Prof. Gasparini - Modelli Statistici
Politecnico di Torino - 16/10/24

Esercizi su alcune distribuzioni notevoli multidimensionali

1. Supponiamo per semplicità che i 40 geni di un genoma possano essere divisi in tre classi disgiunte: classe A di numerosità 20, B e C di numerosità 10 ciascuna. Le tre classi possono per esempio rappresentare funzioni biologiche distinte, o avere origini genetiche diverse.

   (a) si calcoli la probabilità di osservare nessun gene di classe A, due di classe B e tre di classe C in una estrazione casuale di 5 geni;

   (b) come approssimazione, si calcoli la stessa probabilità nel caso di campionamento casuale con reintroduzione;

   (c) si calcolino la probabilità di osservare qualunque campione così estremo o più estremo del precedente rispetto all'ipotesi di campionamento casuale (senza reintroduzione).

   Si noti che la seconda domanda equivale al calcolo del *p-value* per una generalizzazione multivariata di quello che si chiama *test esatto di Fisher* dell'ipotesi nulla di campionamento casuale; a ben guardare, i campioni più estremi sono semplicemente quelli che hanno probabilità più piccole sotto campionamento casuale. Come supporto di calcolo si può usare la libreria `extraDistr` di R, usata anche nella soluzione del primo punto. Si veda infine

   `https://www.youtube.com/watch?v=EF94wPaqXM0`

   per una illustrazione del test di Fisher univariato (una classe per volta) alla *Gene Set Analysis.*

2. Siano Z_1 e Z_2 due normali standard indipendenti.

   (a) Trovare la distribuzione di: $\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1/3 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$

   (b) Trovare la distribuzione di: $\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1/3 & 0 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$
   e calcolare $\mathrm{P}(Y < 3X + 1)$.

   (c) Trovare la distribuzione di: $\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1/3 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix}$
   e scrivere $\mathrm{P}(Y < 3X + 1)$ come integrale doppio.

   (d) Nell'ultimo caso, trovare la distribuzione di $Y|X = x$.

3. Siano $Z_1, Z_2, \ldots Z_3$ i.i.d. normali standard e

$$\begin{pmatrix} X \\ Y \\ S \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} Z_1 \\ Z_2 \\ Z_3 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

(a) Calcolare media e matrice di varianze e covarianze di $(X, Y, S)'$.

(b) Scrivere $\mathrm{P}(X < 0 \cap S < 0)$ come integrale doppio.

(c) Trovare la distribuzione condizionata di S dato $X = x \cap Y = y$.

Soluzioni

- 1.(a) in R: `dmvhyper(c(0,2,3),c(20,10,10),5)` $= 0.008206587$
- 1.(b) in R: `dmnom(c(0,2,3),5,c(1/2,1/4,1/4))` $= 0.009765625$
- 2.(a) densità $f(x, y) = 3/(2\pi) \exp\{-(x^2 + 9y^2)/2\}$
- 2.(b) degenere su $Y = X/3$, $\mathrm{P}(Y < 3X + 1) = 0.6461698$
- 2.(c) densità $f(x, y) = 3/(2\pi) \exp\{-(10x^2 - 18xy + 9y^2)/2\}$
- 2.(d) Normale(x,1/9)
- 3.(a) media $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ e matrice varcov $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 4 & 2 \\ 1 & 2 & 3 \end{pmatrix}$
- 3.(c) Normale($x + y/2 - 1/2$, 1)