

GENERAZIONE DI NUMERI CASUALI



"The generation of random numbers is too important to be left to chance."

..Robert R. Coveyou, Oak Ridge National Laboratory

numero casuale \Leftrightarrow risultato non prevedibile (p.e. lancio di un dado)

- Un **numero casuale** è un numero scelto da un insieme di valori egualmente probabili, cioè un numero estratto da una distribuzione uniforme
- in una **sequenza** di numeri casuali ogni numero estratto deve essere statisticamente indipendente dagli altri

I numeri casuali sono utili in vari casi:

- Generazione di dati cifrati e password
- Simulazione e modellazione di fenomeni complessi
- Selezione di campioni casuali

229

COME GENERARE NUMERI CASUALI?

- È possibile generare numeri casuali con un computer? È più difficile di quanto sembri ...
- ...è difficile far fare qualcosa 'per caso' ad un computer. Un computer segue le istruzioni ciecamente ed è quindi completamente prevedibile (se un computer non segue le istruzioni in questo modo è rotto)
- Ci sono 2 approcci per la generazione di numeri casuali con un computer:
 - Generatori di numeri pseudo-casuali
 - Generatori di numeri casuali veri (da fenomeni fisici, p.e. gli istanti temporali del decadimento radioattivo sono totalmente imprevedibili).

230

GENERAZIONE DI NUMERI PSEUDO-CASUALI

Gli algoritmi per la generazione di numeri pseudo-casuali utilizzano formule matematiche o tabelle pre-calcolate per produrre sequenze di numeri che *sembrano casuali*.

Gli algoritmi oggi disponibili sono buoni e i numeri generati sono come quelli realmente casuali

Un esempio di generatore di numeri pseudo-casuali è il metodo lineare congruenziale.

231

METODO LINEARE CONGRUENZIALE

Il generatore è definito dalla seguente relazione ricorrente:

$$X_{n+1} = (aX_n + c) \bmod m$$

X_n è la sequenza dei valori pseudo-casuali
 a, c , e m sono costanti intere:

$0 < m$ è il modulo

$0 < a < m$ è il "moltiplicatore"

$0 \leq c < m$ l'"incremento" ($c=0$ generatore Park-Miller)

$0 \leq X_0 < m$ il "seme" o "valore iniziale"

Questo metodo è buono ma molto sensibile alla scelta dei valori dei coefficienti c , m , e a .

L'operatore **mod** trova il resto: dati due numeri a (il dividendo) e m (il divisore), **mod** m è il resto della divisione di a per m , p.e. "7 mod 3" vale 1, mentre "9 mod 3" vale 0.

232

PRNG: PREGI E DIFETTI

I generatori di numeri pseudo-casuali (PRNG) sono:

- **efficienti**: possono produrre molti numeri in poco tempo
- **deterministici**: una data sequenza di numeri può essere riprodotta in un momento successivo se il valore iniziale della sequenza è noto
- **periodici**: la sequenza si può ripetere dopo un certo numero di numeri! I generatori attuali hanno un periodo così lungo che può essere ignorato in molte applicazioni.

PRNG utili per generare velocemente molti numeri e quando serve poter replicare la sequenza (p.e. simulazioni e modelli), non vanno bene se i numeri devono essere davvero imprevedibili (p.e. codici per la cifratura dei dati e scommesse)

233

GENERAZIONE DI NUMERI PSEUDO-CASUALI IN SAS

Le funzioni e le routine di SAS per la generazione di numeri casuali producono sequenze di numeri partendo da un valore iniziale detto SEME (*seed*)

Il seme deve essere un intero non negativo minore di $2^{31}-1$

È sempre possibile riottenere la successione di numeri casuali utilizzando lo stesso DATA step.

Se si usa il valore zero come seme è l'orologio di sistema che inizializza la sequenza, in tal caso la sequenza di numeri casuali non è replicabile

Il seme può essere una costante intera o una variabile che contiene la costante intera

La variabile seed deve essere inizializzata prima della 1^a esecuzione della funzione o della CALL routine.

234

GENERAZIONE DI NUMERI PSEUDO-CASUALI IN SAS

Distribuzione uniforme in (0,1).

RANUNI(seme)

seme intero $< 2^{31}-1$. Se $seme \leq 0$, la sequenza di numeri casuali viene inizializzata utilizzando come seme l'ora dell'orologio interno.

Utilizza un generatore a modulo moltiplicativo con $m=2^{31}-1$ e $a=397204094$

$$X_{n+1} = (aX_n) \bmod (2^{31} - 1)$$

- $seme = X_0$;
- X_0 resta invariato per tutta l'esecuzione;
- Per variare X_0 usare CALL.
- Esempi: generala.sas-generalc.sas

235

FUNZIONI SAS PER LA GENERAZIONE DI NUMERI PSEUDO-CASUALI

NORMAL normale standard
RANNOR normale standard
RANBIN binomiale
RANCAU Cauchy
RANEXP esponenziale standard
RANGAM gamma standard
RANPOI Poisson
RANTBL distribuzione discreta
RANTRI distribuzione triangolare
RANUNI uniforme (0,1)
UNIFORM uniforme (0,1)

236

TRASFORMAZIONE DI PROBABILITÀ INVERSA

Metodo per la generazione di numeri casuali da una distribuzione di probabilità qualunque: se X è una v.c. continua, con funzione di ripartizione F_X , e $Y=F_X(X)$, allora $Y \sim U[0,1]$.

Trasformazione di probabilità inversa

se $Y \sim U[0,1]$ e X è definita come

$$X = F_X^{-1}(Y),$$

allora $X \sim F_X$.

il metodo è applicabile ma può essere troppo oneroso computazionalmente per alcune distribuzioni

La trasformazione di Box-Muller è un algoritmo meno generale ma più efficiente computazionalmente.

237

ALGORITMO TRASFORMAZIONE DI PROBABILITÀ INVERSA

. Sia $X \sim F(x)$, $F(x)$ invertibile

1. Generare un numero casuale $u \sim U[0,1]$
2. Calcolare il valore di x tale che $F(x)=u$ cioè $X = F^{-1}(U)$ chiamiamo x_{scelto} questo valore
3. Consideriamo x_{scelto} come il numero casuale estratto dalla distribuzione F .

238

PRNG NORMALE E GAMMA

Distribuzione normale

RANNOR(*seme*)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-(x-\mu)^2 / 2\sigma^2\right]$$
$$\mu = 0, \sigma^2 = 1$$

Trasformazione Box-Muller da uniforme.

Distribuzione gamma

RANGAM(*seme, a*)

$$f(x) = \left(\frac{x}{b}\right)^{a-1} \frac{\exp(-x/b)}{b\Gamma(a)} I_{(0,\infty)}(x)$$
$$b > 0, a > 0$$

- per $b=1 \rightarrow$ Gamma standard
 - per $2*a$ intero \rightarrow chi-quadro con $2*a$ g.l.
- Metodo di accettazione/rifiuto (Cheng, 1977; Fishman, 1978) (genera5.sas)

239