Statistica - Foglio IV

Ana Maria Draghici 2101044

Introduzione

In questa relazione si affronta il problema del calcolo numerico di volumi in più dimensioni mediante il metodo dell'integrazione Monte Carlo.

Il metodo Monte Carlo è particolarmente efficace per problemi in dimensione elevata o con domini di integrazione complessi, dove i metodi deterministici risultano inefficienti o difficili da applicare. In questo contesto, si utilizza l'integrazione Monte Carlo per stimare:

- la costante π, come area del quarto di cerchio unitario;
- il volume della sfera tridimensionale di raggio uno;
- il volume di un corpo tridimensionale definito da una disuguaglianza su $x^2 + y^2$ dipendente dalla coordinata z.

In ciascuno dei tre casi, si determina il numero di campioni Monte Carlo N tale da garantire, con probabilità maggiore del 95%, un errore assoluto inferiore a $\varepsilon = 0.01$ rispetto al valore esatto. Questo è reso possibile sfruttando le proprietà statistiche della media campionaria e l'uso della disuguaglianza di Chebyshev o, più precisamente, dell'approssimazione normale tramite il Teorema del Limite Centrale.

Punto (a) – Calcolo della costante π con il metodo Monte Carlo

Per stimare il valore della costante π tramite integrazione Monte Carlo, consideriamo il fatto che l'area di un quarto di cerchio di raggio unitario (nel primo quadrante) è pari a $\pi/4$.

L'idea consiste nel campionare punti uniformemente all'interno del quadrato unitario $[0, 1] \times [0, 1]$, e contare la frazione di punti che cadono all'interno del quarto di cerchio, cioè che soddisfano la disuguaglianza:

$$x^2 + y^2 \leq 1.$$

Indichiamo con $f(x, y)$ la funzione indicatrice del dominio:

$$f(x,y) = \begin{cases} 1 & \text{se } x^2 + y^2 \leq 1, \\ 0 & \text{altrimenti.} \end{cases}$$

La stima Monte Carlo dell'integrale (cioè dell'area del quarto di cerchio) è data da:

$$\hat{I}_N = \frac{1}{N} \sum_{i=1}^{N} f(x_i, y_i),$$

dove (x_i, y_i) sono N punti indipendenti e uniformemente distribuiti in $[0, 1]^2$.

Il valore di π può quindi essere stimato come:

$$\hat{\pi}_N = 4 \cdot \hat{I}_N.$$

Scelta del numero di campioni N

Poiché la variabile casuale $f(x, y)$ è Bernoulliana, la sua varianza è massima quando $\mathbb{E}[f] = 0.5$, e vale al massimo:

$$\sigma^2 = \mathrm{Var}(f) \leq 0.25.$$

Per ottenere un errore assoluto inferiore a $\varepsilon = 0.01$ con confidenza almeno del 95%, imponiamo:

$$1.96 \cdot \frac{\sigma}{\sqrt{N}} \leq \varepsilon.$$

Sostituendo:

$$1.96 \cdot \frac{0.5}{\sqrt{N}} \leq 0.01 \quad \Rightarrow \quad N \geq \left(\frac{1.96 \cdot 0.5}{0.01} \right)^2 = 9604.$$

Implementazione

L'esperimento Monte Carlo consiste nei seguenti passi:

1. Generare $N = 9604$ coppie di numeri casuali $(x_i, y_i) \in [0, 1]^2$.
2. Calcolare $f(x_i, y_i)$ per ciascun punto.
3. Stimare π con $\hat{\pi}_N = 4 \cdot \left(\frac{1}{N} \sum_{i=1}^{N} f(x_i, y_i) \right)$.

Conclusione

Il metodo fornisce una stima semplice ma efficace di π, dimostrando la potenza dell'approccio probabilistico anche per costanti matematiche fondamentali.

Punto (b) – Volume della sfera unitaria con Monte Carlo

In questo esercizio si vuole stimare il volume della sfera tridimensionale unitaria, cioè del corpo:

$$B = \left\{ (x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 \leq 1 \right\}.$$

Il volume vero della sfera di raggio 1 in $\mathbb{R}^3$ è noto e vale:

$$V = \frac{4}{3}\pi \approx 4.18879.$$

Metodo Monte Carlo

Per stimare questo volume con il metodo Monte Carlo, campioniamo punti uniformemente nel cubo $[-1,1]^3$, che contiene interamente la sfera. Definiamo la funzione indicatrice:

$$f(x,y,z) = \begin{cases} 1 & \text{se } x^2+y^2+z^2 \leq 1, \\ 0 & \text{altrimenti.} \end{cases}$$

Generando N punti $(x_i, y_i, z_i) \in [-1,1]^3$, indipendenti e uniformemente distribuiti, possiamo stimare il volume della sfera come:

$$\hat{V}_N = \frac{8}{N} \sum_{i=1}^{N} f(x_i, y_i, z_i),$$

dove il fattore 8 è il volume del cubo $[-1,1]^3$.

Scelta di N

Poiché $f(x,y,z)$ è una variabile casuale di tipo Bernoulli, con varianza massima pari a 0.25, imponiamo che l'errore assoluto sia minore di $\varepsilon = 0.01$ con probabilità maggiore del 95%. Usiamo il Teorema del Limite Centrale per determinare il numero minimo di campioni N, imponendo:

$$1.96 \cdot \frac{0.5}{\sqrt{N}} \cdot 8 \leq 0.01 \quad \Rightarrow \quad N \geq \left(\frac{1.96 \cdot 0.5 \cdot 8}{0.01} \right)^2 = 61465.6.$$

Pertanto, scegliamo:

$$N = 61466.$$

Implementazione

L'algoritmo si articola nei seguenti passi:

1. Generare $N = 61466$ punti casuali $(x_i, y_i, z_i) \in [-1,1]^3$.
2. Valutare $f(x_i, y_i, z_i)$ per ciascun punto.
3. Calcolare la stima del volume:

$$\hat{V}_N = 8 \cdot \left(\frac{1}{N} \sum_{i=1}^{N} f(x_i, y_i, z_i) \right).$$

Conclusione

Questo esercizio conferma la versatilità del metodo Monte Carlo per l'integrazione in più dimensioni, mantenendo una buona accuratezza anche nel caso tridimensionale.

Punto (c) – Volume del corpo tridimensionale A con Monte Carlo

In questo esercizio si vuole calcolare il volume del seguente corpo tridimensionale:

$$A = \left\{(x, y, z) \in \mathbb{R}^3 : z \in [-2, 2], \quad x^2 + y^2 \leq 1 + \sin(\pi z)\right\}.$$

Osserviamo che, per ogni valore di $z \in [-2, 2]$, la sezione (x, y) del corpo è un disco centrato nell'origine, con raggio $r(z) = \sqrt{1 + \sin(\pi z)}$. Il raggio dipende da z, ed essendo $\sin(\pi z) \in [-1, 1]$, otteniamo:

$$1 + \sin(\pi z) \in [0, 2].$$

Quindi il raggio varia tra 0 e $\sqrt{2}$, e le sezioni sono sempre dischi (o degenerati in un punto).

Metodo Monte Carlo

Per stimare il volume del corpo A, usiamo l'integrazione Monte Carlo nel cubo:

$$(x, y, z) \in [-\sqrt{2}, \sqrt{2}] \times [-\sqrt{2}, \sqrt{2}] \times [-2, 2],$$

che è un dominio parallelepipedo di volume:

$$V_{\text{box}} = (2\sqrt{2}) \cdot (2\sqrt{2}) \cdot 4 = 16 \cdot 4 = 64.$$

Definiamo la funzione indicatrice:

$$f(x, y, z) = \begin{cases} 1 & \text{se } x^2 + y^2 \leq 1 + \sin(\pi z), \\ 0 & \text{altrimenti.} \end{cases}$$

Campionando N punti uniformi nel dominio di volume 64, possiamo stimare il volume del corpo A come:

$$\hat{V}_N = \frac{64}{N} \sum_{i=1}^{N} f(x_i, y_i, z_i).$$

Scelta di N

Come nei punti precedenti, $f(x,y,z)$ è una variabile Bernoulliana, con valore massimo della varianza pari a 0.25. Richiediamo che l'errore assoluto sia inferiore a $\varepsilon = 0.01$ con probabilità superiore al 95%. Imponiamo:

$$1.96 \cdot \frac{0.5}{\sqrt{N}} \cdot 64 \leq 0.01 \quad \Rightarrow \quad N \geq \left(\frac{1.96 \cdot 0.5 \cdot 64}{0.01}\right)^2 = 98596.$$

Arrotondiamo a:

$$N = 98596.$$

Implementazione

L'algoritmo si compone dei seguenti passi:

1. Generare N punti casuali $(x_i, y_i, z_i) \in [-\sqrt{2}, \sqrt{2}]^2 \times [-2, 2]$.
2. Valutare $f(x_i, y_i, z_i)$ per ciascun punto, verificando se il punto ricade nel corpo.
3. Calcolare la stima:

$$\hat{V}_N = 64 \cdot \left(\frac{1}{N}\sum_{i=1}^{N} f(x_i, y_i, z_i)\right).$$

Conclusione

Questo esempio mostra l'efficacia del metodo Monte Carlo per calcolare volumi complessi e dipendenti da funzioni non lineari.

Pseudocodice per l'integrazione Monte Carlo

In questa sezione si presenta il pseudocodice completo per l'approssimazione numerica tramite metodo Monte Carlo dei tre problemi proposti: la stima di π, il volume della sfera unitaria e il volume del corpo tridimensionale A.

-Calcolo di N

Scelta del numero di campioni N. Per ciascuno dei tre casi considerati, il numero di campioni N è stato determinato teoricamente in modo da garantire un errore assoluto inferiore a $\varepsilon = 0.01$ con probabilità superiore al 95%. La procedura di calcolo è stata descritta in precedenza. I valori ottenuti e utilizzati nell'implementazione sono i seguenti:

(a) Stima di π: $N = 9604$

(b) Volume della sfera unitaria: $N = 61466$

(c) Volume del corpo A: $N = 98596$

Punto (a): Stima della costante π

- Definire il dominio: quadrato $[0,1] \times [0,1]$, volume $= 1$
- $N = 9604$
- Inizializzare `contatore` $= 0$
- Per $i = 1$ a N, fare:
  - Genera punto (x, y) uniforme in $[0,1]^2$
  - Se $x^2 + y^2 \leq 1$, incrementa il contatore
- Calcola stima_area $= \frac{\text{contatore}}{N}$
- Calcola $\pi \approx 4 \times$ stima_area

Punto (b): Volume della sfera unitaria

- Definire il dominio: cubo $[-1,1]^3$, volume $= 8$
- $N = 61466$
- Inizializzare `contatore` $= 0$
- Per $i = 1$ a N, fare:
  - Genera punto (x, y, z) uniforme in $[-1,1]^3$
  - Se $x^2 + y^2 + z^2 \leq 1$, incrementa il contatore
- Calcola stima_volume $= \frac{\text{contatore}}{N} \times 8$

Punto (c): Volume del corpo tridimensionale A

- Definire il dominio:
  - $x \in [-\sqrt{2}, \sqrt{2}]$
  - $y \in [-\sqrt{2}, \sqrt{2}]$
  - $z \in [-2, 2]$
  - Volume totale $= (2\sqrt{2})^2 \times 4 = 64$
- $N = 98596$
- Inizializzare `contatore` $= 0$
- Per $i = 1$ a N, fare:
  - Genera punto (x, y, z) uniforme nel dominio
  - Calcola soglia: soglia $= 1 + \sin(\pi z)$
  - Se $x^2 + y^2 \leq$ soglia, incrementa il contatore
- Calcola stima_volume $= \frac{\text{contatore}}{N} \times 64$

Motivazione delle tre serie di realizzazioni Monte Carlo

Nel metodo di integrazione Monte Carlo, la stima della quantità di interesse si basa su una media di variabili casuali indipendenti. Poiché la generazione dei punti casuali avviene tramite campionamenti indipendenti, ogni esecuzione produce una stima leggermente diversa a causa della variabilità intrinseca del metodo. Per questo motivo, in ciascun caso consideriamo tre serie di realizzazioni indipendenti, ossia tre sequenze di campioni generate in modo indipendente e non correlate tra loro. Questo permette di osservare come lo stimatore Monte Carlo si comporta in diverse realizzazioni del processo casuale, evidenziando la stabilità e la convergenza della stima al crescere del numero di campioni.

Osservazioni finali sui grafici. In ciascuno dei tre grafici è mostrata l'evoluzione dello stimatore Monte Carlo in funzione del numero di campioni, per tre diverse realizzazioni indipendenti. Nei primi due casi, in cui il valore atteso è noto (π per il primo e $\frac{4}{3}\pi$ per il secondo), è tracciata una linea rossa tratteggiata che rappresenta il valore teorico. Si osserva come, al crescere del numero di campioni, le stime tendano a stabilizzarsi attorno al valore atteso, con fluttuazioni sempre più contenute, a conferma della convergenza prevista dalla legge dei grandi numeri. Il terzo grafico, relativo al corpo tridimensionale A, mostra una dinamica simile, ma in assenza del valore esatto di riferimento, il confronto è puramente qualitativo. L'accordo tra le diverse realizzazioni e la loro stabilità suggerisce comunque una buona accuratezza della stima. In tutti i casi, il numero $N_{\max}$ di campioni è stato scelto in modo da garantire, con probabilità maggiore del 95%, un errore inferiore a $\varepsilon = 0.01$.