Classi di rischio e Personalizzazione del Premio

Esempi

Prof. Cerchiara Rocco Roberto

Materiale e Riferimenti

1. **Lucidi distribuiti in aula**
2. **Elementi di Matematica delle Assicurazioni – Pitacco 2002 (pag.112 -122)**
3. **Seminario SIFA 10-11 aprile 2002 di Gigante, Sigalotti, Picech**

Nelle lezioni precedenti il problema del calcolo del premio è stato affrontato immaginando di riferirsi a rischi "qualitativamente omogenei"

↓

Significa dire che l'assicuratore assicurerebbe tali rischi allo stesso tasso di premio (potendo variare l'esposizione, cioè l'importo del premio).

LA PERSONALIZZAZIONE A PRIORI

In realtà, rischi pur tra loro sufficientemente simili, possono differenziarsi per alcune caratteristiche che possono indurre l'assicuratore ad adottare diversi livelli di tassi di premio.

E' importante quindi che l'assicurato paghi un premio commisurato alla sua sinistrosità.

E' necessario quindi differenziare i rischi all'interno della collettività. Si procede quindi alla suddivisione dell'aggregato di rischi in sottocollettività➔ CLASSI DI RISCHIO che siano caratterizzate da una elevata

omogeneità all'interno di esse (rischi "analoghi"). Si pensi ad es. all'assicurazione Incendio dividendo gli edifici in base alla destinazione d'uso (ufficio, abitazione, ecc.).
La definizione delle classi di rischio è fondata su:

1. **Individuazione di un certo numero di FATTORI DI RISCHIO (ad es. destinazione, tipo di costruzione, potenza autoveicolo, ecc.)**

2. **ogni fattore di rischio può assumere più determinazioni raggruppabili in classi o livelli che vengono detti MODALITA' (qualitative, ad es. professione ad alto rischio o a basso rischio, e quantitative, ad es. fino a 8 CV, tra 8 e 10 CV, da 10 CV in poi, ecc.)**

Una volta individuati alcuni fattori di rischio e rispettive modalità, si pone il problema di selezionare quelli più significativi, in modo da ripartire l'aggregato in classi omogenee (a cui attribuire lo stesso premio) in base a tali fattori.

Variabili tariffarie = fattori selezionati (in base ad uno dei metodi, ad es., Forward o Backward, che si differenziano semplicemente dall'analizzare la significatività di una variabile introducendone una alla volta o estraendone una alla volta dal complesso delle variabili).

Classi tariffarie = le classi in cui è ripartito l'aggregato (attraverso ad es. la cluster analysis – rif. a Metodo di Ward).

A questo punto si utilizza una funzione, detta <u>MODELLO TARIFFARIO</u>, che ad ogni classe associa il premio corrispondente.

Tale funzione sarà dipendente da alcuni parametri, detti <u>RELATIVITA'</u>, che vengono stimati dai dati.

IN TAL MODO SI OTTIENE LA TARIFFA

PROBLEMI:

1. **Scelta del modello tariffario**
2. **Scelta delle relatività**
3. **controllo della bontà della tariffa**

ESEMPIO

IL CASO DI DUE VARIABILI TARIFFARIE

Supponiamo di avere ripartito la collettività in classi di rischio, mediante due variabili tariffarie, la prima con I modalità e la seconda con J modalità.

La collettività è quindi ripartita in I * J classi.

La generica coppia (i,j) individua la classe in cui il primo fattore ha modalità i, il secondo modalità j.

Operativamente al momento della stipulazione di un nuovo contratto, il rischio è classificato, cioè assegnato ad una

particolare classe di rischio e conseguentemente gli è attribuito il tasso di premio.

Supponiamo di aver osservato la collettività e siano

c(i,j) = risarcimento globale osservato per i rischi della classe (i,j)

r(i,j) = numero totale dei rischi/anno della classe (i,j)

Allora il rapporto

$$Q = c(i,j) / r(i,j)$$

è la quota danni per i rischi della classe: stima "grezza" del premio equo per i rischi della classe.

Esempio numerico (RCA)

Età	< 25 (1)	< 25 (1)	>=25 (2)	>=25 (2)	Totale
Tipo veicolo	Basso Rischio (1)	Alto Rischio (2)	Basso Rischio (1)	Alto Rischio (2)	
Polizze Anno	3.570	1.622	5.826	1.281	12.299
N. sinistri	739	452	880	248	2.319
Costo medio per sinistro	2.194.000	2.826.000	2.040.000	2.972.000	2.342.000

Senza tenere conto dei fattori di rischio, la Q calcolata sui dati dell'intera collettività è:

Q = 2.319 / 12.299 * 2.342.000 = 441.589

mentre le Quote danni per le diverse classi di rischio (età e tipo veicolo) sono:

Q(1,1) = 739 / 3.570 * 2.194.000 = 454.164

Q(1,2) = 787.517

Q(2,1) = 308.136

Q(2,2) = 575.375

Si introduce ora un <u>modello tariffario</u> per perequare le stime grezze perché:

1. **a classi vicine non competano premi troppo diversi;**
2. **i dati possono essere affetti da errori di osservazione (di archiviazione, ecc.) o da perturbazioni accidentali (elevata sinistrosità nel periodo di osservazione, ecc.)**
3. **alcune classi possono essere poco numerose**
4. **si vuole un effetto di <u>solidarietà</u> (che alcuni paghino un premio maggiore di quello che spetterebbe loro, altri minore).**

I modelli tariffari

Sia Q la quota danni ottenuta per l'intera collettività.

Si considera

f(i,j) = Q + λ(i) + μ(j), modello additivo

f(i,j) = Q * α(i) * β(j), modello moltiplicativo (tipico nella RCA)

I parametri α, β, μ e λ sono le relatività e saranno valutati dai dati in modo da "accostare" i dati grezzi:

$$\hat{Q}(i, j) = Q + \hat{\lambda}(i) + \hat{\mu}(j) \cong Q(i, j)$$
$$\hat{Q}(i, j) = Q * \hat{\alpha}(i) * \hat{\beta}(j) \cong Q(i, j)$$

Si noti che utilizzando un modello tariffario (additivo o moltiplicativo) abbiamo che:

1. **devo stimare I + J parametri (relatività) del modello contro gli I*J tassi specifici (Quote Danni)**
2. **si trascura l'effetto combinato dei due fattori**

Esempio.
Rischi RCA (continua)

Età	**< 25 (1)**	**< 25 (1)**	**>=25 (2)**	**>=25 (2)**	**Totale**
Tipo veicolo	**Basso Rischio (1)**	**Alto Rischio (2)**	**Basso Rischio (1)**	**Alto Rischio (2)**	
Polizze Anno	**3.570**	**1.622**	**5.826**	**1.281**	**12.299**
N. sinistri	**739**	**452**	**880**	**248**	**2.319**
Costo medio per sinistro	**2.194.000**	**2.826.000**	**2.040.000**	**2.972.000**	**2.342.000**

Q = 441.589
Q(1,1) = 454.164
Q(1,2) = 787.517
Q(2,1) = 308.136
Q(2,2) = 575.375

Se utilizzo un modello tariffario moltiplicativo, si vuole che

$\hat{Q}(i,j) = Q * \hat{\alpha}(i) * \hat{\beta}(j)$ **si avvicini a Q(i,j)**

Come procedere?

Ad esempio con

a) METODO DEI MINIMI QUADRATI

$$\sum_{i,j}(valoreOsservato - valoreStimato)^2 = \sum_{i,j}\left(Q(i,j) - Q\alpha(i)\beta(j)\right)^2$$

b) METODO DEI MINIMI QUADRATI <u>PONDERATI</u>

(per tenere conto dell'esposizione, rischi/anno, nelle varie classi):

scelgo i parametri che rendono minima la funzione

$$\sum_{i,j}\left((valoreOsservato - valoreStimato)^2\right) * numRischi = \sum_{i,j} r(i,j) * \left(Q(i,j) - Q\alpha(i)\beta(j)\right)^2$$

Nell'esempio la funzione è

$$\begin{aligned} &\mathbf{3.570 * (454.164 - 441.589\, \alpha(1)\, \beta(1))^2 +} \\ &\mathbf{+ 1.622 * (787.517 - 441.589\, \alpha(1)\, \beta(2))^2 +} \\ &\mathbf{+ 5.826 * (308.136 - 441.589\, \alpha(2)\, \beta(1))^2 +} \\ &\mathbf{+ 1.281 * (575.375 - 441.589\, \alpha(2)\, \beta(2))^2} \end{aligned}$$

Il punto di minimo è

$$\hat{\alpha}(1) = 1{,}29509 \qquad \hat{\alpha}(2) = 0{,}91033$$

$$\hat{\beta}(1) = 0{,}78181 \qquad \hat{\beta}(2) = 1{,}39226$$

Quindi la tariffa è

P(1,1) = 447.114;
P(1,2) = 796.230;
P(2,1) = 314.282;
P(2,2)= 559.680

mentre

Q(1,1) = 454.164;
Q(1,2) = 787.517;
Q(2,1) = 308.136;
Q(2,2)= 575.375.

OSSERVAZIONE

Si può procedere in modo analogo a quanto illustrato, separatamente per il numero sinistri e per il risarcimento per sinistro e ottenere i premi come prodotto delle due componenti.

ALLEGARE A PARTE 9

I LIVELLI TARIFFARI PER OGNI FATTORE DI RISCHIO

Obiettivo del raggruppamento delle determinazioni in livelli è la riduzione del numero delle classi tariffarie che può essere giustificata da motivi commerciali o da motivi di applicabilità della metodologia statistica

I livelli con la Cluster Analysis. Metodo Ward

Si aggregano determinazioni di un fattore di rischio in modo da ottenere classi omogenee rispetto ad una variabile risposta

L'omogeneità è ottenuta con la minima varianza totale interna

Variabile risposta: frequenza sinistri

Fattore di rischio con N determinazioni da ripartire in K livelli

Abbiamo un file di dati in cui per ogni individuo abbiamo il tempo di esposizione f_i, la determinazione h del fattore di rischio ed il numero di sinistri n_i

Possiamo ottenere le esposizioni complessive $g_1, \ldots, g_N$:

$g_h = \sum_i g_{ih}$ dove la somma è estesa a tutti gli individui con determinazione h del fattore di rischio

Per l'i-esimo individuo della classe con determinazione h la frequenza sinistri è: $f_{ih} = \dfrac{n_{ih}}{g_{ih}}$

con n_{ih} e g_{ih} numero dei sinistri ed esposizione

La frequenza sinistri del gruppo con determinazione h è:

$$f_h = \sum_{i=1}^{m_h} \frac{g_{ih}}{g_h} f_{ih},$$

e la varianza interna:

$$\sigma_h^2 = \sum_{i=1}^{m_h} \frac{g_{ih}}{g_h} (f_{ih} - f_h)^2 .$$

La varianza totale interna è

$$_N\sigma^2 = \sum_{h=1}^{N} \sigma_h^2 \frac{g_h}{g} .$$

Il metodo di Ward è di tipo aggregativo ed il generico passo del procedimento è il seguente:

supposto di avere una ripartizione in K classi, con varianza totale interna $_K\sigma^2 = \sum_{h=1}^{K} {_K\sigma_h^2} \frac{_K g_h}{g}$, $_K g_h$ essendo l'esposizione nella h-esima classe della ripartizione, si passa a K-1 classi scegliendo l'aggregazione di due classi che rende minima la varianza totale interna $_{K-1}\sigma^2$, con $_{K-1}\sigma^2 = \sum_{h=1}^{K-1} \frac{_{K-1}g_h}{g} {_{K-1}\sigma_h^2}$.

Il numero finale K di clusters può essere prefissato o individuato tramite i valori di una funzione g(K) che misura la perdita di informazione connessa al raggruppamento delle determinazioni della variabile tariffaria in K cluster.

Di solito si prende

$$g(K) = \frac{_{(K)}\sigma^2}{\sigma^2}$$

con $_{(K)}\sigma^2 = \sigma^2$

e si sceglie il massimo K tale che $g(K) \geq 0.95$.

Con due variabili risposta A e B indipendenti, dopo averle standardizzate, si considera la varianza totale interna $_K\sigma_A^2 + {_K\sigma_B^2}$ e si passa da K a K-1 classi raggruppando le due classi che danno il minimo di $_{K-1}\sigma_A^2 + {_{K-1}\sigma_B^2}$.

Consideriamo ad esempio la frequenza sinistri ed il costo medio.

E' necessario procedere preventivamente ad una standardizzazione delle variabili risposta.

La frequenza sinistri standardizzata è data

$$f_i' = \frac{f_i - \bar{f}}{\sigma_f} \qquad \text{dove } \sigma_f^2 = \sum_{i=1}^{N} (f_i - \bar{f})^2 \frac{g_i}{g} ,$$

mentre i costi dei sinistri standardizzati sono

$$c_i' = \frac{c_i - \bar{c}}{\sigma_c} \qquad \text{dove } \sigma_c^2 = \sum_{i=1}^{N} (c_i - \bar{c})^2 \frac{n_i}{n} .$$

La varianza totale interna si definisce come somma delle varianze totali interne misurate per le singole variabili risposta.

Usando come variabile risposta il costo medio, il procedimento di aggregazione è fortemente influenzato da eventuali sinistri eccezionali. Occorre ridurre il loro effetto.

L'esigenza di ottenere classi omogenee per sinistrosità può essere mediata da esigenze di tipo commerciale. Ad esempio quella di ottenere livelli con determinazioni vicine (es. età 18-25)

Per forzare il procedimento a raggruppare determinazioni contigue si può decidere di inserire tra le variabili risposta la variabile esplicativa stessa.

BIBLIOGRAFIA

Anderberg M.R. (1973), Cluster Analysis for Applications, Academic Press, New York.

Dickmann H. (1978), Einsatz der clusteranalyse bei klassifikationsproblemen in der versicherungswirtschaft, Blatter der Deutschen Gesellschaft fur Versicherungsmathemitik, Vol. XIII, n. 4.

Van Eeghen J., Greup E.K., Nijssen J.A. (1983), Rate Making, Surveys of Actuarial Studies, 2, Nationale Nederlande N.V.

Ward J.H. (1963), Hierachical grouping to optimise an objective function, J. Am. Statist. Ass., 58.

1,3

ANALISI STATISTICHE PRELIMINARI

Sui dati di portafoglio

Mettere in relazione la sinistrosità con i fattori di rischio

Dapprima su base univariata:

Ad esempio: esprimere la frequenza sinistri per le varie età

Successivamente su base multivariata

Ad esempio: considerati due fattori di rischio correlati (in base all'esperienza) come età e potenza del veicolo si esprime la frequenza sinistri per le varie età e per ogni livello di potenza

VARIABILI QUALITATIVE, NUMERICHE E VARIABILI CON DETERMINAZIONI SUDDIVISE IN CLASSI (LIVELLI)

Variabili (o fattori di rischio) quali sesso, professione, stato civile, zona di residenza,.., sono dette nominali o qualitative.

Non hanno una determinazione numerica "naturale".

Consideriamo una codifica numerica

Esempio. Possiamo codificare

$$\text{sesso}=\begin{cases}1 & \text{se femmina}\\ 0 & \text{se maschio}\end{cases}$$

Esempio. Stato civile con le quattro determinazioni di single, coniugato/a, vedovo/a, divorziato/a. Definiamo

$$X_1=\begin{cases}1 \text{ se single}\\ 0 \text{ altrimenti}\end{cases},\quad X_2=\begin{cases}1 \text{ se coniugato/a}\\ 0 \text{ altrimenti}\end{cases},$$

$$X_3=\begin{cases}1 \text{ se vedovo/a}\\ 0 \text{ altrimenti}\end{cases},\quad X_4=\begin{cases}1 \text{ se divorziato/a}\\ 0 \text{ altrimenti}\end{cases}.$$

Dato che $X_1 + X_2 + X_3 + X_4 = 1$, possono bastare le determinazioni di X_1, X_2, X_3 per determinare X_4. Le terne (1,0,0), (0,1,0), (0,0,1), (0,0,0) identificano rispettivamente la qualifica di single, coniugato/a, vedovo/a, divorziato/a.

In generale, data una variabile nominale con k determinazioni, possiamo descriverla con k-1 variabili 0-1, nel modo indicato nei due esempi precedenti.

Le variabili numeriche possono essere considerate con le loro determinazioni naturali (si parla di variabili continue)

Oppure, soprattutto per ragioni commerciali, si raggruppano le determinazioni in classi dette livelli.

Una volta che le determinazioni sono ripartite in livelli la variabile può essere descritta con variabili 0-1 come per le variabili qualitative. La descrizione con variabili 0-1 può essere fatta in diversi modi come mostra l'esempio seguente.

Esempio . Consideriamo la variabile età con la seguente suddivisione in livelli.

Età
18-29
30-39
40-54
55-68
69-88

Per descriverla possiamo usare le quattro variabili linearmente indipendenti

$$X_1 = \begin{cases} 1 \text{ se } 18 \le x \le 29 \\ 0 \text{ altrimenti} \end{cases}, \quad X_2 = \begin{cases} 1 \text{ se } 30 \le x \le 39 \\ 0 \text{ altrimenti} \end{cases},$$

$$X_3 = \begin{cases} 1 \text{ se } 40 \le x \le 54 \\ 0 \text{ altrimenti} \end{cases}, \quad X_4 = \begin{cases} 1 \text{ se } 55 \le x \le 68 \\ 0 \text{ altrimenti} \end{cases}$$

ma anche le

$$X_1 = \begin{cases} 1 \text{ se } x \le 29 \\ 0 \text{ altrimenti} \end{cases}, \quad X_2 = \begin{cases} 1 \text{ se } x \le 39 \\ 0 \text{ altrimenti} \end{cases},$$

$$X_3 = \begin{cases} 1 \text{ se } x \le 54 \\ 0 \text{ altrimenti} \end{cases}, \quad X_4 = \begin{cases} 1 \text{ se } x \le 68 \\ 0 \text{ altrimenti} \end{cases}$$

APPENDICE

UN METODO DI CLUSTER ANALYSIS

Per la determinazione del numero di classi ottimali si è fatto ricorso alla "cluster analysis" in cui l'aggregazione delle classi è stata ottenuta tramite il metodo della minima varianza (o metodo di Ward).

L'obiettivo è minimizzare l'incremento di varianza interna totale tramite un procedimento iterativo.

Il processo si arresta sulla base di un criterio operativo opportunamente scelto: ad esempio limitando ad un dato valore la perdita di informazione dovuta al progressivo accorpamento di classi distinte.

Una possibile definizione della perdita di informazione in presenza di k classi può essere:

$$l(k) = \frac{\sigma^2(k)}{\sigma^2}$$

dove:

σ^2 è la varianza totale

$\sigma^2(k) = \sum_{i=1}^{k} \sigma_i^2 \frac{n_i}{n}$ è la varianza interna totale fissate k classi

con σ_i^2 varianza interna dell'i-sima classe.

Si vedano anche le fotocopie distribuite in aula.

LA PERSONALIZZAZIONE A POSTERIORI DEL PREMIO

LA PERSONALIZZAZIONE PASSA TIPICAMENTE ATTRAVERSO DUE FASI.

In una prima fase detta di *personalizzazione a priori* viene determinato un premio sulla base delle determinazioni assunte da un insieme di variabili tariffarie osservabili a priori. Tale fase ha essenzialmente due obiettivi:

- **suddividere i rischi in classi omogenee, cioè le classi tariffarie**
- **stimare i premi della classe**

Nonostante l'utilizzo di molte variabili tariffarie, all'interno di ogni classe rimane notevole eterogeneità nei comportamenti di guida e quindi nella sinistrosità.

A partire dalla metà degli anni '50, l'idea di consentire un aggiustamento del premio, a posteriori, dopo aver osservato la storia di sinistrosità individuale.

Si è passati così alla TARIFFAZIONE IN BASE ALL'ESPERIENZA: si penalizza un assicurato responsabile di uno o più sinistri con un aggravio di premio (malus) e si premia un assicurato che non ha provocato sinistri con uno sconto (bonus).

Questa è la seconda fase in cui viene realizzata una *personalizzazione in base all'esperienza* nella quale si opera una calibrazione del premio a posteriori, progressivamente affinabile nel tempo mediante l'esperienza di sinostrosità relativa al contratto.

In tal modo si ha un passaggio da un premio COLLETTIVO ad un premio INDIVIDUALE.

In tale ambito si considerano tecniche di tipo "dinamico" che consentono di "tariffare" partendo da un premio iniziale e determinando su diverse scadenze una sequenza di premi via via adeguati all'esperienza di sinistrosità acquisita sul portafoglio di contratti (o a ciascuna classe di rischio).

Si precisa che esistono tecniche di tariffazione in base all'esperienza COLLETTIVA (si veda Pitacco per la *formula di credibilità*). In questo corso consideremo il caso della tariffazione in base all'esperienza INDIVIDUALE, con particolare riferimento ai sistemi BM.

Il sistema Bonus Malus (di seguito indicato con la sigla BM) è il sistema più frequentemente adottato per la personalizzazione del premio in base all'esperienza, sebbene in mercati assicurativi diversi da quello italiano siano diffusamente praticati, a tale scopo, anche altri modelli di tariffazione (ad esempio il sistema No-Claim Discount).

Come è ben noto in letteratura, i sistemi BM rappresentano forme di tariffazione in base alle quali:

- **gli assicurati sono distribuiti, secondo prefissate regole evolutive, in un numero finito di classi di merito $j = 1, 2, \ldots, M$;**

- **la classe di merito assegnata ad un assicurato per un certo periodo di assicurazione, generalmente un anno, è determinata unicamente dalla classe di appartenenza e dal numero di sinistri registrati durante il periodo precedente. A ciascuna classe corrisponde un determinato coefficiente di premio β_j; tipicamente $\beta_1 \leq \ldots \leq \beta_M$ ed esiste una classe h tale che $\beta_h = 1$.**

Dato un portafoglio di polizze RCA omogeneo a priori, per calcolare, in ogni istante t, l'ammontare dei premi, quello dell'i-esimo sinistro e la distribuzione degli assicurati nelle M classi di merito, è sufficiente:

1. **conoscere la struttura del sistema BM;**
2. **particolarizzare le distribuzioni del numero di sinistri annui e dell'ammontare degli stessi $\tilde{k}_t$ e $\tilde{Z}_{i,t}$.**

In merito al punto 1, indicando con *M* il numero massimo di classi e con *s* la classe di ingresso, si riportano le regole evolutive e la successione dei coefficienti di premio del sistema BM italiano (*M* = 18 e *s* = 14[1]):

$$\tilde{Y}_{t+1} = \begin{cases} \max\left(\tilde{Y}_t - 1, 1\right) & \tilde{k}_t = 0 \\ \\ \min(18, \tilde{Y}_t + 3\tilde{k}_t - 1) & \tilde{k}_t = 1,2,3,\ldots.. \end{cases}$$

j	1	2	3	4	5	6	7	8	9
β_j	**0,50**	**0,53**	**0,56**	**0,59**	**0,62**	**0,66**	**0,70**	**0,74**	**0,78**

j	10	11	12	13	14	15	16	17	18
β_j	**0,82**	**0,88**	**0,94**	**1,00**	**1,15**	**1,30**	**1,50**	**1,75**	**2,00**

e di quello svizzero (*M* = 22 e *s* = 10):

$$\tilde{Y}_{t+1} = \begin{cases} \max\left(\tilde{Y}_t - 1, 1\right) & \tilde{k}_t = 0 \\ \\ \min\left(22, \tilde{Y}_t + 4\tilde{k}_t\right) & \tilde{k}_t = 1,2,3,\ldots.. \end{cases}$$

j	1	2	3	4	5	6	7	8	9	10	11
β_j	**0,45**	**0,50**	**0,55**	**0,60**	**0,65**	**0,70**	**0,75**	**0,80**	**0,90**	**1,00**	**1,10**

j	12	13	14	15	16	17	18	19	20	21	22
β_j	**1,20**	**1,30**	**1,40**	**1,55**	**1,70**	**1,85**	**2,00**	**2,15**	**2,30**	**2,50**	**2,70**

Per quanto concerne il punto 2, l'approccio considerato è quello della teoria del rischio secondo l'impostazione collettiva, il cui modello base fu

[1] Per il Sistema Bonus Malus italiano è stata considerata, per tutti gli assicurati, la classe di ingresso 14, ipotizzando quindi rischi assicurati per la prima volta.

presentato da Ph. Lundberg al VI Congresso Internazionale degli Attuari in Vienna nel 1909 e rielaborato successivamente dall'autore stesso e soprattutto da H. Cramer (1930) e da Laurin (1930).

In base a questo approccio viene determinata, in funzione della variabile aleatoria numero di sinistri $\tilde{k}_t$ e della variabile aleatoria ammontare del costo dei sinistri $\tilde{Z}_{i,t}$, la distribuzione della variabile aleatoria danno aggregato.

(1) Numero di sinistri. Il comportamento della variabile aleatoria $\tilde{k}$ numero di sinistri durante un determinato periodo di tempo (ad esempio l'anno o l'unità temporale prescelta) è descritta da una distribuzione di probabilità, che è determinata attraverso le probabilità:

$$\boldsymbol{p_k = Prob\left\{\tilde{k} = k\right\}, \qquad (k = 0,1,2,...)}$$

che esattamente si verifichino *k* sinistri nel periodo considerato.
Se l'ipotesi distribuzionale del numero di sinistri sarà rappresentata dalla Poisson, di parametro ν[2] allora:

$$p_k = \frac{e^{-\upsilon}\upsilon^k}{k!}, \qquad k = 0,1,2,...$$

Adottando tale distribuzione implicitamente si suppone che i rischi siano omogenei, che non sussistano fluttuazioni a breve termine o cicli di lungo periodo.

(2) Il costo sinistri e il danno aggregato. Il danno aggregato, $\tilde{X}$, anch'esso v.a., in quanto somma dei singoli importi aleatori $\tilde{Z}_i$, nel generico anno *t*, è dato da:

$$\tilde{X} = \sum_{i=1}^{\tilde{k}} \tilde{Z}_i$$

[2] Essa possiede buona valenza metodologico-teorica, mostrando varie proprietà; però, nell'assicurazione automobilistica, si è riscontrata grossa disparità tra media e varianza

dove $\tilde{Z}_i$ indica il costo dell'*i*-esimo sinistro verificatosi nel periodo considerato[3]. Logicamente in assenza di sinistri, per $\tilde{k}$=0, allora $\tilde{X}$=0. L'ipotesi di base é che sussista indipendenza stocastica tra $\tilde{Z}_i$ e $\tilde{k}$; si postula quindi che un singolo sinistro di una certa entità non sia influenzato dal numero di sinistri che sono avvenuti, né dalla misura degli altri sinistri. Grazie a tale assunzione il premio puro risulta come prodotto della componente di sinistrosità e della componente di costo, $P = E[\tilde{k}] \cdot E[\tilde{Z}]$.

Inoltre ipotizzando l'indipendenza e somiglianza dei singoli risarcimenti $\tilde{Z}_i$, si può assumere (e quindi attribuire in sede operativa) un costo medio per sinistro uguale per tutti.

Segue che la discriminazione tra guidatori più o meno "bravi" (ai quali far pagare un premio maggiore o minore, commisurato alle singole caratteristiche di rischiosità) viene fatta sul solo numero dei sinistri. Questo ragionamento trova perfetta applicabilità nei sistemi di personalizzazione di tipo BM. L'analisi della distribuzione del costo del sinistro ha sicuramente maggiore importanza nell'altra forma di personalizzazione a posteriori del premio: la tariffa con franchigia[4].

L'evoluzione di un sistema BM, però, è determinata anche da fenomeni che non rientrano nella modellizzazione precedente. Tali nuovi elementi riguardano fondamentalmente due aspetti:

- **i flussi in entrata e uscita dal portafoglio;**
- **la rispondenza tra sinistrosità reale e sinistrosità apparente.**

Il portafoglio non è una collettività chiusa, perché ad ogni istante sono possibili sia nuovi contratti sia uscite, le quali possono avvenire sia per fine della vita assicurativa sia per scelta individuale.

Ai fini della modellizzazione utilizzata sarà considerato un portafoglio aperto ai nuovi ingressi, ma non alle uscite: i nuovi ingressi sono tipicamente la massa di giovani automobilisti che stipulano ex novo un contratto assicurativo. In genere si considera che entrino in assicurazione, ogni anno, una percentuale, pressoché costante nel tempo, della popolazione assicurata all'inizio di ogni periodo. Tale percentuale è stata posta pari al 6%. La conseguenza dell'inserimento di tutti i

[3] In tale ambito si é ipotizzato che l'assicurazione sia a valore intero.
[4] Si veda ad es. Gigante P., [1997], "Un modello per una tariffa RCA bonus-malus con franchigia", GIIA.

nuovi entrati nella classe di ingresso, a cui corrisponde un coefficiente di premio superiore al coefficiente medio, é un innalzamento del livello medio di premio.
Per quanto riguarda la rispondenza tra sinistrosità reale e sinistrosità apparente, nell'ambito di un sistema BM, è influenzata principalmente da tre fenomeni:

- **il moral hazard, cioè l'impossibilità per le imprese di osservare il comportamento dell'assicurato riguardo alla possibilità di incidere sulla probabilità di evitare il sinistro. In sostanza, l'assicuratore deve affidarsi alle dichiarazioni dell'assicurato circa l'esistenza e l'entità di un sinistro, non essendo in grado di osservarle direttamente.**

- **l'autoliquidazione dei sinistri, cioè il sopportare in proprio il costo del danno valutando ciò vantaggioso di fronte a futuri aumenti di premio. Un qualsiasi meccanismo BM costituisce un incentivo all'autoliquidazione. Tale effetto incentivante agisce in modo più efficace quanto il meccanismo penalizza chi denunci sinistri. Ciò pone il problema di distinguere tra sinistri provocati e sinistri denunciati. Il fatto che non tutti i sinistri vengano denunciati, fa aumentare la percentuale di assicurati nelle classi di bonus e quindi determina un aumento dei premi di equilibrio;**

- **la diversa severità delle scale dei coefficienti di merito.**

Nell'ambito di queste lezioni tali problematiche non sono state trattate; è stato pertanto supposto che la frequenza annua dei sinistri denunciati sia la stessa, qualunque polizza sia stata scelta dall'assicurato. Ciò equivale a supporre che gli assicurati, provocato un incidente, non si comportino in modo diverso (denuncia o autoliquidazione) a seconda delle caratteristiche del sistema in cui sono inseriti.

Analizziamo un campione pari al 67% del mercato RC Auto, tratto dalla statistica associativa ANIA "Banca Dati Auto" del settore I "Autovetture ad uso privato", relativa all'anno 2001.

Nella seguente tabella sono riportati, in riferimento alla statistica succitata, il numero di veicoli anno, la distribuzione dei veicoli, la frequenza sinistri e il costo medio sinistri per ciascuna classe di BM.

Classe Bonus/Malus	Veicoli Anno	Distribuzione % Veicoli	Frequenza Sinistri (in %)	Costo Medio Sinistri (in euro)
1	6.685.674	40,81	5,56	3.175
2	774.043	4,73	9,54	3.168
3	995.999	6,08	9,34	3.124
4	660.244	4,03	10,71	3.060
5	764.157	4,67	8,22	3.089
6	766.189	4,68	8,81	3.123
7	639.497	3,90	10,07	3.140
8	700.652	4,28	9,36	3.146
9	695.513	4,25	10,35	3.182
10	607.147	3,71	11,63	3.259
11	626.289	3,82	10,65	3.283
12	683.748	4,17	10,87	3.321
13	758.141	4,63	11,97	3.429
14	898.440	5,48	27,01	3.556
15	73.888	0,45	16,97	3.457
16	33.445	0,20	18,02	3.463
17	8.265	0,05	18,26	3.189
18	9.247	0,06	24,26	2.864
Totale	**16.380.578**	**100,00**	**9,22**	**3.248**

Tabella 1: Veicoli, frequenze sinistri e costo medio ripartiti per classi di BM

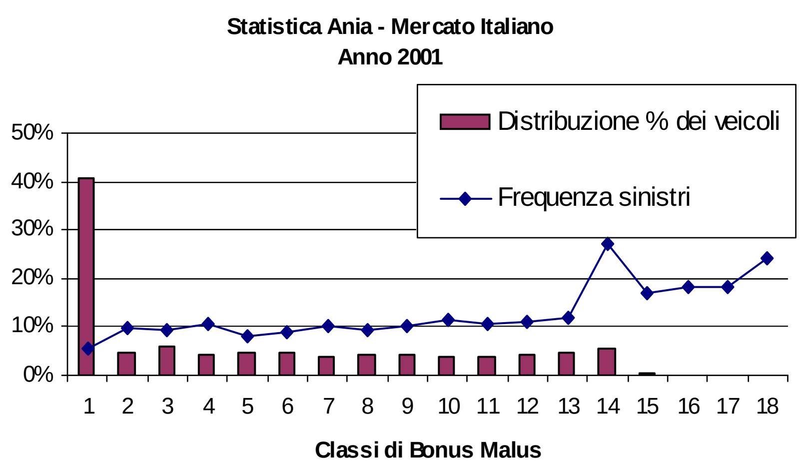

Grafico 1: Veicoli e frequenze sinistri ripartiti per classi di BM – Dati Ania 2001

Nel grafico precedente emerge che la frequenza sinistri osservata nel 2001 ha un andamento non monotono al variare delle classi di BM, con particolare riferimento tra le classi 1 e 4 (il livello osservato nella 14 è motivabile con l'eterogeneità dei rischi tipica della classe iniziale). Tale andamento sembra negare le capacità discriminanti del sistema

BM italiano, ma si deve tener conto che si è preso in considerazione un solo anno di osservazione. Sulla normalizzazione di tale frequenza si rimanda alle sezioni successive.

Si è provveduto a calcolare la frequenza sinistri della popolazione osservata come media ponderata, con pesi dati dal numero di veicoli, delle frequenze sinistri di ciascuna classe. Il parametro ν della distribuzione di Poisson della variabile aleatoria numero sinistri sarà posto pari alla percentuale così ottenuta: 9,22%.

Una conseguenza del funzionamento dei sistemi BM è una diminuzione progressiva del livello medio di premio osservato, dovuto ad una concentrazione degli assicurati nelle classi di maggior sconto e ad una insufficiente penalizzazione dei "cattivi" assicurati.

Questa diminuzione comporta uno squilibrio finanziario del sistema con la conseguente necessità di adeguamenti ricorrenti dei premi. In tal modo, le imprese assicurative applicano agli assicurati una scala di premi differente da quella prevista alla stipula del contratto, contravvenendo così ai principi di trasparenza contrattuale.

I premi di equilibrio costituiscono un valido indicatore ai fini della valutazione di un sistema BM in quanto tengono conto oltre che della ripartizione degli assicurati nelle classi, anche degli esborsi e dei premi attesi. L'analisi dell'evoluzione dei premi di equilibrio permette di valutare gli incrementi di premio di riferimento richiesti per mantenere la stabilità finanziaria del sistema.

Il premio per un assicurato in classe *j*, nell'anno *t*, è ottenuto moltiplicando per il coefficiente di premio β_t un premio di riferimento o premio di base che tiene conto della classe tariffaria:

$$P_t * \beta_j$$

a cui corrisponde il cosiddetto coefficiente medio di premio:

$$b_t = \sum_{j=1}^{M} \beta_j \Pr\left(\tilde{Y}_j = j\right)$$

Il premio di equilibrio P^e, cioè il premio che garantisce la condizione di equilibrio tecnico tra i premi e i risarcimenti nel *t-esimo* anno, è la soluzione della seguente equazione:

$$\sum_{j=1}^{M} P_t \cdot \beta_j \Pr\left(\tilde{Y}_t = j\right) = E\left(\tilde{X}_t\right) \qquad j = 1,\ldots,M$$

Se i coefficienti medi di premio hanno andamento decrescente nel tempo, a parità di premio di riferimento, ad es. di primo anno P_1, nel generico anno t>1 si ha insufficienza dei premi, cioè $P_1 \cdot b_t < E\left(\tilde{X}_t\right)$, e quindi si deve ricorrere ad incrementi del premio di riferimento, manifestando quindi una incapacità del sistema adottato di premiare gli assicurati migliori.

Come è stato evidenziato nella sezione precedente, la popolazione assicurata, particolarmente nell'ipotesi di collettività chiusa, manifesta una rapida convergenza verso le classi di merito comportando quindi una riduzione progressiva dei coefficienti medi di premio[5].

[5] Ai fini del "riequilibrio" del sistema si ricorda ad es. il caso del Belgio nel 1971 (cfr [7] o il rapporto ISVAP "*L'assicurazione RCA in Italia: Analisi e proposte*", 1998).

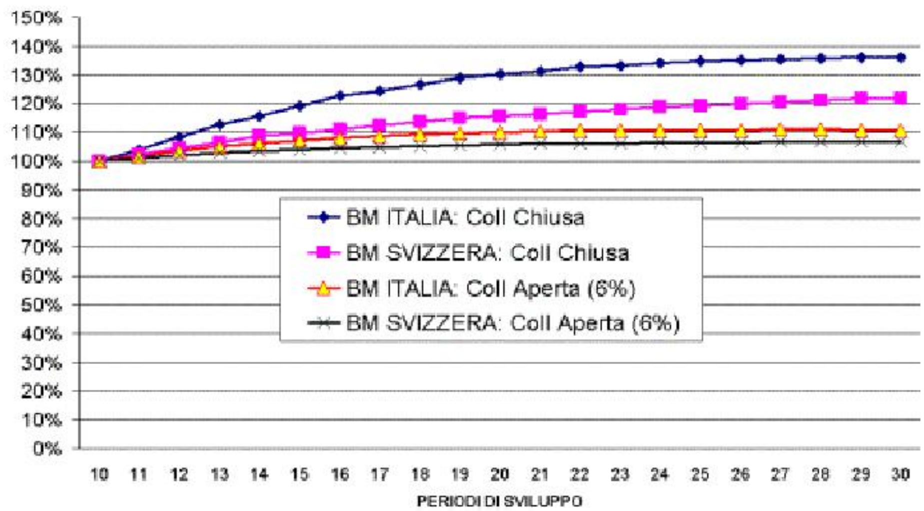

Grafico 1: Numeri indice dei rapporti dei premi di equilibrio tecnico

Per un confronto diretto dei sistemi BM italiano e svizzero è stato seguito un approccio di analisi già utilizzato in letteratura: sono stati considerati i numeri indice dati dai rapporti tra i premi di equilibrio alle epoche da 11 a 30 col premio all'epoca 10 (si veda Grafico 4). E'stata considerata tale epoca come rappresentativa di una possibile situazione iniziale al momento dell'adozione di un nuovo sistema (si ricordi il riferimento alla new company). Il sistema Svizzero presenta incrementi più contenuti rispetto a quello italiano ed è quindi quello che richiede di intervenire di meno sui premi di riferimento e che perciò riesce a premiare meglio gli assicurati migliori. Per le collettività chiuse tali rapporti sono molto più accentuati[6].

Un interessante confronto si ottiene <u>inflazionando</u>, secondo una legge di capitalizzazione composta, <u>il costo medio in base ad un tasso costante annuo del 5%.</u> In tal caso permangono le relazioni precedentemente evidenziate, ma cambia la concavità delle quattro curve dei numeri indice dei premi di equilibrio. Inoltre occorre evidenziare come i numeri indice siano caratterizzati da incrementi annui prossimi al 10% a fronte di uno scenario inflazionistico ipotizzato (5%) contraddistinto da tassi sicuramente inferiori a quelli che si sono registrati storicamente per il costo medio dei sinistri nel ramo RCA.

[6] Se si considerasse l'autoliquidazione, non tutti i sinistri sarebbero denunciati, aumenterebbero quindi le percentuali di assicurati nelle classi di bonus e infine i premi di equilibrio aumenterebbero.

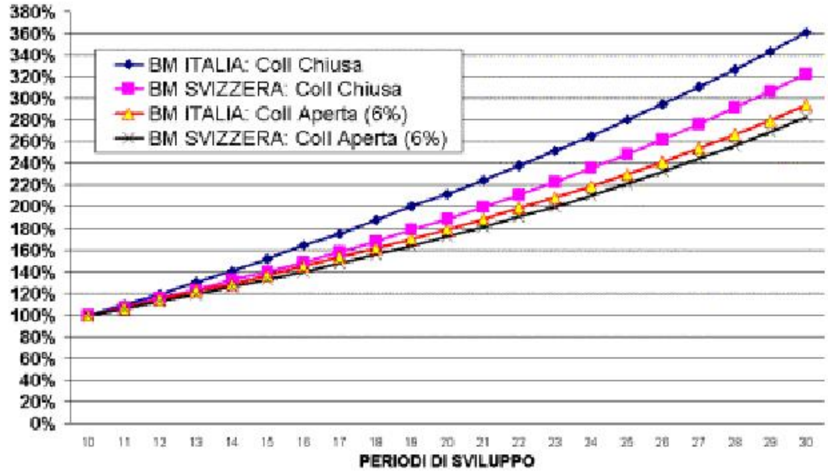

Grafico 2: Numeri indice dei rapporti dei premi di equilibrio tecnico a fronte di un costo medio dei sinistri inflazionato al 5% annuo

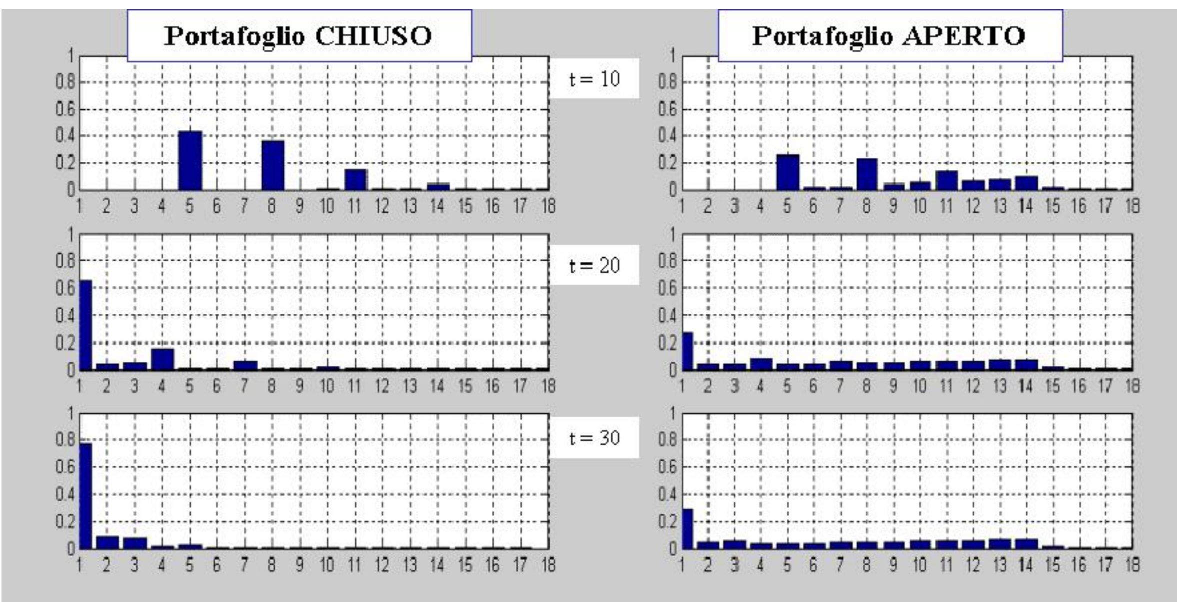

Grafico3 - Sistema Italiano: distribuzione percentuale degli assicurati nelle epoche t =10, 20 e 30; 1000 simulazioni; 100 assicurati