RATIO MATH. 2
(1991), 97 - 110

CONFRONTO FRA INFERENZA BAYESIANA E TEST CLASSICI PER L'ANALISI DI SUCCESSIONI DI NUMERI PSEUDOCASUALI

Giuseppe Di Biase e Antonio Maturo (*)

Sommario - L'impostazione classica della verifica della casualità di successioni di numeri pseudocasuali si propone essenzialmente di scoprire se, ammettendo che l'ipotesi nulla H_0 sia vera, il verificarsi di una successione sia un evento "poco probabile" o "molto probabile". Nel primo caso si è "propensi a rinunciare" all'idea iniziale che H_0 sia vera, nel secondo caso, invece, l'idea viene confermata.

L'impostazione bayesiana, invece, fondata soprattutto su una visione soggettiva del Calcolo delle Probabilità si propone di confrontare H_0 con altre ipotesi possibili e di valutare le probabilità di tali ipotesi sulla base della conoscenza della successione.

Tuttavia, se ci si limita ad effettuare l'analisi bayesiana usuale non si ottengono risultati soddisfacenti. A nostro parere ciò dipende dal fatto che essa può essere considerata sostitutiva del solo test classico delle frequenze e, quindi, come quest'ultimo fornisce delle condizioni necessarie, ma del tutto insufficienti per la verifica di casualità delle successioni. Il lavoro si propone, quindi, di mettere a punto le idee di base per analizzare le successioni pseudocasuali da un punto di vista bayesiano.

Abstract - Classical approach for the check of the "randomness" of a sequence of pseudorandom numbers essentially proposes to evaluate if, in the hypothesis that H_0 (null hypothesis) is true, the occurrence of a sequence is a "not very probable" or a "very probable" result. In the former case it is "inclined to give up" the initial hypothesis that H_0 is true, whereas in the latter case the hypothesis is confirmed. Bayesian approach, on the contrary, is above all based on a subjective view of the Probability Theory, and it proposes to compare H_0 with other possible hypotheses on the basis of the knowledge of the sequence. However, if we limit ourselves to performe the usual Bayesian analysis we shall not obtain satisfying results. In our opinion this is due to the fact that these tests are substitutives of the only frequency test and, then, like this, they give necessary conditions but not sufficient for the check of "randomness" of a sequence. So in this paper we will consider to restate the basic ideas of Bayesian analysis of pseudo-random sequences.

(*) Dipartimento di Scienze, Storia dell'Architettura e Restauro
Viale Pindaro, 42 - 65127 Pescara

1. Successioni di numeri pseudocasuali e loro proprietà matematiche.

Sia N l'insieme dei numeri naturali maggiori di zero e sia $\{X_i\}_{i\in N}$ una successione di variabili casuali somiglianti ed indipendenti con distribuzione uniforme continua, ossia con densità di probabilità

$$f(x) = \begin{cases} 1 \text{ per } x \in (0,1) \\ \\ 0 \text{ per } x \notin [0,1] \end{cases} \qquad (1.1)$$

Per ogni $i\in N$, se x_i è una determinazione di X_i la successione $\{x_i\}_{i\in N}$ si dice successione di numeri casuali.

Sia m un elemento di N «abbastanza grande», di solito dell'ordine di grandezza di almeno 10^8.

Per esigenze di calcolo, dovute al fatto che gli elaboratori usano solo numeri razionali con un dato numero di cifre decimali, generalmente al posto delle X_i, si considerano delle variabili casuali $X_i^{(m)}$, $i\in N$ somiglianti ed indipendenti con distribuzione uniforme discreta normalizzata di parametro m, ossia tali che assumono i valori 0, 1/m, 2/m,..., (m-1)/m, ciascuno con probabilità 1/m.

La possibilità di sostituire le $X_i^{(m)}$ alle X_i, pur essendo variabili casuali di tipo diverso, le une discrete, le altre continue, si basa sul fatto che per ogni $i\in N$ la successione di variabili casuali $\{X_i^{(m)}\}_{m\in N}$ converge in distribuzione verso la variabile casuale X_i.

Infatti, basta osservare che le $X_i^{(m)}$ hanno la funzione caratteristica

$$H_m(u) = (1/m)\,(e^{iu}-1) / (e^{iu/m}-1)$$

e che

$$\lim_{m\to+\infty} H_m(u) = (e^{iu}-1) / iu,$$

funzione caratteristica della variabile casuale uniforme continua.

In pratica una volta fissato m, la successione dei valori assunti dalle $X_i^{(m)}$, $i\in N$, viene sostituita da una successione $\{x_i\}_{i\in N}$ di valori ottenuti, per mezzo dell'elaboratore, con il seguente procedimento.

A partire da una formula ricorrente del tipo

$$y_{i+1} = f(\alpha_1, \alpha_2, ..., \alpha_h; y_i), \quad i\in N, \qquad (1.2)$$

con f funzione della variabile y_i, definita nell'insieme $I_m = \{0, 1, 2, ..., m-1\}$, a valori in I_m e dipendente dai parametri $\alpha_1, \alpha_2, ..., \alpha_h$, si ottiene una successione

$\{y_i\}_{i\in N}$ di elementi dell'insieme I_m.

Successivamente, posto $x_i = y_i/m$ si ottiene una successione di numeri appartenenti al supporto $\{0, 1/m,..., (m-1)/m\}$ delle $X_i^{(m)}$.

Questo procedimento può apparire del tutto fuori di ogni logica. Infatti, invece di numeri a caso si considerano numeri, detti pseudocasuali, ottenuti, in maniera deterministica, da una formula.

Una giustificazione di tale procedimento segue dalla definizione soggettiva di numeri pseudocasuali (cfr. ad esempio Bruno Baldessari (1987), [1], Alfredo Rizzi (1989), [13]) secondo la quale la successione $\{x_i\}_{i\in N}$ si dice pseudocasuale se

(1) un operatore, al quale il generatore (1.2) sia noto, può univocamente generare la successione;
(2) un individuo, al quale il generatore non sia noto, ritiene soggettivamente che la successione $\{x_i\}_{i\in N}$ sia una realizzazione di una successione $\{X_i\}_{i\in N}$ di variabili casuali uniformi discrete normalizzate di parametro m.

Infatti, se la funzione f e i suoi parametri sono assegnati in maniera opportuna il modo in cui si ottiene y_{i+1} a partire da y_i appare il prodotto di un «effetto di casualità» analogo a quello che si ottiene mescolando un mazzo di carte o ruotando la pallina di una roulette. Pensandoci bene, anche in questi casi i percorsi compiuti dalle carte o dalla pallina sono deterministici e la casualità consiste nel fatto che un osservatore non li conosce.

Nella formula (1.2), per ottenere successioni che soddisfino la definizione soggettiva di Baldessari, è necessario che l'operatore che genera la successione conosca bene le proprietà matematiche e statistiche di essa per fare in modo che appaia casuale ad un osservatore.

Ciò si ottiene assegnando f uguale ad una funzione il più possibile semplice. Il caso più utilizzato è quello in cui la (1.2) è del tipo

$$y_{i+1} = (a\, y_i + b) \bmod m, \quad i\in N, \tag{1.3}$$

con a e b elementi di I_m.

Le proprietà matematiche e statistiche della successione $S = \{y_i\}_{i\in N}$ che si ottiene dal generatore (1.3) variano notevolmente al variare dei parametri a, b e del valore iniziale (o seme) y_1.

Per vari motivi (cfr. ad esempio Cera e Maturo (1983), [2] e [3], Maturo (1989), [8]), le scelte più opportune di m sono $m=10^s$, $m=2^s$ con $s\in N$ e m numero primo.

Si dimostra che la successione S è periodica con periodo δ non superiore ad m.

Si dimostra, inoltre, (cfr. Cera e Maturo (1983) e (1990), [2] e [4], Maturo (1989), [8]) che, qualunque sia m, si ottiene il massimo periodo possibile $\delta=m$

se e solo se sono soddisfatte le seguenti condizioni

(a) MCD(a, m) = 1,
(b) p numero primo divisore di m $\Rightarrow$ a mod p = 1,
(c) 4 divisore di m $\Rightarrow$ a mod 4 = 1,
(d) MCD(b, m) = 1.

Queste condizioni non assicurano, però, la validità delle proprietà statistiche di indipendenza ed equidistribuzione. Ad esempio, per a=1 e b=1, esse sono soddisfatte qualunque sia m, ma la successione ottenuta appare tutt'altro che casuale.

Spesso è più conveniente, sia dal punto di vista statistico, sia perchè è particolarmente semplice verificare quali proprietà valgono, porre b=0. Si ottiene allora la seguente formula, detta generatore moltiplicativo,

$$y_{i+1} = (a\, y_i) \bmod m, \quad i \in N. \tag{1.4}$$

Si può dimostrare che la successione S ottenuta a partire dal generatore moltiplicativo ha periodo δ minore o uguale ad un numero $\sigma(m)$ dipendente da m. Il valore $\sigma(m)$ è uguale ad m-1 per m numero primo, mentre è sensibilmente inferiore ad m negli altri casi. Ad esempio, $\sigma(m)=m/4$ per $m=2^s$, $s\geq 4$ e $\sigma(m)=m/20$ per $m=10^s$, $s\geq 5$. Si può inoltre dimostrare (cfr., ad es. Cera e Maturo (1983) e (1990), [2] e [4], Maturo (1989), [8]), che nei casi considerati, condizioni necessarie e sufficienti perchè sia $\delta=\sigma(m)$ sono le seguenti

(I) caso m primo
(IA) a e y_1 sono entrambi primi con m,
(IB) $a^i \bmod m \neq 1$, $\forall\, i \in \{1, 2,..., m-2\}$.

(II) caso $m = 2^s$, con $s \geq 4$.
(IIA) a mod 8 = 3 oppure a mod 8 = 5,
(IIB) y_1 è dispari.

(III) caso $m = 10^s$, con $s \geq 5$.
(IIIA) y_1 è primo con 10,
(IIIB) a mod 200 è uguale ad uno dei seguenti 32 numeri

3, 11, 13, 19, 21, 27, 29, 37, 53, 59, 61, 67, 69, 77, 83, 91, 109,
117, 123, 131, 133, 139, 141, 147, 163, 171, 173, 179, 181, 187, 189, 197.

2. Sull'analisi statistica dei generatori.

Consideriamo una successione $\{y_i\}_{i\in N}$ di elementi di I_m ottenuti da una formula di tipo (1.2) e sia $x_i = y_i/m$.

Da un punto di vista classico la verifica della casualità della successione viene svolta nella maniera seguente.

Fissato un $n\in N$ ci si pone il problema di poter accettare o meno l'ipotesi nulla H_0 che $\{x_1, x_2,..., x_n\}$ è un campione casuale della variabile casuale uniforme continua.

Si sottopone, a tale scopo, la successione finita $\{x_i\}_{i\in\{1,2,...,n\}}$ di lunghezza n ad un insieme di test statistici.

Alcuni dei test più utilizzati, la cui descrizione dettagliata si può trovare nei lavori di Knuth (1969), [7], Maturo (1989), [11], Tausworthe (1965), [18], sono i seguenti

1. Test sulla media dei numeri;
2. Test sulla varianza dei numeri;
3. Test sulla frequenza dei numeri;
4. Test sulla frequenza delle cifre;
5. Test sulla frequenza delle coppie di numeri;
6. Test sulla frequenza delle coppie di cifre;
7. Test sulla correlazione fra i numeri;
8. Test sulla correlazione fra le cifre;
9. Run test;
10. Gap test;
11. Test sulle sequenze complete.

I metodi di controllo, non parametrici, considerati ad esempio in Cera e Maturo (1983) e (1990), [2] e [4], Maturo (1989), [8], si basano essenzialmente sul seguente tipo di procedimento:

(1) si fissa una statistica $D = D(X_1, X_2,..., X_n)$ funzione del campione casuale $\underline{X}=(X_1, X_2,..., X_n)$ di ampiezza n della variabile casuale X con distribuzione uniforme continua tale che $D \geq 0$. La D deve essere costruita in modo da assumere valori tanto più piccoli, quanto più è «accettabile» l'ipotesi H_0 e si dice **misura della discrepanza** o **discrepanza** fra la successione finita $\underline{x}=\{x_1, x_2,..., x_n\}$ e l'ipotesi H_0;

(2) si fissa un numero reale $\delta > 0$. Detto d il valore assunto da D per $\underline{X}$ che assume il valore $\underline{x}$, se $d \geq \delta$ si ritiene eccessiva la discrepanza e si giudica «non accettabile» H_0; se, invece, $d < \delta$ si ritiene sufficientemente piccola la discrepanza e si assume che H_0 sia «accettabile» rispetto alla statistica D.

Il numero δ è assegnato in modo da soddisfare una uguaglianza del tipo

$$\text{prob}\,(D \geq \delta) = \alpha,$$

con α numero reale positivo assegnato a seconda degli scopi da raggiungere. Ad esempio si pone $\alpha=0.1$ oppure $\alpha=0.01$ o $\alpha=0.001$.

Per fissare le idee, limitiamoci a considerare solo alcuni test e precisamente quelli relativi a

(I) media dei numeri;
(II) frequenza dei numeri;
(III) frequenza delle coppie di numeri;
(IV) runs.

(I) Si assume la seguente condizione sulla media: la media M delle x_i deve essere «vicina» a 0.5, media della variabile uniforme continua. La situazione ideale è l'uguaglianza M=0.5.

Ricordiamo che, per un noto teorema, se $X_1, X_2,..., X_n$ sono variabili casuali indipendenti ed equidistribuite, con media m e varianza v, allora, per n «abbastanza grande», la variabile casuale

$$Z = [(X_1, X_2,..., X_n)/n - m]\,(n/v)^{1/2} \qquad (2.1)$$

può considerarsi distribuita come una variabile casuale normale standard N(0, 1).

Il test della media consiste nel fissare come discrepanza la statistica $D=|Z|$.

(II) Si divide l'intervallo $[0, 1[$ in h intervallini di uguale lunghezza con h molto minore dell'ampiezza n del campione e viene assunta la seguente condizione sulle frequenze: il numero di elementi appartenenti a ciascun intervallino deve essere vicino a n/h.

Fissata una successione finita $\{x_i\}$, con $i \in \{1, 2,..., n\}$, di elementi dell'intervallo [0, 1), per un noto teorema, indicato con n_i, $i = 1, 2,..., h$, il numero di elementi della successione appartenenti all'intervallino I_i, per n abbastanza grande si può ritenere che la variabile casuale

$$D = (h/n) \sum_{i=1}^{h} (n/h - n_i)^2 = (h/n) \sum_{i=1}^{h} n_i^2 - h \qquad (2.2)\ .$$

abbia distribuzione chi-quadro ad h-1 gradi di libertà.

Nella situazione ideale, in cui ogni n_i è uguale a n/h, risulta D=0. Il test sulle frequenze dei numeri consiste nell'assumere come discrepanza la statistica data dalla (2.2).

(III) Si divide l'intervallo $[0, 1)$ in h intervallini di uguale lunghezza $I_1, I_2, \ldots, I_h$. Si considera, poi, l'intervallo $[0, 1) \times [0, 1)$ come unione degli h^2 intervallini del piano, di uguale area, $I_{rs} = I_r \times I_s$, con r, s=1, 2,..., h.

Se $\{x_1, x_2, \ldots, x_n\}$ è il campione empirico, si esamina la successione delle coppie (x_i, x_{i+1}), i = 1, 2,..., n-1. Si ammette la seguente condizione sulla frequenza delle coppie: poichè gli intervallini I_{rs} sono h^2 e le coppie sono n-1, la frequenza delle coppie in ciascun intervallino deve essere vicina a $(n-1)/h^2$.

Per n abbastanza grande, detto n_{rs} il numero di elementi della successione appartenenti all'intervallino I_{rs}, si può ritenere che la variabile casuale

$$D = [h^2/(n-1)] \sum_{r=1}^{h} \sum_{s=1}^{h} [(n-1)/h^2 - n_{rs}]^2 \qquad (2.3)$$

abbia una distribuzione chi-quadro a h^2-1 gradi di libertà.

Il test sulle frequenze delle coppie di numeri consiste nell'assumere come discrepanza la statistica D data dalla (2.3).

(IV) Si assume la seguente condizione sui runs: il numero delle sottosuccessioni monotone massimali (dette runs) di una successione con n elementi deve essere «vicino» a $(2n-1)/3$.

La situazione ideale è quella in cui il numero di runs è esattamente uguale a $(2n-1)/3$. Ciò è basato sul fatto che, detta R la variabile casuale «numero di runs», si dimostra che, per n abbastanza grande, si può assumere che la variabile

$$Z = [R - (2n-1)/3] \, |90/(16n-29)|^{\frac{1}{2}} \qquad (2.4)$$

sia normale standard.

Il test sui runs consiste nell'assumere come discrepanza la variabile casuale $D = |Z|$.

Le successioni di tipo (1.2) che si considerano sono, di solito, periodiche di periodo molto elevato, almeno dell'ordine di 10^{10}, mentre l'ampiezza dei campioni considerati nelle precedenti analisi statistiche è piccola rispetto a tali numeri, e va da n=100 fino a n=100.000 in alcuni casi.

Chiamiamo campionari i tests precedentemente esaminati.

Test di natura diversa, che chiamiamo di tipo «globale», sono utilizzati da Knuth (1969), [7], Maturo (1989), [9] e Rizzi (1989), [13]. Essi considerano il comportamento di successioni finite di lunghezza uguale al periodo della successione (1.2).

Anche tali test, tuttavia, sono fondati su un meccanismo di accettazione-rifiuto che si basa sulla distanza da valori ideali di certi valori ottenuti tramite opportune elaborazioni a partire dalla formula (1.2).

3. Le critiche ai test classici ed il problema della ricerca di analisi bayesiane "sostitutive".

Generalizzando le considerazioni svolte nel paragrafo precedente possiamo quindi affermare che da un punto di vista classico l'analisi di casualità delle successioni di numeri pseudocasuali è basata sul seguente procedimento:

(a) Si considerano opportune statistiche $D_k=D_k(X_1, X_2,..., X_n)$, k=1, 2,..., r, che assumono valori reali positivi e che, in certe situazioni ritenute ideali, si annullano. Le D_k hanno, di solito, distribuzione asintotica o di tipo chi-quadro oppure uguale al valore assoluto di una distribuzione simmetrica rispetto allo zero e tabulata, come avviene ad esempio, per la normale standard o la legge di student.

(b) Per ogni k si fissa un numero $\alpha_k \in (0, 1)$ «piccolo», in genere $\alpha_k=0.01$ oppure $\alpha_k=0.001$ o un valore intermedio fra questi, detto livello di significatività, e si determina un numero reale β_k tale che prob $(D_k \geq \beta_k) = \alpha_k$.

La coppia $T_k=(D_k, \beta_k)$ si dice **test statistico** al livello di significatività α_k.

(c) Sia d_k il valore assunto da D_k per $X_i=x_i$, con i=1, 2,..., n. Se $d_k \geq \beta_k$ si respinge l'ipotesi H_0 rispetto al test T_k, in caso contrario si accetta l'ipotesi H_0 rispetto al test T_k.

(d) Si considerano «accettabili», rispetto all'insieme dei test considerati, le successioni per le quali l'ipotesi H_0 è accettata rispetto ad ognuno dei test T_k.

La logica rispetto alla quale viene rifiutata l'ipotesi H_0 se $d_k \geq \beta_k$ si basa su due considerazioni.

(1) **Valutazione probabilistica**: l'evento $E_k=(D_k \geq \beta_k)/H_0$ ha probabilità α_k molto piccola di verificarsi. Poichè si è verificato, ciò induce a pensare che l'ipotesi H_0 sia falsa;

(2) **valutazione psicologica**: se $d_k \geq \beta_k$ allora il valore assunto da D_k è molto lontano da zero, «valore ideale».

Alcuni errori presenti in questo tipo di logica, (cfr. ad esempio Scozzafava (1989), [14], e Maturo (1989), [10]) sono:

(a) la confusione fra E_k e l'evento $A_k=H_0/(D_k \geq \beta_k)$. Nella considerazione (1), dal fatto che la probabilità di E_k è piccola si deduce che è piccola anche quella di A_k.

Conclusioni di questo tipo portano a conseguenze paradossali.

Ad esempio, si può osservare che, qualunque sia la successione finita $\underline{x}$ di lunghezza n, se gli x_i sono numeri con q cifre decimali, allora $prob(x_i/H_0)=10^{-q}$ e quindi risulta $prob(\underline{x}/H_0)=10^{-nq}$.

Poichè, per ogni $\alpha_k>0$, esiste un intero n tale che $10^{-nq}<\alpha_k$, fissato α_k è possibile trovare un intero m tale che, per ogni $n\geq m$, l'evento $\underline{x}/H_0$ ha probabilità minore di α_k.

Allora, ragionando come in (1), siamo indotti a respingere H_0 qualunque sia x, purchè sia $n\geq m$.

(b) L'elemento di disturbo introdotto dalla (2). Il punto di vista psicologico porta a «scartare la coda» della distribuzione di D_k, perchè in essa vi sono valori molto lontani dal valore ideale, 0, di D_k. Ragionando da un punto di vista rigorosamente probabilistico, invece, se a è un qualsiasi numero reale non negativo minore di β_k e b è un qualsiasi numero reale tale che l'evento $E_{ak}=(a\leq D_k\leq b)/H_0$ abbia probabilità α_k, l'evento E_{ak} potrebbe giocare, nel meccanismo di accettazione e rifiuto del test, lo stesso ruolo dell'evento E_k. Ad esempio, ponendo a=0, il test porterebbe a «scartare» la «parte iniziale» della distribuzione.

(c) La mancata considerazione di ipotesi alternative ad H_0. Ammesso che realmente l'evento $A_{0k}=H_0/(D_k\geq\beta_k)$ abbia probabilità molto piccola, ciò non è un motivo sufficiente per respingere H_0. Infatti può capitare che vi siano numerose ipotesi alternative H_i, i=0, 1, 2,..., m e che tutti i vari eventi $A_{ik}=H_i/(D_k\geq\beta_k)$ abbiano probabilità molto piccole. Allora, se si verifica che $d_k\geq\beta_k$ tutte le ipotesi devono essere respinte?

In questo lavoro ci proponiamo di mettere a punto le idee di base per un'analisi bayesiana delle successioni pseudocasuali.

Una analisi bayesiana presenta i seguenti vantaggi:

(a) oltre all'ipotesi H_0 considera un insieme di ipotesi alternative da confrontare con H_0;

(b) fornisce risultati esclusivamente probabilistici senza introdurre, nei ragionamenti, meccanismi estranei al Calcolo delle Probabilità, come quelli di accettazione e rifiuto, ed è quindi esente dai difetti logici del metodo classico.

A nostro parere, il rifiuto da parte di alcuni ricercatori di effettuare la verifica di casualità di una successione pseudocasuale di lunghezza n da un punto di vista bayesiano nasce dall'equivoco di considerare una tale successione semplicemente come un campione di ampiezza n della variabile casuale uniforme continua, senza tener conto dell'ordinamento. Infatti da un tale punto di vista appaiano "migliori" delle altre delle successioni banali e non certo

casuali quali ad esempio quella di termine generale $x_i = (i \bmod m)/m$, con $i \in N$.

In altre parole, un insieme di n numeri reali distinti $x_1, x_2, ..., x_n$, appartenenti all'intervallo [0,1) si può considerare un campione della variabile casuale uniforme continua, ma esistono n! successioni a cui corrisponde tale campione e solo alcune di esse soddisfano al requisito di apparire casuali ad un osservatore.

In definitiva, l'analisi bayesiana che considera la successione come un campione di ampiezza n che chiamiamo, in seguito, per semplicità, analisi bayesiana delle frequenze è sostitutiva unicamente del test classico delle frequenze, ampiamente verificato per praticamente tutte le successioni pseudocasuali che si considerano.

Come avviene per tale test, il fatto che l'ipotesi H_0 sia favorita da queste analisi bayesiane è una condizione necessaria, ma del tutto insufficiente affinchè sia accettabile il fatto che la successione S soddisfi la definizione soggettiva di Baldessari.

Una generalizzazione dell'analisi bayesiana delle frequenze, considerata in Di Biase e Maturo (1990), [5] e [6], è la seguente

(1) si considera l'ipotesi H_0 come un elemento di una famiglia di ipotesi $(H_\alpha)_{\alpha \in I}$, con I insieme di indici, tale che per un certo valore μ di I sia $H_\mu = H_0$ e si assegna, a priori ad ogni ipotesi H_α, in maniera soggettiva, con opportuni criteri, il valore $f(H_\alpha)$ di una pseudodensità, detta **pseudodensità a priori**, ossia di una funzione reale non negativa, definita nell'insieme degli eventi elementari e che tiene conto di una assegnata relazione di preordine fra tali eventi; essa può essere, in particolare, una probabilità o una densità di probabilità (cfr. Scozzafava (1983) e (1989), [15] e [17], e Maturo (1989), [11]);

(2) a partire da una data successione $\underline{x} = \{x_1, x_2, ..., x_n\}$ di numeri pseudocasuali nell'intervallo [0, 1) si calcolano le verosimiglianze $p(\underline{x}/H_\alpha)$, $\alpha \in I$;

(3) con la formula del Teorema di Bayes

$$g(H_\alpha/\underline{x}) = k\, f(H_\alpha)\, p(\underline{x}/H_\alpha), \quad \text{con k funzione solo di } \underline{x} \qquad (3.1)$$

si calcolano i valori di una nuova pseudodensità $g(H_\alpha/\underline{x})$, detta **pseudodensità a posteriori**;

(4) si vede, con criteri convenzionali opportuni, se viene favorita l'ipotesi $\alpha = \mu$, ad esempio controllando se il punto di massimo di g è «sufficientemente vicino» a μ oppure se, fissata una opportuna partizione di I «è favorito» l'insieme della partizione a cui appartiene μ.

Osserviamo che, anche in questo caso generale, dall'ipotesi H_0 di indi-

pendenza ed equidistribuzione delle X_i, segue che ogni successione S' ottenuta come permutazione di $\underline{x}$ dà luogo alle stesse verosimiglianze e, quindi, agli stessi valori della g. Di conseguenza una tale analisi bayesiana fornisce informazioni solo sulle frequenze dei valori $\underline{x}_i$ assunti dalle X_i, e, pertanto, può sostituire solo il test classico delle frequenze.

Per poter eseguire la verifica di casualità delle successioni pseudocasuali da un punto di vista bayesiano è necessario, quindi, a nostro parere, effettuare una analisi bayesiana sostitutiva per ciascuno dei test classici. Ad esempio una verifica di casualità di tipo campionario può essere svolta considerando una analisi bayesiana sostitutiva per ciascuno degli 11 test del paragrafo precedente ed una verifica di casualità di tipo globale considerando una analisi bayesiana sostitutiva per ciascuno dei test di Knuth (1969), [7].

Proponiamo, quindi, il seguente procedimento generale.

Dalla successione finita $\underline{x}=\{x_1, x_2,..., x_n\}$ di numeri pseudocasuali ottenuta come illustrato nel paragrafo 1 se ne deduce un'altra del tipo:

$$T=\{\underline{z}_1, \underline{z}_2,..., \underline{z}_{n-r}\}$$

con r intero positivo opportuno, r<<n, dove le $\underline{z}_i$ sono vettori di uno spazio R^k, con $k\geq 1$, funzioni del vettore $(x_1, x_2,..., x_n) \in R^n$ e determinazioni di variabili casuali $\underline{Z}_i$ equidistribuite ed indipendenti.

Le $\underline{Z}_i$ dipendono da determinati parametri e, per opportuni valori di essi, hanno distribuzione uguale a quelle ideali richieste dal test classico che viene sostituito.

In definitiva si ammette che:

(1) i vettori $\underline{z}_i$ siano una determinazione di variabili casuali $\underline{Z}_i \alpha_1,\alpha_2,...,\alpha_h$ equidistribuite, indipendenti e dipendenti dai parametri $\alpha_1, \alpha_2,..., \alpha_h$;

(2) in condizioni ideali di casualità i parametri $\alpha_1, \alpha_2,..., \alpha_h$ assumono i valori $\mu_1, \mu_2,..., \mu_h$;

(3) si assume una distribuzione a priori di $(\alpha_1, \alpha_2,..., \alpha_h)$, in genere tale che privilegia il valore $(\mu_1, \mu_2,..., \mu_h)$, ad esempio che assume il massimo in corrispondenza a tale valore;

(4) indicando con $f(\underline{\alpha})$ la pseudodensità a priori di $\underline{\alpha}=(\alpha_1, \alpha_2,..., \alpha_h)$ e con $p(\underline{z}/\underline{\alpha})$ la probabilità di $\underline{z}=(z_1, z_2,..., z_n)$ dato $\underline{\alpha}$, si calcola la pseudodensità a posteriori

$$g(\underline{\alpha}/\underline{z}) = k\, f(\underline{\alpha})\, p(\underline{z}/\underline{\alpha}) \qquad (3.2)$$

dove k è una funzione di $\underline{z}$. Se la $g(\underline{\alpha}/\underline{z})$ è una densità di probabilità si ha

$$k = \frac{1}{\int_{R^h} g(\underline{\alpha}, \underline{z})\, d\alpha_1\, d\alpha_2 \ldots d\alpha_h}$$

(5) si guarda se la pseudodensità (3.2) «tende a favorire» il valore $(\mu_1, \mu_2, \ldots, \mu_h)$, dando eventualmente dei criteri operativi, che hanno solo carattere convenzionale, di giudizio sul grado di soddisfacimento delle condizioni ideali.

Alcuni casi particolari sono ad esempio:

(1) analisi bayesiana sulle frequenze; si pone $\underline{z}_i = x_i$, per ogni $i \in \{1, 2, \ldots, n\}$;

(2) analisi bayesiana sulle coppie; si pone $\underline{z}_i = [x_i, x_i+1]'$, per ogni $i \in \{1, 2, \ldots, n-1\}$. Risulta r=1;

(3) analisi bayesiana sulle k-ple; è una generalizzazione dei casi precedenti. Si pone $\underline{z}_i = [x_i, x_{i+1}, \ldots, x_{i+k-1}]'$, per ogni $i \in \{1, 2, \ldots, n-k+1\}$. Risulta r=k-1;

(4) analisi bayesiana sui runs; supponiamo che per ogni i sia $x_i \neq x_{i+1}$, ciò che praticamente è sempre verificato; si pone:

$$\underline{z}_i = \begin{cases} 0 \quad \text{per } i>1 \text{ e } x_i<x_{i+1}<x_{i+2} \text{ oppure } x_i>x_{i+1}>x_{i+2} \\ 1 \quad \text{in caso contrario} \end{cases}$$

per ogni $i \in \{1, 2, \ldots, n-2\}$.

La distribuzione degli z_i non soddisfa, però, alle condizioni di indipendenza. Allora il ragionamento si modifica considerando m successioni $\underline{x}^{(j)} = \{z_{j1}, \ldots, z_{jh}\}$ casuali di ampiezza n e, per ognuna di esse, il numero

$$v_j = \sum_{i=1}^{n} z_{ji}$$

La distribuzione ideale dei v_j (cfr. Rizzi (1977), [12]) è quella normale con media (2n-1)/3 e varianza (16n-29)/90. I parametri su cui eseguire l'inferenza sono la media θ e la varianza σ^2.

I v_j si possono considerare indipendenti e la (3.2) diventa

$$g(\underline{\alpha}/\underline{v}) = k\, f(\underline{a})\, p(\underline{v}/\underline{a}). \qquad (3.3)$$

BIBLIOGRAFIA

1. B. Baldessari (1987), *Aspetti probabilistici della crittografia*, Atti del 1° Simposio Nazionale su "Stato e Prospettive della Ricerca Crittografica in Italia", pp. 9-21.
2. N. Cera e A. Maturo (1983), *Confronto fra alcuni generatori di numeri pseudocasuali*, Facoltà di Architettura, Pescara.
3. N. Cera e A. Maturo (1990), *Generazione di numeri pseudocasuali per mezzo di relazioni di ricorrenza in campi di Galois*, Periodico di Matematica, n. 2, pp. 33-56.
4. N. Cera e A. Maturo (1991), *Analisi della bontà di alcuni generatori di numeri pseudocasuali per la cifratura dei messaggi e la simulazione*, Ratio Mathematica, 2, in corso di stampa.
5. G. Di Biase e A. Maturo (1990), *Analisi bayesiana di successioni pseudorandom*, Atti del XIV Conv. Ann. AMASES, Pescara 13-15 settembre 1990, pp. 293-313.
6. G. Di Biase e A. Maturo (1990), *Su un' analisi bayesiana per la verifica di casualità di successioni pseudorandom: software applicativo ed interpretazione dei risultati*, Atti del Conv. "Classificazione e analisi dei dati, Metodi, Software, applicazioni", Pescara 11-12 ottobre 1990, in corso di stampa.
7. D.E. Knuth (1969), *The art of Computer programming vol. 2*, Addison Wesley, London.
8. A. Maturo (1989), *Numeri pseudocasuali, Libreria dell' Università*, Pescara.
9. A. Maturo (1989), *Analisi di Fourier di successioni di numeri pseudocasuali*, Atti del 2° Simposio Nazionale su "Stato e Prospettive della Ricerca Crittografica in Italia", pp. 188-198.
10. A. Maturo (1989), *Probabilità e statistica con il calcolatore: problematiche di carattere logico ed operativo*, Convegno "Insegnamento integrato di probabilità soggettiva e statistica bayesiana", Teramo.
11. A. Maturo (1989), *Probabilità finitamente additiva a valori nel campo delle serie bilatere*, Rendiconti di Matematica, VII, 9, pp. 67-85.
12. A. Rizzi (1977), *Generazione di distribuzioni statistiche mediante un elaboratore elettronico*, Istituto di Statistica e Ricerca sociale C.Gini, Roma.
13. A. Rizzi (1989), *Verifiche di Pseudo-Casualità in crittografia*, Atti del 2° Simposio Nazionale su "Stato e Prospettive della Ricerca Crittografica in Italia", pp. 3-21.
14. R. Scozzafava (1989), *La probabilità soggettiva e le sue applicazioni*, Editoriale Veschi, Masson, Milano.
15. R. Scozzafava (1983), *Finitely Additive Probabilities and Proper 'Improper' Priors in Bayesian Statistics*, Institute of Statistics Mimeo Series, 1541, Chapel Hill (U.S.A.).

16. R. Scozzafava (1984), *A Survey of Some Common Misunderstandings Concerning the Role and Meaning of Finitely Additive Probabilities in Statistical Inference*, Statistica, 44, pp. 21-45.
17. R. Scozzafava (1982), *Probabilità s-additive e non, Bollettino U.M.I.*, 6, 1-A.
18. R.C. Tausworthe (1965), *Random numbers generated by linear recurrence modulo two*, Mathematics of Computation, 19.