Probabilità e Processi Stocastici (455AA)

Lezione 11

Dario Trevisan

28/10/2024

Section 1

Variabil Gaussiane reali

Intepretazione dei parametri

Sia X una variabile con densità gaussiana

$$p(X = x) \propto \exp\left(ax^2 + bx\right).$$

Allora vale

$$a = -\frac{1}{2\sigma_X^2}, \quad b = \frac{\mathbb{E}[X]}{\sigma_X^2},$$

ossia

$$\operatorname{Var}(X) = \sigma_X^2 = -\frac{1}{2a} \quad \mathbb{E}[X] = -\frac{b}{2a}.$$

Dimostrazione

$$p(X=x) = \exp(ax^2+bx+c) \qquad \underline{a<0}$$

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x \exp(ax^2+bx+c)\,dx = \int_{-\infty}^{+\infty} \underbrace{x \exp(ax^2)}_{=\frac{1}{2a}\frac{d}{dx}\exp(ax^2)} \exp(bx+c)\,dx$$

Integro per part.

$$= \underbrace{\frac{1}{2a} \cancel{\exp(ax^2)\exp(bx+c)}}_{\text{perché } a<0} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \frac{1}{2a} \exp(ax^2)\cdot b\cdot \exp(bx+c)\,dx$$

$$= -\frac{b}{2a} \underbrace{\int_{-\infty}^{+\infty} \exp(ax^2+bx+c)\,dx}_{=1} = -\frac{b}{2a}$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 \exp(ax^2+bx+c)\,dx = \int_{-\infty}^{+\infty} \left(x\exp(ax^2)\right) x \exp(bx+c)\,dx$$

$$= -\frac{1}{2a}\int_{-\infty}^{+\infty} \exp(ax^2)\left[\frac{d}{dx}\, x\exp(bx+c)\right] dx = -\frac{1}{2a} - \frac{b}{2a}\int_{-\infty}^{+\infty} x\exp(ax^2+bx+c)\,dx$$

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 \exp(ax^2+bx+c)\,dx \overset{1}{=} \int_{-\infty}^{+\infty} \left(x \exp(ax^2)\right) x \exp(bx+c)\,dx$$

$$= -\frac{1}{2a}\int_{-\infty}^{+\infty} \exp(ax^2)\left[\frac{d}{dx}\, x \exp(bx+c)\right] dx =$$

$$= -\frac{1}{2a} - \frac{b}{2a}\int_{-\infty}^{+\infty} x \exp(ax^2+bx+c)\,dx = -\frac{1}{2a} + \left(-\frac{b}{2a}\right)\left(-\frac{b}{2a}\right) = -\frac{1}{2a} + \left(\mathbb{E}[X]\right)^2$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - \left(\mathbb{E}[X]\right)^2 = -\frac{1}{2a}$$

$$\boxed{a = -\frac{1}{2\sigma^2}}$$

Trasformazione affine

Sia X una variabile con densità continua $\mathcal{N}(m, \sigma^2)$ e siano $\lambda \neq 0$, $c \in \mathbb{R}$. Allora la variabile $Y = \lambda X + c$ ha densità continua gaussiana, di parametri $\mathcal{N}(\lambda m + c, \lambda^2 \sigma^2)$.

- se X ha densità gaussiana $\mathcal{N}(m, \sigma^2)$, la sua standardizzata

$$\frac{X - m}{\sigma} \quad \text{ha densità continua } \mathcal{N}(0, 1),$$

pertanto detta anche densità **gaussiana standard**, che ha densità

$$\exp\left(-\frac{1}{2}x^2\right)\frac{1}{\sqrt{2\pi}} \quad \text{per } x \in \mathbb{R}.$$

Trasformazione affine

Sia X una variabile con densità continua $\mathcal{N}(m, \sigma^2)$ e siano $\lambda \neq 0$, $c \in \mathbb{R}$. Allora la variabile $Y = \lambda X + c$ ha densità continua gaussiana, di parametri $\mathcal{N}(\lambda m + c, \lambda^2 \sigma^2)$.

- se X ha densità gaussiana $\mathcal{N}(m, \sigma^2)$, la sua standardizzata

$$Z = \frac{X - m}{\sigma} \quad \text{ha densità continua } \mathcal{N}(0, 1),$$

  pertanto detta anche densità **gaussiana standard**, che ha densità

$$\exp\left(-\frac{1}{2}x^2\right)\frac{1}{\sqrt{2\pi}} \quad \text{per } x \in \mathbb{R}.$$

- Se $\lambda = 0$, la variabile $\lambda X + c = c$ è costante. Per uniformare le notazioni, si conviene di considerare anche le variabili costanti come caso *degenere* di una densità gaussiana.

Dimostrazione

$X \sim N(m, \sigma^2)$ $\qquad Y = \lambda X + c \qquad \underline{\underline{\lambda \neq 0}}$

$$p(Y = y) \overset{?}{=} p\left(X = \frac{y-c}{\lambda}\right) \cdot \frac{1}{|g'(g^{-1}(y))|}$$

cambio di variabile $g(x) = \lambda x + c$ $\qquad g^{-1}(y) = \frac{y-c}{\lambda}$

$g'(x) = \lambda$

$$= \exp\Big(\underbrace{a\left(\frac{y-c}{\lambda}\right)^2 + b\left(\frac{y-c}{\lambda}\right)}_{2^\circ \text{ grado in } y} + d\Big)\frac{1}{|\lambda|}$$

2° grado in y $\Longrightarrow$ Y è gaussiana

$\mathbb{E}[Y] = \mathbb{E}[\lambda X + c] = \lambda\,\mathbb{E}[X] + c = \lambda m + c$

$\mathrm{Var}(Y) = \lambda^2\,\mathrm{Var}(X) = \lambda^2 \sigma^2$

CDF

Se X è $N(0,1)$

$$\Phi(t) = CDF_X(t) = \int_{-\infty}^{t} \exp\left(-\frac{x^2}{2}\right) \frac{dx}{\sqrt{2\pi}}$$

La funzione di ripartizione gaussiana (anche nel caso standard) non è esprimibile in termini di funzioni elementari.

- Il comando R per ottenerne i valori è `pnorm()`.

OSS: $CDF_{\lambda X + c}(t) = P(\lambda X + c \leq t) \underset{\lambda > 0}{=} P\left(X \leq \frac{t-c}{\lambda}\right) = CDF_X\left(\frac{t-c}{\lambda}\right)$

In particolare $X \sim N(m, \sigma^2)$ è $X = \sigma Z + m$ con $Z \sim N(0,1)$

$\Rightarrow CDF_X(t) = CDF_{\sigma Z + m}(t) = CDF_Z\left(\frac{t-m}{\sigma}\right) = \Phi\left(\frac{t-m}{\sigma}\right)$

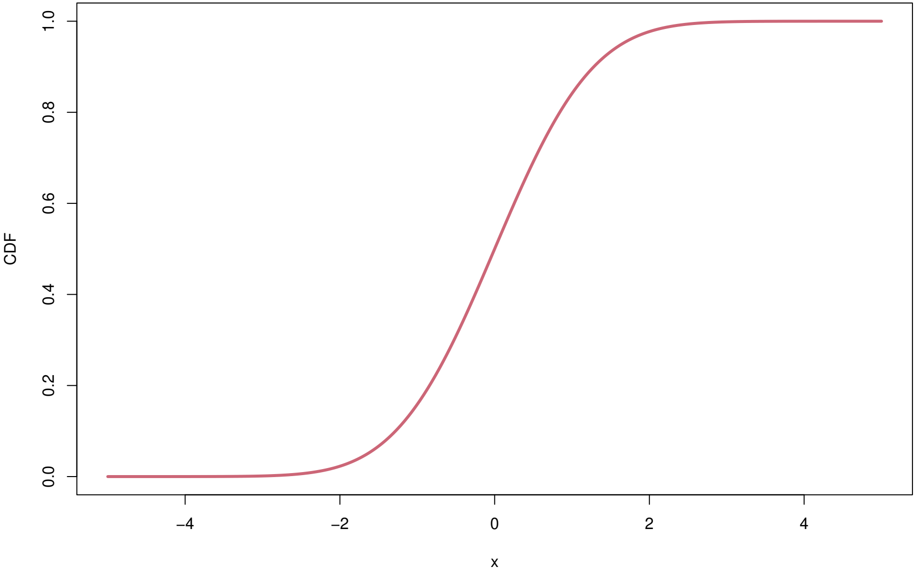


Figure 1: CDF di una variabile gaussiana standard

MGF e funzione caratteristica

Sia X una variabile con densità continua $\mathcal{N}(m, \sigma^2)$. Allora

$$\boxed{\mathrm{MGF}_X(t) = \exp\left(mt + \frac{\sigma^2}{2}t^2\right)},$$

e

$$\varphi_X(\xi) = \exp\left(im\xi - \frac{\sigma^2}{2}\xi^2\right).$$

Dimostrazione

Suppongo $X \sim N(0,1)$ standard

$$MGF_X(t) = \int_{-\infty}^{+\infty} e^{tx} e^{-\frac{1}{2}x^2} \frac{dx}{\sqrt{2\pi}} = \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}x^2 + tx\right) \frac{dx}{\sqrt{2\pi}}$$

$$= \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\left(x^2 - 2tx + t^2\right) + \frac{t^2}{2}\right) \frac{dx}{\sqrt{2\pi}}$$

$$= \int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}(x-t)^2\right) \cdot \exp\left(\frac{t^2}{2}\right) \frac{dx}{\sqrt{2\pi}}$$

$$= \exp\left(\frac{t^2}{2}\right) \cancel{\int_{-\infty}^{+\infty} \exp\left(-\frac{1}{2}\frac{(x-t)^2}{1}\right) \frac{dx}{\sqrt{2\pi}}} = 1$$

densità $N(t,1)$

$$= \exp\left(t^2/2\right)$$

Esercizi

(1) Calcolare $\mathbb{E}[X^3]$ $\mathbb{E}[X^4]$

dove $X \sim N(0,1)$

(2) Calcolare $\mathbb{E}[|X|]$ $X \sim N(0,1)$

Section 2

Problemi su variabili gaussiane reali

Richiami

- X Gaussiana reale se $p(X = x) \propto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)$

Richiami

- X Gaussiana reale se $p(X = x) \propto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)$
- $m = \mathbb{E}[X]$, $\sigma = \sigma_X$

Richiami

- X Gaussiana reale se $p(X = x) \propto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)$
- $m = \mathbb{E}[X]$, $\sigma = \sigma_X$
- X Gaussiana $\to aX + b$ è Gaussiana (ovvia trasformazione dei parametri)

Richiami

- X Gaussiana reale se $p(X = x) \propto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)$
- $m = \mathbb{E}[X]$, $\sigma = \sigma_X$
- X Gaussiana $\to aX + b$ è Gaussiana (ovvia trasformazione dei parametri)
- CDF_X non è esprimibile con funzioni elementari (comando 'pnorm()')

Richiami

- X Gaussiana reale se $p(X = x) \propto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right)$
- $m = \mathbb{E}[X]$, $\sigma = \sigma_X$
- X Gaussiana $\to aX + b$ è Gaussiana (ovvia trasformazione dei parametri)
- CDF_X non è esprimibile con funzioni elementari (comando ‘pnorm()’)
- $\mathrm{MGF}_X(t) = \exp\left(tm + t^2\sigma^2/2\right)$

Problemi

L'orario d'arrivo a lezione degli studenti di ingegneria robotica segue approssimativamente una distribuzione gaussiana di media 8:25 e deviazione standard 5 minuti. Preso uno studente a caso,

1. calcolare la probabilità che arrivi dopo l'inizio delle lezioni (8:30);
2. calcolare il ritardo medio (in minuti).

Esprimere eventualmente i risultati come opportuni integrali o indicare un comando R per il calcolo numerico.

$$X \sim N(8{:}25,\ \sigma = 5\,\text{min}), \quad R = \max\{X - 8{:}30, 0\}$$

$$\boxed{X = \sigma Z + m} \qquad Z \sim N(0,1)$$

$$P(X > 8{:}30) = P(R > 0)$$

$$\parallel$$

$$P(5 \cdot Z + 8{:}25 > 8{:}30)$$

$$\parallel$$

$$P(5 \cdot Z > 5) = \underline{\underline{P(Z > 1)}} = \int_1^{+\infty} \exp\left(-\frac{z^2}{2}\right) \frac{dz}{\sqrt{2\pi}}$$

$$= 1 - P(Z \leq 1)$$

$$= 1 - \text{pnorm}(1)$$

$$X = 5Z + 8{:}25$$

$$\mathbb{E}[R] = \mathbb{E}[\max\{X - 8{:}30, 0\}]$$

$$= \mathbb{E}[\max\{5Z + 8{:}25 - 8{:}30, 0\}]$$

$$= \mathbb{E}[\max\{5(Z-1), 5\cdot 0\}] = 5\,\mathbb{E}[\max\{Z-1, 0\}]$$

$$= 5 \cdot \int_{-\infty}^{+\infty} \max\{z-1, 0\} \exp\left(-\frac{z^2}{2}\right) \frac{dz}{\sqrt{2\pi}}$$

$$= 5 \cdot \int_{1}^{+\infty} (z-1) \exp\left(-\frac{z^2}{2}\right) \frac{dz}{\sqrt{2\pi}}$$

$$= 5\left(\int_1^{+\infty} \underbrace{z \exp\left(-\frac{z^2}{2}\right)}_{-\frac{d}{dz}\exp\left(-\frac{z^2}{2}\right)} \frac{dz}{\sqrt{2\pi}} - \int_1^{+\infty} \exp\left(-\frac{z^2}{2}\right) \frac{dz}{\sqrt{2\pi}}\right)$$

$$= 5\left(\exp\left(-\frac{1}{2}\right) - \left(1 - \Phi(1)\right)\right) \overset{?}{\geq} 0$$

L'altezza degli studenti (maschi) del corso di ingegneria è rappresentata da una distribuzione gaussiana di media 175 cm e deviazione standard 10 cm. L'altezza delle studentesse (femmine) è pure una gaussiana di media 160 cm con deviazione standard 10 cm. Preso uno/a studente a caso, si osserva che è alto/a 165 cm. Dire se è più probabile che sia maschio o femmina,

1. senza conoscere la percentuale di studenti maschi e femmina nel corso;
2. sapendo anche che i maschi rappresentano il 70% degli studenti di ingegneria e il 30% è femmina (solo per semplicità di calcolo escludiamo le persone non identificate in uno dei due generi).

Section 3

Vettori gaussiani

Il caso vettoriale: definizione veloce

Si dice che una variabile aleatoria $X \in \mathbb{R}^d$ ha densità continua gaussiana se vale

$$p(X = x) \propto \exp\left(\sum_{i,j=1}^{d} a_{ij} x_i x_j + \sum_{i=1}^{d} b_i x_i \right), \quad \text{per ogni } x = (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

per degli opportuni parametri $a = (a_{ij})_{i,j=1}^d \in \mathbb{R}^{d \times d}$, $b = (b_i)_{i=1}^d \in \mathbb{R}^d$.

- Il (multi-)parametro $a = (a_{ij})_{i,j=1}^d$ corrisponde ad una matrice e $b = (b_i)_{i=1}^d \in \mathbb{R}^d$ è un vettore.

Il caso vettoriale: definizione veloce

Si dice che una variabile aleatoria $X \in \mathbb{R}^d$ ha densità continua gaussiana se vale

$$p(X = x) \propto \exp\left(\sum_{i,j=1}^{d} a_{ij} x_i x_j + \sum_{i=1}^{d} b_i x_i \right), \quad \text{per ogni } x = (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

per degli opportuni parametri $a = (a_{ij})_{i,j=1}^d \in \mathbb{R}^{d \times d}$, $b = (b_i)_{i=1}^d \in \mathbb{R}^d$.

- Il (multi-)parametro $a = (a_{ij})_{i,j=1}^d$ corrisponde ad una matrice e $b = (b_i)_{i=1}^d \in \mathbb{R}^d$ è un vettore.
- in forma compatta

$$\sum_{i,j=1}^{d} a_{ij} x_i x_j + \sum_{i=1}^{d} b_i x_i = \overset{x^T a x}{\overset{\shortparallel}{x \cdot (ax)}} + \overset{b^T x}{\overset{\shortparallel}{b \cdot x}}.$$

Il caso vettoriale: definizione veloce

Si dice che una variabile aleatoria $X \in \mathbb{R}^d$ ha densità continua gaussiana se vale

$$p(X = x) \propto \exp\left(\sum_{i,j=1}^{d} a_{ij} x_i x_j + \sum_{i=1}^{d} b_i x_i \right), \quad \text{per ogni } x = (x_1, \ldots, x_d) \in \mathbb{R}^d,$$

per degli opportuni parametri $a = (a_{ij})_{i,j=1}^d \in \mathbb{R}^{d \times d}$, $b = (b_i)_{i=1}^d \in \mathbb{R}^d$.

- Il (multi-)parametro $a = (a_{ij})_{i,j=1}^d$ corrisponde ad una matrice e $b = (b_i)_{i=1}^d \in \mathbb{R}^d$ è un vettore.
- in forma compatta

$$a = \left(\frac{a + a^T}{2}\right) + \left(\frac{a - a^T}{2}\right)$$

$$\sum_{i,j=1}^{d} a_{ij} x_i x_j + \sum_{i=1}^{d} b_i x_i = x \cdot (ax) + b \cdot x.$$

- possiamo supporre a simmetrica e definita ~~positiva~~. negativa

Intepretazione dei parametri

$$(\Sigma_X)_{i,j} = \mathrm{Cov}(X_i, X_j) \qquad \mathbb{E}[X] = (\mathbb{E}[X_i])_{i=1\ldots d}$$

Sia $X \in \mathbb{R}^d$ una variabile con densità gaussiana definita come sopra. Allora la matrice delle covarianze Σ_X è definita positiva (quindi invertibile) e vale

$$\boxed{a = -\frac{1}{2}\Sigma_X^{-1},} \quad \boxed{b = \Sigma_X^{-1}\mathbb{E}[X],}$$

ossia

$$\Sigma_X = -\frac{1}{2}a^{-1} \quad \mathbb{E}[X] = -\frac{1}{2}a^{-1}b.$$

Definizione usuale

Si dice che $X \in \mathbb{R}^d$ ha densità continua gaussiana di vettore dei valor medi $m \in \mathbb{R}^d$ e matrice delle covarianze $\Sigma > 0$, e si scrive brevemente $\mathcal{N}(m, \Sigma)$, se vale

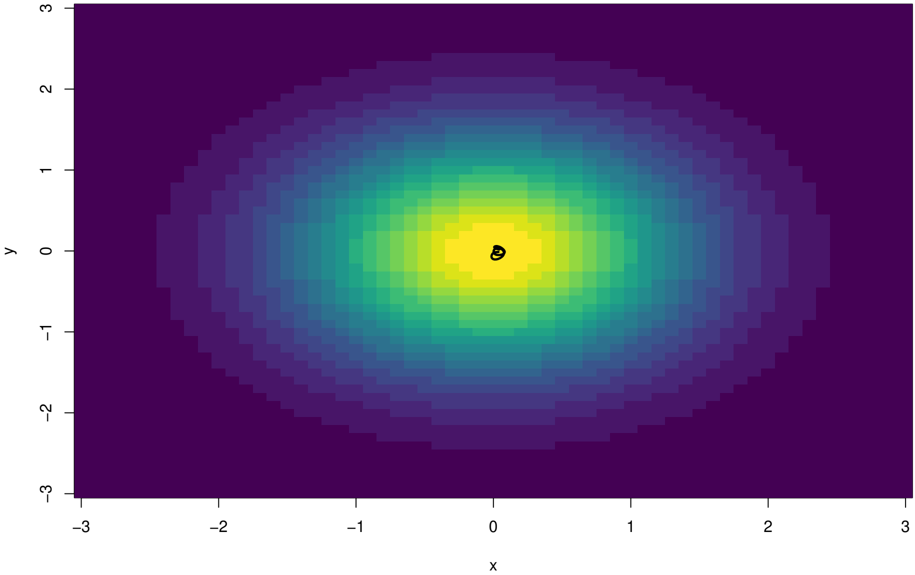
$$\boxed{p(X = x) \propto \exp\left(-\frac{1}{2}\left((x-m)\cdot\Sigma^{-1}(x-m)\right)\right).} \quad \forall x \in \mathbb{R}^d$$

- Più esplicitamente, si può mostrare che vale l'identità

$$p(X = x) = \exp\left(-\frac{1}{2}\left((x-m)\cdot\Sigma^{-1}(x-m)\right)\right)\frac{1}{\sqrt{(2\pi)^d \det(\Sigma)}}.$$

Heatmap

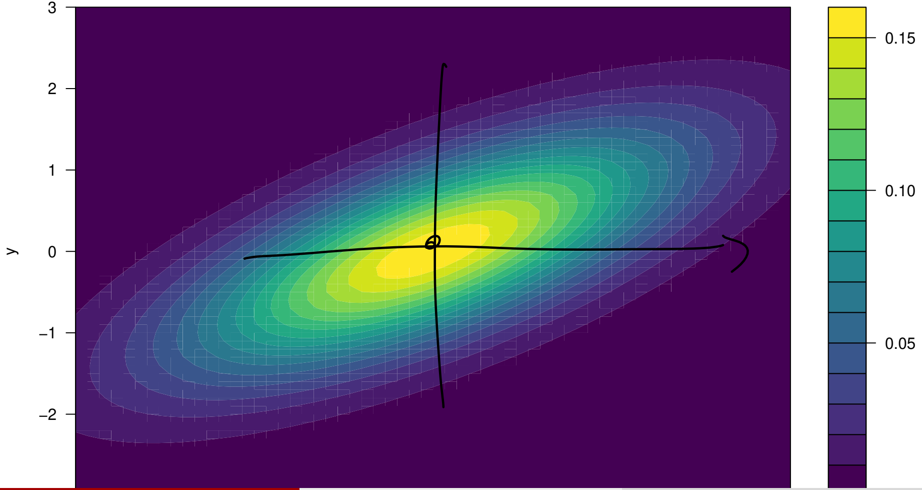
$N\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$



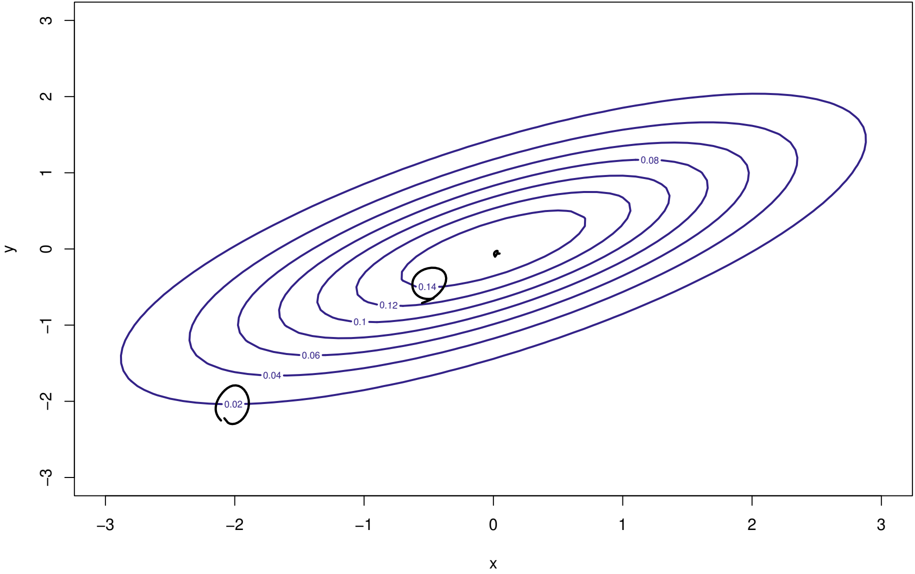
Cosa accade se cambiamo la matrice di covarianza? la figura sotto mostra il caso di

$$\Sigma = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}.$$

$Cov(X,Y) = 1$



Curve di livello



Proprietà: trasformazioni affini

Sia $X \in \mathbb{R}^d$ una variabile con densità gaussiana $\mathcal{N}(m, \Sigma)$ e sia $A \in \mathbb{R}^{k \times d}$, $b \in \mathbb{R}^k$.

- Allora, la variabile $Y = AX + b$ ha densità gaussiana $\mathcal{N}(Am + b, A\Sigma A^T)$, purché $A\Sigma A^T$ sia invertibile (o, il che è lo stesso, definita positiva).

Esempio $Y = 2X$

$$p(Y=y) = p(2X = y)$$
$$= p\left(X = \frac{y}{2}\right) \frac{1}{|\det(2\,\mathrm{Id})|}$$

Conseguenze

$$X_1 = (1, 0, \ldots, 0) \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix}$$

Sia $X \in \mathbb{R}^d$ una variabile con densità continua $\mathcal{N}(m, \Sigma)$. Allora

1. ogni marginale X_i ha densità $\mathcal{N}(m_i, \Sigma_{ii})$,

Conseguenze

Sia $X \in \mathbb{R}^d$ una variabile con densità continua $\mathcal{N}(m, \Sigma)$. Allora

1. ogni marginale X_i ha densità $\mathcal{N}(m_i, \Sigma_{ii})$,
2. per ogni $v \in \mathbb{R}^d$, la variabile $v \cdot X = \sum_{i=1}^d v_i X_i$ ha densità $\mathcal{N}(v \cdot m, v \cdot \Sigma v)$.

Conseguenze

Sia $X \in \mathbb{R}^d$ una variabile con densità continua $\mathcal{N}(m, \Sigma)$. Allora

1. ogni marginale X_i ha densità $\mathcal{N}(m_i, \Sigma_{ii})$,
2. per ogni $v \in \mathbb{R}^d$, la variabile $v \cdot X = \sum_{i=1}^d v_i X_i$ ha densità $\mathcal{N}(v \cdot m, v \cdot \Sigma v)$.

- Ad esempio, se le due marginali X_1, X_2 di X sono non correlate, sa ha che, ponendo $v = (1, 1, 0, 0, \ldots)$ la variabile $X_1 + X_2$ ha densità $\mathcal{N}(m_1 + m_2, \sigma^2_{X_1} + \sigma^2_{X_2})$.

$D = \mathrm{diag}(\lambda_1, \dots, \lambda_d) \qquad \lambda_i > 0 \qquad \sqrt{D} = \mathrm{diag}(\sqrt{\lambda_i})$

Sia $X \in \mathbb{R}^d$ una variabile con densità continua $\mathcal{N}(m, \Sigma)$. Allora la variabile standardizzata

$$Z = \sqrt{D}^{-1} U(X - m) \quad \text{ha densità continua } \mathcal{N}(0, Id),$$

detta anche **gaussiana standard** vettoriale.

- La densità esplicita è

$$P(Z = z | \mathcal{N}(0, Id)) = \exp\left(-\frac{1}{2}\sum_{i=1}^{d} z_i^2\right)\frac{1}{\sqrt{(2\pi)^d}},$$

$$= \left(\frac{\exp\left(-\frac{1}{2}z_1^2\right)}{\sqrt{2\pi}}\right)\left(\frac{\exp\left(-\frac{1}{2}z_2^2\right)}{\sqrt{2\pi}}\right)\cdots\left(\frac{\exp\left(-\frac{1}{2}z_d^2\right)}{\sqrt{2\pi}}\right)$$

se Z è standard
$\Rightarrow (Z_i)_{i=1,\dots,d}$ sono indipendenti

Sia $X \in \mathbb{R}^d$ una variabile con densità continua $\mathcal{N}(m, \Sigma)$. Allora la variabile standardizzata

$$Z = \sqrt{D}^{-1} U(X - m) \quad \text{ha densità continua } \mathcal{N}(0, Id),$$

detta anche **gaussiana standard** vettoriale.

- La densità esplicita è

$$P(Z = z | \mathcal{N}(0, Id)) = \exp\left(-\frac{1}{2} \sum_{i=1}^{d} z_i^2 \right) \frac{1}{\sqrt{(2\pi)^d}},$$

- Le variabili marginali Z_1, Z_2, $\ldots Z_d$ sono *indipendenti*, oltre ad essere non correlate.

Indipendenza

Per le variabili con densità gaussiana l'indipendenza è equivalente alla non correlazione!

- Siano $X \in \mathbb{R}^d$, $Y \in \mathbb{R}^k$ variabili aleatorie indipendenti con densità gaussiane. Allora la variabile congiunta $(X, Y) \in \mathbb{R}^{d+k}$ ha densità gaussiana.

<u>Dim</u>

$$p((X,Y)=(x,y)) = p(X=x)\, p(Y=y)$$

$$= \exp(P_2(x)) \exp(\tilde{P}_2(y))$$

$$= \exp(P_2(x) + \tilde{P}_2(y))$$

Indipendenza

Per le variabili con densità gaussiana l'indipendenza è equivalente alla non correlazione!

- Siano $X \in \mathbb{R}^d$, $Y \in \mathbb{R}^k$ variabili aleatorie indipendenti con densità gaussiane. Allora la variabile congiunta $(X, Y) \in \mathbb{R}^{d+k}$ ha densità gaussiana.
- Viceversa, se la variabile congiunta $(X, Y) \in \mathbb{R}^{d+k}$ ha densità gaussiana e $\operatorname{Cov}(X_i, Y_j) = 0$ per ogni $i \in \{1, \ldots, d\}$, $j \in \{1, \ldots, k\}$, allora X e Y sono indipendenti.

Dimostrazione

Dim nel caso $X \in \mathbb{R}$ $Y \in \mathbb{R}$ $\qquad \Sigma_{XY} = \begin{pmatrix} \sigma_X^2 & 0 \\ 0 & \sigma_Y^2 \end{pmatrix}$

$$\Sigma_{XY}^{-1} = \begin{pmatrix} \sigma_X^{-2} & 0 \\ 0 & \sigma_Y^{-2} \end{pmatrix} \qquad \uparrow \text{diagonale}$$

$$p((X,Y) = (x,y)) \propto \exp\left(-\frac{1}{2} \begin{pmatrix} x - m_x \\ y - m_y \end{pmatrix} \cdot \begin{pmatrix} \sigma_X^{-2} & 0 \\ 0 & \sigma_Y^{-2} \end{pmatrix} \begin{pmatrix} x - m_x \\ y - m_y \end{pmatrix}\right)$$

$$\propto \exp\left(-\frac{1}{2} \begin{pmatrix} x - m_x \\ y - m_y \end{pmatrix} \cdot \begin{pmatrix} (x - m_x)/\sigma_X^2 \\ (y - m_y)/\sigma_Y^2 \end{pmatrix}\right)$$

$$\propto \exp\left(-\frac{1}{2} \frac{(x - m_x)^2}{\sigma_X^2}\right) \exp\left(-\frac{1}{2} \frac{(y - m_y)^2}{\sigma_Y^2}\right)$$

MGF e funzione caratteristica

Sia $X \in \mathbb{R}^d$ una variabile con densità continua gaussiana $\mathcal{N}(m, \Sigma)$. Allora

$$\mathrm{MGF}_X(t) = \exp\left(m \cdot t + \frac{1}{2} t \cdot \Sigma t\right),$$

e

$$\varphi_X(\xi) = \exp\left(i m \cdot \xi - \frac{1}{2} \xi \cdot \Sigma \xi\right).$$

Dimostrazione

Il caso degenere

Se Σ è semidefinita positiva, non necessariamente invertibile, usiamo le espressioni trovate sopra per *definire* una variabile aleatoria vettoriale gaussiana anche nel caso in cui non abbia densità continua.

- L'intepretazione in tali casi è che la densità continua si concentra in un sottospazio affine di dimensione più bassa dello spazio ambiente $\mathbb{R}^d$.

Esiste comunque una legge gaussiana $N(m, \Sigma)$ che è definita tramite la sua MGF

$$\boxed{MGF_X(t) = \exp\left(m \cdot t + \frac{1}{2} t \cdot \Sigma t\right)}$$

Il caso degenere

Se Σ è semidefinita positiva, non necessariamente invertibile, usiamo le espressioni trovate sopra per *definire* una variabile aleatoria vettoriale gaussiana anche nel caso in cui non abbia densità continua.

- L'intepretazione in tali casi è che la densità continua si concentra in un sottospazio affine di dimensione più bassa dello spazio ambiente $\mathbb{R}^d$.
- Per visualizzare cosa accade, diamo un plot nel caso "quasi degenere" ponendo ad esempio

$$\Sigma = \begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}.$$

$$\rho = \frac{\mathrm{cov}(X, Y)}{\sigma_X \sigma_Y}$$

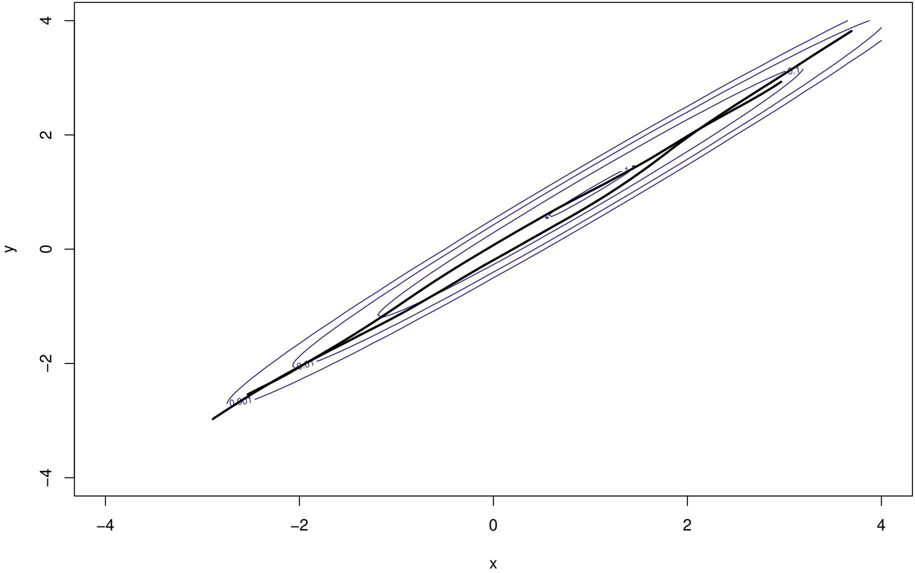


Figure 5: rappresentazione della densità gaussiana vettoriale nel caso vettoriale $d = 2$, $m = (1, 1)$ e Σ definita sopra

Section 4

Stima dei parametri dalle osservazioni di variabili gaussiane

Stima dei parametri da una singola osservazione

Data una variabile aleatoria $X \in \mathbb{R}^d$ con densità gaussiana $\mathcal{N}(m, \Sigma)$, come stimare i parametri basandosi sull'osservazione di X?

1. Punto di vista bayesiano:

Stima dei parametri da una singola osservazione

Data una variabile aleatoria $X \in \mathbb{R}^d$ con densità gaussiana $\mathcal{N}(m, \Sigma)$, come stimare i parametri basandosi sull'osservazione di X?

1. Punto di vista bayesiano:

- i parametri m, Σ sono delle variabili aleatorie con opportune densità *a priori*,

Stima dei parametri da una singola osservazione

Data una variabile aleatoria $X \in \mathbb{R}^d$ con densità gaussiana $\mathcal{N}(m, \Sigma)$, come stimare i parametri basandosi sull'osservazione di X?

1. Punto di vista bayesiano:

- i parametri m, Σ sono delle variabili aleatorie con opportune densità *a priori*,
- si aggiorna la densità avendo osservato $X = x$ tramite Bayes.

Stima dei parametri da una singola osservazione

Data una variabile aleatoria $X \in \mathbb{R}^d$ con densità gaussiana $\mathcal{N}(m, \Sigma)$, come stimare i parametri basandosi sull'osservazione di X?

1. Punto di vista bayesiano:
   - i parametri m, Σ sono delle variabili aleatorie con opportune densità *a priori*,
   - si aggiorna la densità avendo osservato $X = x$ tramite Bayes.
2. Alternativa più diretta, ma meno informativa, è la MLE.

Sia $X \in \mathbb{R}$ una variabile con densità gaussiana $\mathcal{N}(m, \sigma^2)$.

- I parametri non sono noti a priori, e sono stimati sulla base di una osservazione $X = x$.

Sia $X \in \mathbb{R}$ una variabile con densità gaussiana $\mathcal{N}(m, \sigma^2)$.

- I parametri non sono noti a priori, e sono stimati sulla base di una osservazione $X = x$.
- Seguendo il metodo bayesiano si introducono le rispettive variabili aleatorie M per la media e V per la varianza (a valori positivi).

Sia $X \in \mathbb{R}$ una variabile con densità gaussiana $\mathcal{N}(m, \sigma^2)$.

- I parametri non sono noti a priori, e sono stimati sulla base di una osservazione $X = x$.
- Seguendo il metodo bayesiano si introducono le rispettive variabili aleatorie M per la media e V per la varianza (a valori positivi).
- La verosimiglianza, ossia la densità di X, note la media $M = m$ e la varianza $V = v$ è

$$L(m, v; x) = p(X = x | M = m, V = v) = \exp\left(-\frac{1}{2v}(x-m)^2\right)\frac{1}{\sqrt{2\pi v}}$$

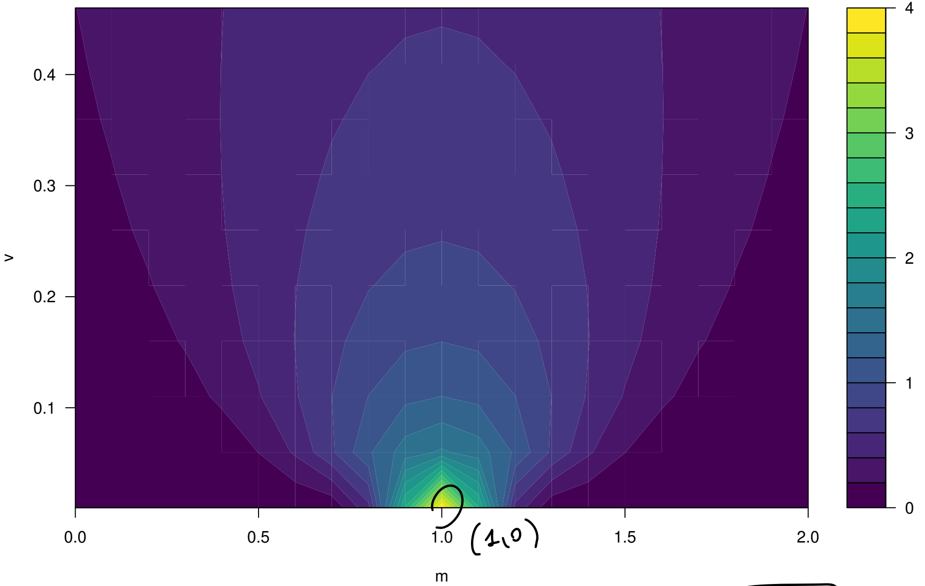


Figure 6: verosimiglianza per M e V avendo osservato $\boxed{X = 1}$.

Stima di massima verosimiglianza

Vediamo che L è massima per $m = x$ ($x = 1$ sopra) e $v = 0$:

- Passando al logaritmo e moltiplicando per -2, invece di massimizzare L basta minimizzare la funzione

$$(m, v) \mapsto -2 \log L(m, v; x) = \frac{1}{v}(x - m)^2 + \log(2\pi v).$$

$$\frac{\partial}{\partial m}\left[\frac{1}{v}(x-m)^2 + \log(2\pi v)\right] = 2\frac{(m-x)}{v} = 0$$

$$\frac{\partial}{\partial v}\left[\quad\right] = -\frac{1}{v^2}(x-m)^2 + \frac{1}{v} = 0$$

Stima di massima verosimiglianza

Vediamo che L è massima per $m = x$ ($x = 1$ sopra) e $v = 0$:

- Passando al logaritmo e moltiplicando per -2, invece di massimizzare L basta minimizzare la funzione

$$(m, v) \mapsto -2 \log L(m, v; x) = \frac{1}{v}(x - m)^2 + \log(2\pi v).$$

- Derivando rispetto a m e imponendo che la derivata si annulli si ottiene, per ogni $v > 0$,

$$\frac{2}{v}(x - m) = 0 \quad \text{da cui} \boxed{m = x,}$$

Stima di massima verosimiglianza

Vediamo che L è massima per $m = x$ ($x = 1$ sopra) e $v = 0$:

- Passando al logaritmo e moltiplicando per -2, invece di massimizzare L basta minimizzare la funzione
$$(m, v) \mapsto -2 \log L(m, v; x) = \frac{1}{v}(x - m)^2 + \log(2\pi v).$$
- Derivando rispetto a m e imponendo che la derivata si annulli si ottiene, per ogni $v > 0$,
$$\frac{2}{v}(x - m) = 0 \quad \text{da cui } \boxed{m = x,}$$
- se deriviamo rispetto a v, tenendo fisso m, si trova
$$-\frac{1}{v^2}(x - m)^2 + \frac{1}{v} = 0, \quad \text{da cui } \boxed{v = (x - m)^2.}$$

Stima di massima verosimiglianza

Vediamo che L è massima per $m = x$ ($x = 1$ sopra) e $v = 0$:

- Passando al logaritmo e moltiplicando per -2, invece di massimizzare L basta minimizzare la funzione
$$(m, v) \mapsto -2 \log L(m, v; x) = \frac{1}{v}(x - m)^2 + \log(2\pi v).$$
- Derivando rispetto a m e imponendo che la derivata si annulli si ottiene, per ogni $v > 0$,
$$\frac{2}{v}(x - m) = 0 \quad \text{da cui } m = x,$$
- se deriviamo rispetto a v, tenendo fisso m, si trova
$$-\frac{1}{v^2}(x - m)^2 + \frac{1}{v} = 0, \quad \text{da cui } v = (x - m)^2.$$
- si trova la coppia $\boxed{m_{\mathrm{MLE}} = x,\ v_{\mathrm{MLE}} = 0.}$

Con calcoli analoghi a quanto fatto sopra si ottiene anche che

1. se il pararametro di varianza $V = v_0 > 0$ è noto (ossia $V = v_0$ è costante rispetto all'informazione a priori), allora la stima di massima verosimiglianza per il parametro di media è

$$m_{\mathrm{MLE}} = x$$

(qualsiasi sia v_0).

Con calcoli analoghi a quanto fatto sopra si ottiene anche che

1. se il pararametro di varianza $V = v_0 > 0$ è noto (ossia $V = v_0$ è costante rispetto all'informazione a priori), allora la stima di massima verosimiglianza per il parametro di media è

$$m_{\mathrm{MLE}} = x$$

(qualsiasi sia v_0).

2. se il parametro di media $M = m_0 \in \mathbb{R}$ è noto (ossia $M = m_0$ è costante a priori), allora la stima di massima verosimiglianza per il parametro di varianza è

$$v_{\mathrm{MLE}} = (x - m_0)^2.$$

Stima dei parametri da osservazioni indipendenti

$$X = (X_1 \dots X_n) \sim N\left((m)_{i=1\dots n}, \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix} \right)$$

- estendiamo i risultati al caso in cui osservano variabili indipendenti $(X_i)_{i=1}^n$, tutte gaussiane con i medesimi parametri (il termine statistico è un *campione* di taglia n, il numero di osservazioni).

Stima dei parametri da osservazioni indipendenti

- estendiamo i risultati al caso in cui osservano variabili indipendenti $(X_i)_{i=1}^n$, tutte gaussiane con i medesimi parametri (il termine statistico è un *campione* di taglia n, il numero di osservazioni).
- Per semplificare la notazione, sia $X = (X_i)_{i=1}^n$ a valori in $\mathbb{R}^n$, con

$$m = \mathbb{E}[X_i] \quad v = \mathrm{Var}(X_i) \quad \text{per ogni } i = 1, \ldots, n.$$

Come nel caso della singola osservazione, si introducono le rispettive variabili aleatorie M per la media, V per la varianza (la seconda a valori positivi).

La rete bayesiana associata è

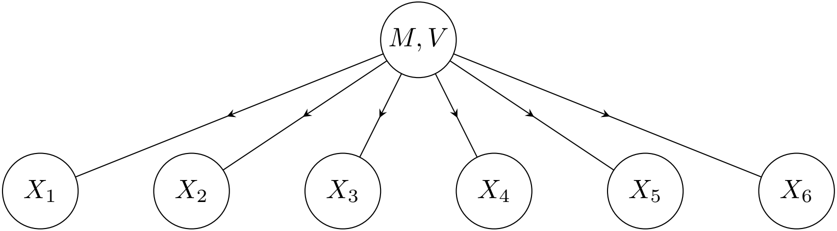


Figure 7: Rete bayesiana tra le variabili (M, V) e le $(X_i)_{i=1}^n$ per $n = 6$

Stima di massima verosimiglianza

L'ipotesi di indipendenza, noti m e v, implica che

$$
\begin{aligned}
L(m, v; x) &= p(X = x | M = m, V = v) = p(X_1 = x_1, \ldots X_n = x_n | m, v) \\
&= p(X_1 = x_1 | m, v) \cdot \ldots \cdot p(X_n = x_n | m, v) \\
&= \prod_{i=1}^{n} \exp\left(-\frac{1}{2v} (x_i - m)^2 \right) \frac{1}{\sqrt{2\pi v}} \\
&\propto \exp\left(-\frac{n}{2} \left(\frac{1}{nv} \sum_{i=1}^{n} (x_i - m)^2 + \log(v) \right) \right),
\end{aligned}
$$

dove nell'ultimo passaggio abbiamo omesso la costante moltiplicativa $(2\pi)^{n/2}$.

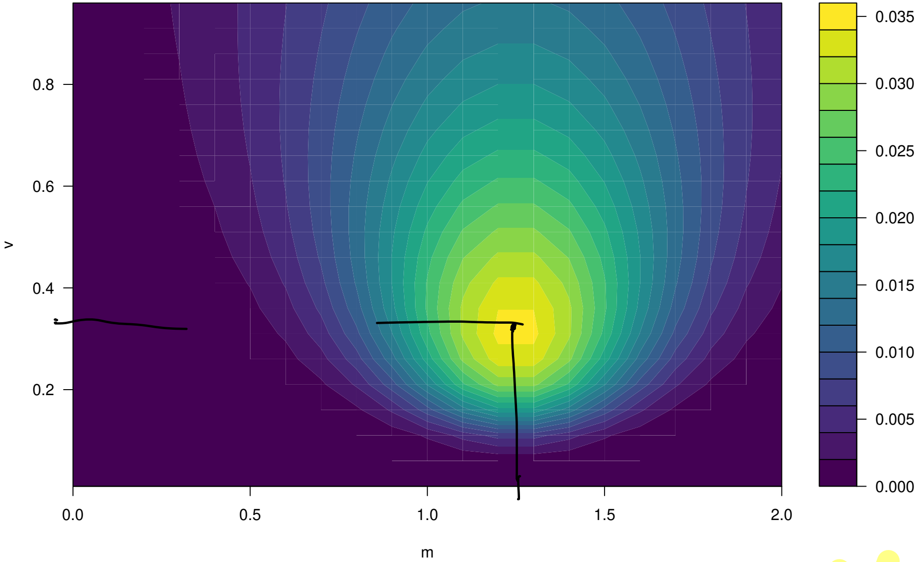


Figure 8: verosimiglianza per m e v, associata alle osservazioni $x = (1, 2, 1.5, 0.5)$.

Passando al logaritmo e cambiando segno, bisogna determinare om_{MLE} e v_{MLE} che *minimizzano* la funzione

$$(m, v) \mapsto \frac{1}{v}\left[\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2\right] + \log(v)$$

$\frac{\partial}{\partial m}$

- Si trova la condizione

$$\boxed{2\sum_{i=1}^{n}(x_i - m) = 0} \quad \text{da cui} \quad \boxed{m = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

  è la media aritmetica delle osservazioni (detta anche **media empirica** o campionaria, in inglese *sample mean*), e indicata anche brevemente con

$$\hat{X}_n = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

- se deriviamo rispetto a v, tenendo fisso m, si trova analogamente al caso della singola osservazione che

$$-\frac{1}{v^2}\frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 + \frac{1}{v} = 0, \quad \text{da cui} \quad \boxed{v = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2.}$$

dovendo massimizzare la funzione delle due variabili si trova quindi la coppia

$$\boxed{m_{\mathrm{MLE}} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i} \quad \boxed{v_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2,}$$

dove l'ultima quantità è detta anche **varianza campionaria** (*sample variance* in inglese).

mean( )

sd( )

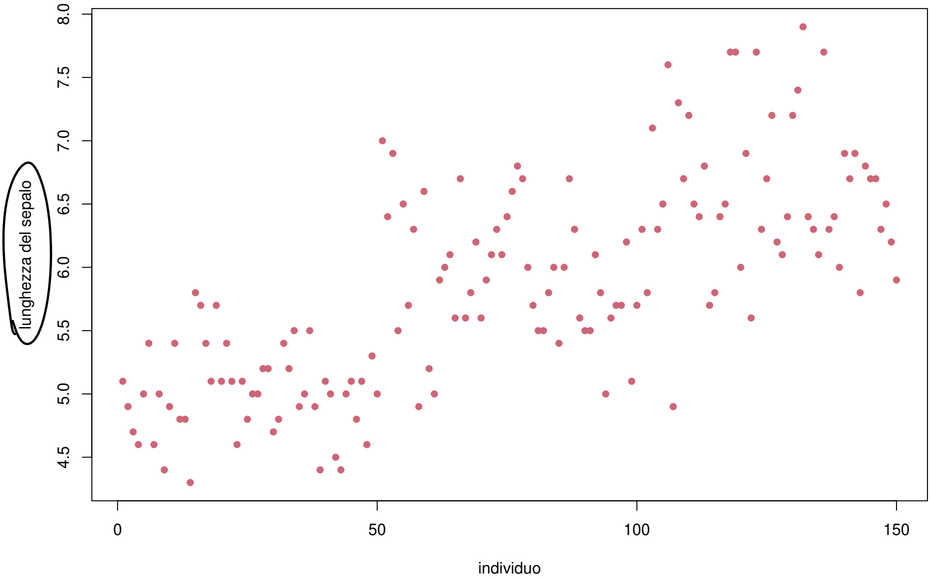
Un esempio, il dataset Iris

Consideriamo il classico dataset `iris`

Table 1: Le prime 10 righe del dataset Iris.

Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa

```
plot(iris$Sepal.Length, pch=16, col=miei_colori[2], xlab="indi
```



```r
m = mean(iris$Sepal.Length)

m
```

```
## [1] 5.843333
```

```r
sd = sd(iris$Sepal.Length)

sd
```

```
## [1] 0.8280661
```

Section 5

Stima bayesiana dei parametri: singola osservazione

Stima bayesiana per la media, varianza nota

Per applicare il metodo bayesiano, bisogna proporre delle densità *a priori* per media M e varianza V.

La rete bayesiana che rappresenta il problema con una singola osservazione X è rappresentata in figura.

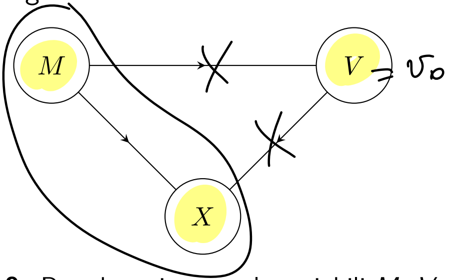


Figure 9: Rete bayesiana tra le variabili M, V e X

- Supponendo che $V = v_0$ sia costante nota, i calcoli sono semplici se si suppone che M sia una variabile gaussiana con parametri (noti) $\mathcal{N}(m_0, \sigma_m^2)$, ossia

$$p(M = m|\Omega) \propto \exp\left(-\frac{1}{2\sigma_m^2}(m - m_0)^2\right).$$

- Supponendo che $V = v_0$ sia costante nota, i calcoli sono semplici se si suppone che M sia una variabile gaussiana con parametri (noti) $\mathcal{N}(m_0, \sigma_m^2)$, ossia

$$p(M = m|\Omega) \propto \boxed{\exp\left(-\frac{1}{2\sigma_m^2}(m - m_0)^2\right)}.$$

- Avendo osservato $X = x$, dalla formula di Bayes segue che

$$\begin{aligned} p(M = m|X = x) &\propto p(M = m|\Omega)\underbrace{p(X = x|M = m, V = v_0)}_{= L(m, v_0; x)} \\ &\propto \exp\left(-\frac{1}{2}\left(\frac{(m - m_0)^2}{\underbrace{\sigma_m^2}_{\text{nota}}} + \frac{(x - m)^2}{v_0}\right)\right). \end{aligned}$$

densità gaussiana (in m)!

- Supponendo che $V = v_0$ sia costante nota, i calcoli sono semplici se si suppone che M sia una variabile gaussiana con parametri (noti) $\mathcal{N}(m_0, \sigma_m^2)$, ossia

$$p(M = m|\Omega) \propto \exp\left(-\frac{1}{2\sigma_m^2}(m - m_0)^2\right).$$

- Avendo osservato $X = x$, dalla formula di Bayes segue che

$$\begin{aligned} p(M = m|X = x) &\propto p(M = m|\Omega)p(X = x|M = m, V = v_0) \\ &\propto \exp\left(-\frac{1}{2}\left(\frac{(m - m_0)^2}{\sigma_m^2} + \frac{(x - m)^2}{v_0}\right)\right). \end{aligned}$$

- una nuova densità gaussiana con

$$\boxed{m_{|X=x} = (1 - \alpha)x + \alpha m_0,} \quad \boxed{\sigma^2_{m|X=x} = \sigma_m^2\alpha,} \le \sigma_m^2$$

dove abbiamo posto, per semplicità,

$$\alpha = \frac{1}{1 + \sigma_m^2/v_0} \in (0, 1).$$

Se $\sigma_m^2 >> v_0 \to \alpha = 0$

Se $\sigma_m^2 << v_0 \to \alpha = 1$

Stima bayesiana per la varianza, media nota

Supponiamo ora che $M = m_0$ sia costante rispetto all'informazione iniziale disponibile e introduciamo una densità a priori per la varianza $\boxed{V \geq 0}$

- densità *esponenziale inversa*, ossia $1/V$ è esponenziale con parametro $\lambda = v_0/2$,

$$\boxed{p(V = v|\Omega) \propto p(1/V = 1/v)\frac{1}{v^2}} \propto \exp\left(-\frac{v_0}{2v}\right)\frac{1}{v^2}.$$

esercizio

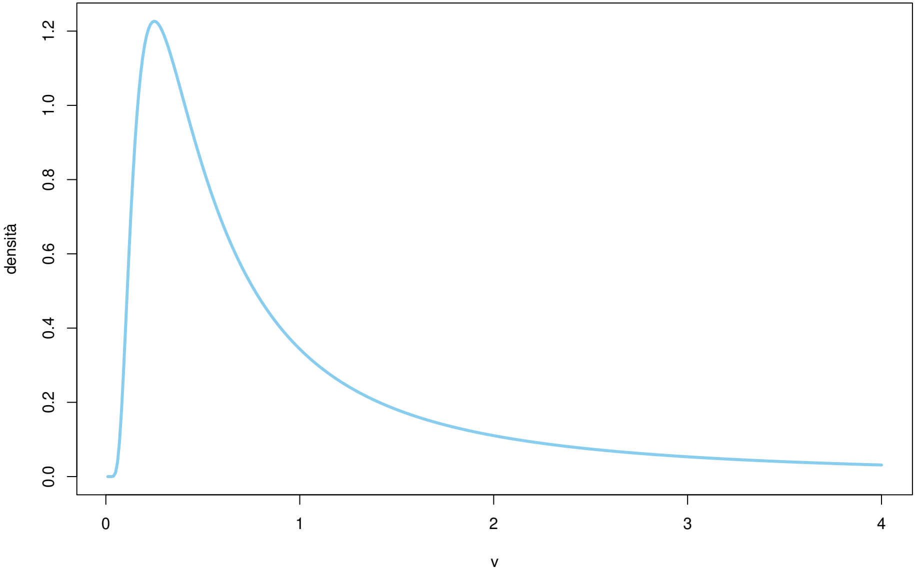


Figure 10: grafico della densità esponenziale inversa con $v_0 = 1$

Con questa scelta, avendo osservato $X = x$, la formula di Bayes è

$$\begin{aligned} p(V = v|X = x) &\propto p(V = v|\Omega)p(X = x|M = m_0, V = v) \\ &\propto \exp\left(-\frac{v_0}{2v}\right)\frac{1}{v^2}\exp\left(-\frac{1}{2v}(x - m_0)^2\right)\frac{1}{\sqrt{v}} \\ &\propto \exp\left(-\frac{(v_0 + (x - m_0)^2)}{2v}\right)\frac{1}{v^{5/2}} \end{aligned}$$

- nuovo parametro (al posto di v_0)

$$v_{|X=x} = v_0 + (x - m_0)^2.$$

Con questa scelta, avendo osservato $X = x$, la formula di Bayes è

$$\begin{aligned} p(V = v|X = x) &\propto p(V = v|\Omega) p(X = x|M = m_0, V = v) \\ &\propto \exp\left(-\frac{v_0}{2v}\right)\frac{1}{v^2}\exp\left(-\frac{1}{2v}(x - m_0)^2\right)\frac{1}{\sqrt{v}} \\ &\propto \exp\left(-\frac{(v_0 + (x - m_0)^2)}{2v}\right)\frac{1}{v^{5/2}} \end{aligned}$$

- nuovo parametro (al posto di v_0)

$$v_{|X=x} = v_0 + (x - m_0)^2.$$

- inoltre al denominatore v^2 diventa $v^{5/2}$.

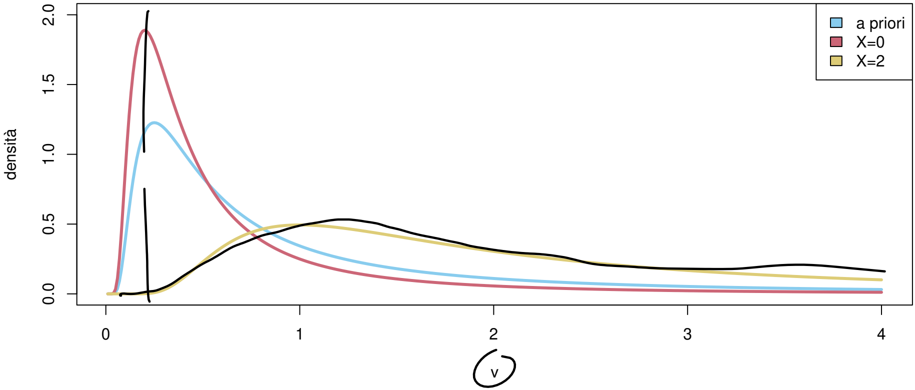


Figure 11: grafico della densità a priori per V, con parametri $v_0 = 1$, $m_0 = 0$, e della densità avendo osservato $X = 0$ (in rosso) oppure $X = 2$ (in blu)

Section 6

Stima bayesiana dei parametri: osservazioni multiple

Stima bayesiana della media, varianza nota

Generalizziamo ad n osservazioni indipendenti $(X_i)_{i=1}^n$ (tutte con gli stessi parametri). Conviene supporre che M a priori abbia una densità gaussiana di parametri $\mathcal{N}(m_0, \sigma_m^2)$, e avendo osservato $X = x = (x_i)_{i=1}^n$, si calcola

$$\begin{aligned} p(M = m|X = x) &\propto p(M = m|\Omega)p(X = x|M = m, V = v_0) \\ &\propto \exp\left(-\frac{1}{2}\left(\frac{(m-m_0)^2}{\sigma_m^2} + \sum_{i=1}^n \frac{(x_i - m)^2}{v_0}\right)\right). \end{aligned}$$

- Si trova quindi, come nel caso $n = 1$, è una nuova densità gaussiana, di cui con passaggi elementari si ricavano i parametri di media e varianza (che dipendono naturalmente dall'osservazione $X = x$)

$$m_{|X=x} = (1-\alpha)\bar{x} + \alpha m_0, \quad \sigma^2_{m|X=x} = \sigma_m^2 \alpha,$$

dove abbiamo posto

$$\alpha = \frac{1}{1 + n\sigma_m^2/v_0} \in (0, 1).$$

Stima bayesiana della varianza, valor medio noto

Supponiamo ora che $M = m_0$ sia costante nota e introduciamo come nella sezione precedente una densità a priori per V di tipo *esponenziale inversa*, dove $v_0 > 0$ è un parametro noto:

$$p(V = v|\Omega) \propto p(1/V = 1/v)\frac{1}{v^2} \propto \exp\left(-\frac{v_0}{2v}\right)\frac{1}{v^2}.$$

Usando la formula di Bayes,

$$\begin{aligned} p(V = v|X = x) &\propto p(V = v|\Omega)p(X = x|M = m_0, V = v) \\ &\propto \exp\left(-\frac{v_0}{2v}\right)\frac{1}{v^2}\exp\left(-\frac{1}{v}\sum_{i=1}^{n}(x_i - m_0)^2\right)\frac{1}{v^{n/2}} \\ &\propto \exp\left(-\frac{\left(v_0 + \sum_{i=1}^{n}(x_i - m_0)^2\right)}{2v}\right)\frac{1}{v^{(4+n)/2}}. \end{aligned}$$

- nuovo parametro (al posto di v_0) è

$$v_{|X=x} = v_0 + \sum_{i=1}^{n}(x_i - m_0)^2,$$

ma vi è anche il termine $v^{(4+n)/2}$ a denominatore, che sempre più rilevante al crescere di n.

- nuovo parametro (al posto di v_0) è

$$v_{|X=x} = v_0 + \sum_{i=1}^{n} (x_i - m_0)^2,$$

ma vi è anche il termine $v^{(4+n)/2}$ a denominatore, che sempre più rilevante al crescere di n.

- il punto di massimo della densità a posteriori (che possiamo ottenere passando al logaritmo, e imponendo la derivata nulla) è dato dall'espressione

$$\frac{v_{|X=x}}{4+n} = \frac{v_0 + \sum_{i=1}^{n}(x_i - m_0)^2}{4+n}$$

che al crescere di n è asintoticamente equivalente alla stima di massima verosimiglianza

$$v_{\mathrm{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (x_i - m_0)^2$$

(supponendo la media m_0 nota).

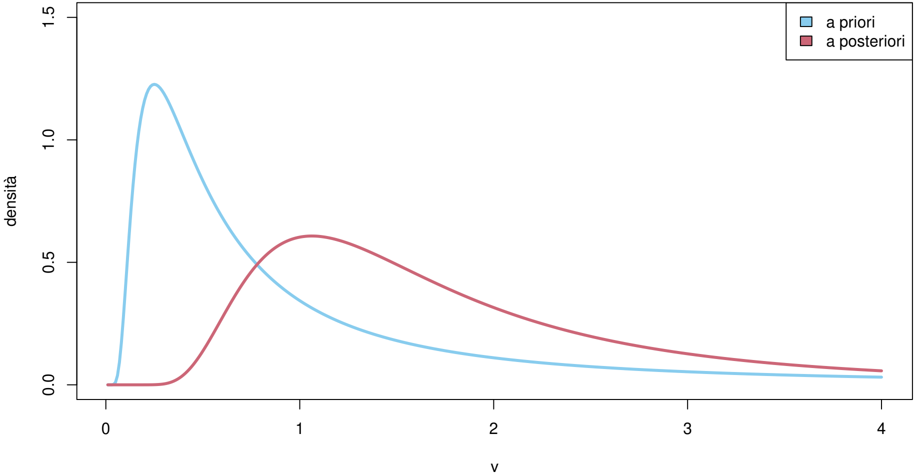


Figure 12: grafico della densità a priori per V, con parametri $v_0 = 1$, $m_0 = 0$, e della densità avendo osservato $x = (1, 2, 1.5, 0.5)$ (la stima di massima verosimiglianza è in questo caso $v_{\mathrm{MLE}} = 1$)

Stime nel caso vettoriale

I risultati della sezione precedente si possono estendere al caso vettoriale, ossia di n osservazioni di variabili aleatorie $X_1, \ldots, X_n \in \mathbb{R}^d$, tutte indipendenti tra loro e ciascuna con densità gaussiana vettoriale di parametri comuni $\mathcal{N}(m, \Sigma)$.

Verosimiglianza

$$L(m, \Sigma; x) \propto \exp\left(-\frac{1}{2}\sum_{i=1}^{n}(x_i - m)\cdot\Sigma^{-1}(x_i - m)\right)\frac{1}{(\det\Sigma)^{n/2}}.$$

- Se la varianza $\Sigma = \Sigma_0$ è nota (rispetto all'informazione prima di osservare le X_i), la stima di massima verosimiglianza per il parametro di media è la media campionaria

$$m_{\mathrm{MLE}} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(qualsiasi sia Σ_0).

- Se il parametro di media $m = m_0 \in \mathbb{R}^d$ è noto (rispetto all'informazione a priori), allora la stima di massima verosimiglianza per la covarianza è

$$\Sigma_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_0)(x_i - m_0)^T,$$

dove T indica l'operazione di trasposizione (quindi il prodotto righe per colonne risulta in una matrice $d \times d$); più esplicitamente, la stima della covarianza tra la componente j e k è

$$(\Sigma_{\mathrm{MLE}})_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_i - m_0)_j(x_i - m_0)_k.$$

La stima (congiunta) di massima verosimiglianza per (m, Σ) è data dalla media e dalla covarianza campionarie:

$$m_{\mathrm{MLE}} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \Sigma_{\mathrm{MLE}} = \frac{1}{n}\sum_{i=1}(x_i - \bar{x})(x_i - \bar{x})^T.$$

- Osserviamo che Σ_{MLE} è una matrice simmetrica e semi-definita positiva. La si può anche interpretare come la matrice di covarianza della variabile aleatoria vettoriale che sceglie uno degli n valori osservati con probabilità uniforme discreta.

Coefficiente di correlazione campionario

È utile anche considerare la matrice delle correlazioni, in cui al posto delle covarianze è calcolato il coefficiente di correlazione campionario,

$$\bar{\rho}_{jk} = \frac{\Sigma_{jk}}{\sqrt{\Sigma_{jj}\Sigma_{kk}}},$$

che è sempre compreso tra -1 e 1 (segue dal fatto che la matrice $\Sigma = \Sigma_{\mathrm{MLE}}$ è semi-definita positiva). Il comando in questo caso è `cor()`.

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Wi
## Sepal.Length         1.00       -0.12         0.87        0
## Sepal.Width         -0.12        1.00        -0.43       -0
## Petal.Length         0.87       -0.43         1.00        0
## Petal.Width          0.82       -0.37         0.96        1
```

Correlogramma

Per visualizzare la correlazione si può usare un *correlogramma*.

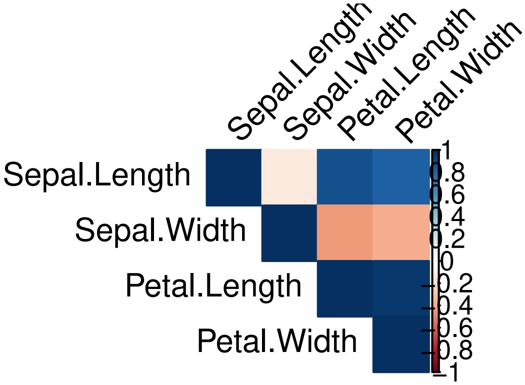


Figure 13: Correlogramma del dataset Iris

Diagramma a dispersione

Il comando `plot()` applicato direttamente ai dati fornisce invece un diagramma a dispersione di tutte le possibili coppie.