# Metodi Montecarlo

Laboratorio Informatico per la Statistica

15/04/2008

## Introduzione

I Metodi Montecarlo sono delle procedure di calcolo basate sull'uso di generatori di numeri casuali per ottenere il risultato.

Hanno diverse applicazioni fra cui:

- integrazione numerica (soprattutto per la soluzione di integrali multipli),
- soluzione numerica di (sistemi di) equazioni differenziali,
- simulazione di processi stocastici,
- modellazione di sistemi fisico-statistici (sono intrinsecamente dei processi stocastici),
- ...

Vedremo alcuni esempi di applicazione di queste tecniche.

## Integrazione Numerica

Un campo di notevole importanza nell'applicazione dei metodi Montecarlo è la risoluzione di problemi di integrazione numerica. Questi metodi sono particolarmente efficienti rispetto alle tecniche tradizionali soprattutto per integrare funzioni non lineari e per risolvere integrali multipli grazie al comportamento probabilistico dell'errore di approssimazione.

Per illustrare il comportamento considereremo il calcolo di un integrale sia con un metodo di integrazione tradizionale (*metodo del trapezio*) che con dei metodi di tipo Montecarlo.

## Il metodo del trapezio

Come tutti i metodi classici di integrazione numerica si basa sull'idea di dividere l'intervallo $[a, b]$ in tanti sottointervalli, non necessariamente della stessa lunghezza, ed in ciascuno di questi calcolare l'area al di sotto della funzione e quindi sommarla su tutti i sottointervalli per ottenere il valore dell'integrale.

Il metodo più semplice si basa sull'idea di integrabilità secondo Riemann (quella che si studia normalmente nei corsi di analisi): l'integrale di una funzione $f$, continua su di un intervallo chiuso finito, esiste sempre ed ha un valore che  ottenuto dai due limiti:

$$\lim_{n\to\infty} L(f : P_n) = \int_a^b f(x)dx = \lim_{n\to\infty} U(f : P_n)$$

in cui $P_1, P_2, \ldots$ è una sequenza di partizioni con la proprietà che la lunghezza del più grande sottointervallo in $P_n$ converge a zero se n tende all'infinito. Scelta una partizione P si possono definire le due funzioni $L$ ed $F$ come segue:

$$L(f : P_n) = \sum_{i=2}^{n} m_i(x_i - x_{i-1})U(f : P_n) = \sum_{i=2}^{n} M_i(x_i - x_{i-1})$$

dove $m_i(x_i - x_{i-1})$ e $M_i(x_i - x_{i-1})$ indicano rispettivamente il valore massimo ed il valore minimo della funzione nel sottointervallo. In parole semplici si approssima la funzione dal basso e dall'alto. Scegliendo una suddivisione dell'intervallo di definizione della funzione sempre più fine si giunge ad avere il valore dell'integrale definito. Il metodo del trapezio suddivide l'intervallo di integrazione $[a, b]$ in sottointervalli e in ognuno di questi si calcola il valore dell'integrale considerando l'area del trapezoide che ha come base la larghezza

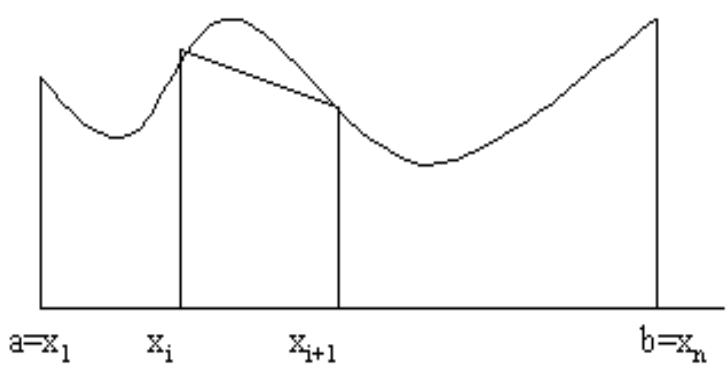


Figure 1: Illustrazione del metodo del trapezio

dell'intervallo $[x_i, x_{i+1}]$ e come punti verticali $f(x_i)$ e $f(x_{i+1})$. La somma delle aree approssima il valore dell'integrale.

Supponendo gli intervalli di egual dimensione, si ha che l'area che approssima la funzione è data da:

$$T(f;P) = \frac{b-a}{n-1}\sum_{i=2}^{n-1} f(x_i) + \frac{b-a}{2\cdot(n-1)}[f(x_1)+f(x_n)] \tag{1}$$

Si dimostra che l'errore $e$ soddisfa:

$$e < \frac{b-a}{12}\left(\frac{b-a}{n-1}\right)^2 \max_{a\le x\le b}\left|f''(x)\right| \tag{2}$$

Da questa formula si può ricavare il valore di $n$ fissato l'errore desiderato $e$:

$$n < \left\lceil \sqrt{\frac{b-a}{12\cdot e}(b-a)^2 \max_{a\le x\le b}|f''(x)|} + 1 \right\rceil \tag{3}$$

Nella pratica il processo di integrazione numerica si applica continuando a suddividere l'intervallo di integrazione in nuovi sottointervalli e valutando ogni volta l'errore commesso. Nel fare questo si usano dei criteri di suddivisione, solitamente quello dicotomico, che permettono di riutilizzare i valori calcolati nel passo precedente. Il processo termina quando uno dei seguenti criteri è verificato:

- si raggiunge una approssimazione sufficiente in base alla formula dell'errore,
- il numero di suddivisioni dell'intervallo ha superato una soglia prefissata.

Si noti, infine, che il metodo presentato ha come unico vincolo che la funzione integranda $f(x)$ sia integrabile nel senso di Riemann e non che sia derivabile. Nei casi in cui la funzione non sia derivabile occorre un criterio di arresto alternativo al primo. In questi casi si adotta la regola che se due iterazioni successive danno lo stesso risultato il valore dell'integrale è sufficientemente approssimato.

In generale le regole di integrazione numerica non probabilistiche hanno una maggiorazione dell'errore che è funzione del numero $n$ di suddivisioni dell'intervallo di integrazione e del numero di derivate possedute dalla funzione integranda. In genere questo valore è dell'ordine[1] $O(n^{-r})$ con $r \ge 1$. Nel caso della regola del trapezio l'errore è dell'ordine di $O(n^{-2})$.

[1] Si ricorda che $f(x) = O(g(x))$ quando $x \to x_0$ sse esiste una costante $c$ tale che:

$$\lim_{x\to x_0} \sup \left|\frac{f(x)}{g(x)}\right| = c$$

Per concludere si riporta il codice per il calcolo del valore dell'integrale usando il metodo del trapezio.

```
#include <stdio.h>
#include <stdlib.h>
#include <math.h>

/*
Funzioni integrande tutte debbono avere forma
         double nomeFunzione(double);
*/
double sinInt(double x)
{
       if (x == 0.0)
          return 1;
       return (sin(x)/x);
}

double para(double x)
{
       return x*x;
}

/*
punti(): restituisce il numero di punti necessari per approssimare
   l'integrale
Parametri
         double inf     limite inferiore integrazione
         double sup     limite superiore integrazione
         double err     valore errore ammesso
         double max     massimo derivata seconda su (inf,sup)
Restituisce
           int  numero punti necessari all'integrazione
*/
int punti(double inf, double sup, double err, double max)
{
    int pt;

    pt = ceil( sqrt(pow(sup-inf,3)*fabs(max)/(12*err)) + 1);

    return pt;
}
```

ovvero $f(x)$ non cresce più velocemente di $g(x)$.

```
/*
integra(): calcola l'integrale di una funzione con il metodo del
           trapezio
Parametri
         double inf     limite inferiore integrazione
         double sup     limite superiore integrazione
         double (*f)(double)   funzione integranda
         int punti      numero punti divisione (inf,sup)
*/
double integra(double inf,double sup,double (*f)(double),int punti)
{
       double dimInt=(sup-inf)/(punti-1), somma;
       int i;

       for(i=1,somma=0.0;i<punti-1;i++)
          somma += (*f)(inf+dimInt*i);

       return    dimInt*somma + dimInt/2 *((*f)(sup)+(*f)(inf));

}

int main()
{   int pun;

    pun = punti(0.0,1.0,1e-10,0.5);
    printf("%d area %.16f\n",pun,integra(0.0,1.0,sinInt,pun));

    system("pause"); return 0;
}
```

Il codice mostra il calcolo del valore della funzione seno integrale fra 0 ed 1. Questa funzione, definita come:

$$Si(x) = \int_x^1 \frac{sin(t)}{t} dt$$

è particolarmente utile nella teoria delle comunicazioni ma non è calcolabile direttamente. Il valore dell'integrale fra

$$\int_0^\infty \frac{sin(x)}{x} dx = \frac{\pi}{2}$$

Il valore del massimo della derivata seconda si può ricavare dallo sviluppo in serie di Taylor. Nell'intervallo $[0, 1]$ si vede che questo valore è inferiore a

$0,5$. Il valore ricercato nell'esempio del codice è

$$\int_0^1 \frac{sin(x)}{x} dx = 0.9460830704$$

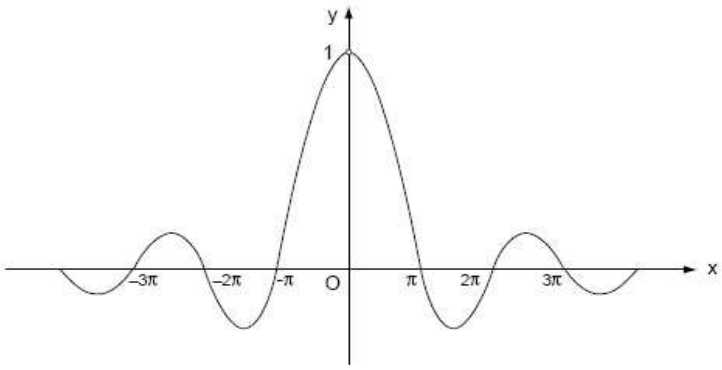


Figure 2: grafico della funzione $\frac{sin(x)}{x}$

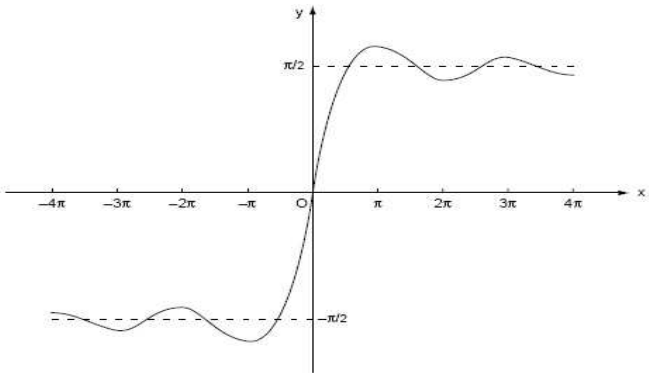


Figure 3: grafico della funzione $\int_0^\infty \frac{sin(x)}{x} dx$

# Integrazione Montecarlo

I metodi Montecarlo per l'integrazione seguono un approccio radicalmente diverso per raggiungere il risultato. Esistono due varianti fondamentali nella risoluzione del problema il primo è detto *metodo accetta-rifiuta*, il secondo

*metodo del valor medio*. Entrambi i metodi sono definiti per l'intervallo di integrazione $[0, 1]$ e si dimostrano particolarmente vantaggiosi nell'integrazione multidimensionale[2].

## Metodo accetta–rifiuta

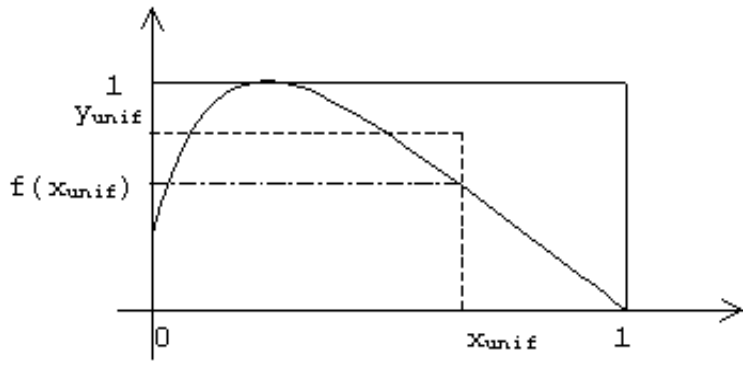


Figure 4: Il metodo accetta-rifiuta

Questo approccio prevede che la funzione integranda $f(x)$ abbia lo stesso segno su tutto l'intervallo di integrazione ed un valore massimo $f_{max}$[3] che supponiamo essere l'intervallo $[0, 1]$.

Quindi, senza perdere di generalità si suppone $x \in (0, 1)$ e $y \in (0, 1)$[4]. Il metodo procede generando $N$ coppie di punti entrambe distribuite secondo una distribuzione uniforme e contando quante coppie $< x_{unif}, y_{unif} >$ verificano $y_{unif} \leq f(x_{unif})$, ovvero il punto si trova al di sotto della funzione e quindi appartiene all'area. Alla fine il rapporto fra i punti che appartengono

[2]Il fatto che l'intervallo di integrazione sia compreso fra 0 ed 1 non fa perdere di generalità infatti:

$$\int_a^b f(x)dx = (b-a)\int_0^1 f(t)dt$$

ponendo $t = \frac{x-a}{b-a}$, $dt = \frac{dx}{b-a}$ si ha $dx = (b-a)dt$ e $x = (b-a) \cdot t + a$.

[3]ci si può riportare a questo caso suddividendo l'intervallo di integrazione in parti nelle quali $f(x)$ ha lo stesso segno e quindi sommando i valori degli integrali. Se $f(x)$ è negativa su tutto l'intervallo si considera $-f(x)$ e si cambia il segno all'area dedterminata.

[4]La funzione può avere una ordinata $f(x) \in (0, 1)$ supposto $0 \leq f(x) \leq f_{max}$ ponendo

$$\frac{f(x)}{f_{max}}$$

.

all'area, $N_{amm}$, ed i punti generati, $N$, fornisce una stima probabilistica del valore dell'area[5].

Questo metodo quindi fornisce una stima dell'area tanto migliore quanto più è grande il numero di coppie di punti considerate.

## Il metodo del valor medio

Un altro approccio al calcolo del valore dell'integrale si basa sul teorema del valor medio dell'integrale secondo il quale, data una funzione $f$ continua su $[a, b]$, esiste un valore $c$ tale che:

$$J = \int_a^b f(x)dx = f(c)(b-a)$$

dove $f(c)$ è il valor medio della funzione nell'intervallo $[a, b]$. Questo approccio non impone condizioni sul segno di $f(x)$ nell'intervallo $[a, b]$ e riduce il problema a quello del calcolo del valor medio di $f(x)$ nell'intervallo $[a, b]$.

Si supponga di campionare la funzione $f(x)$ in $N$ punti indipendenti $X_1, X_2, \cdots, X_N$ secondo una certa distribuzione di probabilità $p(x)$[6]. Una stima di $J$ è data da:

$$\hat{J} = F_N = \frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)}{p(X_i)}$$

dove $F_N$ è una variabile casuale. Si consideri ora:

$$\begin{aligned} E[F_N] = \quad & E\left[\frac{1}{N}\sum_{i=1}^{N}\frac{f(X_i)}{p(X_i)}\right] = \\ & \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\frac{f(x)}{p(x)}p(x)dx = \\ & \int_a^b f(x)\frac{1}{N}\cdot\sum_{i=1}^{N} i = \int_a^b f(x)dx = \\ & J \end{aligned}$$

Ponendo $Y_i = f(X_i)/p(X_i)$ si ha:

$$F_N = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

[5]Se si era partiti da $\int_a^b f(x)dx$ con $0 \leq f(x) \leq f_{max}$ la stima del valore dell'integrale è data da:

$$\frac{N_{amm}}{N}\cdot f_{max}\cdot(b-a).$$

[6]Ad esempio

$$p(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & altrove \end{cases}$$

Se $Y = Y_1$ la varianza diviene:

$$Var[Y] = E[Y^2] - E[Y]^2 = \int_{-\infty}^{+\infty} \frac{f^2(x)}{p(x)} p(x)dx - J^2$$

supponendo che il valore sia finito e che i campioni $Y_i$ siano indipendenti si ha:

$$Var[F_N] = Var\left[\frac{1}{N}\sum_{i=1}^{N} Y_i\right] = \frac{1}{N^2} Var\left[\sum_{i=1}^{N} Y_i\right] = \frac{1}{N^2}\sum_{i=1}^{N} Var\left[Y_i\right] = \frac{1}{N} Var[Y].$$

di conseguenza:

$$\sigma[F_N] = \frac{1}{\sqrt{N}}\sigma(Y)$$

che converge con un tasso $O(N^{-1/2})$. Questo risultato ha diverse conseguenze:

- l'errore quadratico medio decresce con la radice quadrata del numero di punti $N$ indipendentemente dalla forma della funzione ed è legato al valore di $\sigma$;
- il valor medio è uno stimatore non distorto di $J$;

Applicando la diseguaglianza di Chebychev:

$$Pr\left\{|F - E[F]| \geq \left(\frac{Var[F]}{\delta}\right)^{1/2}\right\} \leq \delta$$

verificata per ogni variabile casuale $F$ con varianza $Var[F]$ finita si ha:

$$Pr\left\{|F_N - J| \geq \left(\frac{Var[F_N]}{\delta}\right)^{1/2}\right\} \leq \delta$$

quindi fissata una soglia $\delta$ il valore assoluto dell'errore decresce con un tasso $O(N^{-1/2})$. Infine il teorema del limite centrale ($F_N$ converge alla distribuzione normale quando $N \to \infty$) permette di stabilire questa eguaglianza[7]:

$$Pr\{|F_N - J| \geq t\sigma[F_N]\} = \sqrt{2/\pi}\int_t^{\infty} e^{-x^2/2}dx.$$

[7]

| $t$ | $\sqrt{2/\pi}\int_t^{\infty} e^{-x^2/2}dx$ |
|---|---|
| 1 | $0,6826$ |
| 2 | $0,9546$ |
| 3 | $0,9974$ |
| 4 | $0,9999$ |

# Metodi per la riduzione della varianza

Come si è appena visto l'errore commesso è legato alla varianza, quindi più questa sarà piccola minori saranno i valori da estrarre per ottenere una buona approssimazione del valore dell'integrale. Vi sono diversi metodi per ridurre la varianza del campione usato per il calcolo dell'integrale. Si esaminerà con un certo dettaglio solo uno fra questi.

## Metodo della funzione approssimante

Nel seguito si suppone che il dominio della funzione sia $[0, 1]$. L'idea su cui si basa questo metodo è quella di spezzare la funzione di cui si vuole calcolare l'integrale in due parti, $p(x)$, detta funzione approssimante, di cui si conosce l'integrale con facilità e $e(x)$ che sarà piccola in modulo sull'intervallo desiderato e che verrà calcolata col metodo Montecarlo:

$$f(x) = p(x) + e(x)$$

supponendo che la varianza che deriva dal calcolo di $e(x)$ sia più piccola di quella che deriva dal calcolo della varianza di $f(x)$ con il metodo Montecarlo si è trovato un modo per migliorare le prestazioni del procedimento.
Esempio

$$J_0 = \int_0^1 e^x dx = e - 1 = 1,71828$$

la varianza della variabile casuale che si ottiene è data da:

$$\sigma_0^2 = \int_0^1 e^{2x} dx - \left(\int_0^1 e^x dx\right)^2 = \left[\frac{e^x}{2}\right] - (e-1)^2 = 2e - \frac{3}{2} - \frac{e^2}{2} = 0,24204$$

quindi al $99,99\%$ di probabilitù la stima del valore dell'integrale si discosterà in modulo dal valore reale di $\frac{4\sigma_0}{\sqrt{N}} = \frac{1,96788}{\sqrt{N}}$.

Supponiamo di approssimare $f(x)$ con $p(x) = 1 + x$ e quindi, in base allo sviluppo in serie di Taylor nell'origine della funzione esponenziale, si avrà $e(x) = ex - 1 - x$. Il valore della funzione approssimante è $1,5$; si può applicare il metodo Montecarlo per calcolare il valore di:

$$J_1 = \int_0^1 (e^x - x - 1)\, dx = \int_0^1 (e^x - x)\, dx - 1$$

si può utilizzare come stimatore:

$$Y_1 = \frac{\sum e^{x_i} - x_i}{N} - 1$$

la varianza di questa variabile casuale è data da:

$$\sigma_1^2 = \int_0^1 (e^x - x)^2 \, dx - \left( \int_0^1 (e^x - x)^2 \, dx \right)^2 = 0,0437$$

quindi al $99,99\%$ di probabilitù la stima del valore dell'integrale si discosterà in modulo dal valore reale di $\frac{4\sigma_1}{\sqrt{N}} = \frac{0,836}{\sqrt{N}}$.

Dal punto di vista statistico il metodo si giustifica considerando che si vuole stimare una grandezza $M$ mediante uno stimatore non distorto $Y_1$ per migliorare la stima si usa un secondo stimatore $Y_2$ costruito a partire da una variabile casuale $Y^{'}$ che abbia valor medio $M^{'}$:

$$Y_2 = Y_1 - Y^{'} + M^{'}$$

poiché $E[Y_2] = E[Y_1] = M$ allora anche $Y_2$ è uno stimatore non distorto di $M$. La varianza è data da:

$$Var[Y_2] = Var[Y_1 - Y^{'} + M^{'}] = Var[Y_1] + Var[Y^{'}] - 2cov[Y_1, Y^{'}]$$

quindi, in caso di correlazione positiva e se vale $2 \cdot cov[Y_1, Y^{'}] > Var[Y^{'}]$, si ha:

$$Var[Y_2] < Var[Y_1].$$

Nel caso in esame:

$$\begin{array}{cc} Y_1 = \frac{\sum f(x_i)}{N} & Y^{'} = \frac{\sum p(x_i)}{N} \\ E(Y_1) = J & E(Y^{'}) = \int_0^1 p(x)dx = M^{'} \end{array}$$

campionando $Y_1$ e $Y^{'}$ nello stesso insieme di punti si ottiene il nuovo stimatore:

$$Y_2 = \frac{\sum (f(x_i) - p(x_i))}{N} + M^{'}$$

considerando stime campionarie per la varianza e la covarianza si pu approssimare:

$$Var[Y_2] = \frac{\sum (f(x_i) - J)^2}{N} + \frac{\sum \left(f(x_i) - M^{'}\right)^2}{N} - 2\frac{\sum (f(x_i) - J)\left(p(x_i) - M^{'}\right)}{N}.$$

essendo $p(x)$ molto prossimo a $f(x)$ anche $p(x_i)$ sarà poco diversa da $f(x_i)$ e $M^{'}$ poco diversa da $J$. Da questo segue che:

$$\frac{\sum (f(x_i) - J)\left(p(x_i) - M^{'}\right)}{N} \approx \frac{\sum \left(p(x_i) - M^{'}\right)^2}{N}$$

ovvero $cov[Y_1, Y^{'}] > 0$ ed inoltre $2 \cdot cov[Y_1, Y^{'}] > Var[Y^{'}]$.

### Cenno ad altri metodi

Altri metodi di riduzione della varianza sono diffusi. Alcuni partono dalle idee appena esposte per definire uno stimatore $Y'$ correlato negativamente con $Y_1$ (*metodo delle variate antitetiche*), altri cercano di simmetrizzare la funzione $f(x)$ intorno ad una ascissa (*metodi di simmetrizzazione*). Quelli utilizzati dalle librerie GSL si basano sul

- campionamento stratificato ricorsivo: ovvero si cerca di concentrare i punti dell'integrazione nella regione con la varianza più elevata.
- metodo VEGAS: nel quale si concentrano i punti nelle regioni che danno il maggior contributo al calcolo dell'integrale.

## Esercizi

1. Modificare il programma che realizza il metodo del trapezio in modo che l'algoritmo si fermi quando per due iterazioni successive il valore dell'integrale è uguale o si è superato un numero prefissato di iterazioni.
2. Scrivere il codice necessario a realizzare il metodo Montecarlo accetta–rifiuta.
3. Implementare il metodo del calcolo dell'integrale basato sul valor medio.
4. Stimare l'errore per il metodo precedente. Questo comporta il calcolo del valore di: $\int_a^b f^2(x)p(x)dx$ che può essere stimato, grazie al teorema del limite centrale, con $\frac{1}{N}\sum_{i=1}^N f^2(x_i)$ per cui la stima dell'errore, con una probabilità del $68,26\%$, sarà:

$$err = (b-a)\frac{\frac{1}{N}\sqrt{\sum_{i=1}^N f^2(x_i)} - \left(\frac{1}{N}\sqrt{\sum_{i=1}^N f(x_i)}\right)^2}{N}$$

## Passeggiate a caso nel discreto

I metodi Montecarlo possono trovare applicazione anche nella risoluzione di equazioni differenziali ovvero in quei problemi nei quali si deve determinare una quantità a partire dal suo tasso di variazione.
Non entriamo nel dettaglio del problema ma esaminiamo un problema che può ricondursi alla soluzione di una equazione differenziale molto facile da modellare.

## Il problema della rovina del giocatore

Un giocatore parte con un capitale di $n$ monete e decide di giocare a testa e croce finché non riesce a vincere $s$ monete o perde tutto il capitale. Il gioco ha le seguenti regole ad ogni lancio il giocatore vince una moneta se esce testa (T), perde una moneta se esce croce (C). Si vuole determinare la probabilità di vincita del giocatore in funzione di $n$ ed $s$.
Il problema può essere modellato considerando un asse orientato dove sono definiti solo dei punti interi ed un punto mobile $M$ che inizialmente si trova in posizione $n$:

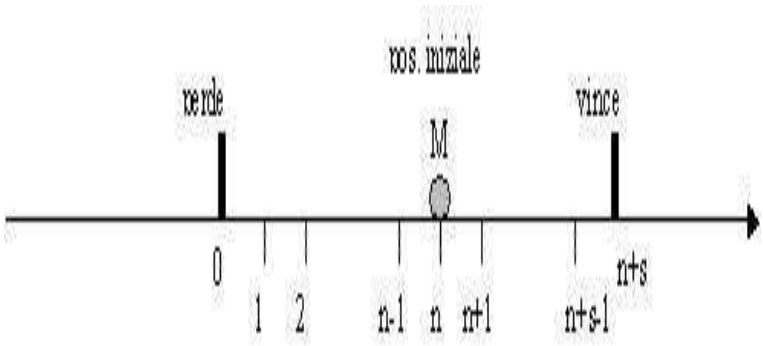


Figure 5: Illustrazione del problema della rovina del giocatore

quindi il modello prevede una passeggiata che inizia in $n$ e prosegue muovendosi ogni volta in avanti o indietro di un passo finché non raggiunge una delle due barriere (0 o $n+s$). Quella che si vuole determinare è la probabilità che $M$ muovendosi arrivi in $n+s$ senza mai passare per 0.
Scegliamo un punto a caso $c$ compreso fra 0 e $n+s$ ed indichiamo con $p(c)$ la probabilità che $M$ arrivi in $n+s$ senza essere passato per 0; dal teorema della probabilità totale si ha che questa probabilità è data dalla somma della:

- probabilità che $M$ passi da $c$ a $c+1$ al prossimo passo e da qui giunga ad $n+s$ senza passare per zero
- probabilità che $M$ passi da $c$ a $c-1$ al prossimo passo e da qui giunga ad $n+s$ senza passare per zero

quindi:

$$p(c) = 0,5p(c+1) + 0,5p(c-1)$$

con le condizioni $p(0) = 0$ e $p(n+s) = 1$[8].

[8]Questo problema si enuncia come un problema alle differenze finite che nel caso particolare è la discretizzazione dell'equazione differenziale $y''(x) = 0$ con $y(0) = 0$ e $y(n+s) = 1$

Come risoluzione secondo il metodo Montecarlo si può procedere effettuando $N$ passeggiate secondo il seguente schema:

- ogni passeggiata avrà inizio nel punto $n$ e proseguirà con successivi spostamenti del punto $M$ in verso positivo o negativo a seconda dell'esito del lancio di una moneta. Il procedimento del lancio di una moneta potrà essere modellato con l'estrazione di un numero casuale compreso fra 0 ed 1 considerando come eventi l'uscita di un numero fra 0 e $0,5$ oppure quella di un numero fra $0,5$ ed 1;

- la passeggiata termina quando $M$ raggiunge 0 o $n+s$. Nel primo caso la passeggiata avrà valore 0, nel secondo valore 1.

La passeggiata è quindi una variabile casuale $X$ che assume due valori 0 od 1. In particolare assume valore 1 con probabilità $p(n) = \frac{n}{n+s}$ e valore 0 con probabilità $1 - p(n)$.
Il valor medio di $X$ sarà dato da $E(X) = 1 \cdot p(n) + 0 \cdot (1 - p(n)) = n/(n+s)$.
Detto $N^{'}$ il numero di passeggiate che terminano in $N$ il rapporto $N^{'}/N$, per la legge dei grandi numeri fornirà una approssimazione abbastanza buona di $n/(n+s)$ per $N$ sufficientemente grande.
Calcoliamo la probabilità che lo scarto fra $N^{'}/N$ ed $n/(n+s)$ superi una quantità data. La varianza di $X$ è data da:

$$\sigma_0^2 = Var(X) = \left[\frac{n}{n+s}\left(1 - \frac{n}{n+s}\right)^2 + \frac{n+s-n}{n+s} \cdot \left(0 - \frac{n}{n+s}\right)^2\right] = \frac{n \cdot s}{(n+s)^2}$$

Considerando la somma di $N$ variabili casuali $X_i$ tutte uguali ad $X$ e la variabile casuale $Y$ data da:

$$Y = \frac{\sum X_i}{N} = \frac{N^{'}}{N}$$

per il teorema del limite centrale avrà approssimativamente una distribuzione normale con media $n/(n+s)$ e varianza $(n \cdot s)/(N(n+s)^2)$ ovvero $\sigma = \sigma_0/\sqrt{N}$ ma essendo $\sigma_0 < 1$ si ha $\sigma < 1/\sqrt{N}$.
Esercizi

- Implementare in C il problema della rovina del giocatore permettendo all'utente di scegliere i valori di n ed s.

- Modificare il problema precedente in modo che l'utente possa scegliere il valore di N indicando l'affidabilità del risultato.