

Probabilità e Processi Stocastici (455AA)

Lezione 11

Dario Trevisan – <https://web.dm.unipi.it/trevisan>

27/10/2025

Vettori gaussiani

Richiami: variabili gaussiane reali

► $X \sim \mathcal{N}(m, \sigma^2)$

Richiami: variabili gaussiane reali

- ▶ $X \sim \mathcal{N}(m, \sigma^2)$
- ▶ $p(X = x) \propto \exp\left(-\frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2\right) \frac{1}{\sigma}$

Richiami: variabili gaussiane reali

- ▶ $X \sim \mathcal{N}(m, \sigma^2)$
- ▶ $p(X = x) \propto \exp\left(-\frac{1}{2} \left(\frac{x-m}{\sigma}\right)^2\right) \frac{1}{\sigma}$
- ▶ $\mathbb{E}[X] = m, \text{Var}(X) = \sigma^2$

Richiami: variabili gaussiane reali

- ▶ $X \sim \mathcal{N}(m, \sigma^2)$
- ▶ $p(X = x) \propto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right) \frac{1}{\sigma}$
- ▶ $\mathbb{E}[X] = m, \text{Var}(X) = \sigma^2$
- ▶ $\text{MGF}_X(t) = \exp\left(mt + \frac{1}{2}(\sigma t)^2\right)$

Richiami: variabili gaussiane reali

- ▶ $X \sim \mathcal{N}(m, \sigma^2)$
- ▶ $p(X = x) \propto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right) \frac{1}{\sigma}$
- ▶ $\mathbb{E}[X] = m, \text{Var}(X) = \sigma^2$
- ▶ $\text{MGF}_X(t) = \exp\left(mt + \frac{1}{2}(\sigma t)^2\right)$
- ▶ $\lambda X + c \sim \mathcal{N}(\lambda m + c, \lambda^2 \sigma^2)$

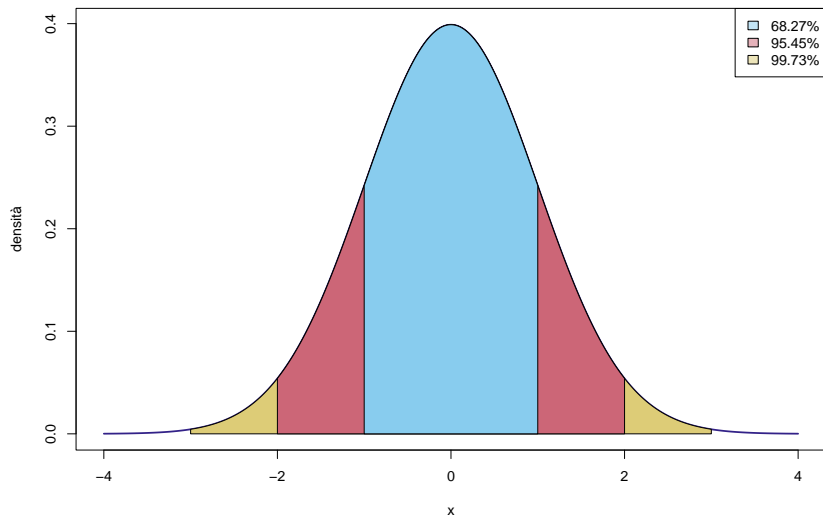
Richiami: variabili gaussiane reali

- ▶ $X \sim \mathcal{N}(m, \sigma^2)$
- ▶ $p(X = x) \propto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right) \frac{1}{\sigma}$
- ▶ $\mathbb{E}[X] = m, \text{Var}(X) = \sigma^2$
- ▶ $\text{MGF}_X(t) = \exp\left(mt + \frac{1}{2}(\sigma t)^2\right)$
- ▶ $\lambda X + c \sim \mathcal{N}(\lambda m + c, \lambda^2 \sigma^2)$
- ▶ $\text{CDF}_X(t) = \text{pnorm}(t, \text{mean} = , \text{sd} =)$

Richiami: variabili gaussiane reali

- ▶ $X \sim \mathcal{N}(m, \sigma^2)$
- ▶ $p(X = x) \propto \exp\left(-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right) \frac{1}{\sigma}$
- ▶ $\mathbb{E}[X] = m, \text{Var}(X) = \sigma^2$
- ▶ $\text{MGF}_X(t) = \exp\left(mt + \frac{1}{2}(\sigma t)^2\right)$
- ▶ $\lambda X + c \sim \mathcal{N}(\lambda m + c, \lambda^2 \sigma^2)$
- ▶ $\text{CDF}_X(t) = \text{pnorm}(t, \text{mean} = , \text{sd} =)$
- ▶ $q_X(\alpha) = \text{qnorm}(\alpha, \text{mean} = , \text{sd} =)$

Regola dei 68-95-99.7



Il caso vettoriale: definizione veloce

Una variabile aleatoria $X \in \mathbb{R}^d$ ha densità continua gaussiana se

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

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

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

Il caso vettoriale: definizione veloce

Una variabile aleatoria $X \in \mathbb{R}^d$ ha densità continua gaussiana se

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

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

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

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

Il caso vettoriale: definizione veloce

Una variabile aleatoria $X \in \mathbb{R}^d$ ha densità continua gaussiana se

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

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

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

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

- ▶ possiamo supporre a simmetrica e definita positiva.

Intepretazione dei parametri

Se $X \in \mathbb{R}^d$ una variabile con densità gaussiana:

- ▶ la matrice delle covarianze Σ_X è definita positiva (invertibile)

Interpretazione dei parametri

Se $X \in \mathbb{R}^d$ una variabile con densità gaussiana:

- ▶ la matrice delle covarianze Σ_X è definita positiva (invertibile)
- ▶ vale

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

Interpretazione dei parametri

Se $X \in \mathbb{R}^d$ una variabile con densità gaussiana:

- ▶ la matrice delle covarianze Σ_X è definita positiva (invertibile)
- ▶ vale

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

- ▶ ossia

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

Parametrizzazione usuale

Una variabile $X \in \mathbb{R}^d$ ha densità continua gaussiana di vettore dei valori medi $m \in \mathbb{R}^d$ e matrice delle covarianze $\Sigma > 0$

$$\mathcal{N}(m, \Sigma)$$

se vale

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

► Più esplicitamente:

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

Heatmap

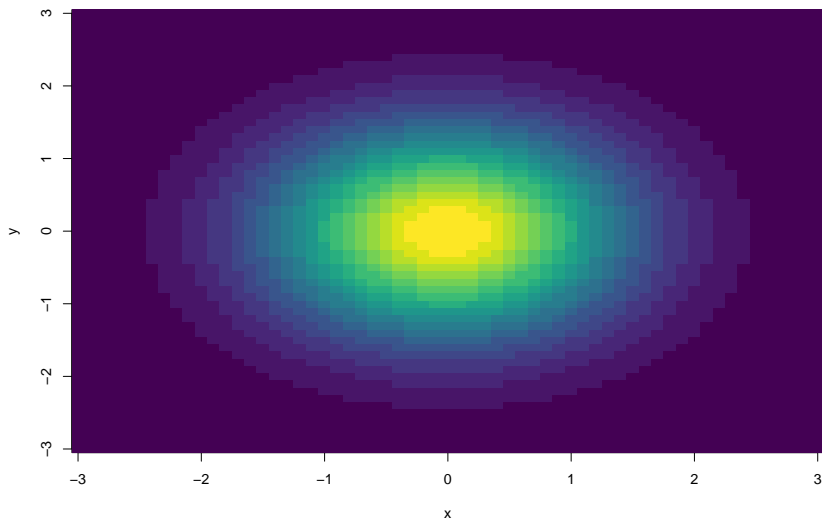


Figure 1: visualizzazione della densità gaussiana vettoriale per $d = 2$, $m = 0$, e $\Sigma = Id$

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

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

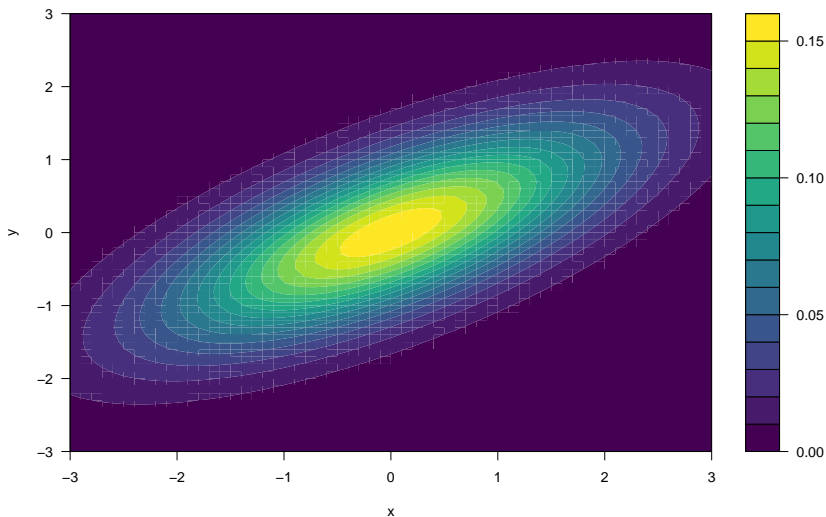


Figura 2: visualizzazione della densità gaussiana vettoriale per $d = 2$

Curve di livello

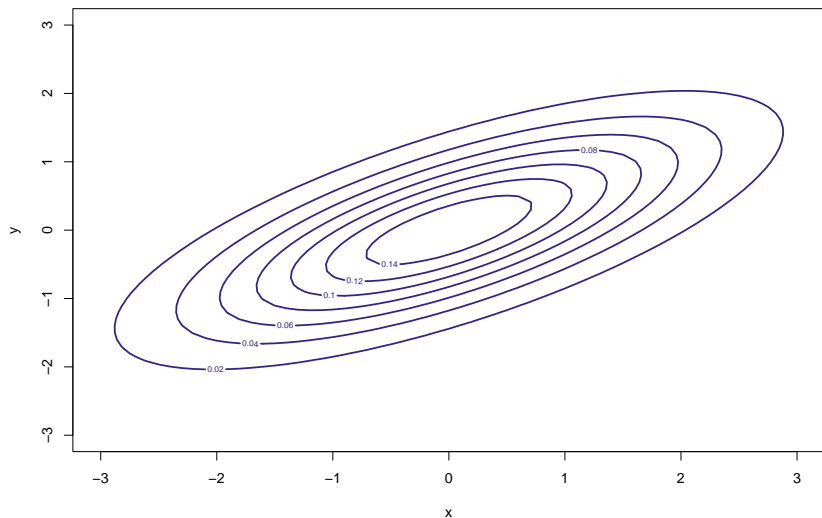


Figure 3: visualizzazione della medesima densità gaussiana ma tramite curve di livello

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).

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 $\mathcal{N}(v \cdot m, v \cdot \Sigma v)$.

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 $\mathcal{N}(v \cdot m, v \cdot \Sigma v)$.
- **Esempio IMPORTANTE**, se le marginali $X = (X_1, X_2)$ sono non correlate, posto $v = (1, 1)$, si trova che la somma $X_1 + X_2$ ha densità $\mathcal{N}(m_1 + m_2, \sigma_{X_1}^2 + \sigma_{X_2}^2)$.

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

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

detta anche **gaussiana standard** vettoriale.

► La densità esplicita è

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

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

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

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, \dots, Z_d sono *indipendenti*, oltre ad essere non correlate.

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.

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 $\text{Cov}(X_i, Y_j) = 0$ per ogni $i \in \{1, \dots, d\}$, $j \in \{1, \dots, k\}$, allora X e Y sono indipendenti.

MGF e funzione caratteristica

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

► la MGF vale

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

MGF e funzione caratteristica

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

- ▶ la MGF vale

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

- ▶ la funzione caratteristica vale

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

Il caso degenere

Se Σ è solo **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'interpretazione in tali casi è che la densità continua si concentra in un sottospazio affine di dimensione più bassa dello spazio ambiente \mathbb{R}^d .

Il caso degenere

Se Σ è solo **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'interpretazione 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

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

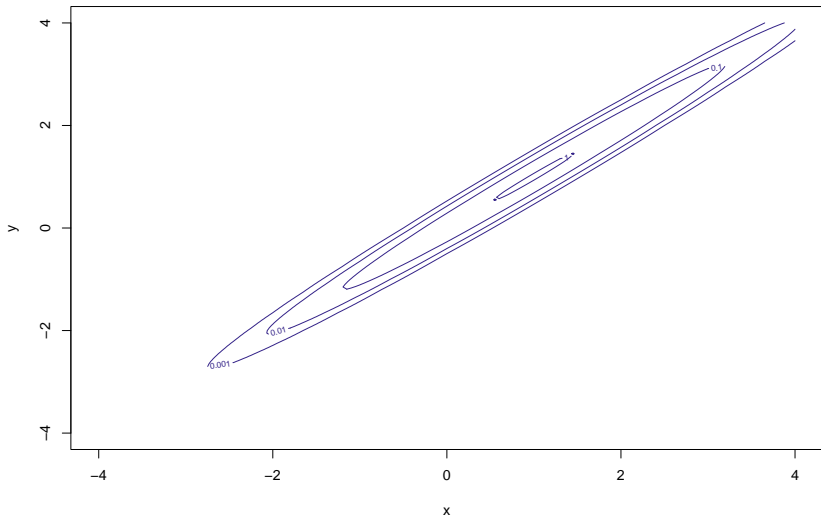


Figure 4: rappresentazione della densità gaussiana vettoriale nel caso vettoriale $d = 2$, $m = (1, 1)$ e Σ definita sopra

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 semplice, ma meno flessibile, 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

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

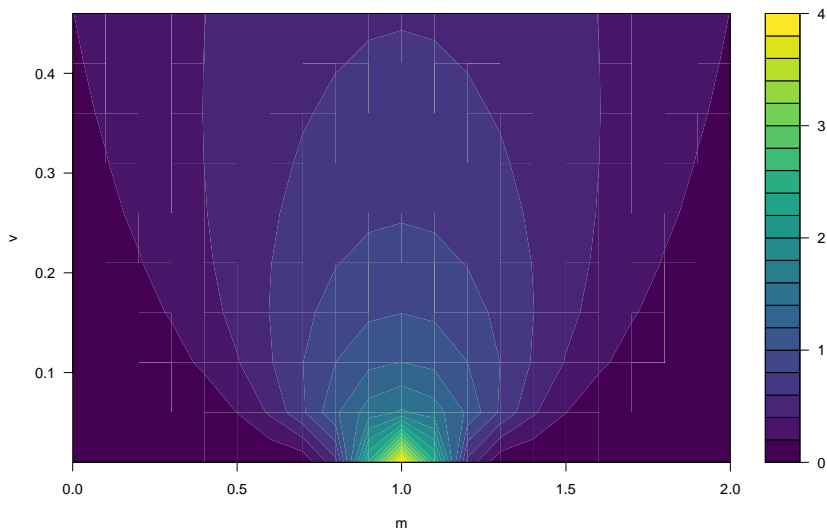


Figure 5: verosimiglianza per M e V avendo osservato $X = 1$.

Stima di massima verosimiglianza

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

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

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

Stima di massima verosimiglianza

Vediamo che L è massima per $m = x$ ($x = 1$ esempio 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 $v > 0$,

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

Stima di massima verosimiglianza

Vediamo che L è massima per $m = x$ ($x = 1$ esempio 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

$$-\frac{1}{v^2}(x - m)^2 + \frac{1}{v} = 0, \quad \text{da cui } v = (x - m)^2.$$

Stima di massima verosimiglianza

Vediamo che L è massima per $m = x$ ($x = 1$ esempio 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 $m_{\text{MLE}} = x$, $v_{\text{MLE}} = 0$.

Con calcoli analoghi si ottiene anche che

1. se varianza $V = v_0 > 0$ è nota (a priori), allora la stima di massima verosimiglianza per la media è

$$m_{\text{MLE}} = \bar{x}$$

(qualsiasi sia v_0).

Con calcoli analoghi si ottiene anche che

1. se varianza $V = v_0 > 0$ è nota (a priori), allora la stima di massima verosimiglianza per la media è

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

(qualsiasi sia v_0).

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

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

Stima dei parametri da osservazioni indipendenti

- ▶ estendiamo i risultati di osservazioni $(X_i)_{i=1}^n$, tutte gaussiane *indipendenti* ma con gli stessi parametri (il termine statistico è un *campione* di taglia n).

Stima dei parametri da osservazioni indipendenti

- ▶ estendiamo i risultati di osservazioni $(X_i)_{i=1}^n$, tutte gaussiane *indipendenti* ma con gli stessi parametri (il termine statistico è un *campione* di taglia n).
- ▶ Per semplificare la notazione poniamo $X = (X_i)_{i=1}^n$ a valori in \mathbb{R}^n , con

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

Stima dei parametri da osservazioni indipendenti

- ▶ estendiamo i risultati di osservazioni $(X_i)_{i=1}^n$, tutte gaussiane *indipendenti* ma con gli stessi parametri (il termine statistico è un *campione* di taglia n).
- ▶ Per semplificare la notazione poniamo $X = (X_i)_{i=1}^n$ a valori in \mathbb{R}^n , con

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

- ▶ Introduciamo variabili M per la media, V per la varianza (la seconda a valori positivi).

La rete bayesiana associata è

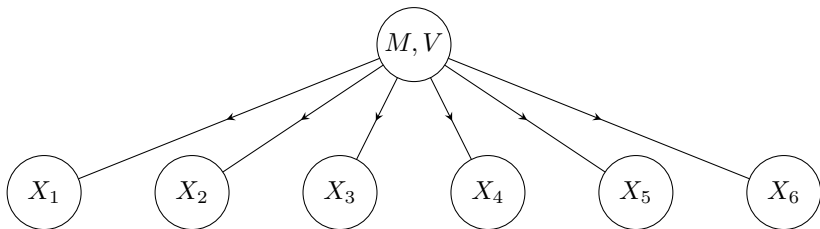


Figure 6: Rete bayesiana tra le variabili (M, V) e le $(X_i)_{i=1}^n$ per $n = 6$

Stima di massima verosimiglianza

L'ipotesi di indipendenza, noti $M = m$ e $V = v$, implica

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

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

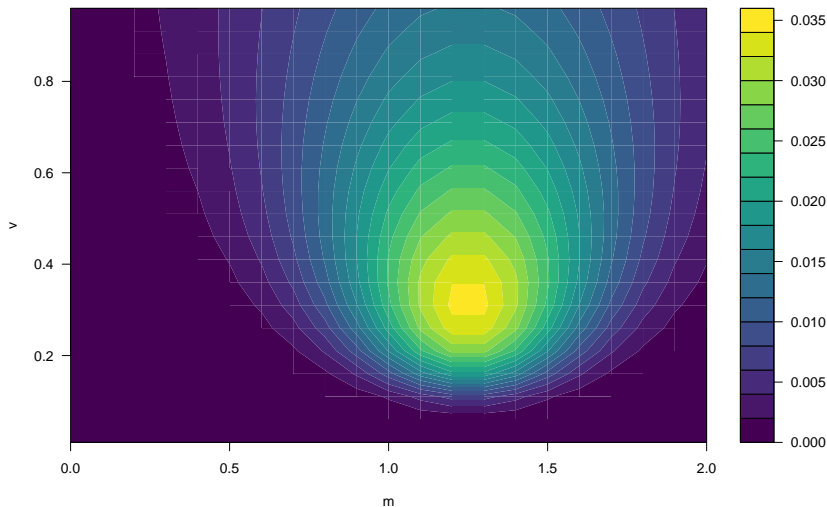


Figure 7: verosimiglianza per m e v , associata alle osservazioni $x = (1, 2, 1.5, 0.5)$.

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

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

► Si trova la condizione

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

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

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

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

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

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

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

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

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

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

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

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

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

- ▶ **Attenzione!** In statistica (R) si usa piuttosto la **varianza campionaria corretta**

$$\text{var}(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad \text{sd}(x) = \sqrt{\text{var}(x)}$$

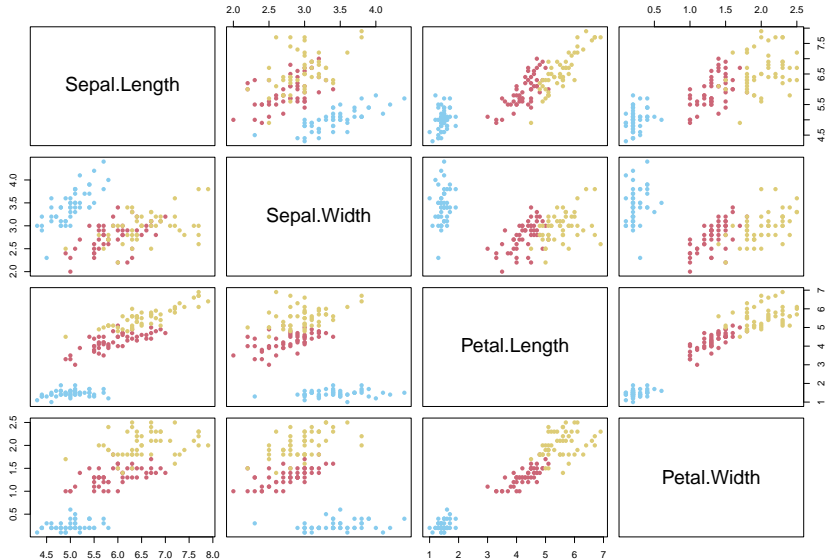
Un esempio, il dataset Iris

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[1:4], pch=16, col=miei_colori[as.numeric(iris$Species)]
```



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

```
m
```

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

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

```
sd
```

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

```
summary(iris)
```

##	Sepal.Length	Sepal.Width	Petal.Length	Petal
##	Min. :4.300	Min. :2.000	Min. :1.000	Min.
##	1st Qu.:5.100	1st Qu.:2.800	1st Qu.:1.600	1st Qu.
##	Median :5.800	Median :3.000	Median :4.350	Median
##	Mean :5.843	Mean :3.057	Mean :3.758	Mean
##	3rd Qu.:6.400	3rd Qu.:3.300	3rd Qu.:5.100	3rd Qu.
##	Max. :7.900	Max. :4.400	Max. :6.900	Max.
##	Species			

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.

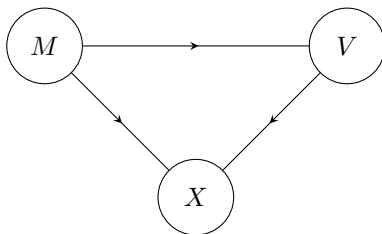


Figure 8: Rete bayesiana tra le variabili M , V e X

- 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

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

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

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

- 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

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

dove abbiamo posto, per semplicità,

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

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 V .

- densità *esponenziale inversa*, ossia $1/V$ è esponenziale con parametro $\lambda = \nu_0/2$,

$$p(V = v|\Omega) \propto p(1/V = 1/v) \frac{1}{v^2} \propto \exp\left(-\frac{\nu_0}{2v}\right) \frac{1}{v^2}.$$

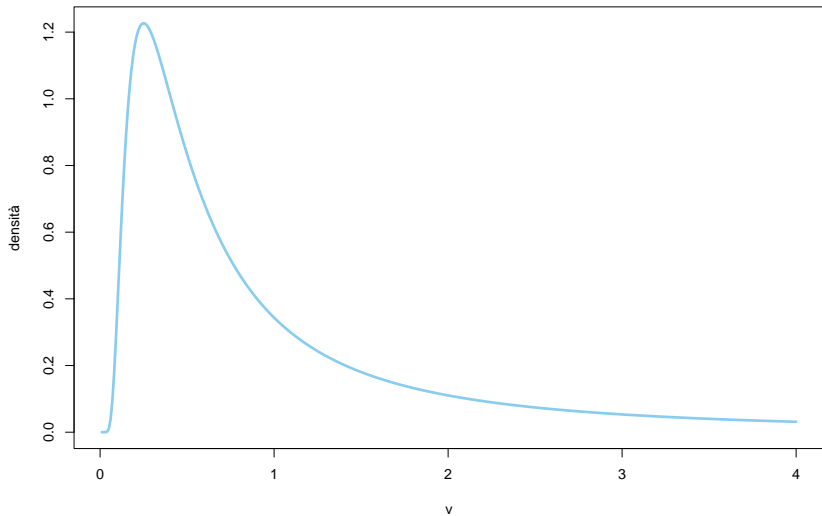


Figure 9: grafico della densità esponenziale inversa con $v_0 = 1$

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 $v^{5/2}$.

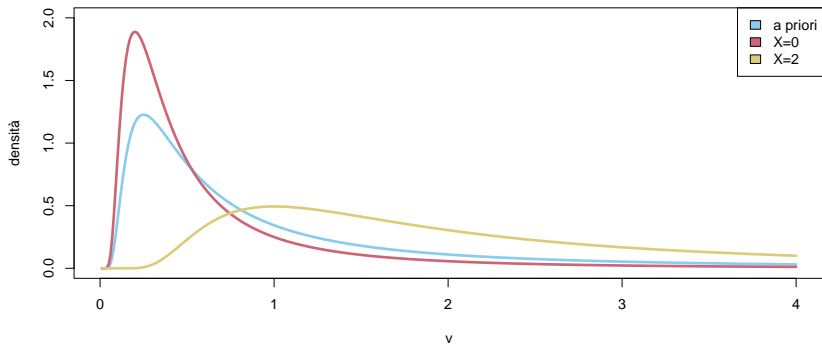


Figure 10: grafico della densità a priori per V , con parametri $v_0 = 1$, $m_0 = 0$, e della densità avendo osservato $X = 0$ (in rosso) oppure $X = 2$ (in blu)

Stima bayesiana dei parametri: osservazioni multiple

Stima bayesiana della media, varianza nota

Generalizziamo a un campione di taglia n : $(X_i)_{i=1}^n$.

- Supponiamo che M a priori abbia una densità gaussiana di parametri $\mathcal{N}(m_0, \sigma_m^2)$.

Stima bayesiana della media, varianza nota

Generalizziamo a un campione di taglia n : $(X_i)_{i=1}^n$.

- ▶ Supponiamo che M a priori abbia una densità gaussiana di parametri $\mathcal{N}(m_0, \sigma_m^2)$.
- ▶ Osservato $X = x = (x_i)_{i=1}^n$, troviamo

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

Stima bayesiana della media, varianza nota

Generalizziamo a un campione di taglia n : $(X_i)_{i=1}^n$.

- ▶ Supponiamo che M a priori abbia una densità gaussiana di parametri $\mathcal{N}(m_0, \sigma_m^2)$.
- ▶ Osservato $X = x = (x_i)_{i=1}^n$, troviamo

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

- ▶ M a posteriori ha densità gaussiana con parametri

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

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

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

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

► Se $\sigma_m \ll v_0/n \rightarrow \alpha \approx 1$, $m_{|X=x} \approx m_0$, $\sigma_{m|X=x}^2 \approx \sigma_m^2$

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

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

- ▶ Se $\sigma_m \ll v_0/n \rightarrow \alpha \approx 1$, $m_{|X=x} \approx m_0$, $\sigma_{m|X=x}^2 \approx \sigma_m^2$
- ▶ Se $\sigma_m \gg v_0/n$, $\rightarrow \alpha \approx 0$, $m_{|X=x} \approx \bar{x}$, $\sigma_{m|X=x}^2 \approx v_0/n$

Intervallo di credibilità (fiducia bayesiano)

- Se $V = v_0$ è nota e $p(M = m) \propto 1$ (ossia $\sigma_m^2 \approx \infty$):

$$P\left(|M - \bar{x}| \leq z_{1-\alpha} \sqrt{\frac{v_0}{n}} \middle| X = x\right) = 1 - \alpha$$

dove $z_{1-\alpha} = q_{1-\alpha/2} = \text{qnorm}(1-\alpha/2)$

$$P(|M - 5| \leq 0.98 | X = x) = 95\%$$

Intervallo di credibilità (fiducia bayesiano)

- ▶ Se $V = v_0$ è nota e $p(M = m) \propto 1$ (ossia $\sigma_m^2 \approx \infty$):

$$P\left(|M - \bar{x}| \leq z_{1-\alpha} \sqrt{\frac{v_0}{n}} \middle| X = x\right) = 1 - \alpha$$

dove $z_{1-\alpha} = q_{1-\alpha/2} = \text{qnorm}(1-\alpha/2)$

- ▶ $z_{.95} = 1.96$, $z_{.99} = 2.575$, $z_{.999} = 3.29$

$$P(|M - 5| \leq 0.98 | X = x) = 95\%$$

Intervallo di credibilità (fiducia bayesiano)

- ▶ Se $V = v_0$ è nota e $p(M = m) \propto 1$ (ossia $\sigma_m^2 \approx \infty$):

$$P\left(|M - \bar{x}| \leq z_{1-\alpha} \sqrt{\frac{v_0}{n}} \middle| X = x\right) = 1 - \alpha$$

dove $z_{1-\alpha} = q_{1-\alpha/2} = \text{qnorm}(1-\alpha/2)$

- ▶ $z_{.95} = 1.96$, $z_{.99} = 2.575$, $z_{.999} = 3.29$
- ▶ **Esempio:** se $n = 16$, $\bar{x} = 5$, $v_0 = 4$ (noto!) e $\alpha = 5\%$:

$$P(|M - 5| \leq 0.98 | X = x) = 95\%$$

Stima bayesiana: media e varianza non note

Vogliamo stimare M , ma ora anche V non è nota.

- ▶ Introduciamo una densità a priori *congiunta* (impropria)

$$p(M = m, V = v) \propto \frac{1}{v}$$

Stima bayesiana: media e varianza non note

Vogliamo stimare M , ma ora anche V non è nota.

- ▶ Introduciamo una densità a priori *congiunta* (impropria)

$$p(M = m, V = v) \propto \frac{1}{v}$$

- ▶ Si può modificare per introdurre informazioni più precise su M e V .

Stima bayesiana: media e varianza non note

Vogliamo stimare M , ma ora anche V non è nota.

- ▶ Introduciamo una densità a priori *congiunta* (impropria)

$$p(M = m, V = v) \propto \frac{1}{v}$$

- ▶ Si può modificare per introdurre informazioni più precise su M e V .
- ▶ La densità (congiunta) a posteriori è

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

Stima bayesiana: media e varianza non note

Vogliamo stimare M , ma ora anche V non è nota.

- ▶ Introduciamo una densità a priori *congiunta* (impropria)

$$p(M = m, V = v) \propto \frac{1}{v}$$

- ▶ Si può modificare per introdurre informazioni più precise su M e V .
- ▶ La densità (congiunta) a posteriori è

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

- ▶ Possiamo ottenere le densità marginali integrando.

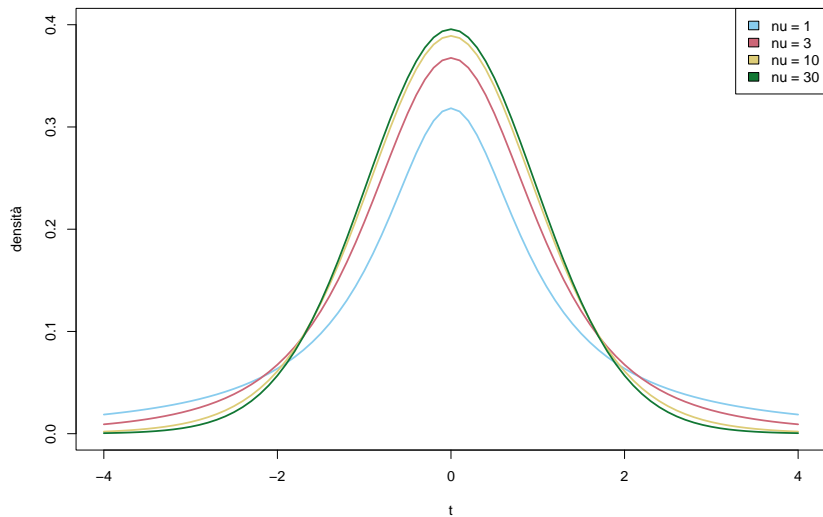
Densità marginale della media (t di Student)

$$\begin{aligned}p(M = m|X = x) &= \int_0^\infty p(M = m, V = v|X = x) dv \\&= \int_0^\infty \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - m)^2}{v}\right) \frac{1}{v^{n/2}} \cdot \frac{dv}{v} \\&= [z = \sum_i (x_i - m)^2 / v] = \int_0^\infty e^{-z/2} z^{n/2-1} dz \left(\sum_{i=1}^n (x_i - m)^2\right)^{-n/2} \\&\propto \left(1 + \frac{1}{n-1} \left(\frac{m - \bar{x}}{\text{sd}(x)/\sqrt{n}}\right)^2\right)^{-n/2}\end{aligned}$$

► $\frac{M - \bar{x}}{\text{sd}(x)/\sqrt{n}}$ ha densità continua **t di Student** con $n - 1$ gradi di libertà

$$p(T_\nu = t) \propto \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}$$

Densità t di Student



Intervallo di credibilità

Dato un campione di taglia n con media e varianza *ignote*, per ogni $\alpha \in (0, 1)$ si ha

$$P\left(|M - \bar{x}| \leq t_{1-\alpha, n-1} \frac{\text{sd}(x)}{\sqrt{n}} \mid X = x\right) = 1 - \alpha$$

dove $t_{1-\alpha, n-1} = \text{qt}(1-\alpha/2, \text{df}=n-1)$.

► **Esempio:** se $n = 16$, $\bar{x} = 5$, $\text{var}(x) = 4$ e $\alpha = 5\%$:

$$P(|M - 5| \leq 1.06 \mid X = x) = 95\%$$

(confrontare con il caso $v_0 = 4$ noto:

$$P(|M - 5| \leq 0.98 \mid X = x) = 95\%)$$

Densità marginale della varianza

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

► Integrando rispetto ad m , troviamo

$$\begin{aligned} p(V = v | X = x) &= v^{-(n+1)/2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - m)^2}{v}\right) \frac{dm}{\sqrt{v}} \\ &= v^{-(n+1)/2} \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2v}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{n(m - \bar{x})^2}{2v}\right) \frac{dm}{\sqrt{v}} \\ &\propto \exp\left(-\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2v}\right) v^{-(n+1)/2}. \end{aligned}$$

Stime nel caso vettoriale

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

► Verosimiglianza

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

Stime nel caso vettoriale

I risultati si possono estendere al caso vettoriale: n osservazioni di variabili aleatorie $X_1, \dots, X_n \in \mathbb{R}^d$, tutte indipendenti, 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 $\Sigma = \Sigma_0$ è nota (a priori), la stima di massima verosimiglianza per il parametro di media è la media campionaria

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

(qualsiasi sia Σ_0).

- Se la media $m = m_0 \in \mathbb{R}^d$ è nota (a priori), allora la stima di massima verosimiglianza per la covarianza è

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

dove T indica l'operazione di trasposizione (quindi il prodotto righe per colonne risulta in una matrice $d \times d$)

- ▶ Se la media $m = m_0 \in \mathbb{R}^d$ è nota (a priori), allora la stima di massima verosimiglianza per la covarianza è

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

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

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

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

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

- Σ_{MLE} è una matrice simmetrica e semi-definita positiva.
Intepretazione: è 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 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 $\Sigma = \Sigma_{MLE}$ è semi-definita positiva). Il comando in questo caso è `cor()`.

##	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
## Sepal.Length	1.00	-0.12	0.87	
## Sepal.Width	-0.12	1.00	-0.43	
## Petal.Length	0.87	-0.43	1.00	
## Petal.Width	0.82	-0.37	0.96	

Correlogramma

Per visualizzare la correlazione si può usare un *correlogramma* (heat map).

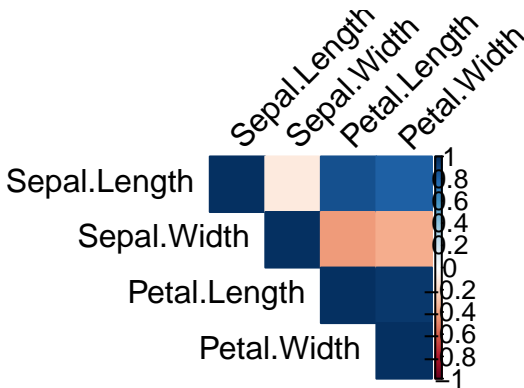


Figure 11: Correlogramma del dataset Iris