PAC - 15 MAGGIO 2008

E 9. Metodo della trasformazione per v.a. continue.

1. *Esponenziale:* Simulare la v.a. continua X con densità

$$f(t) = \lambda\, e^{-\lambda t}\,, \qquad t \in [0, \infty)\,,$$

per diversi valori di λ (per es. $\lambda = 0.1, 1, 10$). Graficare media e varianza empirica al variare del numero n di prove e confrontare con i valori teorici

$$\mathbb{E}[X] = \frac{1}{\lambda}\,, \quad \mathrm{Var}[X] = \frac{1}{\lambda^2}\,.$$

2. *Cauchy:* Si ripeta l'esperimento proposto al punto 1 per la v.a. X con densità

$$f(t) = \frac{1}{\pi(1+t^2)}\,, \qquad t \in (-\infty, \infty)\,.$$

La v.a. X non ha valor medio, ossia $\mathbb{E}|X| = +\infty$. In particolare, si osserverà una minore stabilità delle misure.

E 10. Algoritmo Box–Muller per v.a. Gaussiane

1. Simulare una v.a. Gaussiana $X \sim N(\mu, \sigma^2)$ di media $\mu \in \mathbb{R}$ e varianza $\sigma^2 > 0$ usando l'algoritmo Box–Muller per diversi valori di σ (per es. $(\mu, \sigma^2) = (0, 1)$, e $(\mu, \sigma^2) = (10, 0.5)$). Graficare media e varianza empirica su prove ripetute.

2. Graficare il corrispondente istogramma delle frequenze nell'intervallo $[\mu - 4\sigma, \mu + 4\sigma]$ (che dovrebbe contenere piú del 99,99% dei dati).

3. Per una verifica dell'approssimazione normale (teorema del limite centrale) si confronti l'istogramma ottenuto al punto 2 sopra con l'istogramma ottenuto nell'esercitazione **E4**. Si osservera' che se Z è una variabile binomiale di parametri N e p, allora per valori grandi di N si ha che

$$\widetilde{Z}_N = \frac{Z - pN}{\sqrt{p(1-p)N}}$$

è ben approssimata da una normale $X \sim N(0, 1)$. Equivalentemente si puo' osservare che Z è approssimata da un variabile $X \sim N(\mu, \sigma^2)$ dove $\mu = pN$ e $\sigma^2 = p(1-p)N$.

caputo@mat.uniroma3.it.

Suggerimenti

E 9. Il metodo della trasformazione può essere brevemente illustrato come segue. Sia X una v.a. continua con densità $f(t)$, $t \in [x_1, x_2] \subset \mathbb{R}$. La funzione di distribuzione associata è $F(x) = \mathbb{P}[X \leqslant x] = \int_{x_1}^{x} f(t)dt$. Pertanto si ha

$$\mathbb{P}\big[X \in [a,b]\big] = \int_a^b f(t)dt = F(b) - F(a)\,, \quad x_1 \leqslant a \leqslant b \leqslant x_2\,. \tag{0.1}$$

Se Y è v.a. uniforme in $[0,1]$ allora la (0.1) equivale a $\mathbb{P}\big[Y \in [F(a), F(b)]\big]$. In particolare, se $F : [x_1, x_2] \to [0,1]$ è strettamente crescente allora abbiamo $\{X \in [a,b]\} \iff \{F(X) \in [F(a), F(b)]\}$ e la (0.1) stabilisce che la v.a. $Y = F(X)$ è distribuita uniformemente in $[0,1]$, qualunque sia la distribuzione di X. Invertendo F otteniamo che X è distribuita come $F^{-1}(Y)$. Quindi se si dispone di un'espressione esplicita per F^{-1} il metodo della trasformazione permette di simulare X calcolando $F^{-1}(Y)$ dove Y è uniforme in $[0,1]$. Nei casi proposti:

- se $f(t) = \lambda e^{-\lambda t}$, $t \geqslant 0$ allora $F(x) = \int_0^x f(t)dt = 1 - e^{-\lambda x}$, $x \geqslant 0$. Quindi

$$F^{-1}(y) = -\frac{1}{\lambda}\log(1-y)\,, \qquad y \in (0,1)\,.$$

- se $f(t) = \frac{1}{\pi(1+x^2)}$, $t \in \mathbb{R}$ si ha $F(x) = \int_{-\infty}^{x} f(t)dt = \frac{1}{2} + \frac{1}{\pi}\operatorname{arctg} x$. Invertendo:

$$F^{-1}(y) = \operatorname{tg}[\pi(y - \frac{1}{2})]\,, \qquad y \in (0,1)\,.$$

E 10. Poiché l'integrale Gaussiano

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

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 $Y = R\sin\Theta$, allora X e Y sono due variabili $N(0,\sigma^2)$ *indipendenti.* Ora la variabile Θ si può simulare banalmente come $\Theta = 2\pi U$ con U 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 U_1, U_2 uniformi in $[0,1]$; si pone

$$\Theta = 2\pi U_1\,, \qquad R = \sqrt{-2\sigma^2\log U_2}\,;$$

si calcola infine $X = R\cos\Theta$. Ripetendo M volte la simulazione di X si ottiene nel modo usuale un istogramma delle frequenze.

Infine, per una verifica del teorema del limite centrale si possono rappresentare su uno stesso grafico l'istogramma relativo a Z, e quello relativo a una normale $X = (pN, p(1-p)N)$, per esempio nel caso $(p, N) = (0.5, 1000)$.