PAC - MAGGIO 2004

E 10. Algoritmo Box–Muller per variabili Gaussiane

1. Simulare una v.a. Gaussiana X di media nulla e varianza $\sigma^2 = 1$ usando l'algoritmo Box–Muller. Graficare media e varianza empirica su un numero di prove M e stabilire un numero M_0 a partire dal quale la misura si stabilizza.
2. Graficare il corrispondente istogramma delle frequenze empiriche.

E 11. Teorema del limite centrale.

Siano $Y_1, \ldots, Y_N$ copie indipendenti di una variabile di Bernoulli di parametro p. Si ricordi la formula di De Moivre–Laplace (teorema del limite centrale)

$$\mathbb{P}\left[X_N \in [a,b]\right] \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} \mathrm{d}z \,, \quad N \to \infty \,, \tag{0.1}$$

dove $-\infty < a < b < +\infty$ e

$$X_N = \frac{1}{\sqrt{p(1-p)N}} \sum_{i=1}^{N} (Y_i - p) \,. \tag{0.2}$$

Considerando diversi valori di p, si effettui una verifica del risultato (0.1) graficando un istogramma delle frequenze per la variabile X_N e confrontandolo con l'istogramma ottenuto nell'esercitazione precedente.

Suggerimenti

E 10. Poiché l'integrale Gaussiano

$$F_X(x) = \mathbb{P}[X \leqslant x] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-\frac{t^2}{2\sigma^2}} dt$$

non è noto analiticamente, il metodo della trasformazione non si può applicare direttamente in questo caso. L'algoritmo di Box–Muller si basa sull'osservazione che se R è v.a. continua in $[0, \infty)$ con densità

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}}$$

e Θ è v.a. uniforme in $[0, 2\pi)$, allora $X = R\cos\Theta$ è v.a. $N(0, \sigma^2)$, cioè Gaussiana di media nulla e varianza σ^2. In effetti, tramite un passaggio a coordinate polari nel piano cartesiano si dimostra facilmente che se $X = R\cos\Theta$ e $Y = R\sin\Theta$, allora X e Y sono due variabili $N(0, \sigma^2)$ *indipendenti*. Ora la variabile Θ si può simulare banalmente come $\Theta = 2\pi Z$ con Z uniforme in $[0, 1)$ mentre per la variabile R si può applicare il metodo della trasformazione: Infatti

$$F_R(x) = \frac{1}{\sigma^2} \int_0^x r\, e^{-\frac{r^2}{2\sigma^2}} dr = 1 - e^{-\frac{x^2}{2\sigma^2}} \,, \quad x \geqslant 0 \,.$$

e invertendo si ha $F_R^{-1}(y) = \sqrt{-2\sigma^2 \log(1-y)}$, $y \in (0, 1)$. Dunque, per simulare una Gaussiana $X = N(0, \sigma^2)$ si può utilizzare il seguente algoritmo: si generano 2 v.a. indipendenti Z_1, Z_2 uniformi in $[0, 1]$; si pone

$$\Theta = 2\pi Z_1 \,, \quad R = \sqrt{-2\sigma^2 \log Z_2} \,;$$

si calcola infine $X = R\cos\Theta$. Utilizzando ora M copie indipendenti della X cosi' ottenuta si puo' graficare l'istogramma delle frequenze, per esempio nell'intervallo $(-4\sigma, 4\sigma)$ che dovrebbe contenere piu' del 99.9% dei dati.

E 11. Una volta generate le $\{Y_i\}_{i=1}^N$ con un N fissato (e.g. $N = 10^3$) si calcola la variabile X_N come nella (0.2). Si possono generare poi M copie indipendenti $X_N^1, \ldots, X_N^M$ della variabile X_N e graficare il corrsipondente istogramma delle frequenze, nell'intervallo $(-4, 4)$. Per M grande l'istogramma si avvicina a quello ottenuto nell'esercitazione precedente.

caputo@mat.uniroma3.it.