I principali schemi delle prove ripetute per rappresentare le insolvenze.

Raffaella Calabrese.

25 settembre 2006

1 Introduzione.

Il seguente lavoro si propone di analizzare i principali schemi probabilistici con indipendenza tra le serie. Inizialmente si introduce l'ipotesi di indipendenza tra le prove di ciascuna serie, nel qual caso si considerano gli schemi probabilistici di Bernoulli, Poisson, Lexis e Coolidge, evidenziando in particolare come le differenti assunzioni presenti in tali schemi conducano a caratteristiche di dispersione diverse, rispettivamente binomiale, ipo-binomiale, iper-binomiale e asintoticamente iper-binomiale.
Si suppone, in seguito, che esista una forma di dipendenza (lineare) tra le prove di ciascuna serie, considerando solamente il caso in cui tale dipendenza si manifesti in maniera uniforme. Analizzando il legame che sussiste tra il coefficiente di correlazione lineare e la dispersione di uno schema, si ottiene che, anche in tale contesto, la dispersione puó essere di tipo binomiale, ipo-binomiale e iper-binomiale, a seconda che tra le prove la correlazione sia rispettivamente nulla, negativa e positiva.
Conoscendo la variabilitá delle frequenze relative delle serie é possibile costruire, grazie al quoziente di divergenza di Lexis, un test per verificare l'ipotesi di dispersione binomiale.
Poiché gli schemi probabilistici rappresentano fenomeni dicotomici, é interessante applicare tale metodologia all'analisi delle insolvenze nel rischio di credito, potendo cosí valutare la dispersione di un portafoglio di crediti.

2 Gli schemi probabilistici con indipendenza tra le prove di una serie e indipendenza tra le serie.

Si supponga di essere interessati all'ottenimento di un evento A (successo) in k serie di n_j prove ciascuna con $j = 1, 2, ..., k$. Per i successivi risultati sará fondamentale l'assunzione di indipendenza sia tra le k serie che tra le n_j prove di ciascuna serie. Sia dunque A_{ji} la variabile casuale indicatore associata alla i-ma prova della j-ma serie, con $i = 1, 2, ..., n_j$ e $j = 1, 2, ..., k$

$$A_{ji} = \begin{cases} 1 & \text{se nella } i\text{-ma prova della } j\text{-ma serie si verifica l'evento } A \\ 0 & \text{altrimenti} \end{cases} \tag{2.1}$$

avente le seguenti probabilitá di successo ed insuccesso

$$P\{A_{ji} = 1\} = p_{ji} \qquad P\{A_{ji} = 0\} = 1 - p_{ji} = q_{ji}.$$

Si definiscano, inoltre, le variabile casuali

$$X_j = \sum_{i=1}^{n_j} A_{ji}$$

che indica il numero di volte che l'evento A si verifica nelle n_j prove della j-ma serie e

$$X = \sum_{j=1}^{k} X_j = \sum_{j=1}^{k} \sum_{i=1}^{n_j} A_{ji}$$

che rappresenta il numero di volte che l'evento A si é presentato nelle $n = \sum_{j=1}^{k} n_j$ prove. Per le precedenti assunzioni le n variabili casuali indicatore A_{ji} sono dunque mutuamente indipendenti.
La frequenza relativa dell'evento A nelle n_j prove della j-ma serie puó essere rappresentata mediante la variabile casuale

$$\widehat{p}_j = \frac{X_j}{n_j},$$

mentre la frequenza relativa dell'evento A sul totale delle n prove risulta

$$\widehat{p} = \frac{X}{n} = \frac{1}{n} \sum_{j=1}^{k} \left(\frac{X_j}{n_j} n_j \right) = \frac{1}{n} \sum_{j=1}^{k} \widehat{p}_j n_j, \tag{2.2}$$

che coincide con la media aritmetica ponderata delle frequenze relative delle k serie con pesi pari a n_j. Le variabili cosí definite presentano dunque le seguenti aspettative e varianze

$$\begin{aligned} \mathbb{E}(A_{ji}) = p_{ji} \qquad & \mathbb{V}(A_{ji}) = p_{ji}(1 - p_{ji}) \\ \mathbb{E}(X_j) = \sum_{i=1}^{n_j} \mathbb{E}(A_{ji}) = \sum_{i=1}^{n_j} p_{ji} \qquad & \mathbb{V}(X_j) = \sum_{i=1}^{n_j} \mathbb{V}(A_{ji}) = \sum_{i=1}^{n_j} p_{ji}(1 - p_{ji}). \end{aligned} \tag{2.3}$$

Per il calcolo della varianza di X_j si é impiegata, nella precedente equazione, l'ipotesi di indipendenza tra le prove.

$$\mathbb{E}(X) = \sum_{j=1}^{k} \mathbb{E}(X_j) = \sum_{j=1}^{k} \sum_{i=1}^{n_j} p_{ji} \qquad \mathbb{V}(X) = \sum_{j=1}^{k} \mathbb{V}(X_j) = \sum_{j=1}^{k} \sum_{i=1}^{n_j} p_{ji}(1 - p_{ji}). \tag{2.4}$$

Per determinare la varianza di X si é utilizzata, invece, l'assunzione di indipendenza tra le serie.

$$\mathbb{E}(\widehat{p}_j) = \frac{1}{n_j} \mathbb{E}(X_j) = \frac{1}{n_j} \sum_{i=1}^{n_j} p_{ji} \tag{2.5}$$

$$\mathbb{V}(\widehat{p}_j) = \frac{1}{n_j^2} \mathbb{V}(X_j) = \frac{1}{n_j^2} \sum_{i=1}^{n_j} p_{ji}(1 - p_{ji}) \tag{2.6}$$

$$\mathbb{E}(\widehat{p}) = \sum_{j=1}^{k} \frac{n_j}{n} \mathbb{E}(\widehat{p}_j) = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} p_{ji} \tag{2.7}$$

$$\mathbb{V}(\widehat{p}) = \sum_{j=1}^{k} \left(\frac{n_j}{n} \right)^2 \mathbb{V}(\widehat{p}_j) = \frac{1}{n^2} \sum_{j=1}^{k} \sum_{i=1}^{n_j} p_{ji}(1 - p_{ji}) \tag{2.8}$$

Gli schemi probabilistici con indipendenza sia tra le prove sia tra le serie prevedono dunque l'esecuzione di k serie di n_j prove ciascuna. Tali schemi possono essere classificati a seconda delle condizioni in cui si effettuano tali prove, che influiscono sulle probabilitá di successo p_{ji}:

- nello *schema di Bernoulli* le prove si svolgono nelle medesime circostanze e quindi con le medesime probabilitá $p_{ji} = p$;

- nello *schema di Poisson* si suppone che le probabilitá di successo p_{ji} varino all'interno della stessa serie, mentre si assumono costanti da serie a serie
  - le probabilitá medie
$$\overline{p}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} p_{ji} = \overline{p} \qquad j = 1, 2, ..., k \tag{2.9}$$
  - le varianze
$$\mathbb{V}_j(p_{ji}) = \frac{1}{n_j}\sum_{i=1}^{n_j}(p_{ji} - \overline{p}_j)^2 = \sigma_j^2(p) = \sigma^2(p) \quad j = 1, 2, ..., k; \tag{2.10}$$
- nello *schema di Lexis* le condizioni rimangono costanti in ciascuna serie ma variano da serie a serie, questo significa che $p_{ji} = p_j$, per $i = 1, 2, ..., n_j$ e $j = 1, 2, ..., k$;
- nello *schema di Coolidge* (Pompilj (1967), Zenga (1968)) o *schema misto* (Faleschini (1949)) le probabilitá di successo p_{ji} variano da prova a prova e da serie a serie.

2.1 Lo schema probabilistico di Bernoulli.

Nello schema probabilistico di Bernoulli si assume che la probabilitá di successo sia costante da prova a prova e da serie a serie

$$p_{ji} = p \qquad \text{per} \quad i = 1, 2, ..., n_j \quad \text{e} \quad j = 1, 2, ..., k.$$

Sotto tali condizioni le variabili casuali indicatore A_{ji} sono indipendenti ed identicamente distribuite con parametro comune p e i valori attesi e le varianze delle variabili casuali X_j e X, calcolate nelle equazioni (2.3) e (2.4), diventano

$$\mathbb{E}(X_j) = n_j p \qquad \mathbb{V}(X_j) = \sum_{i=1}^{n_j} \mathbb{V}(A_{ji}) = n_j p(1-p) \tag{2.11}$$
$$\mathbb{E}(X) = np \qquad \mathbb{V}(X) = np(1-p). \tag{2.12}$$

L'aspettativa e la varianza della frequenza relativa $\widehat{p}$, per il cui calcolo conviene determinare la speranza matematica e la varianza di $\widehat{p}_j$, in uno schema di Bernoulli risultano

$$\mathbb{E}(\widehat{p}_j) = p \qquad \mathbb{V}(\widehat{p}_j) = \mathbb{E}\left\{\left(\frac{X_j}{n_j} - p\right)^2\right\} = \frac{p(1-p)}{n_j} \tag{2.13}$$

$$\mathbb{E}(\widehat{p}) = p \qquad \mathbb{V}(\widehat{p}) = \sum_{j=1}^{k}\left(\frac{n_j}{n}\right)^2 \frac{p(1-p)}{n_j} = \frac{p(1-p)}{n^2}\sum_{j=1}^{k} n_j = \frac{pq}{n}. \tag{2.14}$$

Si consideri ora l'aspettativa dello scarto semplice e di quello quadratico tra la frequenza relativa empirica e quella teorica dell'evento A riguardante la serie j-ma, ponderato con peso n_j

$$\mathbb{E}\left[\left(\frac{X_j}{n_j} - p\right) n_j\right] = n_j \mathbb{E}\left(\frac{X_j}{n_j} - p\right) = 0$$
$$\mathbb{E}\left[\left(\frac{X_j}{n_j} - p\right)^2 n_j\right] = n_j \mathbb{V}(\widehat{p}_j) = pq.$$

Ricorrendo alle note proprietá del valore atteso, si deduce dalla precedente equazione il seguente risultato [1]

$$\mathbb{E}\left[\sum_{j=1}^{k}\left(\frac{X_j}{n_j} - p\right)^2 n_j\right] = \sum_{j=1}^{k} n_j \left[\mathbb{E}\left(\widehat{p}_j - p\right)^2\right] = \sum_{j=1}^{k} n_j \mathbb{V}(\widehat{p}_j) = kpq. \tag{2.15}$$

Lo schema di Bernoulli é definito *schema a dispersione normale* (Boldrini (1968)) o *binomiale* (Zenga (2003)).
In tale schema probabilistico X_j é una variabile casuale binomiale che, per n_j elevato, puó essere approssimata ad una normale con media e varianza date dalle equazioni (2.11) (Johnson, Kemp e Kotz (2005), pp. 116). Anche la frequenza relativa $\widehat{p}_j$ della serie j-ma si puó analogamente approssimare con una normale avente media e varianza date dalla (2.13). Questo significa che la seguente variabile casuale

$$\frac{\widehat{p}_j - E(\widehat{p}_j)}{\sigma(\widehat{p}_j)} = \frac{\widehat{p}_j - p}{\sigma(\widehat{p}_j)} = \frac{(\widehat{p}_j - p)\sqrt{n_j}}{\sqrt{pq}}$$

[1] Si noti che la quantitá casuale di cui si calcola l'aspettativa rappresenta una sorta di 'devianza' delle frequenze relative $\widehat{p}_j$.

si puó approssimare, al devergere di n_j, con una normale standard. Ne consegue che la quantitá casuale

$$\frac{(\widehat{p}_j - p)^2}{pq} n_j$$

si approssima, al divergere di n_j, con una Chi-Quadrato con un grado di libertá e quindi

$$\sum_{j=1}^{k} \frac{(\widehat{p}_j - p)^2}{pq} n_j, \tag{2.16}$$

essendo la somma di k variabili casuali indipendenti, per l'assunzione di indipendenza tra le serie, si distribuisce, in modo approssimato, al divergere di n_j, secondo una legge Chi-Quadrato con k gradi di libertá. Considerazioni analoghe applicate alla frequenza relativa $\widehat{p}$, avente aspettativa e varianza date dalle equazioni (2.14), permettono di affermare che la seguente variabile casuale

$$\left[\frac{(\widehat{p} - p)\sqrt{n}}{\sqrt{pq}}\right]^2 = \frac{(\widehat{p} - p)^2}{pq} n \tag{2.17}$$

si puó approssimare, al divergere di n, con una Chi-Quadrato con un grado di libertá. Nelle quantitá casuali definite dalle espressioni (2.16) e (2.17) é presente la probabilitá di successo p, il cui valore é solitamente ignoto. Per tale motivo conviene modificare le suddette quantitá casuali, affinché diventino funzioni di parametri noti.
Dalla scomposizione della devianza si deduce la seguente relazione

$$\sum_{j=1}^{k} (\widehat{p}_j - \widehat{p})^2 n_j = \sum_{j=1}^{k} (\widehat{p}_j - p)^2 n_j - n(\widehat{p} - p)^2.$$

Dividendo ambo i membri della precedente equazione per il fattore pq si ricava

$$\sum_{j=1}^{k} \frac{(\widehat{p}_j - \widehat{p})^2}{pq} n_j = \sum_{j=1}^{k} \frac{(\widehat{p}_j - p)^2}{pq} n_j - \frac{(\widehat{p} - p)^2}{pq} n.$$

Avendo dimostrato in precedenza che la prima quantitá casuale presente nel lato destro della precedente uguaglianza si distribuisce asintoticamente come una Chi-Quadrato con k gradi di libertá e che la seconda quantitá casuale, considerando sempre il lato destro della suddetta equazione, si distribuisce, in modo asintotico, come una Chi-Quadrato con un grado di libertá, allora, per la proprietá associativa della variabile casuale Chi-Quadrato, si deduce che la seguente espressione

$$\sum_{j=1}^{k} \frac{(\widehat{p}_j - \widehat{p})^2}{pq} n_j \tag{2.18}$$

si puó approssimare, al divergere delle numerositá n_j, con una Chi-Quadrato con $(k-1)$ gradi di libertá. Dalla convergenza in probabilitá della frequenza relativa $\widehat{p}$, sul totale delle n prove, al parametro ignoto p, si ricava che il rapporto $(pq)/(\widehat{pq})$ converge in probabilitá ad uno. Tenendo conto di tale risultato ed applicando il teorema di Slutsky [2] alla seguente variabile casuale

$$\sum_{j=1}^{k} \frac{(\widehat{p}_j - \widehat{p})^2}{pq} n_j \frac{pq}{\widehat{pq}} = \sum_{j=1}^{k} \frac{(\widehat{p}_j - \widehat{p})^2}{\widehat{pq}} n_j, \tag{2.19}$$

si ottiene che la quantitá casuale definita nel secondo membro della suddetta equazione, che dipende esclusivamente da parametri noti, tende in distribuzione, al divergere delle numerositá n_j, ad una variabile casuale Chi-Quadrato con $(k-1)$ gradi di libertá.

2.2 Lo schema probabilistico di Poisson.

Nel 1830 Poisson formalizzó lo schema delle prove ripetute in condizioni di indipendenza con probabilitá di successo p_{ji} variabili da prova a prova nell'ambito della stessa serie. Lo schema probabilistico che da tale autore ha preso il nome, come giá accennato in precedenza, considera costanti da serie a serie sia le medie parziali $\overline{p}_j = \overline{p}$, con $j = 1, 2, ..., k$, sia le varianze $\sigma_j^2(p) = \sigma^2(p)$ fra le probabilitá delle prove di ciascuna serie [3], con $j = 1, 2, ..., k$, grandezze che sono state definite in precedenza nelle equazioni (2.9) e (2.10). Nello schema di Poisson il numero casuale di volte in cui si verifica l'evento A

[2] Siano $\{X_n\}$, $\{Y_n\}$ due successioni di variabili casuali e sia θ una costante reale. Se $\{X_n\}$ tende in distribuzione alla variabile casuale X e $\{Y_n\}$ converge in probabilitá a θ, allora $(X_n Y_n) \xrightarrow{d} \theta X$ e $(X_n \pm Y_n) \xrightarrow{d} (X \pm \theta)$. Per la dimostrazione di tale teorema si veda Cramer (1996) pp. 254-255 e Rohatgi (1976) pp. 253-254.

[3] Le caratteristiche dello schema probabilistico di Poisson coincidono con quelle di uno schema di campionamento stratificato in cui si estrae una sola unitá campionaria da ciascuno strato della popolazione (Pompilj (1967), Cap. V).

nella serie j-ma ha aspettativa e varianza che, dalle equazioni (2.3), risultano essere

$$\mathbb{E}(X_j) = n_j\overline{p} \qquad \mathbb{V}(X_j) = \sum_{i=1}^{n_j} p_{ji} - \sum_{j=1}^{n_j} p_{ji}^2. \tag{2.20}$$

Per il calcolo della suddetta varianza conviene considerare lo scarto λ_{ji} fra la probabilitá di sucesso della prova i-ma della j-ma serie e la media generale

$$\lambda_{ji} = p_{ji} - \overline{p} \qquad i = 1, 2, ..., n_j \quad \text{e} \quad j = 1, 2, ..., k$$

cosí risulta che

$$p_{ji} = \overline{p} + \lambda_{ji}.$$

Sostituendo il precedente risultato nell'ultimo termine dell'equazione (2.20) si ottiene

$$\begin{aligned}\sum_{i=1}^{n_j} p_{ji}^2 &= \sum_{i=1}^{n_j} (\overline{p} + \lambda_{ji})^2 = \sum_{i=1}^{n_j} (\overline{p}^2 + \lambda_{ji}^2 + 2\overline{p}\lambda_{ji}) \\ &= n_j\overline{p}^2 + \sum_{i=1}^{n_j} \lambda_{ji}^2 + 2\overline{p}\sum_{i=1}^{n_j} \lambda_{ji} = n_j\overline{p}^2 + n_j\sigma^2(p) + 2\overline{p}\sum_{i=1}^{n_j} \lambda_{ji},\end{aligned}$$

nell'ultimo passaggio si é tenuto conto dell'assunzione di uguaglianza delle varianze $\sigma_j^2(p) = \sigma^2(p)$ fra le probabilitá delle prove di ciascuna serie. Dalla ulteriore ipotesi che le probabilitá medie $\overline{p}_j$ delle k serie siano costanti, si deduce che l'ultimo termine $\sum_{i=1}^{n_j} \lambda_{ji}$ della precedente equazione é nullo, sostituendo cosí tale risultato nel secondo membro della (2.20) si ricava

$$\mathbb{V}(X_j) = n_j\overline{p}(1 - \overline{p}) - n_j\sigma^2(p). \tag{2.21}$$

Dalle equazioni (2.20) e (2.21) si deduce, dunque, l'aspettativa e la varianza del numero casuale di volte in cui si verifica l'evento dicotomico A in un totale di n prove

$$\mathbb{E}(X) = n\overline{p} \qquad \mathbb{V}(X) = n\overline{p}(1 - \overline{p}) - n\sigma^2(p),$$

per il calcolo della varianza si é considerata l'ipotesi di indipendenza tra le serie.

Dai precedenti risultati é possibile determinare l'aspettativa e la varianza della frequenza relativa $\widehat{p}$ dei successi nelle n prove complessive

$$\mathbb{E}(\widehat{p}) = \overline{p} \qquad \mathbb{V}(\widehat{p}) = \frac{\overline{p}(1 - \overline{p})}{n} - \frac{\sigma^2(p)}{n}$$

Confrontando i precedenti risultati con le rispettive grandezze (2.14) calcolate in precedenza nello schema di Bernoulli con n prove e probabilitá costante pari a $\overline{p}$, si comprende che nello schema di Poisson si ha la stessa aspettativa della frequenza relativa $\widehat{p}$, mentre la varianza risulta minore.
Analogamente allo schema di Bernoulli, si calcola l'aspettativa della somma ponderata degli scarti al quadrato tra le frequenze relative $\widehat{p}_j$ e la probabilitá media complessiva $\overline{p}$, con pesi pari alle numerositá n_j

$$\begin{aligned} \mathbb{E}\left[\sum_{j=1}^{k}\left(\frac{X_j}{n_j}-\overline{p}\right)^2 n_j\right] &= \sum_{j=1}^{k} n_j\left[\mathbb{E}\left(\widehat{p}_j-\overline{p}\right)^2\right] \\ &= \sum_{j=1}^{k} n_j\left(\frac{1}{n_j^2}\right)\mathbb{V}(X_j) = k\overline{p}(1-\overline{p}) - k\sigma^2(p). \end{aligned} \tag{2.22}$$

Confrontando tale risultato con quello ottenuto (2.15) nello schema di Bernoulli con probabilitá di successo costante pari a $\overline{p}$, si comprende perché lo schema di Poisson é definito *schema a dispersione subnormale* (Boldrini (1968)) o *ipo-binomiale* (Zenga (2003)). La dispersione dello schema di Poisson dipende dunque dalla variabilitá $\sigma^2(p)$ fra le probabilitá di una serie, in particolare piú questa risulta elevata, tanto minore sará l'aspettativa della 'devianza' delle frequenze relative $\widehat{p}_j$.
Prima di analizzare il successivo schema probabilistico é interessante considerare la generalizzazione proposta da Faleschini (1949) dello schema di Poisson, il quale dimostra che, affinché uno schema abbia dispersione ipobinomiale, é necessario che le probabilitá di successo p_{ji} varino da prova a prova nell'ambito della stessa serie e che le probabilitá medie $\overline{p}_j$ delle diverse serie siano uguali, per $j = 1, 2, ..., k$, ma al contrario non é necessaria l'ipotesi di uguaglianza delle varianze $\sigma_j^2(p)$ delle diverse serie. Tale modello é definito da Faleschini 'schema di Poisson generalizzato'.

2.3 Lo schema probabilistico di Lexis.

Lexis propose nel 1876 il seguente schema probabilistico, che da lui prese il nome, nel quale le probabilitá di successo p_{ji} rimangono costanti nell'ambito della stessa serie $p_{ji} = p_j$, con $i = 1, 2, ..., n_j$ e $j = 1, 2, ..., k$, ma variano da serie a serie[4]. Sia dunque

$$\overline{p} = \sum_{j=1}^{k} p_j \frac{n_j}{n}$$

la probabilitá media di successo, ottenuta come media aritmetica ponderata delle probabilitá di successo delle singole serie con pesi pari alle numerositá n_j e sia

$$\sigma^2(p_j) = \sum_{j=1}^{k} (p_j - \overline{p})^2 \frac{n_j}{n}$$

la varianza fra le probabilitá delle diverse serie.
Dalle ipotesi che caratterizzano lo schema probabilistico di Lexis si deduce che X_j é una variabile casuale binomiale avente parametri p_j e n_j, per $j = 1, 2, ..., k$. Si determina dunque facilmente sia l'aspettativa che la varianza della frequenza relativa $\widehat{p}_j$ della j-ma serie

$$\mathbb{E}(\widehat{p}_j) = \frac{1}{n_j}\mathbb{E}(X_j) = p_j \qquad \mathbb{V}(\widehat{p}_j) = \frac{1}{n_j^2}\mathbb{V}(X_j) = \frac{p_j(1-p_j)}{n_j}. \tag{2.23}$$

Essendo

$$\mathbb{E}(\widehat{p}) = \frac{1}{n}\sum_{j=1}^{k} \mathbb{E}(\widehat{p}_j) n_j = \overline{p} \qquad \mathbb{V}(\widehat{p}) = \frac{1}{n^2}\sum_{j=1}^{k} \mathbb{V}(\widehat{p}_j) n_j^2 = \frac{1}{n^2}\sum_{j=1}^{k} n_j p_j (1-p_j),$$

si deduce che, analogamente allo schema di Poisson, anche lo schema di Lexis presenta dunque la stessa aspettativa di uno schema di Bernoulli, nel quale la probabilitá di successo é costante e pari a $\overline{p}$ durante l'esecuzione di tutte le n prove.
Per analizzare la dispersione che caratterizza lo schema di Lexis conviene considerare l'aspettativa della seguente quantitá casuale

$$\begin{aligned}\mathbb{E}\left[\left(\frac{X_j}{n_j} - \overline{p}\right)^2 n_j\right] &= n_j \mathbb{E}\left[(\widehat{p}_j - p_j + p_j - \overline{p})^2\right] \\ &= n_j \mathbb{E}\left[(\widehat{p}_j - p_j)^2\right] + n_j (p_j - \overline{p})^2 + 2n_j (p_j - \overline{p})\mathbb{E}(\widehat{p}_j - p_j).\end{aligned} \tag{2.24}$$

[4]Uno schema probabilistico di Laplace-Lexis, secondo la definizione di Pompilj, rappresenta un particolare schema di campionamento a due stadi (Pompilj (1967), Cap. VI).

L'ultima aspettativa della precedente equazione risulta nulla ed, inoltre, dai risultati (2.23) si deduce che

$$\mathbb{E}\left[(\widehat{p}_j - p_j)^2\right] = \mathbb{E}\left\{[\widehat{p}_j - \mathbb{E}(\widehat{p}_j)]^2\right\} = \mathbb{V}(\widehat{p}_j) = \frac{p_j(1-p_j)}{n_j},$$

ottenendo infine

$$\mathbb{E}\left[\left(\frac{X_j}{n_j} - \overline{p}\right)^2 n_j\right] = p_j(1-p_j) + n_j(p_j - \overline{p})^2. \tag{2.25}$$

Per un possibile confronto con la dispersione di uno schema di Bernoulli con probabilitá di successo costante pari a $\overline{p}$, é necessario scomporre il fattore $p_j(1-p_j)$. Considerando lo scarto tra le probabilitá di successo della serie j-ma e la probabilitá media generale

$$\lambda_j = p_j - \overline{p} \qquad j = 1, 2, ..., k,$$

si ricava la seguente relazione

$$p_j(1-p_j) = (\overline{p} + \lambda_j)(1 - \overline{p} - \lambda_j) = \overline{p}(1-\overline{p}) + \lambda_j(1 - 2\overline{p}) - \lambda_j^2,$$

che, sommata rispetto all'indice j per tutte le k serie, conduce al seguente risultato

$$\sum_{j=1}^{k} p_j(1-p_j) = k\overline{p}(1-\overline{p}) + (1-2\overline{p})\sum_{j=1}^{k}\lambda_j - \sum_{j=1}^{k}\lambda_j^2.$$

Tenendo in considerazione l'equazione (2.25) si ricava infine

$$\mathbb{E}\left[\sum_{j=1}^{k}\left(\frac{X_j}{n_j} - \overline{p}\right)^2 n_j\right] = k\overline{p}(1-\overline{p}) + (1-2\overline{p})\sum_{j=1}^{k}\lambda_j + \sum_{j=1}^{k}(p_j - \overline{p})^2(n_j - 1).$$

Definendo

$$\widetilde{p} = \frac{1}{k}\sum_{j=1}^{k} p_j$$

la media semplice (non ponderata) delle probabilitá di successo delle diverse serie, é possibile riscrivere la precedente equazione

$$\mathbb{E}\left[\sum_{j=1}^{k}\left(\frac{X_j}{n_j} - \overline{p}\right)^2 n_j\right] = k\overline{p}(1-\overline{p}) + k(1-2\overline{p})(\widetilde{p} - \overline{p}) + \sum_{j=1}^{k}(p_j - \overline{p})^2(n_j - 1). \tag{2.26}$$

Al divergere delle numerositá n_j, l'ultima sommatoria tende all'infinito. Adesso diviene possibile confrontare il risultato ottenuto con quello (2.15) ricavato in precedenza nello schema di Bernoulli con probabilitá di successo costante pari a $\overline{p}$. Lo schema di Lexis presenta dunque una dispersione *supernormale* (Boldrini (1968)) o *iper-binomiale* (Zenga (2003)).
Tale conclusione appare evidente se si considerano k serie aventi numerositá costante pari ad m ($n_j = m$ per $j = 1, 2, ..., k$): coincidono le medie $\widetilde{p} = \overline{p}$ e il precedente risultato (2.26) diventa quindi

$$\mathbb{E}\left[\sum_{j=1}^{k}\left(\frac{X_j}{m} - \overline{p}\right)^2 m\right] = k\overline{p}(1-\overline{p}) + (m-1)\sum_{j=1}^{k}(p_j - \overline{p})^2.$$

2.4 Lo schema probabilistico di Coolidge.

Si consideri infine lo schema probabilistico proposto da Coolidge nel 1921, che rappresenta una generalizzazione degli schemi di prove ripetute analizzati in precedenza, dal momento che le probabilitá di successo p_{ji} sono libere di variare sia da prova a prova che da serie a serie.
Per ricavare le caratteristiche della variabile casuale X associata allo schema di Coolidge conviene procedere nel modo proposto da Zenga (1968): si associa inizialmente ad ogni serie la variabile casuale X_j dello schema probabilistico di Poisson e si procede in seguito al miscuglio delle k variabili individuate con pesi pari alle numerositá n_j di ciascuna serie. Tale procedimento permette dunque di utilizzare alcuni dei risultati ricavati in precedenza, ai quali bisogna apportare alcune modifiche.
Dalle equazioni iniziali (2.3) risulta

$$\mathbb{E}(X_j) = \sum_{i=1}^{n_j} p_{ji} = n_j\overline{p}_j \qquad \mathbb{V}(X_j) = \sum_{i=1}^{n_j} p_{ji}(1-p_{ji}) = \sum_{i=1}^{n_j} p_{ji} - \sum_{i=1}^{n_j} p_{ji}^2 \quad (2.27)$$

con $j = 1, 2, ..., k$. Analogamente al procedimento seguito nello schema probabilistico di Poisson, si definisce

$$\lambda_{ji} = p_{ji} - \overline{p} \qquad i = 1, 2, ..., n_j \text{ e } j = 1, 2, ..., k$$

lo scarto tra la probabilitá di successo della i-ma prova e j-ma serie e la probabiltá media totale. Ricavando dalla precedente uguaglianza p_{ji} si ottiene

$$\begin{aligned}\sum_{i=1}^{n_j} p_{ji}^2 = \sum_{i=1}^{n_j} (\overline{p} + \lambda_{ji})^2 = \sum_{i=1}^{n_j} (\overline{p}^2 + \lambda_{ji}^2 + \\ n_j \overline{p}^2 + \sum_{i=1}^{n_j} \lambda_{ji}^2 + 2\overline{p} \sum_{i=1}^{n_j} \lambda_{ji}.\end{aligned} \tag{2.28}$$

Si analizzano i singoli addendi della precedente equazione

$$\begin{aligned}\sum_{i=1}^{n_j} \lambda_{ji}^2 &= \sum_{i=1}^{n_j} (p_{ji} - \overline{p})^2 = \sum_{i=1}^{n_j} (p_{ji} - \overline{p}_j + \overline{p}_j - \overline{p})^2 \\ &= \sum_{i=1}^{n_j} (p_{ji} - \overline{p}_j)^2 + \sum_{i=1}^{n_j} (\overline{p}_j - \overline{p})^2 = \sum_{i=1}^{n_j} (p_{ji} - \overline{p}_j)^2 + n_j (\overline{p}_j - \overline{p})^2.\end{aligned}$$

Nello sviluppo del precedente quadrato si trascura il doppio prodotto in quanto nullo.

$$\sum_{i=1}^{n_j} \lambda_{ji} = \sum_{i=1}^{n_j} (p_{ji} - \overline{p}) = \sum_{i=1}^{n_j} p_{ji} - n_j \overline{p} = n_j (\overline{p}_j - \overline{p}).$$

Sostituendo i precedenti risultati nella (2.28) si ottiene

$$\sum_{i=1}^{n_j} p_{ji}^2 = n_j \overline{p}^2 + \sum_{i=1}^{n_j} (p_{ji} - \overline{p}_j)^2 + n_j (\overline{p}_j - \overline{p})^2 + 2\overline{p} n_j (\overline{p}_j - \overline{p}).$$

Riportando tale risultato nella seconda equazione della (2.27) si ricava

$$\mathbb{V}(X_j) = n_j \overline{p}_j - n_j \overline{p}^2 - \sum_{i=1}^{n_j} (p_{ji} - \overline{p}_j)^2 - n_j (\overline{p}_j - \overline{p})^2 - 2\overline{p} n_j (\overline{p}_j - \overline{p}).$$

Analogamente agli schemi analizzati in precedenza si calcola l'aspettativa della seguente quantitá casuale

$$
\begin{aligned}
&\mathbb{E}\left[\left(\frac{X_j}{n_j}-\overline{p}\right)^2 n_j\right]=\mathbb{E}\left[\left(\frac{X_j}{n_j}-\overline{p}_j+\overline{p}_j-\overline{p}\right)^2 n_j\right]\\
&=\mathbb{E}\left[\left(\frac{X_j}{n_j}-\overline{p}_j\right)^2 n_j\right]+n_j(\overline{p}_j-\overline{p})^2\\
&=\frac{1}{n_j}\mathbb{E}\left\{[X_j-\mathbb{E}(X_j)]^2\right\}+n_j(\overline{p}_j-\overline{p})^2=\frac{1}{n_j}\mathbb{V}(X_j)+n_j(\overline{p}_j-\overline{p})^2\\
&=\overline{p}_j-\overline{p}^2-\frac{\sum_{i=1}^{n_j}(p_{ji}-\overline{p}_j)^2}{n_j}-2\overline{p}(\overline{p}_j-\overline{p})+(\overline{p}_j-\overline{p})^2(n_j-1)\\
&=\overline{p}_j-\overline{p}^2-\sigma_j^2(p)-2\overline{p}(\overline{p}_j-\overline{p})+(\overline{p}_j-\overline{p})^2(n_j-1).
\end{aligned}
\tag{2.29}
$$

Nella seconda equazione della (2.29) si é trascurato il doppio prodotto in quanto nullo.
A questo punto si calcola l'aspettativa del miscuglio, rappresentato dalla somma degli scarti elevati al quadrato tra la frequenza relativa di ogni serie e la probabilitá media totale $\overline{p}$, considerando come pesi le numerositá n_j di ciascuna serie

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{j=1}^{k}\left(\frac{X_j}{n_j}-\overline{p}\right)^2 n_j\right]=\sum_{j=1}^{k}\left\{\mathbb{E}\left[\left(\frac{X_j}{n_j}-\overline{p}\right)^2 n_j\right]\right\}\\
&=\sum_{j=1}^{k}\overline{p}_j-k\overline{p}^2-\sum_{j=1}^{k}\sigma_j^2(p)+\sum_{j=1}^{k}(\overline{p}_j-\overline{p})^2(n_j-1)-2\overline{p}\sum_{j=1}^{k}(\overline{p}_j-\overline{p}).
\end{aligned}
\tag{2.30}
$$

A seconda delle varie assunzioni dei diversi schemi probabilistici, la precedente espressione contiene come casi particolari i risultati ricavati per lo schema di Bernoulli, di Poisson e di Lexis. Dalla assunzione di uguaglianza delle probabilitá medie di ciascuna serie $\overline{p}_j$ dello schema di Poisson generalizzato proposto da Faleschini (1949), si deduce che le ultime due sommatorie della precedente equazione sono nulle, e dunque anche tale schema probabilistico, analogamente a quello di Poisson (semplice), presenta dispersione ipo-binomiale, come anticipato in precedenza.
Per poter effettuare delle considerazioni sul precedente risultato conviene considerare uno schema di Coolidge composto da k serie avente tutte numerositá costante pari a m, in tale circostanza l'ultima sommatoria dell'equazione

(2.30) diventa nulla e quindi risulta

$$\mathbb{E}\left[\sum_{j=1}^{k}\left(\frac{X_j}{m}-\overline{p}\right)^2 m\right]=k\overline{p}(1-\overline{p})+(m-1)\sum_{j=1}^{k}(\overline{p}_j-\overline{p})^2-\sum_{j=1}^{k}\sigma_j^2(p).$$

Tale risultato coincide con quello ottenuto da Faleschini (1949). Nell'esame della precedente espressione si trova conferma della spiegazione fornita da Boldrini (1968) alla diffusa convinzione secondo cui i fenomeni empirici presentano per la maggior parte dispersione iper-binomiale.
Nello schema di Coolidge gli ultimi due addendi della precedente equazione sono entrambi diversi da zero, ma essendo le due sommatorie $\sum_{j=1}^{k}(\overline{p}_j-\overline{p})^2$ e $\sum_{j=1}^{k}\sigma_j^2(p)$ dello stesso ordine di grandezza, al divergere di m, la componente positiva prevale su quella negativa, ottenendo dunque uno schema di dispersione iper-binomiale, anche nel caso in cui non siano soddisfatte le ipotesi dello schema probabilistico di Lexis. Di conseguenza, affinché si manifesti un fenomeno con dispersione ipo-binomiale, devono essere soddisfatte entrambe le assunzioni dello schema di Poisson, cioé le probabilitá devono variare all'interno della medesima serie, ma le probabilitá medie $\overline{p}_j$ e le varianze $\sigma_j^2(p)$ devono rimanere costanti da serie a serie. Per rilevare dispersione iper-binomiale non occorre invece che le probabilitá di successo rimangano costanti da prova a prova in ciascuna serie, purché esse varino da serie a serie.
Dal momento che, a livello empirico, raramente le probabilitá medie $\overline{p}_j$ e le varianze $\sigma_j^2(p)$ sono costanti da serie a serie, si comprende perché un minor numero di fenomeni presenta dispersione ipo-binomiale, caratteristica dello schema di Poisson, rispetto a quelli con dispersione iper-binomiale, che per la maggior parte seguono lo schema probabilistico di Coolidge e solo in piccola parte quello di Lexis.

3 Schemi probabilistici con dipendenza fra le prove di una serie ma indipendenza tra le serie.

In uno schema probabilistico, nel quale si é sempre interessati all'ottenimento di un evento A (successo) in k serie di n_j prove ciascuna con $j=1,2,...,k$, si introduce a questo punto l'ipotesi di dipendenza tra le n_j prove di ciascuna

serie, mantenendo peró l'assunzione di indipendenza tra le k serie.
Poiché la seguente analisi si concentra sulle relazioni di dipendenza tra le variabili, si suppone per semplicitá che la probabilitá di successo p sia costante da prova a prova e da serie a serie, in altri termini si associa alla variabile casuale indicatore A_{ji}, definita nella (2.1), le seguenti probabilitá

$$P\{A_{ji}=1\}=p \qquad P\{A_{ji}=0\}=1-p=q \quad j=1,2,...,k \text{ e } i=1,2,...,n_j.$$

Si consideri il caso in cui la dipendenza (lineare) tra ogni coppia di variabili casuali A_{ji} e A_{jl}, con $i \neq l$, della j-ma serie, si manifesta in modo uniforme

$$r(A_{ji},A_{jl})=\rho \quad i\neq l; \quad i,l=1,2,...,n_j \text{ e } j=1,2,...,k,$$

indicando con ρ il valore assunto dal coefficiente di correlazione lineare tra ogni coppia di variabili casuali di una data serie. Dall'ipotesi di indipendenza tra le serie si deduce che

$$r(A_{ji},A_{sl})=0 \quad j\neq s; \quad j,s=1,2,...,k \text{ e } i=1,2,...,n_j;\ l=1,2,...,n_s.$$

É possibile sintetizzare tale schema probabilistico con un indicatore bivariato[5]

$A_{ji}\backslash A_{jl}$	1	0	
1	p_{11}	p_{10}	p
0	p_{01}	p_{00}	$1-p$
	p	$1-p$	1

Da questa rappresentazione risulta

$$\begin{aligned} r(A_{jl},A_{ji}) &= \frac{Cov(A_{jl},A_{ji})}{\sqrt{\mathbb{V}(A_{jl})}\sqrt{\mathbb{V}(A_{ji})}} = \frac{\mathbb{E}(A_{jl}A_{ji})-\mathbb{E}(A_{jl})\mathbb{E}(A_{ji})}{\sqrt{p(1-p)}\sqrt{p(1-p)}} \\ &= \frac{p_{11}-p^2}{p(1-p)} = \rho. \end{aligned} \tag{3.1}$$

Dalla precedente relazione si deduce

$$Cov(A_{jl},A_{ji})=r(A_{jl},A_{ji})(1-p)p=\rho(1-p)p.$$

[5]La seguente rappresentazione é stata proposta da Zenga (2003).

Analogamente al caso di indipendenza tra le prove, sia X_j la variabile casuale che indica il numero di volte che l'evento A si manifesta nella j-ma serie, la quale presenta in tale contesto aspettativa e varianza pari a

$$
\begin{aligned}
\mathbb{E}(X_j) &= \mathbb{E}\left[\sum_{i=1}^{n_j} A_{ji}\right] = \sum_{i=1}^{n_j} \mathbb{E}(A_{ji}) = \sum_{i=1}^{n_j} p = n_j p \qquad\qquad j = 1, 2, ..., k \\
\mathbb{V}(X_j) &= \mathbb{E}\left[X_j - \mathbb{E}(X_j)\right]^2 = \mathbb{E}\left[\sum_{i=1}^{n_j} A_{ji} - n_j p\right]^2 = \mathbb{E}\left[\sum_{i=1}^{n_j} (A_{ji} - p)\right]^2 \\
&= \mathbb{E}\left[\sum_{i=1}^{n_j}\sum_{s=1}^{n_j} (A_{ji} - p)(A_{js} - p)\right] = \sum_{i=1}^{n_j}\sum_{s=1}^{n_j} \mathbb{E}\left[(A_{ji} - p)(A_{js} - p)\right] \\
&= \sum_{i=1}^{n_j}\sum_{s=1}^{n_j} Cov(A_{ji}, A_{js}) = \sum_{i=1}^{n_j} \mathbb{V}(A_{ji}) + \sum_{i=1}^{n_j}\sum_{s=1}^{n_j}{}_{i \neq s} \quad Cov(A_{ji}, A_{js}) \\
&= n_j p(1-p) + n_j(n_j - 1)\rho p(1-p) \qquad j = 1, 2, ..., k.
\end{aligned}
\tag{3.2}
$$

Si consideri ora l'aspettativa della seguente quantitá casuale

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{X_j}{n_j} - p\right)^2 n_j\right] = n_j \mathbb{V}\left(\frac{X_j}{n_j}\right) &= n_j \frac{1}{n_j^2}[n_j p(1-p) + n_j(n_j - 1)\rho p(1-p)] \\
&= pq + \rho pq(n_j - 1).
\end{aligned}
$$

Ne consegue che

$$
\mathbb{E}\left[\sum_{j=1}^{k}\left(\frac{X_j}{n_j} - p\right)^2 n_j\right] = kpq + \rho pq \sum_{j=1}^{k}(n_j - 1) = kpq + \rho pq(n - k). \tag{3.3}
$$

Confrontando tale risultato con quelli determinati in precedenza nei diversi schemi probabilistici con indipendenza tra le prove, si ottiene la seguente relazione tra il coefficiente di correlazione lineare ρ e la dispersione dello schema probabilistico considerato:

- se $\rho > 0$ la dispersione é iper-binomiale, comportamento analogo alla schema di Lexis;
- se $\rho = 0$ la dispersione é binomiale, comportamento analogo allo schema di Bernoulli;

- se $\rho < 0$ la dispersione é ipo-binomiale, comportamento analogo allo schema di Poisson.

Per determinare uno stimatore del coefficiente di correlazione lineare ρ, conviene isolare inizialmente ρ dall'equazione (3.3)

$$\rho = \frac{\mathbb{E}\left[\sum_{j=1}^{k}\left(\frac{X_j}{n_j} - p\right)^2 n_j\right] - kpq}{pq(n-k)},$$

il cui stimatore é

$$\widehat{\rho} = \frac{\sum_{j=1}^{k}\left[\left(\frac{X_j}{n_j} - \widehat{p}\right)^2 \frac{n_j}{n}\right] - \frac{k\widehat{pq}}{n}}{\widehat{pq}(1-\frac{k}{n})}$$

con $\widehat{p}$ definito nell'equazione (2.2). Se il numero delle prove n é molto grande rispetto a k, si puó approssimare la precedente equazione come segue[6]

$$\widehat{\rho} \simeq \frac{\sum_{j=1}^{k}\left[\left(\frac{X_j}{n_j} - \widehat{p}\right)^2 \frac{n_j}{n}\right]}{\widehat{pq}}.$$

Si noti che il numeratore di questo rapporto rappresenta la variabilitá delle frequenze relative $\widehat{p}_j$, mentre il denominatore consiste nella variabilitá della variabile casuale indicatore A_{ji} nello schema probabilistico di Bernoulli con probabilitá di successo costante pari a p.

4 Il quoziente di divergenza di Lexis.

Per comprendere se un esperimento soddisfa le ipotesi dello schema di Bernoulli oppure degli altri schemi probabilistici analizzati in precedenza sono stati indicati diversi criteri[7].

Il procedimento piú utilizzato é stato proposto da Lexis nel 1876, che si basa sulla seguente statistica

$$\frac{\sum_{j=1}^{k}\left[\left(\frac{X_j}{n_j} - \widehat{p}\right)^2 \frac{n_j}{n}\right]}{(k\widehat{pq})n^{-1}} = \frac{\sum_{j=1}^{k}\left[\left(\frac{X_j}{n_j} - \widehat{p}\right)^2 n_j\right]}{k\widehat{pq}}, \tag{4.1}$$

[6]Il seguente risultato coincide con quello ricavato da Resti (2001), pp. 127.

[7]Per una approfondita analisi dei diversi procedimenti proposti in letteratura, con relativi pregi e difetti, ed interessanti esempi si veda Boldrini (1968), pp.1201.

nota come *quoziente di divergenza di Lexis*. Come é stato giá accennato in precedenza, il numeratore del quoziente di divergenza rappresenta la variabilitá delle frequenze relative $\widehat{p}_j$, e, come si vede dal risultato (2.15), nello schema probabilistico di Bernoulli l'aspettativa del numeratore e del denominatore coincidono. Questo significa che, se il quoziente di divergenza di Lexis é prossimo all'unitá, allora l'esperimento considerato soddisfa le assunzioni dello schema probabilistico di Bernoulli[8]. Poiché dalle equazioni (2.22) e (3.3) ricavate in precedenza, l'aspettativa del numeratore del quoziente di divergenza di Lexis risulta minore del denominatore, si deduce che, se tale rapporto risulta significativamente minore di uno, si propende invece per uno schema probabilistico con dispersione ipo-binomiale, quindi di Poisson o per uno schema con correlazione negativa uniforme tra le variabili casuali indicatore di ciascuna serie. Se infine tale rapporto risulta significativamente maggiore di uno, per le relazioni (2.26) e (3.3), si predilige uno schema con dispersione iper-binomiale, dunque di Lexis oppure uno schema con correlazione positiva uniforme tra le prove di ciascuna serie. Si precisi che nell'ultimo caso, in cui si rileva un valore del quoziente di divergenza di Lexis significativamente maggiore di uno, si potrebbe considerare anche lo schema di Coolidge, dal momento che presenta approssimativamente dispersione iper-binomiale al divergere del numero di prove di ciascuna serie.
Per poter costruire degli intervalli di confidenza per il quoziente di divergenza di Lexis si osservi che tale statistica é una trasformazione di scala della grandezza casuale definita nell'equazione (2.19), la quale, si é dimostrato in precedenza, che in uno schema di Bernoulli, al divergere delle numerositá n_j, tende in distribuzione ad una variabile casuale Chi-Quadrato con $(k-1)$ gradi di libertá. Fissando quindi un livello di significativitá pari ad α, si ricava che se il valore assunto dalla statistica test

$$L = \sum_{j=1}^{k} \frac{\left(\frac{X_j}{n_j} - \widehat{p}\right)^2}{\widehat{p}\widehat{q}} n_j$$

é compreso tra i valori che assumono i quantili di ordine $\alpha/2$ e $(1-\alpha/2)$ di una variabile casuale Chi-Quadrato con $(k-1)$ gradi di libertá, allora si accetta l'ipotesi nulla che l'esperimento considerato aderisce allo schema di

[8]Si precisi che in tale caso si potrebbe considerare anche uno schema probabilistico con dipendenza ed incorrelazione ($\rho = 0$) tra le prove, questo significa che tra le variabili casuali indicatore di ciascuna serie sussiste un legame di dipendenza di tipo non lineare, ma vista la raritá del caso si preferisce trascurare tale eventualitá.

Bernoulli. Se il valore assunto dalla statistica test risulta invece maggiore del quantile di ordine $(1-\alpha/2)$ di una variabile casuale Chi-Quadrato con $(k-1)$ gradi di libertá, allora si accetta l'ipotesi alternativa e si sceglie o uno schema di Lexis oppure uno schema con correlazione positiva (uniforme) tra le prove di ciascuna serie. Nell'ultimo caso in cui il valore assunto dalla suddetta statistica test risulti minore del quantile di ordine $\alpha/2$ di una variabile casuale Chi-Quadrato con $(k-1)$ gradi di libertá, si accetta sempre l'ipotesi alternativa che consiste peró nello schema di Poisson oppure in uno schema con correlazione negativa (uniforme) tra le prove di ciascuna serie.
Faleschini (1949) ha mostrato come é possibile avere utili informazioni sulla natura della variabile casuale X esaminando la sua inversa X'. Considerando k serie di m prove ciascuna, la variabile casuale X' rappresenta il numero di volte che un evento A si verifica in m serie di k prove ciascuna.
Se la variabile casuale X presenta dispersione binomiale, anche la variabile casuale X' presenta dispersione binomiale. Se la variabile casuale X presenta dispersione ipo-binomiale o iper-binomiale, anche la variabile X' ha dispersione non binomiale, che puó fornire utili informazioni sulle caratteristiche della dispersione della variabile casuale X. In particolare se X rappresenta il numero totale di successi in uno schema di Poisson, allora la sua inversa X' indica il numero totale di successi di uno schema probabilistico di Lexis.
Sia dunque x_j il numero rilevato di successi che si sono verificati nelle m prove della j-ma serie e x'_i il numero osservato di successi nelle k prove della i-ma serie, se in altri termini i valori $x_1, ..., x_j, ..., x_k$ costituiscono una serie di Poisson, allora $x'_1, ..., x'_i, ..., x'_m$ costituiscono una serie di Lexis.

prob. delle prove	I serie	...	j-ma serie	...	k-ma serie	Serie di Lexis
p_1	1	...	0	...	1	x'_1
...	...	...	...	...	...	...
p_i	0	...	1	...	1	x'_i
...	...	...	...	...	...	...
p_m	1	...	0	...	0	x'_m
Serie di Poisson	x_1	...	x_j	...	x_k	$\sum_j x_j = \sum_i x'_i$

5 Conclusioni.

Nell'analisi di un portafoglio di crediti é consuetudine raggruppare i clienti che presentano caratteristiche simili, secondo l'appartenenza alla medesima classe di rating, come il modello CreditMetrics (J.P. Morgan (1997)), oppure in base a perdite simili in caso di insolvenza, come il modello CreditRisk+ (Credit Suisse Financial Products (1997)). Questo significa che considerando un portafoglio composto da n crediti, le k serie rappresentano le suddette classi. La variabile casuale indicatore A_{ji} indica dunque lo stato di insolvenza o solvenza dell'i-ma posizione creditizia appartenente alla j-ma classe.
Il quoziente di divergenza di Lexis, definito nell'equazione (4.1), permette di valutare se il portafoglio in esame presenta una dispersione binomiale, ipo-binomiale oppure iper-binomiale, nel caso in cui esista indipendenza tra le insolvenze appartenenti a fasce diverse.
Dal primo risultato si deduce che le probabilitá di insolvenza p_{ji} sono costanti sia da classe a classe sia da credito a credito in ciascuna classe e le insolvenze sono indipendenti in ogni fascia (schema di Bernoulli).
Una dispersione iper-binomiale evidenzia invece due possibili situazioni: le probabilitá di insolvenza sono costanti da credito a credito in ciascuna classe, ma variano da fascia a fascia e le insolvenze sono indipendenti in ciascuna fascia (schema di Lexis) oppure le probabiliá di insolvenza sono costanti da credito a credito e da classe a classe, ma esiste una correlazione lineare (uniforme) positiva tra le insolvenze della medesima fascia.
Infine nel caso in cui la dispersione risulta ipo-binomiale, si considerano due diverse circostanze: sia le medie sia le varianze delle probabilitá di insolvenza sono costanti da fascia a fascia, permettendo peró a tali probabilitá di variare all'interno di ciascuna classe e le insolvenze sono indipendenti in ciascuna fascia (schema di Poisson) oppure le probabilitá di insolvenza sono costanti da credito a credito e da classe a classe, ma esiste una correlazione lineare (uniforme) negativa tra le insolvenze della medesima classe.

6 Bibliografia.

Boldrini M. (1968). *Statistica: Teoria e Metodi.* Giuffré, Milano, cap. XVIII.

Credit Suisse Financial Products (1997). *CreditRisk+. A Credit Risk Management Framework.* Technical Document, Londra.

Cramer H. (1996). *Mathematical Methods of Statistics.* Princeton University Press, Princeton.

Faleschini L. (1949). *Sullo schema generale del problema delle prove ripetute con probabilitá indipendente.* Estratto dalla 'Rivista Italiana di demografia e Statistica', Vol **III**, pp. 1-25.

Feller W. (1968). *An introduction to Probability Theory and Its Applications.* Vol. **I** John Wiley & Sons, New York.

Johnson N. L., Kemp A. W. e Kotz S. (2005). *Univariate Discrete Distributions.* Wiley, New York.

J. P. Morgan (1997). *CreditMetrics.* Technical Document, New York.

Johnson N. L., Kemp A. W. e Kotz S. (1969). *Discrete Distributions.* Houghton Mifflin, Boston.

Kendall S. (1994) *The Advanced Theory of Statistics.* Vol. **I**. Hafner Publishing Company, New York.

Pompilj G. (1967). *Teoria dei Campioni.* Libreria Eredi di Virgilio Veschi, Roma.

Resti A. (2001). *Misurare e gestire il rischio di credito nelle banche: una guida metodologica.* Fondo Interbancario di Tutela dei Depositi, Roma.

Rohatgi V. K. (1976). *An Introduction to Probability Theory and Mathematical Statistics.* Wiley & Sons, New York.

Zenga M. (2003). *Schemi probabilistici con indipendenza fra le prove di una serie e dipendenza fra le prove di una serie.* Seminario tenuto presso il Dipartimento di Metodi Quantitativi per le Scienze Economiche ed Aziendali, Universitá degli Studi di Milano-Bicocca.

Zenga M. (1968). *Schema Probabilistico Bivariato con Probabilitá Variabili da Prova e Prova e da Serie a Serie.* Estratto da 'La Scuola in Azione', **8**. E.N.I.- Scuola Enrico Mattei di studi superiori idrocarburi, San Donato Milanese