GENERATORI DI NUMERI PSEUDO-CASUALI

I generatori di numeri casuali sono impiegati in molti campi, ma quelli che a noi interessano in particolar modo sono, nell'ambito della simulazione, le tecniche di Montecarlo e il campionamento statistico.

Esistono molti processi fisici che possono essere considerati generatori di numeri casuali e una lunga storia di costruzione di macchine per generare "numeri a caso" per la simulazione di sistemi stocastici.
Le prime macchine di un certo rilievo furono costruite per studiare la risposta delle centrali telefoniche alla variazione della domanda; altre furono costruite in seguito per studi riguardanti la congestione del traffico.
Ma la "macchina" più caratteristica è certamente la roulette del casinò di Montecarlo (dal quale deriva il nome dei suddetti metodi). Infatti, le macchine di tipo meccanico si basano, proprio come la roulette, sul principio di un disco diviso in settori uguali e numerati, che veniva fatto ruotare da un motore e arrestato dopo un tempo arbitrario, accettando come numero casuale il numero del settore in corrispondenza di un indice.
Le macchine di tipo elettronico, invece, sfruttano un generatore di impulsi pilotato da una sorgente di rumore (agitazione termoionica, sorgente radioattiva, ecc…) in modo da avere impulsi di periodo casuale; gli impulsi vengono contati ciclicamente da 0 a 9 e il contatore viene letto ad intervalli fissati. Usando n sorgenti di rumore in parallelo, e altrettanti contatori, è possibile ottenere numeri random a n cifre decimali.
Un metodo più classico è quello dell'urna contenente palline numerate le quali, prima di ogni estrazione, vengono mescolate; ogni pallina estratta viene reinserita nell'urna per l'estrazione successiva.
Esistono poi delle tavole contenenti milioni di numeri casuali, ma come può essere facile immaginare, la loro gestione è certamente onerosa in termini di tempo e di spazio, anche se realizzata mediante l'uso di computer.
Perciò si sono sviluppati dei metodi per la generazione di sequenze casuali basate sull'uso di formule matematiche.

I fattori che determinano l'accettabilità di un metodo sono essenzialmente i seguenti:

1. I numeri della sequenza generata devono essere uniformemente distribuiti (cioè devono avere la stessa probabilità di presentarsi);
2. I numeri devono risultare statisticamente indipendenti;
3. La sequenza deve poter essere riprodotta;
4. La sequenza deve poter avere un periodo di lunghezza arbitraria;
5. Il metodo deve poter essere eseguito rapidamente dall'elaboratore e deve consumare poco spazio di memoria.

Il progresso tecnologico, che ha consentito di aumentare in modo considerevole la memoria di un elaboratore, ha ovviamente reso inutile il requisito 5.
Nel caso di generazione mediante formule matematiche, i numeri vengono calcolati in modo completamente deterministico e, essendo sempre necessario impostare un numero iniziale (seme), avviene che a parità di seme il metodo fornisca la stessa sequenza soddisfacendo quindi il requisito 3 (ciò è molto importante, poiché talvolta nella simulazione si rende necessario poter ripetere un esperimento nella sua totalità e a parità di condizioni). Tuttavia, la generazione mediante formule matematiche invalida il requisito 2 data la natura deterministica del modello (formula matematica) e del mezzo impiegato per applicarlo (computer).

Autori: Morena De Poli - Roberto Rossi

Quello però che si può fare è generare sequenze di valori le cui caratteristiche siano il più possibile assimilabili a quelle della variabile uniforme (prima fra tutte la caratteristica dell'indipendenza tra le determinazioni della variabile), verificandole mediante l'applicazione di opportuni test, tra i quali ricordiamo il test chi-quadro.
Tali sequenze allora si chiameranno pseudo-casuali e, anche laddove ci riferiremo a tali sequenze con l'aggettivo "casuali" intenderemo sempre dire "pseudo-casuali".

METODO DELLA CONGRUENZA LINEARE

Tale metodo permette, dato un valore iniziale x_0 detto seme, di ottenere una sequenza di numeri pseudo-casuali mediante l'applicazione ripetuta della seguente formula:

$$x_{i+1} = (a * x_i + c) \quad (\text{MOD } m)$$

dove:

a è un coefficiente intero strettamente positivo detto moltiplicatore

c è un coefficiente intero non negativo detto incremento

m è un coefficiente intero strettamente positivo detto modulo

x i è il generico numero della sequenza

MOD è l'operazione modulo, cioè a (MOD b) rappresenta il resto della divisione intera tra a e b.

Il metodo prende il nome dalla seguente definizione:
due numeri x e y si dicono congrui modulo m, e scriveremo $x \equiv y \pmod m$, se essi differiscono per un multiplo intero di m, ossia se $x \pmod m \equiv y \pmod m$.
Nel nostro caso x_{i+1} sarà congruo modulo m a $(a * x_i + c)$.

Il metodo è detto moltiplicativo se c=0, misto se $c \neq 0$; se a=1 il metodo è detto additivo.

Facciamo un esempio banale ipotizzando le seguenti assegnazioni:

- a=3
- c=5
- m=11

Se x_0=3 la sequenza che si ottiene applicando la formula della congruenza modulo m è la seguente:

3, 3, 3, 3, …, cioè una sequenza assolutamente non casuale.

Le cose cambiano se scegliamo x_0=1; la sequenza ottenuta è allora la seguente:

1, 8, 7, 4, 6, 1, 8, 7, 4, 6, 1, …

Possiamo notare che i primi 5 numeri vengono riprodotti interamente!

Se x_0=2 si ottiene

2, 0, 5, 9, 10, 2, 0, 5, 9, 10, 2, …;

ancora una sequenza di 5 numeri ripetuta.

Se modifichiamo invece a assegnandogli il valore 12, e poniamo x_0=1 allora si ottiene:

1, 6, 0, 5, 10, 4, 9, 3, 8, 2, 7, 1, 6, 0, 5, …

una sequenza di lunghezza 11 e cioè pari a m.

E' da sottolineare che in ogni caso i numeri ottenuti sono compresi tra 0 e 10 e cioè tra 0 e m-1.

(Se si desiderano numeri compresi tra 0 e 1 sarà sufficiente dividere il numero x_i per m ad ogni passaggio di riapplicazione della formula)

Da tutto ciò si possono ricavare le seguenti osservazioni:

- La lunghezza massima raggiungibile dalla sequenza generata, senza ripetizione, vale m;
- Particolari scelte di a e c possono ridurre notevolmente la lunghezza utile della sequenza;
- Il valore di x_0 può essere determinante nella lunghezza della sequenza.

E' allora necessario individuare dei criteri per assegnare ad a, c, m e al seme dei valori in modo che la sequenza riprodotta sia la più lunga possibile.

Alcuni studiosi hanno approfondito tale aspetto e hanno individuato i seguenti criteri necessari e sufficienti che garantiscono l'ottimalità del metodo:

1. I parametri c e m devono essere coprimi cioè MCD(c,m) = 1
2. ogni divisore primo di m deve dividere (a-1)
3. se m è multiplo di 4, anche (a-1) lo deve essere.

Alcuni studiosi hanno individuato quindi i seguenti valori nel rispetto dei suddetti criteri:

KNUTH $m = 2^{31}$; $a = int(\pi * 10^8)$; c = 453806245

GOODMAN e MILLER $m = 2^{31}-1$; $a = 7^5$; c = 0

GORDON $m = 2^{31}$; $a = 5^{13}$; c = 0

LEORMONT e LEWIS $m = 2^{31}$; $a = 2^{16} + 3$; c = 0

In ogni caso, per rendere più aleatorio il processo, il seme viene fissato in modo hardware, prelevandone il valore da un contatore interno al computer usato normalmente per altri scopi, oppure ne viene richiesto il valore all'inizio del processo di generazione.

In laboratorio è interessante, prima di cimentarsi nella costruzione di un generatore di numeri pseudo-casuali mediante l'applicazione della congruenza modulo m, studiare la bontà dei generatori messi a disposizione da pacchetti applicativi (ad es. EXCEL, CALC) o linguaggi di programmazione (C, Turbo Pascal, ecc…).
Verrà allora generata una sequenza di numeri pseudo-casuali e costruita la tabella di frequenza raggruppando i dati in classi e conteggiando le frequenze assolute; tale distribuzione empirica verrà messa a confronto con la distribuzione teorica della variabile uniforme mediante l'applicazione del test chi-quadro.