


Sim
UniPisa
LaSpezia

Generazione di numeri pseudocasuali


Simulazione – Lezione n. 9
Corso di Laurea in Informatica Applicata
Università di Pisa, sede di La Spezia

Giovanni A. Cignoni – Simulazione – www.di.unipi.it/~giovanni 1/21 

Sim
UniPisa
LaSpezia

Contenuti


- Valori casuali e valori pseudocasuali
- Valutazione di un generatore
- Algoritmi di generazione
- Generazione di distribuzioni non uniformi
- I metodi dell'inversa e di Box-Muller

Giovanni A. Cignoni – Simulazione – www.di.unipi.it/~giovanni 2/21 

Sim
UniPisa
LaSpezia

Generazione di numeri casuali


- Casuale, come per variabile casuale
 - Valore impossibile da prevedere
 - Ma che rispetta una distribuzione di probabilità
 - Per esempio uniforme: ogni valore è equiprobabile
- Utilizzi
 - Macchine da gioco oneste
 - Crittografia, generazione di chiavi
- Strumenti
 - Generatori hardware
 - Da fenomeni macroscopici: monete, dadi, ruote ...
 - ... o microscopici: rumori, fenomeni ottici e quantistici

Giovanni A. Cignoni – Simulazione – www.di.unipi.it/~giovanni 3/21 

Sim
UniPisa
LaSpezia

Storia e curiosità


- Necessità di tabelle di numeri casuali
 - Parte di una lunga storia di tabelle matematiche calcolate
 - Prime esigenze intorno agli anni '20
 - Poi legata alla diffusione dei metodi Monte Carlo
- RAND Corp.
 - A Million Random Digits with 100,000 Normal Deviates
 - Tabelle realizzate in 8 anni, dal 1947 al 1955
 - Ruota da roulette elettronica e filtri software
- Commodore C64
 - Generatore hardware di numeri casuali
 - Oscillatore sonoro programmabile: PEEK(54299)

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 4/21 

Sim
UniPisa
LaSpezia

Generatori di valori pseudocasuali


- Produzione di valori nella simulazione
 - Per le variabili casuali che abbiamo identificato
 - Da usare negli esperimenti di simulazione
- Sequenze di valori
 - Definite e controllate
 - Cicliche, per non incappare in una fine della sequenza
 - Utilizzabili in esperimenti ripetibili
- Strumenti deterministici
 - Tabelle, come quelle della RAND
 - In realtà algoritmi di generazione

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 5/21 

Sim
UniPisa
LaSpezia

Un compito non facile


- Von Neumann *middle-square*
 - Generatore deterministico, usato sull'ENIAC (1949)
 - Esempio su 4 cifre decimali:
 $x_i = 1234, 1234^2 = 01522756 \rightarrow x_{i+1} = 5227$
 - Diversi problemi, ma al tempo considerato sufficiente
- RANDU, IBM 360 Scientific Subroutine Package
 - Un generatore congruenziale ...
 - ... ma una pessima scelta dei parametri
 - "... its very name *RANDU* is enough to bring dismay into the eyes and stomachs of many computer scientists!" (Donald Knuth)

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 6/21 

Sim
UniPisa
LaSpezia

Valutazione di un generatore

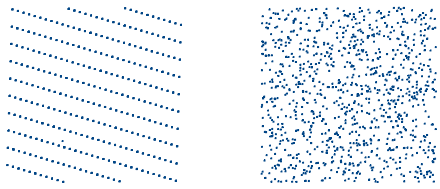
- Sequenze cicliche
 - Rispettose di una distribuzione, di solito uniforme
 - Con numeri “sufficientemente” grandi (o precisi)
 - Di periodo “sufficientemente” grande
 - Con punto di partenza dato (eventualmente non predicibile)
- Requisiti per i generatori usati in simulazione
 - Disponibilità, praticità d'uso
 - Generazione efficiente
- Prove di casualità
 - Strumenti statistici
 - Per verificare la “bontà” delle sequenze


Giovanni A. Cignoni – Simulazione – www.di.unipi.it/~giovanni 7/21 

Sim
UniPisa
LaSpezia

Scatter plot

- Disposizione di punti su un n -spazio
 - Valori “successivi” come coordinate
 - Ricerca di regolarità




Giovanni A. Cignoni – Simulazione – www.di.unipi.it/~giovanni 8/21 

Sim
UniPisa
LaSpezia

Diehard test


- Batteria di prove
 - Uno dei benchmark più riconosciuti
 - G. Marsaglia, dal 1995 a oggi
- Riconoscimento di distribuzioni derivate
 - Poisson (birthday test)
 - Esponenziale (min distance, random spheres)
 - Normale (parking lot)
- Riconoscimento di schemi
 - Sequenze di bit (count the 1's)
 - Sovrapposizione di parole (monkey test)

Giovanni A. Cignoni – Simulazione – www.di.unipi.it/~giovanni 9/21 

Sim
UniPisa
LaSpezia

Generatori congruenziali


- Linear Congruential Generator
 - $X_{i+1} = (aX_i + c) \bmod m$
 - Intervallo generato $[0 .. m - 1]$, X_0 valore iniziale o seme
 - Periodo m e tutti i valori generati se e solo se:
 - m e c primi fra loro;
 - $a - 1$ divisibile da tutti i fattori primi di m ;
 - $a - 1$ è multiplo di 4 se m è multiplo di 4.
- Generatori moltiplicativi: $c = 0$
- Generatori noti
 - Il congruenziale con $m = 2^{32}$ di Knuth & Lewis (quick & dirt)
 - Il *Minimal Standard* di Park & Miller ($c = 0$)

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 10/21 

Sim
UniPisa
LaSpezia

Generatori lagged-Fibonacci


- Serie di Fibonacci
 - $F_n = F_{n-1} + F_{n-2}$, $F_0 = 0$, $F_1 = 1$
 - Numerose proprietà della serie e dei numeri di Fibonacci
 - Comune in natura, dalle conchiglie ai petali dei fiori
- Generalizzazione: $S_n = S_{n-h} \circ S_{n-k}$ $0 < h < k$
- Generatore: $S_n = (S_{n-h} \circ S_{n-k}) \bmod m$
 - Operatore binario: somma, moltiplicazione, xor bit a bit, ...
 - Scelta mirata dell'operatore e di h , k e m
 - Generatori sensibili all'inizializzazione

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 11/21 

Sim
UniPisa
LaSpezia

Generatori di Marsaglia

- Proposti in un post su sci.stat.math, nel '99
 - Variazioni su temi noti (congruenziale, 4-lagged-Fibonacci)
 - Composizione
- Multiply with carry
 - Un congruenziale in cui c varia: $X_i = (aX_{i-r} + c_{i-r}) \bmod m$
- Triple shift
 - Permutazioni sui bit del valore corrente (shift e xor)
- KISS
 - Multiply with carry + triple shift + congruential

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 12/21 

Sim
UniPisa
LaSpezia

Mersenne Twister

- Proposto da M. Matsumoto, T. Nishimura, nel '97
 - Variazione sul tema del Lagged-Fibonacci
 - Una tabella di 624 elementi, in due parti (pivot a 397, primo)
 - Valore dalla tabella previa perturbazione con shift
 - Ricalcolata quando esaurita
- Ottime caratteristiche
 - Periodo lunghissimo: $2^{19937}-1$ (primo di Mersenne)
 - Passa tutti i test conosciuti
 - Ragionevolmente efficiente
- Generalmente considerato lo stato dell'arte

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 13/21

Sim
UniPisa
LaSpezia

Generare distribuzioni non uniformi

- Generare valori di una VC discreta
 - Nota $F_X(x) = P(X \leq x) = \sum_{x_i \leq x} f_X(x_i)$
 - A partire da un generatore uniforme
- Ricostruire un campione "esteso"
 - I valori della VC da generare sono rappresentati da gettoni
 - Replicati proporzionalmente alla probabilità dei valori
- Es. il lancio di due dadi
 - Occorrono 36 gettoni
 - $1 \times 2^2, 2 \times 3^2, 3 \times 4^2, 4 \times 5^2, 5 \times 6^2, 6 \times 7^2, 5 \times 8^2, 4 \times 9^2, \dots, 1 \times 12^2$
- Estrarre uniformemente con reiserimento

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 14/21

Sim
UniPisa
LaSpezia

Il metodo del cappello

- Algoritmo replicabile
 - Generare y uniforme in $[1 \dots 36]$, riportare su F_X per ricavare x
 - I valori x che si ottengono sono distribuiti secondo F_X

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 15/21

Sim
UniPisa
LaSpezia

Il metodo dell'inversa

- Estensione alle distribuzioni continue
 - Purché F_x esista e sia invertibile

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 16/21

Sim
UniPisa
LaSpezia

Distribuzioni invertibili (1)

- Esponenziale

$$F_x(x) = 1 - e^{-\lambda x} \quad F_x^{-1}(u) = -\frac{\ln(1-u)}{\lambda}$$
- Weibull

$$F_x(x) = 1 - e^{-(x/\lambda)^k} \quad F_x^{-1}(u) = \lambda(-\ln(u))^{1/k}$$
- Cauchy

$$F_x(x) = \frac{1}{\pi} \arctan\left(\frac{x-x_0}{y}\right) + \frac{1}{2} \quad F_x^{-1}(u) = x_0 + y \tan\left(\pi\left(u - \frac{1}{2}\right)\right)$$

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 17/21

Sim
UniPisa
LaSpezia

Distribuzioni invertibili (2)

- Triangolare

$$F_x(x) = \begin{cases} \frac{(x-a)^2}{(b-a)(c-a)} & a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & c \leq x < b \end{cases}$$

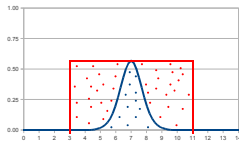
$$F_x^{-1}(u) = \begin{cases} a + \sqrt{u(b-a)(c-a)} & a \leq x \leq c \\ b - \sqrt{(1-u)(b-a)(b-c)} & c \leq x < b \end{cases}$$

Giovanni A. Cignoni - Simulazione - www.di.unipi.it/~giovanni 18/21

Sim
UniPisa
LaSpezia

Metodo degli scarti

- Scartare i valori “fuori” della distribuzione
 - Data $f_X(x)$ definita in $[a, b]$ e limitata $c = 1/\max(f_X(x))$
 - Generare due valori pseudocasuali uniformi u_1 e $u_2 \in [0, 1]$
 - $x = a + u_1(b-a)$ è accettato se $u_2 \leq cf_X(x)$
 - L'efficienza dipende da quanto è “in scatolata” $f_X(x)$



Giovanni A. Cignoni – Simulazione – www.di.unipi.it/~giovanni 19/21

Sim
UniPisa
LaSpezia

Distribuzioni normali

- Come generarle?
 - $F_X(x)$ non è invertibile (non è neanche definita)
 - $f_X(x)$ non è inscatolabile (a meno di costose approssimazioni)
- Metodo di Box-Muller per generare $N(0, 1)$
 - Generare due valori uniformi $u_1, u_2 \in (-1, 1)$
 - Eliminare quelli fuori del cerchio unitario (m. degli scarti)
 - Scalarli rispetto alla distanza dall'origine
 - I valori x_1 e x_2 sono $N(0, 1)$

$$R = u_1^2 + u_2^2 < 1 \quad r = \sqrt{\frac{-2 \ln(R)}{R}} \quad x_1 = u_1 r, \quad x_2 = u_2 r$$

Giovanni A. Cignoni – Simulazione – www.di.unipi.it/~giovanni 20/21

Sim
UniPisa
LaSpezia

Riferimenti

- G. Marsaglia, *Sci.Stat.Math post*, 1999.01.12
- P. Hellekalek, *Good RNG are (not so) easy to find*, in *Mathematics & Computers in Simulation*, n. 48, 1998
- G. Gallo, *Note di Simulazione*, cap. 4.4
- M. Pidd, *Computer Simulation in Management Science*, Cap. 11.2-11.7, J. Wiley & Sons, 1998

Giovanni A. Cignoni – Simulazione – www.di.unipi.it/~giovanni 21/21

