Ratio Math. 2
(1991), 83 - 96

ANALISI DELLA BONTÀ DI ALCUNI GENERATORI DI NUMERI PSEUDOCASUALI PER LA CIFRATURA DEI MESSAGGI E LA SIMULAZIONE (*)

Nisida Cera (**) Antonio Maturo (**)

1. INTRODUZIONE

Indichiamo con il termine "messaggio" sia una comunicazione inviata da un individuo ad altri sia un promemoria messo da parte da una persona per poterlo consultare in un secondo tempo.

Utilizzando un computer, un messaggio del primo tipo consiste in un documento inviato attraverso un canale elettronico; un messaggio del secondo tipo, consiste in un file di dati immessi in una opportuna area di memoria e richiamabili per mezzo di un programma.

In entrambi i casi, un messaggio si presenta come una successione finita $\mathbf{M} = \{m_i\}_{i\in\{1,2,\dots,h\}}$ di cifre in una certa base b. Il numero h si dice **lunghezza**, in base b, del messaggio.

In genere si assume b=2 oppure b=10.

Si può porre il problema di voler rendere accessibile un messaggio solo a particolari individui e non ad altri.

A tale scopo, si considera un seconda successione finita $\mathbf{K} = \{k_i\}_{i\in\{1,2,\dots,h\}}$ di simboli in base b, anch'essa di lunghezza h, conosciuta solo dalle persone destinatarie del messaggio e, al posto di **M**, s'invia il messaggio **C=M+K**, definito come la successione di termine generale

(*) Lavoro svolto nell'ambito della ricerca M.P.I. 40% 1988 Modelli Economici Discreti, Ottimizzazione, Crittografia.

(**) Dipartimento di Scienze e Storia dell'Architettura, Università degli Studi "G. D'Annunzio" viale Pindaro 42, 65127 PESCARA.

$$c_i = (m_i + k_i) \mod b, \qquad i=1,2,...,h.$$

Poiché $(c_i - k_i) \mod b = m_i$, un individuo che conosce **K** può risalire, dalla conoscenza di **C**, al messaggio **M**.

Il procedimento consistente nel sostituire **C** ad **M** si dice di **cifratura** o **codifica** del messaggio, quello che fa risalire ad **M** a partire da **C** si dice di **decodifica**.

La successione **K** si dice **chiave del messaggio**.

Due esigenze fondamentali a cui deve soddisfare la chiave **K** sono:

(1) **segretezza** o **sicurezza**: dev'essere massimizzata la difficoltà di risalire a **K** dalla conoscenza di **C**;

(2) **semplicità**: dev'essere massimizzata la semplicità nella formulazione della chiave e nei procedimenti di codifica e decodifica.

Le due esigenze sono solitamente contrastanti e la scelta della chiave da utilizzare va fatta tenendo conto dei problemi che ci si pongono e delle circostanze in cui si opera.

Si può capire ciò ricorrendo ad una analogia con la vita comune.

Alcuni oggetti che ci interessano, in casa o in ufficio, vengono chiusi a chiave, con una chiave comune, in un cassetto.

Il sistema chiave comune - cassetto non offre certamente un grado di sicurezza pari ad esempio a quello del sistema formato da una cassaforte.

Tuttavia si usa il sistema chiave comune - cassetto poiché, per l'importanza degli oggetti ed una valutazione obiettiva dei rischi, si ritiene sufficiente la sicurezza ottenuta e soddisfacente le semplicità di acquisizione e di uso di tale sistema di sicurezza.

Un procedimento di codifica che ha il massimo requisito di sicurezza è quello di porre **K** uguale ad una **successione casuale** di lunghezza h di cifre in una data base b.

Esso, tuttavia, non ha il requisito della semplicità, in quanto, se h è elevato, ad esempio superiore a 50, è certamente un'operazione molto lunga ottenere la successione **K** ed è difficoltoso trasmetterla e utilizzarla per operazioni di codifica e decodifica.

Si ottiene, invece, il massimo requisito di semplicità, paragonabile a quella del sistema chiave comune - cassetto, ponendo **K** uguale ad un segmento h di una **successione pseudocasuale** di cifre in base b.

Per quanto riguarda il requisito della sicurezza, esso dipende dal tipo della successione pseudocasuale utilizzata, precisamente dalle sue **proprietà matematiche e statistiche**.

2. SUCCESSIONI OTTENUTE DA GENERATORI CONGRUENZIALI

Sia **m** un intero positivo maggiore o uguale a 2 ed indichiamo con **I(m)** l'insieme {0, 1,....., m-1}.

Introduciamo, in **I(m)**, le operazioni (+) e (•) ponendo, per definizione, per ogni coppia (x,y) di elementi di **I(m)**,

$$x\ (+)\ y = (x + y) \bmod m,\ x\ (\bullet)\ y = (x \bullet y) \bmod m.$$

È facile verificare che la terna (**I(m)**, (+), (•)) è un anello, isomorfo all'anello $(Z_m, +, \bullet)$ degli interi modulo m.

Sia **n** un intero positivo ed indichiamo con $[\mathbf{I(m)}]^n$ l'insieme delle n-ple di elementi di **I(m)**. Chiamiamo, per semplicità, **vettori** tali n-ple e **scalari** gli elementi di **I(m)**.

Per ogni elemento **x** di $[\mathbf{I(m)}]^n$, sia $(x)_i$ la sua componente **i-esima**.

Introduciamo, in $[\mathbf{I(m)}]^n$, le operazioni (+) e * ponendo, per definizione, per ogni coppia (**x, y**) di elementi di $[\mathbf{I(m)}]^n$ e per ogni $a \in \mathbf{I(m)}$,

$$(\mathbf{x}\ (+)\ \mathbf{y})_i = (\mathbf{x})_i + (\mathbf{y})_i,\ (a * \mathbf{x})_i = a\ (\bullet)\ (\mathbf{x})_i.$$

È facile verificare che ($[\mathbf{I(m)}]^n$, (+), *) è un modulo su **I(m)**.

È noto che, se m è primo, **I(m)** è un campo e quindi $[\mathbf{I(m)}]^n$ è uno **spazio vettoriale su I(m)**.

2.1 Definizione Siano **n** ed **s** interi positivi ed **m** un intero maggiore di 1. Diciamo **generatore congruenziale di dimensione n, ordine s e modulo m ogni funzione**

$$g: \{[\mathbf{I(m)}]^n\}^s \longrightarrow \mathbf{I(m)} \qquad (2.1)$$

non costante rispetto al vettore prima coordinata. ***

Fissando **s** vettori $\mathbf{x}_0, \mathbf{x}_1, \ldots\ldots, \mathbf{x}_{s-1}$, a partire dalla (2.1), ponendo, per ogni $h \geq s$,

$$\mathbf{x}_h = g(\mathbf{x}_{h-s}, \mathbf{x}_{h-s+1}, \ldots\ldots, \mathbf{x}_{h-1}), \qquad (2.2)$$

si ottiene una successione $\mathbf{X} = \{x_n\}$, con $h \in N_0$, di elementi di $[\mathbf{I(m)}]^n$.

2.2 Definizione Diciamo **funzione di supporto** in $[\mathbf{I(m)}]^n$ ogni **epimorfismo** Φ fra gli **I(m)**-moduli $[\mathbf{I(m)}]^n$ e **I(m)**. (Considerando **I(m)** modulo su **I(m)**) ***

2.3 Definizione Diciamo **successione in I(m) associata al generatore g, alla**

funzione di supporto Φ e alla scelta dei vettori iniziali x_0, x_1,.....,x_{s-1}, la successione **Y** di termine generale $y_h = \Phi(x_h)$, con $h \in N_0$. ***

Se la successione **Y** soddisfa in maniera "sufficiente" ad alcuni requisiti matematici e statistici, essa può essere assunta come atta a simulare una successione casuale e si dice **successione pseudocasuale**.

Da una successione pseudocasuale si può ottenere una chiave **K** formata da k cifre in una data base **b** con vari criteri, di cui i più comuni sono i seguenti:

(1) caso $\mathbf{m = b}$

Fissato un elemento y_a qualsiasi di **Y** si assume:

$$\mathbf{K} = \{y_a, y_{a+1}, \ldots\ldots, y_{a+k-1}\}.$$

(2) caso $\mathbf{m = b^s}$, con s intero maggiore di 1.

Ogni elemento y_h si può scrivere con s cifre in base b.

Allora, fissato un elemento y_a qualsiasi di **Y** e, detti q ed r il quoziente ed il resto della divisione di k per s, si assume **K** uguale alla successione delle cifre, in base **b**, di y_a, y_{a+1},..... y_{a+q-1}, nell'ordine, e delle prime r cifre, sempre in base b, y_{a+q}.

(3) caso $\mathbf{m = b^{1/s}}$, con s intero maggiore di 1.

Ogni elemento di **K** si ottiene scrivendo in base **b** un numero avente come cifre, in base m, s elementi consecutivi di **Y**.

Fissando un elemento y_a qualsiasi di **Y**, si assume **K** uguale alla successione delle cifre in base **b** ottenute scrivendo, in base **b**, i numeri già scritti in base m nella forma

$$y_{a+rs}\, y_{a+rs+1} \ldots\ldots y_{a+rs+s-1}$$

per r = 0, 1, 2,....,k-1.

(4) caso in cui non valgono le condizioni precedenti o comunque si vogliono ottenere numeri con s cifre in una data base fissata **b** e risulta $m > b^s$.

Sostituiamo ad ogni elemento y_h della successione **Y** il numero

$$u_h = [y_h b^s / m], \qquad (2.3)$$

indicando, per ogni numero reale x, con [x] la sua parte intera nella base b

fissata.

Ogni u_h si scrive con s cifre in base b, per cui ci si riconduce al caso (2), sostituendo le $\mathbf{u}_h$ alle y_h.

In pratica, specie nell'ultimo caso, è opportuno scegliere le basi b = 2 oppure b = 10. Per altre scelte della base, dato che i calcolatori usano o il sistema binario o quello decimale, la (2.3) può presentare errori di arrotondamento o troncamento, per cui per essere sicuri che sono verificate le necessarie proprietà matematiche e statistiche deve essere preso s in modo che sia $m>>b^s$. Ciò però ha come conseguenza una notevole perdita nella velocità di acquisizione delle cifre.

3. PROPRIETÀ MATEMATICHE E STATISTICHE DEI GENERATORI CONGRUENZIALI LINEARI

Un generatore congruenziale **g** di dimensione **n**, ordine **s** e modulo **m** si dice **lineare** se la (2.2) assume la forma

$$\mathbf{x}_h = a_0 * \mathbf{x}_{h-s} (+)\, a_1 * \mathbf{x}_{h-s+1} (+) \ldots\ldots (+)\, a_{s-1} * \mathbf{x}_{h-1} (+)\, \mathbf{b}, \qquad (3.1)$$

con $a_0 \neq 0$, a_1, a_2,... a_{s-1} elementi di $\mathbf{I(m)}$ e **b** elemento di $[\mathbf{I(m)}]^n$.

I generatori congruenziali più studiati nella letteratura e più usati sia per la loro semplicità e sia per la possibilità di individuarne con chiarezza alcune fondamentali proprietà matematiche e statistiche sono quelli lineari di dimensione 1 e di ordine 1. Essi assumono la forma

$$x_h = a * x_{h-1} (+) b, \qquad (3.2)$$

con $a \neq 0$ e b elementi di $\mathbf{I(m)}$.

Limitiamoci a considerare il caso particolare dei generatori (3.2) con la funzione di supporto Φ uguale alla funzione identica in $\mathbf{I(m)}$. Allora la successione in $\mathbf{I(m)}$ associata al generatore (3.2) e alla Φ assume la forma:

$$y_h = (a y_{h-1} + b) \mod m. \qquad (3.3)$$

Usando la terminologia classica, diversa da quella usata in questo lavoro, in cui vengono introdotte le funzioni g e Φ, chiamiamo, da ora in poi, "**generatore lineare**" senza ulteriori specificazioni il generatore (3.3) e "**generatore moltiplicativo**" un generatore lineare con b=0.

Rinviando ai lavori di A. Rizzi [10], A. Maturo [6], N.Cera e A.Maturo [7], [8] e altri (cfr. ad es.[1], [11], [12]), per uno studio dettagliato delle proprietà "a priori" dei generatori lineari e per le dimostrazioni, riportiamo di seguito

alcuni dei risultati più significativi.

Sia $\{y_h\}$, con $h\varepsilon N_0$, la successione associata alla (3.3) e alla scelta di un fissato valore iniziale y_0.

3.1 **Risultato** La successione in **I(m)** $\{y_h\}$, con $h\varepsilon N_0$, è periodica. Se μ è l'antiperiodo e δ è il periodo, risulta $\mu+\delta\leq m$. ***

3.2 **Risultato** La successione in **I(m)** $\{y_h\}$, con $h\varepsilon N_0$, ha il massimo periodo possibile $\delta=m$ se e solo se sono soddisfatte le seguenti condizioni:

(L1) M.C.D.(a,m) = 1;
(L2) p numero primo divisore di m $\Rightarrow$ a mod p = 1;
(L3) 4 divisore di m $\Rightarrow$a mod 4 = 1;
(L4) M.C.D.(b,m) = 1. ***

Osserviamo che, se $\delta=m$, la successione di termine generale y_h è, a meno di una traslazione, indipendente dal valore iniziale y_0. Inoltre, se $p_1, p_2,.....,p_s$ sono i fattori primi distinti di m le **(L1)**, **(L2)**, **(L3)**, si possono esprimere scrivendo:

$$a \text{ mod}(p_1 \cdot p_2 \cdot \cdots \cdot p_s) = 1, \text{ se 4 non divide m;} \tag{3.4}$$

$$a \text{ mod}(2p_1 \cdot p_2 \cdot \cdots \cdot p_s) = 1, \text{ se 4 divide m.} \tag{3.5}$$

La **(L4)** non è una condizione molto restrittiva. Invece le **(L1)**, **(L2)**, **(L3)**, conducono a successioni molto "regolari" e quindi senza "apparenza di casualità" se **m** è primo oppure è il prodotto di 2^r, con $r\varepsilon\{0, 1, 2\}$ e di numeri primi dispari distinti. Infatti, in tali circostanze le nostre condizioni implicano che a = 1.

In vari lavori (ad es. [4], [6], [10]) è stata messa in evidenza l'opportunità che, ai fini di ottenere buone proprietà statistiche di casualità, nella scomposizione in fattori primi di m, compaiano non più di due fattori primi, ciascuno con un esponente "abbastanza grande" (ad es. non inferiore ad 8).

Si ritiene inoltre opportuno che tali fattori primi sia piccoli, preferibilmente formati da una sola cifra decimale.

Fra i moduli che soddisfano a tali condizioni vi sono $m=2^r$ e $m=10^r$, fondamentali per l'invio di messaggi con cifre, rispettivamente, in base 2 ed in base 10.

Il generatore più semplice è quello moltiplicativo. Esso, dato che non è soddisfatta la **(L4)**, non dà luogo a successioni di periodo m. Tuttavia, nei lavori citati, è stato dimostrato che si possono ottenere, con tale generatore, successioni con periodi di lunghezze accettabili e con **"apparenza di casualità"** spesso superiore a quella dei generatori lineari di massimo periodo.

Alcuni dei risultati più significativi relativi al generatore moltiplicativo

sono i seguenti:

3.3 **Risultato** Sia $m = p^r$ con $r \geq 1$ e p numero primo. La successione $\{y_h\}$, $h \in \mathbf{N}_0$, in **I(m)**, associata al generatore

$$y_h = a\, y_{h-1} \quad \text{mod } m \tag{3.6}$$

è periodica e priva di antiperiodo. Si ottiene il massimo periodo $\delta = m(p-1)/p$ se e solo se sono verificate le seguenti condizioni:

(M1) a è radice primitiva modulo m;

(M2) M.C.D.(y,m) = 1. ***

3.4 **Risultato** Sia $m = 2^r$, con $r > 3$. La successione in **I(m)** associata al generatore (3.6) è periodica e priva di anti-periodo. Il periodo massimo ottenibile è $\delta = m/4$ e si ha se e solo se valgono le seguenti condizioni:

(P1) a mod 8 = 3 oppure a mod 8 = 5;

(P2) y_0 è dispari. ***

Se **m** non è una potenza di un numero primo, le condizioni per avere il massimo periodo sono più complesse ed il massimo periodo ottenibile è molto inferiore ad m.

Per il caso in cui $m = 10^r$ si ha il seguente

3.5 **Risultato** Sia $m = 10^r$, con $r > 4$. La successione in **I(m)** associata al generatore (3.6) è periodica e priva di antiperiodo. Si ottiene il periodo massimo $\delta = m/20$ se e solo se valgono le seguenti condizioni:

(D1) 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;

(D2) y_0 è primo con 10. ***

Osserviamo che, nei casi in cui $m = 2^r$ o $m = 10^r$ c'è una vistosa diminuzione del periodo massimo rispetto a quello dei generatori lineari non moltiplicativi. Tuttavia il periodo non è l'unico elemento da considerare, nè il più importante. Una volta che ci siamo assicurati che il periodo è sufficientemente lungo, almeno dell'ordine di 10^8-10^{10}, dobbiamo preoccuparci soprattutto che venga-

no soddisfatte le proprietà statistiche di "apparenza di casualità" e della velocità di elaborazione. Quest'ultima è massima se m è una potenza della base b in cui sono considerate le cifre nel computer utilizzato.

4. ANALISI STATISTICHE

Consideriamo una successione $\mathbf{Y} = \{y_h\}$, con $h \varepsilon \mathbf{N}_0$, di elementi di $\mathbf{I(m)}$ ottenuta da un qualsiasi generatore congruenziale modulo m.

Indichiamo con $\mathbf{Z} = \{z_h\}$, con $h \varepsilon \mathbf{N}_0$, la successione di termini generale $z_h = y_h/m$ e con $\mathbf{U} = \{u_h\}$, con $h \varepsilon \mathbf{N}_0$, una successione di cifre in una data base b ottenuta, con uno dei criteri descritti nel paragrafo 2, a partire dalla **Y**.

Sia r un intero positivo assegnato. Si ammette che **Z** sia una **successione di numeri pseudocasuali rispetto ai campioni di ampiezza r e rispetto a dei fissati test statistici** se, per un generico segmento $\{z_{h+1}, z_{h+2}, \ldots\ldots, z_{h+r}\}$, di lunghezza r della successione, i testi statistici dati portano ad "accettare" o meglio a "non rifiutare" l'ipotesi nulla:

H_0: $\{z_{h+1}, z_{h+2}, \ldots\ldots, z_{h+r}\}$ è un campione casuale della variabile casuale UC, uniforme continua in [0,1].

Si ammette che **U** sia una **successione di cifre pseudocasuali in base b rispetto ai campioni di ampiezza r e rispetto a dei fissati test statistici** se, per un generico segmento $\{u_{h+1}, u_{h+2}, \ldots\ldots, u_{h+r}\}$, di lunghezza r della successione, i testi statistici dati portano ad "accettare" l'ipotesi nulla:

H_1: $\{u_{h+1}, u_{h+2}, \ldots\ldots, u_{h+r}\}$ è un campione casuale della variabile casuale UD (b), uniforme discreta di parametro b.

A seconda del valore dell'ampiezza r possiamo distinguere le analisi statistiche per la verifica della casualità delle successioni **Z** ed **U** in

(i) analisi "globali", per r uguale al periodo δ della successione considerata;

(ii) analisi campionarie, per r "molto minore" del periodo δ, ma "abbastanza grande" perché il campione sia significativo.

Abbiamo eseguito sia delle analisi di tipo (i), sia delle analisi di tipo (ii) per vari tipi di generatori.

Per quanto riguarda le analisi globali, basandoci sulle teorie svolte nei lavori di Coveyou-Macpherson [3], Knuth [5], Maturo [9], abbiamo elaborato un complesso programma in basic per la verifica della casualità delle succes-

sioni. Gli output più significativi sono, per ogni successione, dei numeri positivi σ_j, $j\in\{1, 2,....., 7\}$, indicanti il "grado di vicinanza" della distribuzione delle j-ple di elementi consecutivi della successione da una distribuzione uniforme j-dimensionale.

Precisamente, tenuto conto delle considerazioni svolte nei lavori citati, abbiamo considerato tale grado di vicinanza

(a) buono, per $\sigma_j\geq 1$;

(b) sufficiente, per $0.1<\sigma_j<1$;

(c) insufficiente, per $\sigma_j\leq 0.1$.

Abbiamo inoltre giudicato l'importanza di avere un alto grado di vicinanza decrescente al crescere di j. In pratica, per $j>4$ le informazioni date dai valori σ_j sono di modesta rilevanza.

Riportiamo i risultati ottenuti per 9 generatori particolari nella tabella della pagina seguente. In essa non sono riportati i valori di σ_1, sempre buoni.

I 9 generatori sono stati scelti per confrontare le varie e spesso discordanti proposte di autorevoli matematici e statistici. Precisamente:

(1) per i generatori 1, 5, 8, 9 si è fatto in modo che a sia dell'ordine di grandezza di $[(\sqrt{5} - 1)/8]m$;

(2) per i generatori 2 e 6 si è seguita l'indicazione di Greenberger di porre $a\approx\sqrt{m}$;

(3) per i generatori 3 e 7 sono stati sperimentati moltiplicatori vicini, rispettivamente, al doppio e alla metà di quelli relativi al primo criterio;

(4) per il generatore 4 abbiamo voluto sperimentare il comportamento con il modulo 10^{10} di un moltiplicatore che aveva dato un buon esito con il modulo 2^{35}.

Tabella dei valori σj per j=2, 3, ..., 7.

n° d'ordine	term. b	modulo m	moltipl.a	σ_2	σ_3	σ_4	σ_5	σ_6	σ_7
(b,m)=1	1	10^{10}	1545084981	3.14	0.32	0.26	1.30	1.07	1.32
(b,m)=1	2	10^{10}	100001	3.14	6.2E-9	7.9E-9	1.7E-8	3.3E-8	6.0E-8
(b,m)=1	3	10^{10}	3141592421	2.68	0.34	0.54	0.59	1.42	0.75
(b,m)=1	4	10^{10}	5308871541	2.82	2.84	0.09	3.82	1.53	0.56
(b,m)=1	5	2^{35}	5308871541	3.30	2.93	0.69	1.39	2.44	2.54
(b,m)=1	6	2^{35}	185365	3.14	1.1E-4	0.01	1.24	3.27	1.69
(b,m)=1	7	2^{35}	2718281821	2.59	1.16	1.75	4.25	2.26	0.74
b=0	8	10^{11}-23	28^7	0.74	1.02	0.41	0.66	2.28	0.05
b=0	9	2^{35}	5308871551	1.25	4.44	1.86	2.93	1.56	1.69

Dalla tabella della pagina precedente si vede che, dal punto di vista delle proprietà statistiche globali, i generatori 2 e 6 appaiono piuttosto scadenti, soprattutto per quanto riguarda il comportamento delle terne e delle quaterne. Dal punto di vista crittografico ciò fa pensare che si presentino delle "regolarità" che permettono ad un esperto di decodificare facilmente un messaggio che usa tali generatori come chiave.

Il generatore 4 presenta qualche debolezza solo per il comportamento delle quaterne, mentre gli altri appaiono tutti sufficienti. Particolarmente buoni appaiono i generatori 7 e 9, il primo con m uguale ad una potenza di 2 e il cui moltiplicatore a non è fra quelli consigliati nella letteratura, ma è uno di quelli sperimentati da noi con il criterio (3) ed il secondo moltiplicativo, con m numero primo.

Un "buon comportamento globale" di un generatore di numeri pseudocasuali non garantisce però che non ci siano delle "regolarità" nei segmenti brevi della successione. Calcolando una chiave ottenuta da 200 numeri per un messaggio relativamente breve, corrispondente (per i valori considerati di m) all'emissione di circa 6000-7000 cifre binarie ed una chiave di 1000 numeri pseudocasuali per un messaggio più lungo, abbiamo quindi ritenuto necessario eseguire delle analisi statistiche campionarie per r=200 ed r=1000 sui generatori risultati sufficienti nell'analisi globale.

A tale scopo abbiamo elaborato un programma in basic con il quale sottoporre ad 11 test statistici un qualsiasi generatore lineare ed abbiamo fatto girare il programma, per ognuno dei 9 generatori considerati, valutando i risultati, ai livelli di significatività del 10% e del 1% con 200 replicazioni per r=200 e con 100 replicazioni per r=1000.

I test utilizzati, la cui descrizione dettagliata si può trovare nei nostri lavori [6], [8] ed in [10], 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 tests;
10. Gap tests;
11. Tests sulle sequenze complete.

In generale, la logica seguita per tali test è la seguente:
data una successione $W = \{w_h\}$ di numeri reali ed assegnata una variabile

casuale X, ci poniamo il problema di verificare se il segmento generico $\{w_h, w_{h+1},...,w_{h+r-1}\}$ di lunghezza r di **W** soddisfa l'ipotesi nulla:

H: $\{w_h, w_{h+1},...,w_{h+r-1}\}$ è un campione casuale di ampiezza r di X.

I metodi di controllo, non parametrici, considerati in questo lavoro si basano essenzialmente nel seguente tipo di procedimento:

(1) si fissa una statistica $D = D(X_1, X_2,..., X_r)$ funzione del campione casuale $X = (X_1, X_2,..., X_r)$ di ampiezza r di X tale che $D \geq 0$. La D dev'essere costruita in modo da assumere, generalmente, valori tanto più piccoli, quanto più è "accettabile" l'ipotesi H e si dice **misura della discrepanza o discrepanza** fra un generico segmento di lunghezza r della successione W e l'ipotesi H;

(2) si fissa un numero reale $\delta > 0$ e, detto d il valore assunto da D per $X = (w_h, w_{h+1},..., w_{h+r-1})$, se $d \geq \delta$ si ritiene eccessiva la discrepanza e quindi si giudica "non accettabile" H; se, invece, $d < \delta$ si ritiene "accettabile" la discrepanza e quindi H.

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

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

con α numero reale positivo, di solito minore o uguale a 0.1.

Nel nostro lavoro sono state considerate 11 diverse misure di discrepanza, di cui 6 associate a statistiche riguardanti i numeri pseudocasuali e 5 a statistiche collegate alla distribuzione delle cifre di tali numeri. Prevalentemente, abbiamo assunto come discrepanze o i valori di distribuzioni di tipo chi-quadro oppure le distanze, in valore assoluto, dei valori di distribuzioni simmetriche dalla media.

La notevole massa dei dati raccolti ci ha imposto di cercare un "test riassuntivo" per i risultati ottenuti, per ogni test, nelle varie replicazioni. Ciò allo scopo di confrontare, rispetto a ciascun test, la bontà dei generatori considerati. Abbiamo ritenuto opportuno, per tali motivi, modificare il procedimento di accettazione o rifiuto dell'ipotesi nulla, nella convinzione che valori eccessivamente bassi della discrepanza siano in disaccordo con l'ipotesi di indipendenza ed equidistribuzione così come quelli molto alti.

Abbiamo allora considerato due numeri σ e τ non negativi tali che:

$$\text{prob}(D \geq \sigma) = 0.1;\ \text{prob}(D \geq \tau) = 0.01$$

e diviso l'insieme dei valori possibili di D nei tre intervalli

$$I = [0, \sigma),\ J = [\sigma, \tau),\ L = [\tau, +\infty).$$

Dati m campioni casuali di ampiezza r della variabile casuale X, abbiamo

ritenuto di poterci attendere, in condizioni ideali, in media, che circa 0.9 m volte D assuma un valore d appartenente ad I, 0.09 m volte un valore in J e 0.01 m volte un valore in L.

Abbiamo allora pensato di valutare la discrepanza dei generatori dall'ipotesi H, rispetto ad un dato test ripetuto m volte, con un "test riassuntivo" di tipo chi-quadro applicato alla distribuzione dei valori assunti da D sulla partizione {I, J, L} di $[0, +\infty)$.

Abbiamo quindi ottenuto, per ciascuno dei nove generatori considerati, 11 tests riassuntivi per r=200 ed 11 per r=1000.

L'analisi campionaria ha confermato, in linea di massima, i risultati ottenuti con l'analisi globale. Alcuni esiti degni di rilievo sono i seguenti:

- i generatori con modulo 10^{10} risultati sufficienti nell'analisi globale risultano sufficienti, ma non eccezionali, nell'analisi campionaria. In particolare, nei test sulla frequenza delle coppie delle cifre e, in misura inferiore, nei test sulla frequenza delle cifre i valori della discrepanza assumono con una frequenza troppo elevata valori bassi.
- i generatori con modulo 2^{35} apparsi accettabili nell'analisi globale, risultano buoni dal punto di vista delle analisi campionarie. L'unica notazione degna di rilievo appare la percentuale relativamente elevata, del 3%, di esiti negativi, al livello di significatività dell'1%, del generatore 5 nel test delle varianze.
- per i generatori con modulo primo l'analisi campionaria dà dei buoni risultati, con qualche eccezione per i gap tests. Per qualche cifra, infatti, la percentuale di esiti negativi, al livello di significatività dell'1%, supera il 3%. Si arriva ad un massimo di esiti negativi del 4.5% per la cifra 7 nel generatore 9.

BIBLIOGRAFIA

1. J.H.Ahrens-U.Dieter-A.Grube, *Pseudo-random numbers. A new proposal for the choice of multiplicators*, Computing 6°. 1970
2. G.Ascoli, *Lezioni di Algebra*, Ed. Tirrenia. Torino. 1965
3. R.R.Coveyou-R.Macpherson, *Fourier Analysis of uniform random generators*, Journal of the Association for Computing Machinery 14°. 1967
4. M.Greenberger, *An a priori determination of serial correlation in Computer generated random numbers*, Mathematics of Computation 15. 1961
5. D.E.Knuth, *The art of computer programming*, Vol:2. Seminumerical Algorithms. Addison-Wesley. London. 1969
6. A.Maturo, *Numeri pseudocasuali*, Montefeltro Ed. Urbino. 1982
7. A.Maturo-N.Cera, *Generazione di numeri pseudocasuali per mezzo di relazioni di ricorrenza su campi di Galois*. Facoltà di Architettura. Pescara. 1983
8. A.Maturo-N.Cera, *Confronto fra alcuni generatori di numeri pseudocasuali*,

Facoltà di Architettura, Pescara. 1983

9. A.Maturo, *Analisi di Fourier di successioni di numeri pseudocasuali: considerazioni e proposte*, Facoltà di Architettura.1983
10. A. Rizzi, *Generazione di distribuzioni statistiche mediante un elaboratore elettronico*, Istituto di Statistica e ricerca sociale G.Gini. Roma. 1977
11. C.S.Smith, *Multiplicative pseudo-random numbers generators with prime modulus*, Journal of the Association for Computing Machinery. 18°. 1971
12. R.C.Tausworthe, *Random numbers generated by linear recurrence modulo two*, Mathematics of Computation 19°. 1965