TECNICHE DI SIMULAZIONE

Generazione di Variabili Aleatorie

Francesca Mazzia

Dipartimento di Matematica
Università di Bari

a.a. 2004/2005

Generazione di Variabili Aleatorie

- Supponiamo che una distribuzione sia stata completamente specificate.
- Analizziamo come generare variabili appartenenti a questa distribuzione da usare come input per un modello di simulazione.
- Tutte le tecniche suppongono che si abbia a disposizione una routine per generare numeri casuali uniformemente distribuiti in $(0, 1)$.

Tecnica della trasformazione inversa

- Analizziamo la distribuzione esponenziale.
- pdf :

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & altrimenti \end{cases}$$

- cdf:

$$F(x) = \begin{cases} 0 & x < 0 \\ \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}, & x \geq 0 \end{cases}$$

- λ è interpretato come il numero medio di occorrenze per unità di tempo (per esempio il numero medio di arrivi per unità di tempo).

Tecnica della trasformazione inversa

- La tecnica di trasformazione inversa può essere utilizzata per qualsiasi distribuzione.
- Passo 1: calcola la cdf della variabile aleatoria desiderata X. Per l'esponenziale $F(X) = 1 - exp(-\lambda x), x \geq 0$.
- Passo 2: Poni $F(X) = R$ nel range di X. $1 - exp(-\lambda X) = R$, $x \geq 0$. Poichè X è una variabile aleatoria, R è una variabile aleatoria. Si dimostra che R ha una distribuzione uniforme in $(0, 1)$.
- Passo 3. Risolvi l'equazione $F(X) = R$ per X in termini di R. Per la distribuzione esponenziale: $X = -\frac{1}{\lambda} ln(1 - R)$.
- Passo 4. Genera numeri casuali unformi $R_1, R_2, ...$e calcola X_i usando la relazione determinata al passo 3.

Distribuzione Uniforme

- X distribuita uniformemente in $[a, b]$.
- $X = a - (b - a)R$
- $F(x) = \frac{x-a}{b-a}, a \leq x \leq b$, 0 altrimenti.
- $F(X) = R$ implica $X = a - (b - a)R$

Distribuzione di Weibull

- $$f(x) = \begin{cases} \frac{\beta}{\alpha}(\frac{x}{\alpha})^{\beta-1} exp(-(\frac{x}{\alpha})^{\beta}), & x \geq 0 \\ 0 & altrimenti \end{cases}$$
- passo 1: cdf $F(X) = 1 - exp(-(x/\alpha)^{\beta})$, $x \geq 0$
- passo 2: $F(X) = 1 - exp(-(X/\alpha)^{\beta}) = R$
- passo 3: Risolvere per X in termini di R:
  $X = \alpha(-ln(1-R))^{1/\beta}$

Tecnica della trasformazione inversa

- Se la distribuzione continua non permette di calcolare facilmente X in funzione di R si può usare una inversa approssimata della cdf.

Distribuzioni continue empiriche

- Se non si è riusciti a trovare una distribuzione teorica che sia un buon modello per i dati di input, bisogna usare distribuzioni empiriche.
- Esempio: abbiamo a disposizione n=5 osservazioni dei tempi di risposta di una squadra anti incendi agli allarmi: 2.76 1.83 0.80 1.45 1.24 (minuti).
- Prima di collezionare nuovi dati si vuole sviluppare un modello preliminare basato su queste 5 osservazioni.
- Si assume che il tempo di risposta ha un range compreso fra 0 e c, con c incognito.
- Stimiamo c come il massimo dei tempi osservati 2.75.
- ordiniamo i dati dal più piccolo al più grande $x_{(1)}, ...x_{(5)}$ e definiamo $x_{(0)} = 0$.

Distribuzioni continue empiriche

- assegnamo una probabilità di 1/n ad ogni intervallo $x_{(i-1)} \leq x \leq x(i)$, quindi la probabilità cumulativa sarà $cp = (0, 0.2, 0.4, 0.6, 0.8, 1.0)$, quindo $cp_i = i/n$.
- Calcoliamo la cdf empirica usando una tecnica di interpolazione lineare. In ogni intervallino $[x_{(i-1)}, x_{(i)}]$ tracciamo una retta che passa per i punti $(x_{(i-1)}, cp_{i-1}), (x_{(i)}, cp_i)$.

Distribuzioni continue empiriche

- Essendo interessati a calcolare i valori di x conosciuti i valori di cp che appartengono ad una distribuzione uniforme, calcoliamo il polinomio di primo grado in ogni intervallo che passa per $(cp_{(i-1)}, x_{(i-1)}), (cp_i, c_{(i)})$, usando l'interpolazione di Newton:

$$x = x_{(i-1)} + \frac{x_{(i)} - x_{(i-1)}}{i/n - (i-1)/n}(R - (i-1)/n)$$

per $(i-1)/n < R \leq i/n$.

Distribuzioni continue empiriche

- Se i dati sono tanti è più conveniente eseguire prima una distribuzione delle frequenza in un numero inferiore di intervalli e quindi costruire la cdf empirica basandosi su questa distribuzione.

Distribuzioni discrete empiriche

- Alla fine del giorno, il numero di spedizioni è contato come 0,1,2 con frequenza relativa osservata di 0.50, 0.30 e 0.20. I dirigenti hanno chiesto di sviluppare un modello per migliorare l'efficienza delle operazioni di carico e scarico.
- X che rappresenta il numero di spedizioni viene modellata come una variabile casuale discreta.
- $p(0) = 0.50$, $P(1) = 0.30$, $P(2) = 0.20$

Distribuzioni discrete empiriche

- cdf

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.5 & 0 \leq x < 1 \\ 0.8 & 1 \leq x < 2 \\ 1.0 & 2 \leq x \end{cases}$$

- Per generare i valori dobbiamo verificare in quale intervallo cade R_i e quindi prendere il valore della x corrispondente:

$$X = \begin{cases} 0 & R \leq 0.5 \\ 1 & 0.5 < R \leq 0.8 \\ 2 & 0.8 < R \leq 1.0 \end{cases}$$

Distribuzione discrete uniforme

- Consideriamo una distribuzione discreta in $\{1, 2, ..., k\}$ con $p(x) = 1/k$;
- $$F(x) = \begin{cases} 0, & x < 1 \\ i/k, & i/k \leq x < (i+1)/k, \ i = 1, .., k-1 \\ 1. & k \leq x \end{cases}$$
- se $(i-1)/k < R \leq i/k$ allora $X = i$.
- quindi $i - 1 < Rk \leq i$ e $Rk \leq i < Rk + 1$ e i è uguale al più piccolo intero maggiore di Rk.
- $R_1 = 0.78$, $X_1 = 8$, $R_2 = 0.03$, $X_2 = 1$ ...

Metodo di convoluzione

- La distribuzione di probabilità della somma di due o più variabili aleatorie è chiamata una convoluzione delle distribuzioni delle variabili originali.
- Questa tecnica si può applicare per ottenere la distribuzione di Erlang.
- Ha importanza la relazione con altre variabili che possono essere generate facilmente.

La distribuzione di Erlang

- Una variabile aleatoria di Erlang con parametri (K, θ) è la somma di K variabili aleatorie esponenziali X_i con media $1/(k\theta)$.
- Una variabile appartenente alla distribuzione di Erlang può essere generata come

$$X = \sum_{i=1}^{K} -\frac{1}{K\theta} ln R_i = -\frac{1}{K\theta} ln(\prod_{i=1}^{K} R_i)$$