Variabili (vettori e matrici) casuali

Variabile casuale (o aleatoria): Variabile che può assumere un insieme di valori ognuno con una certa probabilità

La variabile aleatoria rappresenta la *popolazione* in esame

$\star$ X var. discreta $\Rightarrow$ valori $X_1, X_2, \ldots, X_N$ con prob. $(p_1, p_2, \ldots, p_N)$

$\star$ X var. continua$\Rightarrow$ valori in (a, b) con *funzione di densità di prob.* $f{=}f(X)$

Vettore (o matrice) casuale (o aleatorio): vettore (o matrice) i cui elementi sono variabili aleatorie

---

Il Valore atteso di una *variabile* aleatoria discreta X che può assumere i valori $x_1, \ldots, x_n$ è

$$E[X] = \sum_{i=1}^{n} P_i x_i \qquad P_i = P(X = x_i)$$

(stessa cosa se *matrice* di variabili aleatorie)

Media, Varianza

Proprietà di $E[X]$: $E[X+Y] = E[X] + E[Y], \quad E[AXB] = AE[X]B$

Sia $X = [X_1, \ldots, X_p]$ vettore casuale

Media: $\mu_i = E[X_i], \quad i = 1, \ldots p, \qquad \boldsymbol{\mu}^T = [\mu_1, \ldots, \mu_p]$

Varianza: (per componenti)

$$\sigma_{i,i} \equiv \sigma_i^2 = E[(X_i - \mu_i)^2] = \sum_{j=1}^{p} P(X_i = x_{i,j})(x_{i,j} - \mu_i)^2, \quad i = 1, \ldots p$$

Covarianza tra le due variabili casuali X_i, X_j:

$$\begin{aligned} \sigma_{i,j} &= E[(X_i - \mu_i)(X_j - \mu_j)], \qquad i, j = 1, \ldots, p \\ &= \sum_{X_i, X_j} (x_i - \mu_i)(x_j - \mu_j) p_{i,j}(x_i, x_j) \end{aligned}$$

$p_{i,j}$ funzione di probabilità congiunta

In forma matriciale: matrice di varianza/covarianza della popolazione

$$\Sigma = Cov(X) = \begin{bmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,p} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \sigma_{2,p} \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_{p,1} & \sigma_{p,2} & \cdots & \sigma_{p,p} \end{bmatrix} = E[(X-\boldsymbol{\mu})(X-\boldsymbol{\mu})^T]$$

Σ è sim. definita positiva: $\quad \mathbf{a}^T \Sigma \mathbf{a} > 0$ per ogni $\mathbf{a} \neq 0$

(autovalori di Σ sono reali e strettamente positivi)

Matrice di Correlazione della popolazione:

$$\boldsymbol{\rho} = (\rho_{i,j}) \qquad \rho_{i,j} = \frac{\sigma_{i,j}}{\sqrt{\sigma_{i,i}}\sqrt{\sigma_{j,j}}} \qquad \text{simmetrica}$$

⋆ X_1, X_2 sono indipendenti $\Rightarrow \quad \rho_{1,2} = 0$

⋆ $\boldsymbol{\rho}$ misura la quantità di associazione *lineare*

Sia $\{\mathbf{X}_1^T, \mathbf{X}_2^T, \ldots, \mathbf{X}_n^T\}$ campione di n osservazioni di vettori casuali

$\star$ Campione casuale: Se $\mathbf{X}_1^T, \mathbf{X}_2^T, \ldots, \mathbf{X}_n^T$ sono osservazioni indipendenti di una stessa funzione densità composta $f(\mathbf{x}) = f(x_1, x_2, \ldots, x_p)$

$\Rightarrow$ le misure da prove diverse non devono influenzarsi reciprocamente

$\star$ Stimatore corretto: Una statistica $\mathbf{a}$ è uno stimatore corretto (o *unbiased*) della variabile aleatoria $\boldsymbol{\alpha}$ se $E(\mathbf{a}) = \boldsymbol{\alpha}$

Supp. $\mathbf{X}_1^T, \mathbf{X}_2^T, \ldots, \mathbf{X}_n^T$ osservazioni indipendenti di una popolazione con media $\boldsymbol{\mu}$ e matrice di covarianza Σ

Statistiche campionarie:

- $\bar{\mathbf{X}}$ è uno stimatore corretto di $\boldsymbol{\mu}$: $E(\bar{\mathbf{X}}) = \mu$
- La Matrice di covarianza di $\bar{\mathbf{X}}$ è: $Cov(\bar{\mathbf{X}}) = \frac{1}{n}\Sigma$
- S_n stimatore distorto (biased) di Σ: $E(S_n) = \frac{n-1}{n}\Sigma$

$\Rightarrow$ $\frac{n}{n-1}S_n$ stimatore corretto (unbiased) di Σ : $E(\frac{n}{n-1}S_n) = \Sigma$

$$S = \frac{n}{n-1}S_n$$

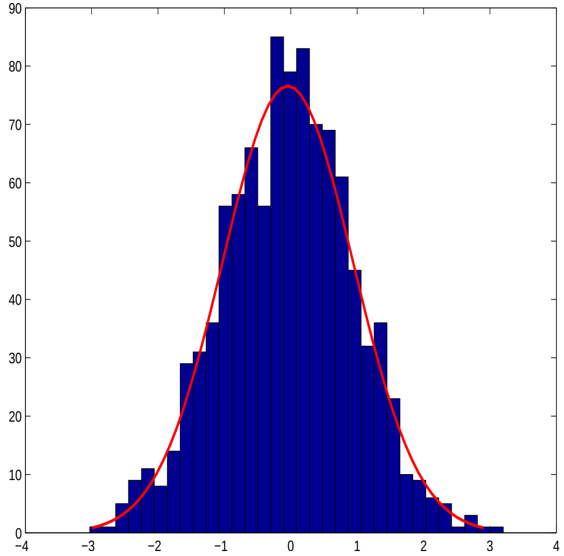
La distribuzione normale multivariata

Richiamo dal caso univariato: funzione di densità

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad -\infty < x < +\infty$$

$\frac{1}{\sqrt{2\pi\sigma^2}}$ costante di normalizzazione (per Probabilità=1)

μ media, σ^2 varianza della popolazione



$\mathcal{N}(\mu, \sigma^2)$: insieme delle variabili casuali che seguono tale distribuzione

Estensione al caso multidimensionale

Osservazione:

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

Nel caso di $\mathbf{x}$ vettore casuale:

$$(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})^T \qquad \boldsymbol{\mu} \in \mathbb{R}^p \text{ valore atteso di } \mathbf{x}$$

Fattore di normalizzazione per funzione di densità: $\dfrac{1}{\sqrt{(2\pi)^p|\Sigma|}}$

$(|\Sigma| = \det\Sigma)$

$$\Rightarrow \quad f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^p|\Sigma|}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})^T} \qquad -\infty < x_i < \infty$$

$\mathcal{N}_p(\boldsymbol{\mu}, \Sigma)$ insieme delle variabili con f come funzione densità

Alcune proprietà

variabili casuali normali X_1, X_2 *non* sono correlate (Σ diagonale)

$$\Rightarrow \qquad f(\mathbf{x}) = f(x_1) \cdot f(x_2)$$

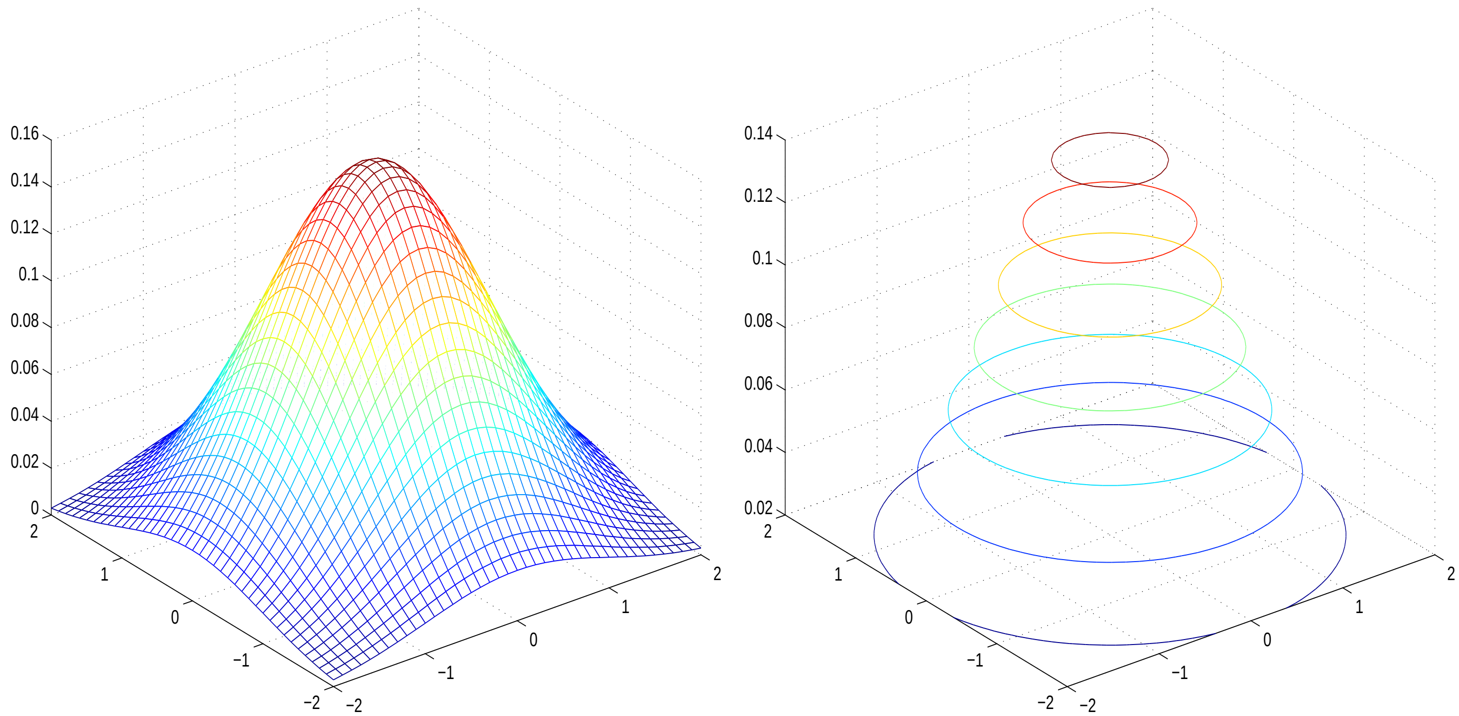
OSS: L'insieme delle $\mathbf{x}$ tali che

$$(\mathbf{x} - \boldsymbol{\mu})\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})^T = c^2$$

con c costante è ellissoide centrato in $\boldsymbol{\mu}$ ed assi $\pm c\sqrt{\lambda_i}\mathbf{v}_i$

es. $\sigma_{1,2} = 0 \quad \Rightarrow \quad \frac{(x_1-\mu_1)^2}{\sigma_{1,1}} + \frac{(x_2-\mu_2)^2}{\sigma_{2,2}} = c^2$

OSS: Se $\sigma_{1,1} = \sigma_{2,2}$ allora il valore di $\sigma_{1,2}$ cambia la forma del grafico


0.16
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
2
1
0
−1
−2
−2
−1
0
1
2
0.14
0.12
0.1
0.08
0.06
0.04
0.02
2
1
0
−1
−2
−2
−1
0
1
2