UNIVERSITÀ DEGLI STUDI DI TRIESTE

Dipartimento di scienze economiche, aziendali, matematiche e statistiche "Bruno de Finetti"

Statistica (c.p.)

Richiami di Calcolo delle Probabilità

Francesco Pauli

DEAMS

Università di Trieste

A.A. 2016/2017

Richiami di algebra matriciale

- una matrice quadrata A di dimensione n si dice **simmetrica** se $a_{ij} = a_{ji}\ \forall i, j = 1, \dots, n$
- il **rango** di una matrice A è il numero di righe (colonne) linearmente indipendenti di A
- la **traccia** di una matrice A è $\text{tr}(A) = \sum_i a_{ii}$
- una matrice si dice **idempotente** se $A = A^2$
- una matrice si dice **ortogonale** se $A^T = A^{-1}$
- se A è idempotente allora $\text{tr}(A) = \text{rango}(A)$
- se A, di ordine k, è semidefinita postiva di rango m $(< k)$, allora m autovalori sono positivi e $k - m$. sono nulli.

Teorema di scomposizione spettrale

Teorema di scomposizione spettrale

Sia A matrice simmetrica di ordine k, allora esistono $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ (detti autovalori di A) e Q ortogonale (la cui j-ma colonna è detta autovettore di A) tali che

$$A = Q\Lambda Q^T$$

dove $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_k)$.

Il teorema di decomposizione spettrale permette di definire la radice quadrata di una matrice, nelle condizioni del teorema, infatti, si può scrivere $A = BB^T$ con $B = Q\Lambda^{1/2}$ dove in $\Lambda^{1/2}$ si intende che la radice è applicata elemento per elemento.

Forme quadratiche

Definizione: forma quadratica

Si dice **forma quadratica** una funzione $f : \mathcal{R}^n \to \mathcal{R}$ definita da $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \sum a_{ij}x_ix_j$ dove A è una matrice simmetrica

- una forma quadratica si dice **semidefinita positiva** se $f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} \geq 0 \ \forall\mathbf{x}$

Indice

Variabili aleatorie multiple

Una variabile aleatoria multipla è un vettore di v.a. $X = (X_1, \ldots, X_d)^T$ definite sullo stesso spazio di probabilità.

Si definisce il valore medio di X come il vettore dei valori medi delle componenti di X

$$E(X) = (E(X_1), \ldots, E(X_d))^T$$

e la matrice di varianze e covarianze

$$\text{var}(X) = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) & \ldots & \text{cov}(X_1, X_d) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) & \ldots & \text{cov}(X_2, X_d) \\ \vdots & \vdots & & \vdots \\ \text{cov}(X_d, X_1) & \text{cov}(X_d, X_2) & \ldots & \text{var}(X_d) \end{bmatrix}$$

assumendo che tutti gli elementi esistano. Si noti che essa è ovviamente simmetrica dato che $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$.

Alcune proprietà

- Se $E(X) = \mu$ e $\text{var}(X) = V$ e detta $Y = AX + b$ con A matrice $k \times d$ e $b \in \mathbb{R}^k$ allora
$$E(Y) = A\mu + b;$$
$$\text{var}(Y) = AVA^T.$$
- Qualunque sia X, $\text{var}(X)$ è semidefinita positiva.

Indice

Distribuzione normale multivariata

Distribuzione normale multivariata

Una v.a. X d-variata è distribuita secondo una normale multivariata se la sua funzione di densità è

$$f(x) = f(x_1, \ldots, x_d) = (2\pi)^{-d/2} |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right\}$$

dove $\mu \in \mathbb{R}^d$ e Σ è una matrice $d \times d$ simmetrica e semidefinita positiva. Sinteticamente si scrive $X \sim \mathcal{N}(\mu, \Sigma)$ (o $X \sim \mathcal{N}_d(\mu, \Sigma)$ o $X \sim MVN_d(\mu, \Sigma)$).

I parametri μ e Σ sono rispettivamente la media e la matrice di varianze e covarianze di X.

Normale bivariata

Se $d = 2$

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

ponendo $\rho = \sigma_{12}(\sigma_1\sigma_2)^{-1}$ (è il coefficiente di correlazione), si ottiene

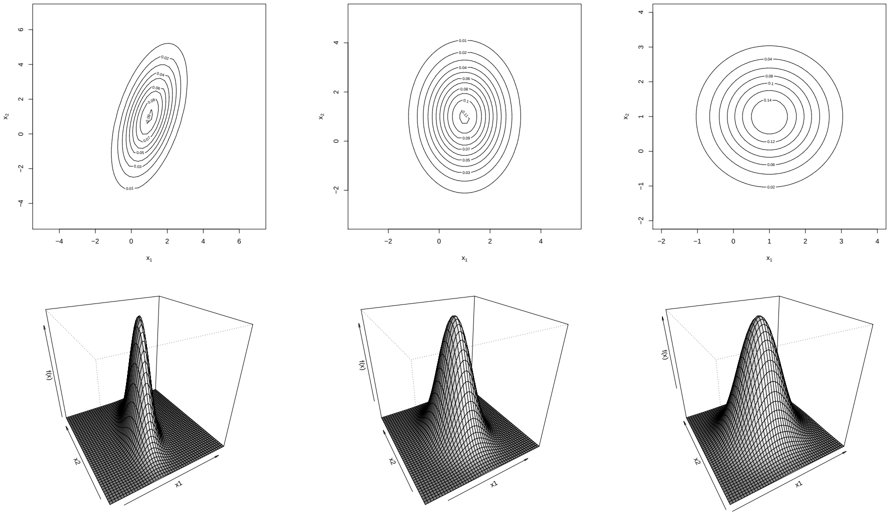
$$f(x) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right\}.$$

Si nota allora che le linee di livello della funzione hanno equazione

$$c = \left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2,$$

cioè sono delle ellissi centrate in μ, le lunghezze degli assi dipendono da σ_1 e σ_2 mentre l'inclinazione degli stessi dipende da ρ. In figura **??** sono riportati alcuni esempi.

Normale bivariata



$\sigma_1^2 = 1,\ \sigma_2^2 = 2,\ \rho = 0.5 \quad \sigma_1^2 = 1,\ \sigma_2^2 = 2,\ \rho = 0 \quad \sigma_1^2 = 1,\ \sigma_2^2 = 1, \rho = 0$

Normale bivariata

Notiamo poi che, se $\rho = 0$ – cioè se le variabili sono non correlate – la funzione di densità diventa

$$
\begin{aligned}
f(x) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left(\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right)\right\} \\
&= \frac{1}{2\pi\sigma_1} \exp\left\{-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right\} \frac{1}{2\pi\sigma_2} \exp\left\{-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right\} \\
&= f(x_1)f(x_2),
\end{aligned}
$$

si fattorizza cioè nel prodotto di due densità, questo dimostra che si ha indipendenza tra X_1 e X_2. La distribuzione normale è, in effetti, l'unica distribuzione per la quale la non correlaizone implica l'indipendenza.

Normale multivariata

Sia d qualunque, assumiamo $\mu = 0$ e $\Sigma = I$, si ha allora

$$f(x) = (2\pi)^{-d/2} \exp\left\{-\frac{1}{2}x^T x\right\} = (2\pi)^{-d/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{d} x_i^2\right\}$$

si può mostrare allora che se A è invertibile

$$Y = AX + b \sim \mathcal{N}(b, AA^T)$$

infatti possiamo scrivere $X = A^{-1}(Y - b)$ e quindi

$$f_y(y) = f_x(x(y)) \left|\frac{\partial X_i}{\partial Y_j}\right| = (2\pi)^{-d/2} \exp\left\{-\frac{1}{2}(Y - b)^T (A^{-1})^T A^{-1}(Y - b)\right\} |A|^{-1}$$

si ha poi, $(A^{-1})^T A^{-1} = (AA^T)^{-1}$ e inoltre è $|AA^T| = |A|^2$, scrivendo quindi $|A| = \sqrt{|A|^2} = \sqrt{|AA^T|}$

$$f_y(y) = (2\pi)^{-d/2}|AA^T|^{-1/2} \exp\left\{-\frac{1}{2}(Y - b)^T (AA^T)^{-1}(Y - b)\right\}$$

che è appunto l'espressione della f.d.d. per una $\mathcal{N}(b, AA^T)$.

Normale multivariata

Sia d qualunque, assumiamo $\mu = 0$ e $\Sigma = I$, si ha allora

$$f(x) = (2\pi)^{-d/2} \exp\left\{-\frac{1}{2}x^T x\right\} = (2\pi)^{-d/2} \exp\left\{-\frac{1}{2}\sum_{i=1}^{d} x_i^2\right\}$$

si può mostrare allora che se A è invertibile

$$Y = AX + b \sim \mathcal{N}(b, AA^T)$$

La proprietà appena mostrata suggerisce anche un modo alternativo di definire la distribuzione normale multivariata, si può infatti definire la $\mathcal{N}(0, I)$ dando l'espressione (13) e affermando che una v.a. Y è $\mathcal{N}(\mu, \Sigma)$ se è distribuita come $Y' = AZ + \mu$ dove $Z \sim \mathcal{N}(0, I)$ e $AA^T = \Sigma$.

Proprietà della normale multivariata

- Se $X \sim \mathcal{N}(\mu, \Sigma)$ e $Y = Ax + b$ allora $Y \sim \mathcal{N}(A\mu + b, A\Sigma A^T)$.
- Scomponiamo il vettore X, il vettore μ e la matrice Σ in componenti di dimensioni d_1 e d_2

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}, \Sigma^{-1} = \begin{bmatrix} (\Sigma^{-1})_{11} & (\Sigma^{-1})_{12} \\ (\Sigma^{-1})_{12} & (\Sigma^{-1})_{22} \end{bmatrix}$$

- le distribuzioni marginali sono ancora normali, cioè

$$X_1 \sim \mathcal{N}(\mu_1, \Sigma_{11})$$

- le distribuzioni condizionate sono ancora normali e

$$(X_1 | X_2 = x_2) \sim \mathcal{N}(\mu_1 + \Sigma_{12}(\Sigma^{-1})_{22}(x_2 - \mu_2), \Sigma_{11} - \Sigma_{12}(\Sigma^{-1})_{22}\Sigma_{21}).$$

- la funzione caratteristica, se $X \sim \mathcal{N}(\mu, \Sigma)$ è

$$E\left(\exp\left\{it^T X\right\}\right) = \exp\left\{it^T\mu - \frac{1}{2}t^T\Sigma t\right\}.$$

Indice

χ^2

La distribuzione χ^2 è in realtà un caso particolare della distribuzione gamma, è tuttavia utile illustrarne separatamente la genesi, che deriva dalla normale multivariata.

Posto $X \sim \mathcal{N}_d(0, I_d)$ si definisce χ^2 con d gradi di libertà, χ^2_d, la distribuzione di

$$Y = X^T X = \sum_{i=1}^{d} X_i^2$$

Il χ^2_1 è una Gamma$(1/2, 1/2)$, infatti, posto $y \geq 0$,

$$F_Y(y) = P(Y \leq y) = P(X_1^2 \leq y) = P(|X_1| \leq \sqrt{y}) = P(-\sqrt{y} \leq X_1 \leq \sqrt{y}) = 2(\Phi(\sqrt{y}) - 1/2)$$

$$f_Y(y) = \frac{d}{dy} 2\Phi(\sqrt{y}) = 2\phi(\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-t/2}$$

che è la densità di una Gamma$(1/2, 1/2)$ (si ricordi che $\Gamma(1/2) = \sqrt{2\pi}$).

χ^2

Per come è stato definito, poi, un χ^2_d è la somma di d variabili distribuite secondo un χ^2_1 ovvero una Gamma$(1/2, 1/2)$, esso è pertanto una Gamma$(d/2, 1/2)$.

Questo ci permette allora di calcolare media e varianza del χ^2_d: $E(Y) = d$, $\text{var}(Y) = 2d$. Ancora, possiamo affermare che, se $Y \sim \chi^2_d$ e $V \sim \chi^2_k$ e Y è indipendente da V allora $Y + V \sim \chi^2_{d+k}$.

χ^2 non centrale

Sia $(Z_1, \dots, Z_d) \sim \mathcal{N}_d(\mu, I_d)$, la v.a.

$$Y = Z^T Z = \sum_i Z_i^2$$

ha distribuzione χ^2 non centrale con d gradi di libertà e parametro di non centralità $\mu^T \mu$ e si indica $Y \sim \chi^2_k(\mu^T \mu)$.

- X e Y indipendenti e $X \sim \chi^2_{k_X}(\delta_X)$, $Y \sim \chi^2_{k_Y}(\delta_Y)$, allora $X + Y \sim \chi^2_{k_X+k_Y}(\delta_X + \delta_Y)$
- se $X \sim N_d(\mu, \Sigma)$ con Σ definita positiva allora

$$H = X^T V^{-1} X \sim \chi^2_d(\mu^T V^{-1} \mu).$$

χ^2 non centrale

La dimostrazione passa per la scomposizione $V = BB^T$ (teorema di decomposizione spettrale e conseguenze), con ciò si definisce $Y = B^{-1}X$ che ha matrice di varianze e covarianze

$$B^{-1}BB^T(B^{-1})^T = B^{-1}B(B^TB^{-1})^T = I$$

e quindi è $Y \sim \mathcal{N}(B^{-1}\mu, I)$ e di conseguenza

$$Y^TY \sim \chi^2((B^{-1}\mu)^TB^{-1}\mu)$$

dove $(B^{-1}\mu)^TB^{-1}\mu = \mu^T(BB^T)^{-1}\mu = \mu^TV^{-1}\mu$
D'altra parte si ha

$$Y^TY = (B^{-1}X)^TB^{-1}X = X^T(BB^T)^{-1}X = X^TV^{-1}X$$

e quindi si ha la tesi.

t di Student

La distribuzione t di Student con n gradi di liberta si ottiene come rapporto tra una normale standard e un χ^2_n indipendenti tra loro. Siano allora

$$Z \sim \mathcal{N}(0,1), \quad W \sim \chi^2_n$$

con Z e W indipendenti, allora

$$T = \frac{Z}{\sqrt{\frac{W}{n}}}$$

è distribuito come una t con n gradi di libertà e si indica $T \sim t_n$.
L'esempio più classico di statistica con distribuzione t di Student è dato dalla media campionaria standardizzata utilizzando la radice della varianza campionaria nel caso di un campione da una popolazione normale.

F di Fisher

Infine, il rapporto tra due χ^2 indipendenti definisce la distribuzione F di Fisher (o di Snedecor).
Siano V e W indipendenti con $V \sim \chi^2_m$ e $W \sim \chi^2_n$ la legge di probabilità di

$$T = \frac{\frac{V}{m}}{\frac{W}{n}}$$

è detta F di Snedecor con m, n gradi di liberta, in breve $T \sim F_{m,n}$.
Si nota l'ovviaproprietà per cui se $T \sim F_{m,n}$ allora $1/T \sim F_{n,m}$.

Indice

Teorema di Fisher-Cochran

Teorema di Fisher-Cochran

Sia

- $Y \sim \mathcal{N}_k(\mu, I_k)$
- $A_1, \ldots, A_m$ matrici
  - (a) semidefinite positve,
  - (b) di rango rispettivamente $r_1, \ldots, r_m$,
  - (c) tali che $A_1 + \ldots + A_m = I_k$,

allora le seguenti affermazioni sono equivalenti

- (i) le forme quadratiche $Q_j = Y^T A_j Y$, $j = 1, \ldots, m$ sono $\chi^2_{r_j}(\mu^T A_j \mu)$ indipendenti;
- (ii) $r_1 + \ldots + r_m = k$.

Dimostrazione $(i) \Rightarrow (ii)$

Ricaveremo la distribuzione di $Y^T Y$ in due modi diversi, eguagliando i parametri delle due espressioni trovate si ottiene la tesi.

Da una parte si ha

$$\begin{aligned} Y^T Y &= Y^T I_k Y \\ &= Y^T (A_1 + \ldots + A_m) Y \\ &= Y^T A_1 Y + \ldots + Y^T A_m Y \\ &= Q_1 + \ldots Q_m \sim \chi^2_{r_1 + \ldots r_m} \left(\sum_j \mu^T A_j \mu \right) \end{aligned}$$

D'altra parte essendo $Y \sim \mathcal{N}_k(\mu, I_k)$ si ha anche

$$Y^T Y \sim \chi^2_k(\mu^T \mu)$$

pertanto $r_1 + \ldots r_m = r = k$.

Dimostrazione $(ii) \Rightarrow (i)$

Esprimiamo le forme quadratiche Q_j come somme di v.a. gaussiane indipendenti.
In base al teorema di decomposizione spettrale possiamo scrivere

$$A_j = Q_j \Lambda_j Q_j^T = (Q_j \Lambda_j^{1/2})(\Lambda_j^{1/2} Q_j^T) = B_j B_j^T$$

dove $B_j = \bar{Q}_j \bar{\Lambda}_j^{1/2}$ è la matrice di dimensione $k \times r_j$ che si ottiene cancellando gli autovalori nulli da Λ_j e togliendo le corrispondenti colonne di Q_j (essendo A_j di rango r_j, $k - r_j$ autovalori sono nulli).
Definiamo

$$B = (B_1, \ldots, B_m) \quad (k \times (r_1 + \ldots + r_m = k))$$

La matrice B è ortogonale poiché

$$BB^T = B_1 B_1^T + \ldots + B_m B_m^T = A_1 + \ldots + A_m = I_k.$$

Dimostrazione $(ii) \Rightarrow (i)$

Dunque la variabile

$$Z = B^T Y \sim \mathcal{N}_k(B^T\mu, I_k)$$

quindi

- le v.a. $Z_j = B_j^T Y$, $i = 1, \ldots, m$ sono indipendenti tra loro
- $Z_j \sim \mathcal{N}_{r_j}(B_j^T\mu, I_{r_j})$

e pertanto

$$Q_j = Y^T A_j Y = Y^T B_j B_j^T Y = Z_j^T Z_j$$

sono indipendenti e Q_j è un $\chi^2_{r_j}$ non centrale con parametro di non centralità $\mu^T B_j B_j^T \mu = \mu^T A_j \mu$ c.v.d.

Conseguenza del teorema di Fisher-Cochran

Teorema

Sia $Y \sim \mathcal{N}_n(\mu 1_n, \sigma^2 I_k)$, e si considerino le statistiche

$$T_n = \sum_{j=1}^{n}(Y_j - \bar{Y})^2 \quad V_n = \frac{1}{n}\left(\sum_{j=1}^{n} Y_j\right)^2 = n\bar{Y}^2$$

dove $\bar{Y} = \frac{1}{n}\sum_{j=1}^{n} Y_j$, allora T_n e V_n sono indipendenti e

$$T_n \sim \sigma^2\chi^2_{n-1} \quad V_n \sim \sigma^2\chi^2_1(n\mu^2/\sigma^2)$$

Dimostrazione

Si ha $T_n = Y^T A_1 Y$ con $A_1 = I_n - \frac{1}{n} 1_n 1_n^T$:

$$
\begin{aligned}
Y^T A_1 Y &= \begin{bmatrix} Y_1 & \dots & Y_n \end{bmatrix} \begin{bmatrix} 1-\frac{1}{n} & -\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ -\frac{1}{n} & 1-\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} \\ \dots & \dots & \dots & \dots & \dots \\ -\frac{1}{n} & -\frac{1}{n} & \dots & -\frac{1}{n} & 1-\frac{1}{n} \end{bmatrix} \begin{bmatrix} Y_1 \\ \dots \\ Y_n \end{bmatrix} \\
&= \begin{bmatrix} Y_1 & \dots & Y_n \end{bmatrix} \begin{bmatrix} Y_1 - \bar{Y} \\ \dots \\ Y_n - \bar{Y} \end{bmatrix} \\
&= Y_1^2 - Y_1 \bar{Y} + Y_2^2 - Y_2 \bar{Y} + \dots + Y_n^2 - Y_n \bar{Y} \\
&= \sum_i Y_i^2 - \bar{Y} \sum_i Y_i = \sum_i Y_i^2 - n\bar{Y}^2 = \sum_i (Y_i - \bar{Y})^2
\end{aligned}
$$

La matrice A_1 è idempotente,

$$
\begin{aligned}
A_1^T A_1 &= \left(I_n - \frac{1}{n} 1_n 1_n^T \right)^T \left(I_n - \frac{1}{n} 1_n 1_n^T \right) \\
&= I_n^T I_n + \frac{1}{n^2} (1_n 1_n^T)^T (1_n 1_n^T) - \frac{1}{n} I_n^T (1_n 1_n^T) - \frac{1}{n} (1_n^T 1_n) I_n = I_n^T I_n - \frac{1}{n} (1_n^T 1_n) = A_1
\end{aligned}
$$

quindi $\text{rango}(A) = \text{tr}(A) = \sum_{j=1}^n (1 - \frac{1}{n}) = n - 1.$

Dimostrazione

Per quanto riguarda V_n si ha $V_n = Y^T A_2 Y$ con $A_2 = \frac{1}{n} 1_n 1_n^T$, che ha rango 1.
Infine, che T_n e V_n siano semidefinite positive è evidente se si guarda alla loro forma non matriciale.
Con ciò, siamo nelle ipotesi del teorema, per cui T_n e V_n sono indipendenti e T_n/σ^2 è un χ^2_{n-1} con parametro di non centralità

$$\mu 1_n^T A_1 \mu 1_n = \mu^2 1_n^T 0_n = 0$$

mentre V_n/σ^2 è un χ^2 con 1 g.d.l. e parametro di non centralità

$$\mu 1_n^T A_2 \mu 1_n = \mu^2 1_n^T 1_n 1_n^T 1_n = n\mu^2/\sigma^2.$$