Stima di pi greco con l'uso di metodi stocastici.

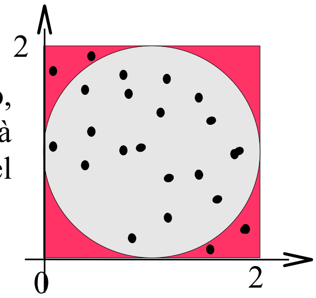
4C, maggio 2012

Parte prima

Con la funzione `rand()` otteniamo un numero intero casuale compreso fra 0 e la costante predefinita `RAND_MAX`. Per ottenere un numero x reale compreso fra 0 e 2 dovremo quindi scrivere:

```
x=2.*(float)rand()/(float)RAND_MAX;        (1)
```

Le due istruzioni `(float)` nella precedente servono per convertire quanto sta a destra ad una variabile di tipo float e vengono dette "cast", ovvero conversione di tipo. Senza tali istruzioni la divisione dell'istruzione precedente risulterebbe essere una divisione fra interi (quali sono appunto l'output della funzione "rand" e la costante predefinita `RAND_MAX`), e quindi il risultato verrebbe troncato all'intero, anche se la variabile x è stata definita come float.



Proviamo quindi a calcolare l'area di un cerchio sparando dei punti a caso, compresi nel quadrato circoscritto al cerchio. È chiaro che la probabilità di trovare un punto all'interno del cerchio è pari al rapporto fra l'area del cerchio e quella del quadrato.

Campioneremo quindi un numero di punti N_{TOT} sufficientemente grande, di coordinate (x,y), con x e y calcolati come dalla precedente formula.

Per ogni punto verificheremo se è interno alla circonferenza, ovvero se la distanza del punto (x,y) dal centro del cerchio è minore uguale al raggio.

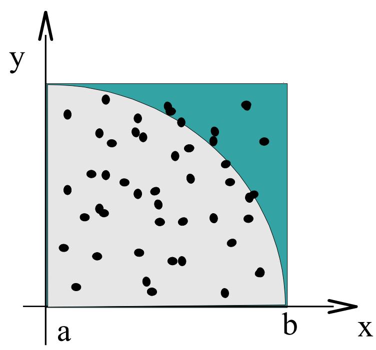
Il rapporto del numero dei punti interni N_{INT} alla circonferenza sul numero dei punti totali N_{TOT} soddisferà la seguente:

$$\lim_{N_{TOT}\to\infty} \frac{N_{INT}}{N_{TOT}} = \frac{\pi r^2}{4r^2} = \frac{\pi}{4} \qquad (2)$$

Limitandoci ad un N_{TOT} finito otterremo una stima approssimata di π

$$\pi \simeq 4\frac{N_{INT}}{N_{TOT}} \qquad (3)$$

Parte seconda



Con una piccola modifica vediamo come migliorare l'algoritmo. Considera la figura seguente corrispondente ad un quarto di cerchio.

L'arco di circonferenza rappresentato può essere visto come una funzione di equazione

$$y = f(x) = \sqrt{r^2 - x^2} \qquad \text{per } 0 < x < r \qquad (4)$$

Il rapporto fra le due aree sarà sempre pari a $\pi/4$. Come sopra potremmo campionare un insieme di punti (x,y) distribuiti uniformemente nel quadrato e proseguire allo stesso modo.

Possiamo notare però che in questo caso non è più necessario campionare la y, e si può migliorare l'efficienza dell'algoritmo. Per una data x la probabilità che il punto sia interno alla circonferenza (ovvero sotto la curva di equazione (4)) sarà proporzionale al valore della funzione in quel punto. Ovvero, indicando con $\{x_i\}$ l'insieme delle x campionate, con $1 < i < N_{TOT}$, avremo che

$$\frac{\pi}{4} = \lim_{N_{TOT}\to\infty} \sum_{i=1}^{N_{TOT}} \frac{f(x_i)(b-a)}{N_{TOT}} \qquad (5)$$

Parte terza

Siamo partiti con l'intenzione di stimare semplicemente il valore di π. Se osserviamo però la formula (5) ci accorgiamo come questa si possa usare per il calcolo dell'area sottesa ad una generica funzione nell'intervallo (a,b).

Volendo, per esempio, calcolare l'area sottesa al grafico della funzione

$$g(x) = \exp(x) - 1 \qquad (6)$$

nell'intervallo (a=2,b=5) posiamo stimarla semplicemente campionando un numero N_{TOT} di variabili stocastiche x_i uniformemente distribuite nell'intervallo (a,b), ovvero:

```
xi=(float)rand()/(float)RAND_MAX * (b-a) + a;   (7)
```

e, allo stesso modo dell'equazione (5), stimare l'area con la seguente:

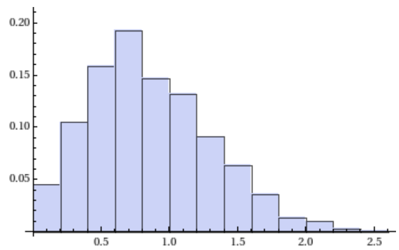
$$A \simeq \sum_{i=1}^{N_{TOT}} \frac{g(x_i)(b-a)}{N_{TOT}} \qquad (8)$$

che in questo caso risulta uguale 138.024.

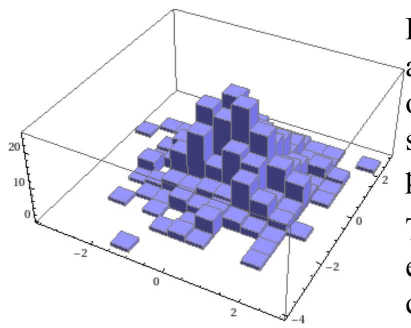
Curiosità

Come probabilmente avrai notato, il metodo dei "rettangoli" risulta essere più accurato a parità di costo computazionale (tempo impiegato) per calcolare l'area. L'approccio appena visto sembrerebbe per tanto una inutile complicazione. Così però non è se si considera un problema in più dimensioni, ovvero se si considera una funzione di più variabili, f(x,y,z....).

Facciamo un esempio.



Per il calcolo con una certa accuratezza dell'area sottesa ad una funzione f(x) si necessita di un campionamento di almeno N_{TOT} punti, ovvero del calcolo dell'area di N_{TOT} rettangolini.



Per analogo calcolo con analoga accuratezza del volume(e non l'area, ora c'è una dimensione in più!) sotteso alla curva di una funzione f(x,y) si necessiterà di un campionamento di almeno $(N_{TOT} * N_{TOT}) = N_{TOT}^2$ punti, ovvero del calcolo del volume di N_{TOT}^2 parallelepipedi.

Ti accorgi così che il numero di parallelepipedi da considerare va esponenzialmente con il numero di variabili della funzione f. Ovvero calcolare l'iper-volume sotteso ad una funzione $f(x_1,x_2,...x_K)$ costerà il calcolo di N^K iper-volumetti. Al contrario col metodo stocastico appena visto questo non è altrettanto vero, e per problemi a più dimensioni risulta pertanto essere molto più efficiente.