella forma da noi comunemente impiegata) si avrebbe:

$$\lambda(x) = \frac{\alpha}{x}$$

L'indice, cioè, è strettamente decrescente da $\lambda(x_0) = \frac{\alpha}{x_0}$ a $\lambda(\infty) = 0$. In sostanza le distribuzioni possono essere classificate a seconda che l'indice sia crescente (distribuzione con coda corta) o decrescente (a coda lunga).

A. A. Klugman, H.H. Panjer e G.E. Willmot introducono anche l'indice trattato da G. Benktander e C. O. Segerdhal nella forma:

$$e_x = \frac{\int_x^\infty (t-x) f(t) dt}{S(x)}$$

Intuitivamente, si dice, che se questa funzione, denominata durata media di vita residua, assume valori elevati per valori grandi di x , allora la distribuzione ha una coda lunga. La funzione e_x viene posta a confronto con $\lambda(x)$, tuttavia manca la minuziosa analisi che permette di definire le distribuzioni della classe di Pareto come le più pericolose e asimmetriche.

Inoltre ritengo che il tasso di estinzione nella forma definita da G. Benktander, cioè $\mu(x) = k \cdot x^p$ sia estremamente comprensibile e permette con estrema "leggerezza" di passare a definire anche, come abbiamo visto le distribuzioni esterne alla classe e potenzialmente "pericolose".

4.7. Considerazioni sul VaR

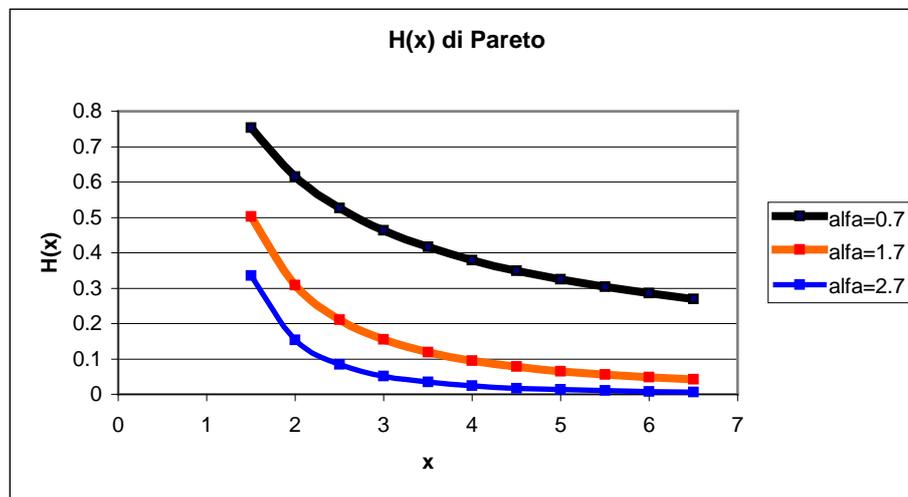
La conoscenza delle caratteristiche della distribuzione che descrive la dimensione dei sinistri si rivela indispensabile quando si è alla ricerca dei "peggiori scenari" e nello stesso tempo si usano delle "misure di rischio" che non danno alcuna informazione sulla distribuzione stessa. Una di queste misure è il VaR (che probabilisticamente rappresenta un quantile della distribuzione). Senza voler entrare sulla coerenza del VaR¹³, occorre chiedersi: nella ricerca dei peggiori scenari siamo autorizzati a considerare distribuzioni con media infinita? Senza ricorrere alle distribuzioni definite da G. Benktander come distribuzioni esterne alla classe di Pareto, calcoliamo il VaR (Value at Risk) al 99,9% di una distribuzione paretiana, cioè il

valore $1 - \frac{x_0}{x^\alpha} = 0,999$ per vari valori di α ponendo per semplicità $x_0 = 1$:

α	Note sulla distribuzione	VaR
2,7	Media e varianza finite	13
1,7	Media finita e varianza infinita	58
0,7	Media e varianza infinite	19 306

¹³ Si veda a tale proposito il capitolo sesto di A.J. McNeil, R. Frey, P. Embrechts - Quantitative Risk Management - Princeton 2005

Le distribuzioni con momenti infiniti sono tangibili e rappresentabili graficamente, però la loro utilità ai fini assicurativi è molto discutibile. Fornisco la rappresentazione grafica delle tre curve prese in considerazione per il calcolo del VaR.



A questo proposito è interessante riportare alcune osservazioni fatte da P. Embrechts¹⁴ e dai suoi collaboratori a proposito dell'uso di medie infinite. Si parte dalla definizione della variabile casuale rappresentativa del sinistro posta nella forma

$$L = \sum_{k=1}^d L_k$$

e si suppone che la coda della distribuzione sia approssimativamente di tipo parietano fortemente asimmetrica, cioè

$$P(L_k > x) = x^{-\alpha_k} h_k(x) \quad k=1, \dots, d$$

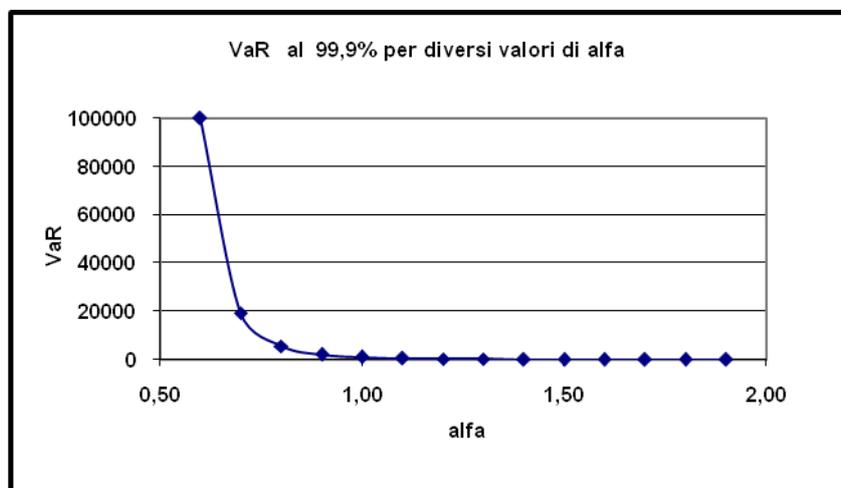
dove $h_k(x)$ è una funzione che varia lentamente.

In queste condizioni se $0 < \alpha < 1$ allora $E(L_k) = \infty$, mentre per $1 < \alpha < 2$ $E(L_k) < \infty$ ma $var(L_k) = \infty$. Tuttavia, avverte P. Embrechts, i risultati della teoria dei valori estremi, non vanno usati ciecamente. Il calcolo del VaR può condurre a oneri ridicolamente alti per il valore del capitale. In maniera provocatoria P. Embrechts afferma che i modelli aventi media infinita dovrebbe essere banditi dagli strumenti usati dagli specialisti.

Riconosce che tutte le distribuzioni del tipo di Pareto hanno momenti infiniti e, nonostante questo, il loro uso è molto frequente. Tuttavia avverte che se la media è infinita, le distribuzioni vanno trattate con molta cautela. Teoricamente è sempre possibile calcolare il VaR anche con distribuzioni con media infinita, tuttavia quando questo accade occorre concedersi una pausa di riflessione e analizzare in dettaglio le cause che hanno portato a considerare tali distribuzioni.

¹⁴ J. Nešlehová – P. Embrechts – V. Chaves-Demoulin – Infinite Mean Models and LDA for Operational Risk – Embrechts Website

Nel grafico che segue vengono mostrati i valori del VaR (al 99,9%) in corrispondenza di vari valori di α , alcuni dei quali ($\alpha < 1$) G. Benktander definirebbe non interessanti ai fini assicurativi o che P. Embrechts vorrebbe bandire dagli strumenti usati dagli specialisti.



Bibliografia essenziale:

- G. Ferrara – *Distributions des sinistres selon leur coût*. ASTIN Bulletin – vol. VI
- S. A. Klugman – H.H. Panjer- G.E. Willmot – *Loss Models* – Wiley – 1998
- Harald Cramér “*Mathematical Methods of Statistics*” – Princeton University Press
- J. Aitchisson & J.C. Brown “*The Lognormal Distribution*” – University Press Cambridge

PARTE V

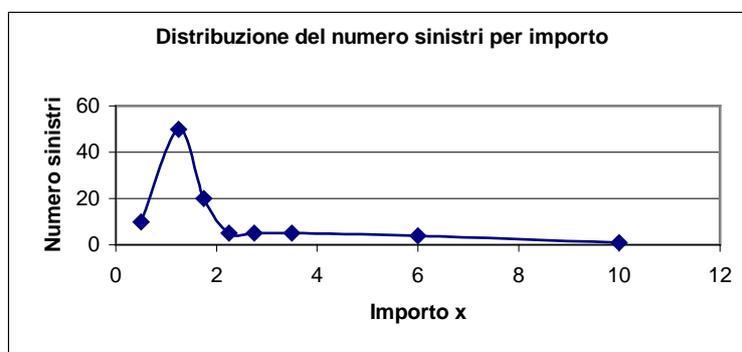
ALCUNE INDICAZIONI PER LA RAPPRESENTAZIONE GRAFICA

Una efficace rappresentazione grafica è basilare per una buona presentazione dei risultati. Per fornire qualche indicazione supponiamo di dover presentare i risultati raccolti nella tavola 5.1

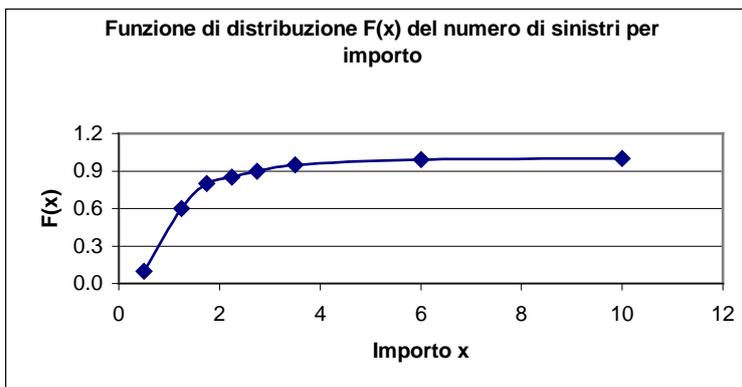
Tavola 5.1: Distribuzione del numero dei sinistri per classi d'importo

Classi d'importo	Valore centrale (z)	Numero dei sinistri	Frequenza $s(z)$	$S(z)$
Da 0 a 1	0,50	10	0,100	0,100
Da 1 a 1,5	1,25	50	0,500	0,600
Da 1,5 a 2	1,75	20	0,200	0,800
Da 2 a 2,5	2,25	5	0,050	0,850
Da 2,5 a 3	2,75	5	0,050	0,900
Da 3 a 4	3,50	5	0,050	0,950
Da 4 a 8	6,00	4	0,040	0,990
Da 8 a 12	10,00	1	0,010	1,000
Totali		100	1,000	

La distribuzione precedente è puramente indicativa. Inoltre, ai fini applicativi supporremo che la frequenza sia attribuibile al valore centrale della classe. In alternativa potremmo calcolare, nell'ambito della classe un opportuno valore medio. Per meglio presentare i dati potremmo pensare di rappresentare graficamente la funzione di frequenza supponendo che essa sia concentrata nel valore centrale delle classi d'importo.



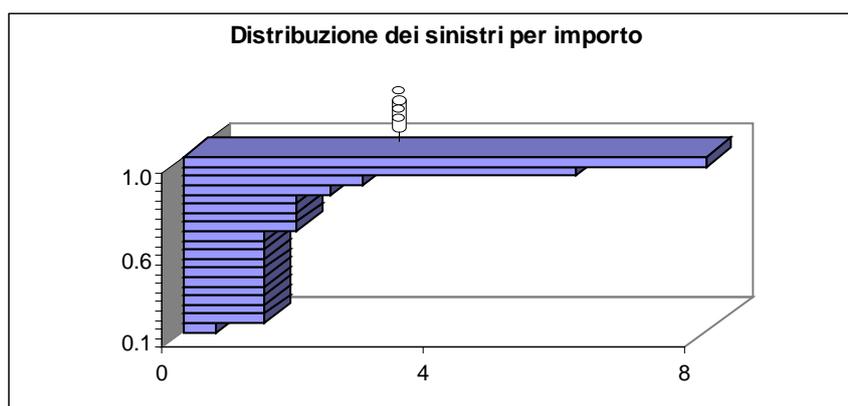
A volte, risulta molto più utile rappresentare la funzione di distribuzione:



Tuttavia se dobbiamo presentare questi grafici a “non attuari” dovremmo sottolineare la “drammaticità” del rischio. Una rappresentazione grafica che può aiutare è quella proposta dalle pubblicazioni della SWISS RE ¹⁵. Il punto di partenza è la funzione di distribuzione, per la quale i valori sono espressi in una scala costante della funzione di distribuzione $S(z)$, cioè si parte, ad esempio, dalla seguente tabella:

z	$S(z)$
0,500	0,1000
1,250	0,1500
1,250	0,2000
1,250	0,2500
1,250	0,3000
1,250	0,3500
1,250	0,4000
1,250	0,4500
1,250	0,5000
1,250	0,5500
1,250	0,6000
1,750	0,6500
1,750	0,7000
1,750	0,7500
1,750	0,8000
2,250	0,8500
2,750	0,9000
6,000	0,9500
8,000	0,9900

Questa regolarità dei valori in ascissa permette di rappresentare la funzione di distribuzione come una serie di cataste (barre) sovrapposte che rammentano l’instabilità insita nel portafoglio. Il grafico è completato, negli studi della SWISS Re, da una figurina stilizzata in cima alle distribuzioni che sottolinea lo squilibrio sottostante.



¹⁵ SWISS RE – Exposure rating

RIFLESSIONI FINALI

Al termine di questo mio percorso vorrei trasmettere alcuni messaggi agli studiosi che iniziano o proseguono il loro cammino nel campo attuariale.

Il primo messaggio riguarda la necessità di risalire alle fonti originarie della materia da trattare: domande come quella che mi sono posta nell'espone, ad esempio, la formula di Pareto dovrebbero portare sempre a riaccostarsi alla vera natura degli argomenti, +prima che essi siano stati sedimentati dalle interpretazioni di posteriori studiosi. Solo successivamente ci si deve inoltrare, con senso critico, nel campo degli ulteriori sviluppi. So per esperienza che questo non sempre è possibile, ma quando lo è, occorre fare questo sforzo iniziale

Vorrei, d'altra parte, trasmettere la mia curiosità per tutti gli studi effettuati da attuari che hanno operato o operano nel campo assicurativo. Molto possiamo apprendere dalle loro difficoltà e dubbi. E' lampante, ad esempio, che un attuario, nella sua prassi quotidiana, ha sempre bisogno di definire una distribuzione, di cui possa stimare almeno la media e la varianza.

Infine vorrei comunicare il mio disagio circa la considerazione delle medie infinite. L'infinito è una pura invenzione del nostro intelletto: nell'immanente, e quindi anche nell'assicurazione, tutto è finito. Lo sconforto, manifestato da Galileo Galilei, è anche mio e potrei dire con lui¹⁶: " queste son di quelle difficoltà che derivano dal discorrer che noi facciamo col nostro intelletto finito intorno a gl'infiniti dandogli quelli attributi che noi diamo alle cose finite e terminate".

¹⁶ Galileo Galilei – Discorsi e dimostrazioni matematiche intorno a due nuove scienze – Einaudi 1990