MS 16/10/24

ALGEBRA DEI MOMENTI PER VETTORI ALEATORI

Se $X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}$ è un vettore aleatorio

allora

$$E(X) = \left[\begin{pmatrix} E(X_i) \end{pmatrix}\right] = \mu_X$$

il valore atteso di un vettore è il vettore dei valori attesi

Se X e Y sono vettori della stessa dimensione

$$E(aX + bY) = aE(X) + bE(Y)$$

Si può inoltre definire, per $\underset{m_X \times 1}{X}$ e $\underset{m_Y \times 1}{Y}$ di dimensioni qualunque m_X e m_Y, la matrice di covarianze incrociate

$$\text{CrossCov}(X, Y) = \left[(\text{Cov}(X_i, Y_j) \right]$$

si vede facilmente che

$$\text{CrossCov}(X, Y) = \left[\left(E(X_i - \mu_{X_i})(Y_j - \mu_{Y_j}) \right) \right]$$

$$= E\left[((X_i - \mu_{X_i})(Y_j - \mu_{Y_j})) \right]$$

$$= E\left((X - \mu_X)(Y - \mu_Y)' \right)$$

trasposta

(una prima proprietà)

In particolare (seconda proprietà)

$$\text{CrossCov}(X, X) = E[(X-\mu_X)(X-\mu_X)']$$

$$= \text{VarCov}(X) =$$ matrice di varianza e covarianza di X stesso

Un'altra proprietà è che

$$\text{CrossCov}(AX, BY) = E[(AX - A\mu_X)(BY - B\mu_Y)']$$

$$= A\,\text{CrossCov}(X, Y)\,B'$$

e quindi per una sola X

$$\text{VarCov}(AX) = \text{CrossCov}(AX, AX)$$

$$= A\,\text{CrossCov}(X, X)\,A'$$

$$= A\,\text{VarCov}(X)\,A'$$

Come abbiamo già usato nella costruzione della normale multivariata.

Se X e Y hanno la stessa dimensione

$$\mathrm{VarCov}(X+Y) = \mathrm{VarCov}\left((1\ 1)\begin{pmatrix} X \\ Y \end{pmatrix}\right)$$

$$= (1\ 1)\ \mathrm{VarCov}\begin{pmatrix} X \\ Y \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

perché X e Y stessa dimensione

$$= (1\ 1)\begin{pmatrix} \mathrm{VarCov}(X), & \mathrm{CrossCov}(X,Y) \\ \mathrm{CrossCov}(X,Y) & \mathrm{VarCov}(Y) \end{pmatrix}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$= \mathrm{VarCov}(X) + 2\,\mathrm{CrossCov}(X,Y) + \mathrm{VarCov}(Y)$$

Nota: $\mathrm{CrossCov}(X,Y)=0$ se X e Y indipendenti

ESEMPIO

$$X \sim \mathrm{Multinomiale}(n, p_1 \dots p_d)$$

$$p_D = p_{d+1} = 1-\sum_{i=1}^{d} p_i$$

per calcolare media e matrice di varianza e covarianza di X, scriviamo

$$X = \sum_{j=1}^{n} Y_j$$

dove

$Y_1 \dots Y_n$ sono iid $\mathrm{Multinomiale}(1, p_1 \dots p_d)$ perché le prove sono indipendenti con le stesse probabilità delle diverse classi

chiamata $\mathrm{Multinoulli}(p_1 \dots p_d)$ oppure categorica $(p_1 \dots p_d)$

	$d=1$	$d>1$
$n=1$	Bernoulli	Multinoulli
$n>1$	Binomiale	Multinomiale

ESEMPIO

$$
\begin{array}{cccccc}
 & a & b & c & d & \\
(& 0 & 1 & 0 & 0 &) \quad Y_1 \\
(& 1 & 0 & 0 & 0 &) \quad Y_2 \\
 & & \vdots & & & \\
(& 0 & 1 & 0 & 0 &) \quad Y_n \\
\hline
(& 1 & 2 & 0 & 0 &) \quad \sum Y_j = X
\end{array}
$$

quindi:

$$\mathrm{VarCov}(X) = \mathrm{VarCov}\left(\sum Y_j\right)$$

$$= \sum \mathrm{VarCov}(Y_j)$$

(non ci sono i termini CrossCov per l'indipendenza

la singola matrice $\mathrm{VarCov}(Y_j)$ e' facile da calcolare:

$$E(Y_{ji}) = p_i$$

$$\mathrm{Var}(Y_{ji}) = p_i(1-p_i)$$

sempre uguale a 0

$$\mathrm{Cov}(Y_{ji}, Y_{jk}) = E(Y_{ji} Y_{jk}) - E(Y_{ji})E(Y_{jk})$$

$$= 0 - p_i p_k = -p_i p_k$$

quindi $\forall j$

$$\mathrm{VarCov}(Y_j) = \begin{pmatrix} p_1(1-p_1) & & -p_i p_k \\ & p_2(1-p_2) & \\ -p_i p_k & & \ddots \\ & & & p_d(1-p_d) \end{pmatrix}$$

quindi:

$$\mathrm{VarCov}(X) = \begin{pmatrix} np_1(1-p_1) & \cdots & -np_1 p_d \\ -np_1 p_2 & np_2(1-p_2) & \vdots \\ \vdots & \ddots & \\ -np_1 p_d & & np_d(1-p_d) \end{pmatrix}$$

CONDIZIONAMENTO PER NORMALE MULTIVARIATA

$$\underset{n\times 1}{X} \sim N_n(\underset{n\times 1}{\mu}, \underset{n\times n}{\Sigma})$$

supponiamo di partizionarla, così possiamo studiarne le distribuzioni condizionate.

$$\underset{n\times 1}{X} = \begin{pmatrix} \underset{n_1\times 1}{X_1} \\ \underset{n_2\times 1}{X_2} \end{pmatrix} \qquad \mu = \begin{pmatrix} \underset{n_1\times 1}{\mu_1} \\ \underset{n_2\times 1}{\mu_2} \end{pmatrix} \qquad \begin{matrix} \mu_1 = E(X_1) \\ \mu_2 = E(X_2) \end{matrix}$$

$$\text{VarCov}(X) = \underset{n\times n}{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

$$\Sigma_{11} = \text{VarCov}(X_1)$$
$$\Sigma_{22} = \text{VarCov}(X_2)$$
$$\Sigma_{12} = \text{CrossCov}(X_1, X_2) = \text{CrossCov}(X_2, X_1)' = \Sigma_{21}'$$

La funzione caratteristica si può scrivere

$$\varphi(t) = \exp\left\{ it'\mu - \frac{1}{2} t'\Sigma t \right\} \quad \text{(visto la volta scorsa)}$$

$$= \exp\left\{ i\,(t_1' \; t_2')\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} - \frac{1}{2}(t_1' \; t_2')\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}\begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \right\}$$

$$= \dots = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2) \exp\left\{ -\frac{1}{2}\left(t_1'\Sigma_{12}t_2 + t_2'\Sigma_{21}t_1 \right) \right\}$$

quindi $\varphi_X(t) = \varphi_{X_1}(t_1)\varphi_{X_2}(t_2)$ sse $\Sigma_{12} = \Sigma_{21} = 0$

Cioè x_1 e x_2 sono indipendenti sse le crosscov sono nulle

Quindi:

normalità + cov nulla $\iff$ indipendenza

Vediamo il caso particolare $n=2$, con le ovvie notazioni

$$\begin{pmatrix}X\\Y\end{pmatrix} \sim N_2\left(\begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & Cov(X,Y)\\ Cov(X,Y) & \sigma_Y^2\end{pmatrix}\right)$$

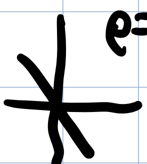
Ricordiamo la definizione di coefficiente di correlazione lineare

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

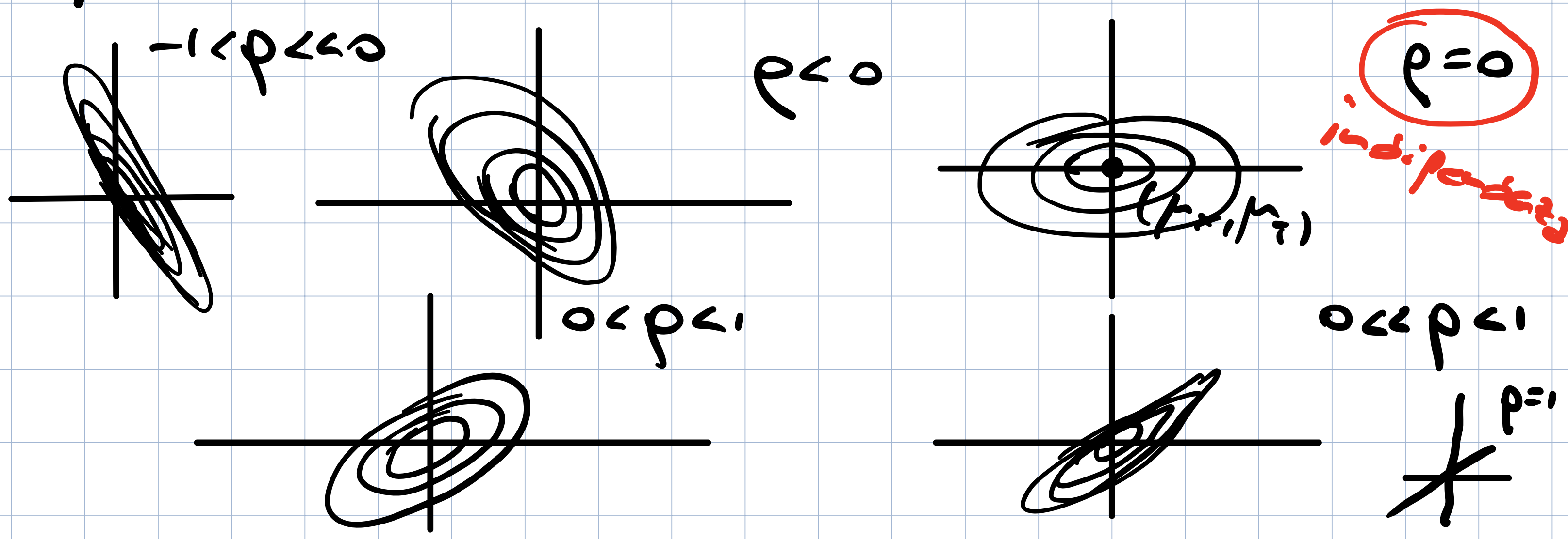
dev. std. di x

Per Cauchy Schwarz, $-1 \le \rho \le 1$.

Si può anche scrivere



$$\begin{pmatrix}X\\Y\end{pmatrix} \sim N_2\left(\begin{pmatrix}\mu_X\\\mu_Y\end{pmatrix}, \begin{pmatrix}\sigma_X^2 & \rho\sigma_X\sigma_Y\\ \rho\sigma_X\sigma_Y & \sigma_Y^2\end{pmatrix}\right)$$



(vedremo un esempio dei casi estremi: $\rho = \pm 1$ negli esercizi)

Nel caso generale ($n > 2$) si dimostra che, se Σ_{11} e' invertibile (nonsingolare)

$$X_2 \mid X_1 = x_1 \sim N_{n_2}\left(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1),\ \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)$$

(ovvio scambio per $X_1 \mid X_2 = x_2$)

Per la normale bivariata ($n=2$), questo diventa

$$Y \mid X = x \sim N\left(\mu_Y + \rho \cancel{\sigma_X}\sigma_Y \frac{1}{\sigma_X^{\cancel{2}}}(x - \mu_X),\ \sigma_Y^2 - \rho \cancel{\sigma_X}\sigma_Y \frac{1}{\cancel{\sigma_X^2}} \rho \cancel{\sigma_X}\sigma_Y\right)$$

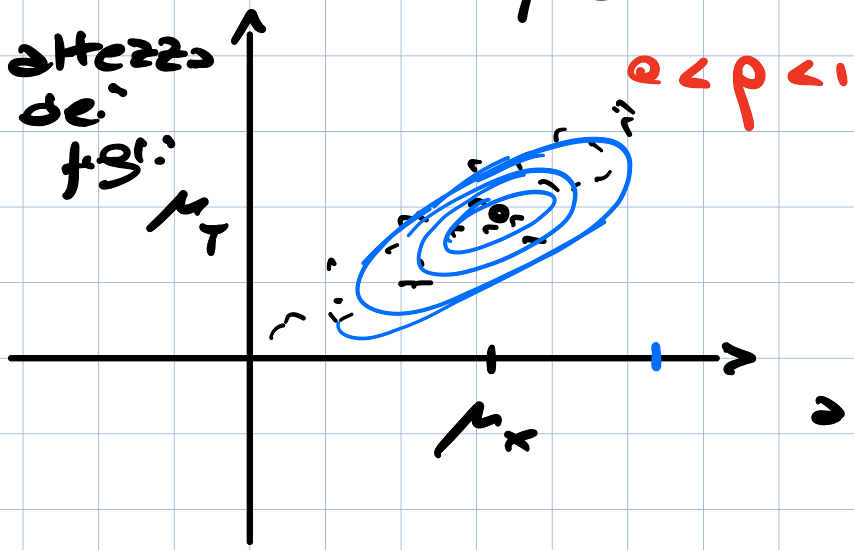
$$= N\left(\underbrace{\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)}_{E(Y|X=x)},\ (1-\rho^2)\sigma_Y^2\right)$$

NOTA STORICA:

$$E(Y|X=x) = \mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

$$\underbrace{\frac{E(Y|X=x) - \mu_Y}{\sigma_Y}}_{\text{quanto } E(Y|X=x) \text{ si discosta da } \mu_Y \text{ in termini di } \sigma_Y} = \underbrace{\rho}_{\text{ma c'e' un } \rho \text{ di mezzo!}} \underbrace{\frac{x - \mu_X}{\sigma_X}}_{\text{di quante deviaz std di } X\ x \text{ si discosta da } \mu_X}$$

$E(Y|X=x)$ è una previsione di quanto sia Y in media quando $X=x$



Galton metà XIX

Effetto di regressione: la previsione di Y per $X=x$ eccezionalmente grande (o piccolo) non è così eccezionale perché mediata da $\rho<1$.