

Probabilità e Processi Stocastici (455AA)

Lezione 13

Dario Trevisan

7/11/2024

Section 1

Intepretazione probabilistica del metodo dei minimi quadrati

Metodo dei minimi quadrati e MLE

Intepretiamo il metodo come una stima di massima verosimiglianza sotto opportune ipotesi (**residui gaussiani indipendenti**). Supponiamo

- $X \in E$, $Y \in E = \mathbb{R}^{d'}$, $U \in \mathbb{R}^k$ tali che

$$Y = g(X; U) + W,$$

↑
parametro

Metodo dei minimi quadrati e MLE

Intepretiamo il metodo come una stima di massima verosimiglianza sotto opportune ipotesi (residui gaussiani indipendenti). Supponiamo

- $X \in E$, $Y \in E = \mathbb{R}^{d'}$, $U \in \mathbb{R}^k$ tali che

$$Y = g(X; U) + W,$$

- W è una variabile (il residuo) con densità gaussiana vettoriale $\mathcal{N}(0, \nu I_d)$ (dove $\nu > 0$ è un parametro).

Metodo dei minimi quadrati e MLE

Intepretiamo il metodo come una stima di massima verosimiglianza sotto opportune ipotesi (residui gaussiani indipendenti). Supponiamo

- $X \in E$, $Y \in E = \mathbb{R}^{d'}$, $U \in \mathbb{R}^k$ tali che

$$Y = g(X; U) + W,$$

- W è una variabile (il residuo) con densità gaussiana vettoriale $\mathcal{N}(0, \nu I_d)$ (dove $\nu > 0$ è un parametro).
- Le variabili X , U e W siano tra loro indipendenti.

Metodo dei minimi quadrati e MLE

Intepretiamo il metodo come una stima di massima verosimiglianza sotto opportune ipotesi (residui gaussiani indipendenti). Supponiamo

- $X \in E$, $Y \in E = \mathbb{R}^{d'}$, $U \in \mathbb{R}^k$ tali che

$$Y = g(X; U) + W,$$

- W è una variabile (il residuo) con densità gaussiana vettoriale $\mathcal{N}(0, \nu Id)$ (dove $\nu > 0$ è un parametro).
- Le variabili X , U e W siano tra loro indipendenti.
- Anche senza ipotesi sulla densità a priori di X , scriviamo la verosimiglianza come segue:

$$Y = g(x, U) + W$$

$$\begin{aligned}
 L(u; x, y) &= p(X = x, Y = y | U = u) \\
 &= p(Y = y | U = u, X = x) p(X = x | U = u) \quad \text{indep} \\
 &= p(Y - g(x; u) = y - g(x, u) | U = u, X = x) p(X = x) \\
 &= p(W = y - g(x, u) | U = u, X = x) p(X = x) \\
 &= \exp\left(-\frac{1}{2v} |y - g(x, u)|^2\right) \frac{1}{\sqrt{2\pi v}} p(X = x).
 \end{aligned}$$

densità gaussiana

- La stima di massima verosimiglianza per U equivale a minimizzare

$$u \mapsto |y - g(x, u)|^2,$$

ossia il minimo residuo quadratico.

Se si dispone di n variabili (X_i, W_i) tutte indipendenti tra loro (e dalla U) tali che

$$Y_i = g(X_i; U) + W_i,$$

allora la verosimiglianza di U associata alle osservazioni $x = (x_i)_{i=1}^n$, $y = (y_i)_{i=1}^n$,

$$\begin{aligned} L(u; x, y) &= p(X = x, Y = y | U = u) \\ &= \exp\left(-\frac{1}{2v} \sum_{i=1}^n |y_i - g(x_i, u)|^2\right) \frac{1}{\sqrt{(2\pi v)^n}} \prod_{i=1}^n p(X_i = x_i). \end{aligned}$$

- La stima di massima verosimiglianza consiste nel minimizzare

$$u \mapsto \sum_{i=1}^n |y_i - g(x_i, u)|^2,$$

$$\boxed{u_{KLE} = u_{OLS}}$$

Se si dispone di n variabili (X_i, W_i) tutte indipendenti tra loro (e dalla U) tali che

$$Y_i = g(X_i; U) + W_i,$$

allora la verosimiglianza di U associata alle osservazioni $x = (x_i)_{i=1}^n$, $y = (y_i)_{i=1}^n$,

$$\begin{aligned} L(u; x, y) &= p(X = x, Y = y | U = u) \\ &= \exp\left(-\frac{1}{2v} \sum_{i=1}^n |y_i - g(x_i, u)|^2\right) \frac{1}{\sqrt{(2\pi v)^n}} \prod_{i=1}^n p(X_i = x_i). \end{aligned}$$

- La stima di massima verosimiglianza consiste nel minimizzare

$$u \mapsto \sum_{i=1}^n |y_i - g(x_i, u)|^2,$$

- Massimizzando anche in v si ottiene anche l'errore quadratico medio

$$v_{\text{MLE}} = \text{MSE} = \frac{1}{n} \sum_{i=1}^n |y_i - g(x_i; u)|^2.$$

Approccio bayesiano

La derivazione del metodo come MLE suggerisce un approccio bayesiano per raffinare il metodo.

- Supponiamo che sia noto a priori che U non si discosti troppo da un parametro noto u_0 , ad esempio con una variabilità dell'ordine di $\sigma_u > 0$ (lungo ciascuna componente) e che le componenti di U siano indipendenti tra loro.

$$U \sim \mathcal{N}(u_0, \sigma_u^2 \text{Id})$$

Approccio bayesiano

La derivazione del metodo come MLE suggerisce un approccio bayesiano per raffinare il metodo.

- Supponiamo che sia noto a priori che U non si discosti troppo da un parametro noto u_0 , ad esempio con una variabilità dell'ordine di $\sigma_u > 0$ (lungo ciascuna componente) e che le componenti di U siano indipendenti tra loro.
- Si pone U a priori $\mathcal{N}(u_0, \sigma_u^2 Id)$, e dalla formula di Bayes

$$\begin{aligned}
 p(U = u | X_i = x_i, Y_i = y_i, \forall i = 1, \dots, n) \\
 &\propto \exp\left(-\frac{1}{2\sigma_u^2}|u - u_0|^2\right) L(u; x, y) \\
 &\propto \exp\left(-\frac{1}{2}\left(\frac{1}{v}\sum_{i=1}^n |y_i - g(x_i; u)|^2 + \frac{1}{\sigma_u^2}|u - u_0|^2\right)\right)
 \end{aligned}$$

- Il massimo della densità a posteriori (stima MAP) si ottiene minimizzando

$$u \mapsto \frac{1}{v} \sum_{i=1}^n |y_i - g(x_i; u)|^2 + \frac{1}{\sigma_u^2} |u - u_0|^2.$$

- Il massimo della densità a posteriori (stima MAP) si ottiene minimizzando

$$u \mapsto \sum_{i=1}^n |y_i - g(x_i; u)|^2 + \frac{\lambda}{\sigma_u^2} |u - u_0|^2.$$

- È stato quindi introdotto un termine di *regolarizzazione* (o penalizzazione) alla somma dei residui, che diventa rilevante se u è troppo lontano dal parametro u_0 .

- Il massimo della densità a posteriori (stima MAP) si ottiene minimizzando

$$u \mapsto \frac{1}{v} \sum_{i=1}^n |y_i - g(x_i; u)|^2 + \frac{1}{\sigma_u^2} |u - u_0|^2.$$

- È stato quindi introdotto un termine di *regolarizzazione* (o penalizzazione) alla somma dei residui, che diventa rilevante se u è troppo lontano dal parametro u_0 .
- L'introduzione di questi ed altre funzioni è spesso utile per regolarizzare appunto la soluzione fornita dal semplice metodo dei minimi quadrati (queste tecniche hanno diversi nomi a seconda del tipo di termini che si aggiungono, ad esempio *ridge*, *weight decay*, LASSO, ecc.).

Approccio bayesiano e modelli lineari

- Supponendo ulteriormente che $g(x; u) = x \cdot u$, la densità a posteriori per U diventa

$$p(U = u | X_i = x_i, Y_i = y_i, \forall i = 1, \dots, n) \\ \propto \exp \left(-\frac{1}{2} \left(\frac{1}{v} \sum_{i=1}^n |y_i - x_i \cdot u|^2 + \frac{1}{\sigma_u^2} |u - u_0|^2 \right) \right),$$

che è una densità gaussiana vettoriale (essendo un esponenziale di polinomio di secondo grado rispetto alla variabile u).

Approccio bayesiano e modelli lineari

- Supponendo ulteriormente che $g(x; u) = x \cdot u$, la densità a posteriori per U diventa

$$p(U = u | X_i = x_i, Y_i = y_i, \forall i = 1, \dots, n) \\ \propto \exp \left(-\frac{1}{2} \left(\frac{1}{v} \sum_{i=1}^n |y_i - x_i \cdot u|^2 + \frac{1}{\sigma_u^2} |u - u_0|^2 \right) \right),$$

che è una densità gaussiana vettoriale (essendo un esponenziale di polinomio di secondo grado rispetto alla variabile u).

- Si trova che U ha come nuovi parametri, il vettore dei valor medi

$$\underline{u}_{|X=x, Y=y} = \left(x^T x + (v/\sigma_u^2) Id \right)^{-1} \left(\underline{x^T y} + \underbrace{\left(\frac{v}{\sigma_u^2} \right)}_{\lambda} u_0 \right)$$

e la matrice delle covarianze

$$\Sigma_{U|X=x, Y=y} = v \left(x^T x + (v/\sigma_u^2) Id \right)^{-1}.$$

- La deviazione standard della componente U_j del vettore dei parametri U , si ottiene dal termine diagonale della matrice,

$$\begin{aligned}\sigma_{U_j|X=x, Y=y} &= \sqrt{\text{Var}(U_j|X=x, Y=y)} \\ &= \sqrt{v(x^T x + (v/\sigma_u^2)Id)_{jj}^{-1}}.\end{aligned}$$

- La deviazione standard della componente U_j del vettore dei parametri U , si ottiene dal termine diagonale della matrice,

$$\begin{aligned}\sigma_{U_j|X=x, Y=y} &= \sqrt{\text{Var}(U_j|X=x, Y=y)} \\ &= \sqrt{v(x^T x + (v/\sigma_u^2)Id)_{jj}^{-1}}.\end{aligned}$$

- Nel limite $v \ll \sigma_u^2$ dalle formule sopra si recuperano la stima del metodo classico dei minimi quadrati per il modello lineare

$$u_{OLS} = (x^T x)^{-1} x^T y$$

e (avendo posto $v = v_{MLE} = MSE$ la stima di massima verosimiglianza) gli errori standard dei parametri, per $j \in \{1, \dots, k\}$,

$$\sigma_j = \sqrt{v(x^T x)_{jj}^{-1}}.$$

Altre funzioni obiettivo

Perché minimizzare i quadrati dei residui? in effetti ci sono altre scelte possibili e ragionevoli (ma non si trovano formule esplicite).

- Ad esempio il valore assoluto (*least absolute deviation* in inglese) darebbe

$$u_{\text{LAD}} \in \arg \min_{u \in \mathbb{R}^k} \sum_{i=1}^n |y_i - g(x_i; u)|.$$

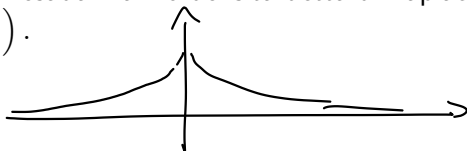
Altre funzioni obiettivo

Perché minimizzare i quadrati dei residui? in effetti ci sono altre scelte possibili e ragionevoli (ma non si trovano formule esplicite).

- Ad esempio il valore assoluto (*least absolute deviation* in inglese) darebbe

$$u_{\text{LAD}} \in \arg \min_{u \in \mathbb{R}^k} \sum_{i=1}^n |y_i - g(x_i; u)|.$$

- L'interpretazione è che i residui hanno densità detta di Laplace $p(W = w) \propto \exp\left(-\frac{|w|}{b}\right)$.



Section 2

Sull'ipotesi di gaussianità

Giustificare l'ipotesi

- L'introduzione di opportune variabili gaussiane permette di interpretare metodi come la PCA o i minimi quadrati in termini probabilistici (stime di massima verosimiglianza).

Giustificare l'ipotesi

- L'introduzione di opportune variabili gaussiane permette di interpretare metodi come la PCA o i minimi quadrati in termini probabilistici (stime di massima verosimiglianza).
- Questo permette di chiarire le ipotesi sottostanti per garantire, almeno in teoria, una corretta applicazione del metodo.

Giustificare l'ipotesi

- L'introduzione di opportune variabili gaussiane permette di interpretare metodi come la PCA o i minimi quadrati in termini probabilistici (stime di massima verosimiglianza).
- Questo permette di chiarire le *ipotesi* sottostanti per garantire, almeno in teoria, una corretta applicazione del metodo.
- Ad esempio, nel caso dei minimi quadrati, i residui dovrebbero avere densità gaussiane.

Giustificare l'ipotesi

- L'introduzione di opportune variabili gaussiane permette di interpretare metodi come la PCA o i minimi quadrati in termini probabilistici (stime di massima verosimiglianza).
- Questo permette di chiarire le *ipotesi* sottostanti per garantire, almeno in teoria, una corretta applicazione del metodo.
- Ad esempio, nel caso dei minimi quadrati, i residui dovrebbero avere densità gaussiane.
- Come argomentare che n osservazioni $(x_i)_{i=1}^n$ di dati dalla realtà possano essere ragionevolmente modellizzate tramite eventi del tipo $X_i = x_i$, dove le variabili aleatorie X_i siano gaussiane indipendenti?

Giustificare l'ipotesi

- L'introduzione di opportune variabili gaussiane permette di interpretare metodi come la PCA o i minimi quadrati in termini probabilistici (stime di massima verosimiglianza).
- Questo permette di chiarire le *ipotesi* sottostanti per garantire, almeno in teoria, una corretta applicazione del metodo.
- Ad esempio, nel caso dei minimi quadrati, i residui dovrebbero avere densità gaussiane.
- Come argomentare che n osservazioni $(x_i)_{i=1}^n$ di dati dalla realtà possano essere ragionevolmente modellizzate tramite eventi del tipo $X_i = x_i$, dove le variabili aleatorie X_i siano gaussiane indipendenti?
- Discutiamo due approcci (qualitativi e quantitativi).

Approcci qualitativi

In questo caso si sfruttano **teoremi limite**, come la legge dei grandi numeri (che vedremo) per garantire che le osservazioni $(x_i)_{i=1}^n$ si avvicinano per n grande alle quantità teoriche in un certo senso.

- In questa categoria rientrano:

Approcci qualitativi

In questo caso si sfruttano teoremi limite, come la legge dei grandi numeri (che vedremo) per garantire che le osservazioni $(x_i)_{i=1}^n$ si avvicinano per n grande alle quantità teoriche in un certo senso.

- In questa categoria rientrano:
 - 1 il confronto tra l'istogramma dei valori osservati e la densità gaussiana stimata.

Approcci qualitativi

In questo caso si sfruttano teoremi limite, come la legge dei grandi numeri (che vedremo) per garantire che le osservazioni $(x_i)_{i=1}^n$ si avvicinano per n grande alle quantità teoriche in un certo senso.

- In questa categoria rientrano:
 - 1 il confronto tra l'istogramma dei valori osservati e la densità gaussiana stimata.
 - 2 il confronto tra la funzione quantile empirica (ossia della variabile uniforme discreta sugli n valori osservati) con la funzione quantile di una opportuna gaussiana e quella teorica (QQ-plot).

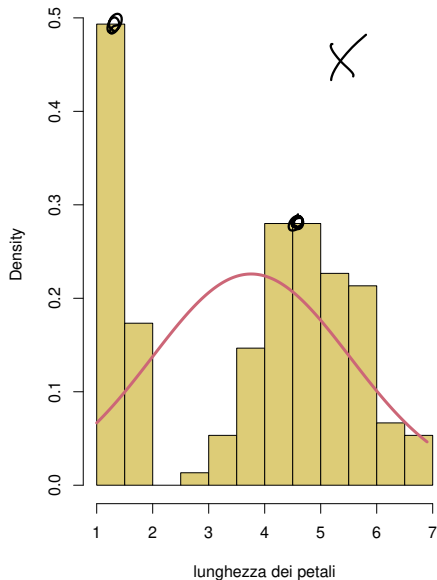
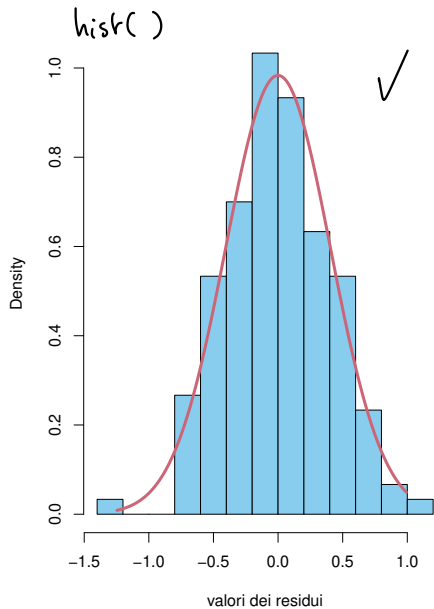
$$CDF_{\text{empirica}}(t) = \frac{\#\{x_i : x_i \leq t\}}{n} \approx CDF_N(t)$$

q

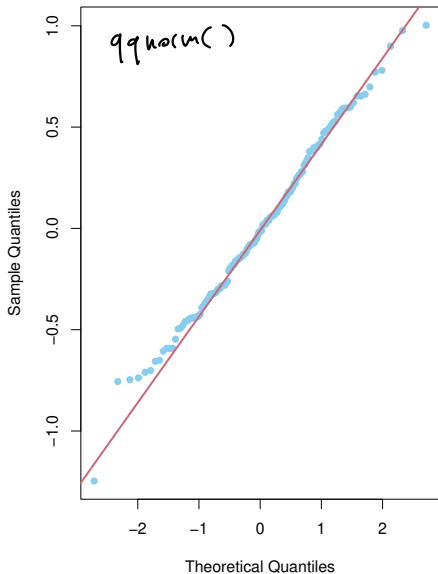
Approcci qualitativi

In questo caso si sfruttano teoremi limite, come la legge dei grandi numeri (che vedremo) per garantire che le osservazioni $(x_i)_{i=1}^n$ si avvicinano per n grande alle quantità teoriche in un certo senso.

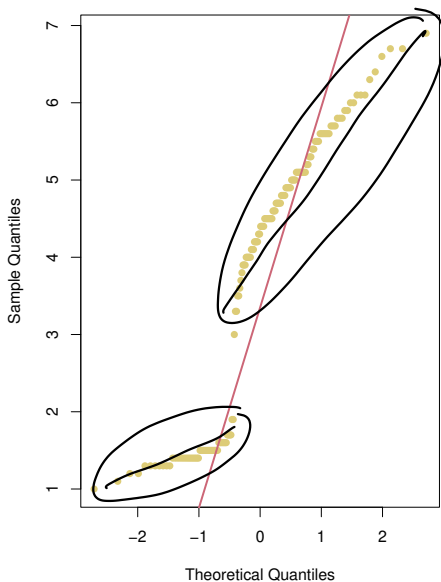
- In questa categoria rientrano:
 - 1 il confronto tra l'istogramma dei valori osservati e la densità gaussiana stimata.
 - 2 il confronto tra la funzione quantile empirica (ossia della variabile uniforme discreta sugli n valori osservati) con la funzione quantile di una opportuna gaussiana e quella teorica (QQ-plot).
- In entrambi i casi si valuta graficamente la somiglianza tra le quantità teoriche e quelle empiriche.



Normal Q-Q Plot



Normal Q-Q Plot



Generalità sui test statistici

- **Obiettivo:** scartare (rifiutare) alcune alternative A_i sulla base dell'osservazione di B .

Generalità sui test statistici

- **Obiettivo:** scartare (rifiutare) alcune alternative A_i sulla base dell'osservazione di B .
- **Intuizione:** eliminare quelle con probabilità a posteriori nulla (ovvio) o bassa.

$$P(A_i | B) < \varepsilon \implies \text{scarta } A_i$$

$$|L(A_i; B)| < \delta \implies \text{scarta } A_i$$

Generalità sui test statistici

- **Obiettivo:** scartare (rifiutare) alcune alternative A_i sulla base dell'osservazione di B .
- **Intuizione:** eliminare quelle con probabilità a posteriori nulla (ovvio) o bassa.
- Nella pratica ci si appoggia alla verosimiglianza invece delle probabilità a posteriori (simile se a priori sono uniformi).

Generalità sui test statistici

- **Obiettivo:** scartare (rifiutare) alcune alternative A_i sulla base dell'osservazione di B .
- **Intuizione:** eliminare quelle con probabilità a posteriori nulla (ovvio) o bassa.
- Nella pratica ci si appoggia alla verosimiglianza invece delle probabilità a posteriori (simile se a priori sono uniformi).
- Ipotesi nulla \mathcal{H}_0 (da rifiutare) e alternativa \mathcal{H}_1 .

Generalità sui test statistici

- **Obiettivo:** scartare (rifiutare) alcune alternative A_i sulla base dell'osservazione di B .
- **Intuizione:** eliminare quelle con probabilità a posteriori nulla (ovvio) o bassa.
- Nella pratica ci si appoggia alla verosimiglianza invece delle probabilità a posteriori (simile se a priori sono uniformi).
- Ipotesi nulla \mathcal{H}_0 (da rifiutare)) e alternativa \mathcal{H}_1 .
- **Valore p (p -value)** di Fisher:

$$p = \max_{A_i \in \mathcal{H}_0} P(B|A_i) = \max_{A_i \in \mathcal{H}_0} L(A_i; B).$$

Generalità sui test statistici

- **Obiettivo:** scartare (rifiutare) alcune alternative A_i sulla base dell'osservazione di B .
- **Intuizione:** eliminare quelle con probabilità a posteriori nulla (ovvio) o bassa.
- Nella pratica ci si appoggia alla verosimiglianza invece delle probabilità a posteriori (simile se a priori sono uniformi).
- Ipotesi nulla \mathcal{H}_0 (da rifiutare)) e alternativa \mathcal{H}_1 .
- **Valore p** (p -value) di Fisher:

$$p = \max_{A_i \in \mathcal{H}_0} P(B|A_i) = \max_{A_i \in \mathcal{H}_0} L(A_i; B).$$

- Più piccolo è il valore p , minore è la probabilità che B sia vero rispetto a l'ipotesi \mathcal{H}_0 , e quindi, invocando Bayes, anche che A_i sia vero sapendo B .

Generalità sui test statistici

- **Obiettivo:** scartare (rifiutare) alcune alternative A_i sulla base dell'osservazione di B .
- **Intuizione:** eliminare quelle con probabilità a posteriori nulla (ovvio) o bassa.
- Nella pratica ci si appoggia alla verosimiglianza invece delle probabilità a posteriori (simile se a priori sono uniformi).
- Ipotesi nulla \mathcal{H}_0 (da rifiutare)) e alternativa \mathcal{H}_1 .
- **Valore p** (p -value) di Fisher:

$$p = \max_{A_i \in \mathcal{H}_0} P(B|A_i) = \max_{A_i \in \mathcal{H}_0} L(A_i; B).$$

- Più piccolo è il valore p , minore è la probabilità che B sia vero rispetto a l'ipotesi \mathcal{H}_0 , e quindi, invocando Bayes, anche che A_i sia vero sapendo B .

Test di gaussianità

- l'ipotesi nulla è l'evento $\mathcal{H}_0 =$ “le osservazioni provengono da n variabili gaussiane indipendenti, tutte con gli stessi parametri”

Test di gaussianità

- l'ipotesi nulla è l'evento $\mathcal{H}_0 =$ “le osservazioni provengono da n variabili gaussiane indipendenti, tutte con gli stessi parametri”
- l'alternativa semplice \mathcal{H}_1 è la sua negazione.

Test di gaussianità

- l'ipotesi nulla è l'evento $\mathcal{H}_0 =$ “le osservazioni provengono da n variabili gaussiane indipendenti, tutte con gli stessi parametri”
- l'alternativa semplice \mathcal{H}_1 è la sua negazione.
- La descrizione specifica di questi test è troppo lunga (vediamo solo i comandi R)

$\textcircled{B} \leftarrow ?$

Test di gaussianità

- l'ipotesi nulla è l'evento $\mathcal{H}_0 =$ “le osservazioni provengono da n variabili gaussiane indipendenti, tutte con gli stessi parametri”
- l'alternativa semplice \mathcal{H}_1 è la sua negazione.
- La descrizione specifica di questi test è troppo lunga (vediamo solo i comandi R)
- Il valore p , è calcolato rispetto alla probabilità condizionata alla validità dell'ipotesi nulla.

Test di gaussianità

- l'ipotesi nulla è l'evento $\mathcal{H}_0 =$ “le osservazioni provengono da n variabili gaussiane indipendenti, tutte con gli stessi parametri”
- l'alternativa semplice \mathcal{H}_1 è la sua negazione.
- La descrizione specifica di questi test è troppo lunga (vediamo solo i comandi R)
- Il valore p , è calcolato rispetto alla probabilità condizionata alla validità dell'ipotesi nulla.
- Più piccolo il valore p , maggiore sarà il “grado di fiducia” che il test attribuisce nel rifiutare l'ipotesi nulla.

Test di gaussianità

- l'ipotesi nulla è l'evento $\mathcal{H}_0 =$ “le osservazioni provengono da n variabili gaussiane indipendenti, tutte con gli stessi parametri”
- l'alternativa semplice \mathcal{H}_1 è la sua negazione.
- La descrizione specifica di questi test è troppo lunga (vediamo solo i comandi R)
- Il valore p , è calcolato rispetto alla probabilità condizionata alla validità dell'ipotesi nulla.
- Più piccolo il valore p , maggiore sarà il “grado di fiducia” che il test attribuisce nel rifiutare l'ipotesi nulla.
- Se usiamo il test per “confermare” l'ipotesi di gaussianità, il test sarà tanto più utile quanto più *grande* (ossia vicino ad 1) è il valore p .

Esempi in R

Un test di gaussianità è dovuto a Shapiro (e Wilk).

```
shapiro.test(iris_res)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  iris_res  
## W = 0.99298, p-value = 0.6767
```

```
shapiro.test(iris_petal)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  iris_petal  
## W = 0.87627, p-value = 7.412e-10
```

```
ks.test(iris_res, "pnorm", mean=mean(iris_res), sd=sd(iris_re
```

```
##  
## Asymptotic one-sample Kolmogorov-Smirnov test  
##  
## data: iris_res  
## D = 0.040916, p-value = 0.9632  
## alternative hypothesis: two-sided
```

```
ks.test(iris_petali, "pnorm", mean=mean(iris_petali), sd=sd(i
```

```
##  
## Asymptotic one-sample Kolmogorov-Smirnov test  
##  
## data: iris_petali  
## D = 0.19815, p-value = 1.532e-05  
## alternative hypothesis: two-sided
```

Section 3

Approssimazione di Laplace

Un metodo euristico per approssimare con gaussiane

Data una variabile $X \in \mathbb{R}^d$ con densità non gaussiana, è possibile approssimarla con una gaussiana?

- In questo modo gli strumenti sviluppati per le variabili gaussiane si potrebbero applicare (tenendo conto dell'errore di approssimazione).

Un metodo euristico per approssimare con gaussiane

Data una variabile $X \in \mathbb{R}^d$ con densità non gaussiana, è possibile approssimarla con una gaussiana?

- In questo modo gli strumenti sviluppati per le variabili gaussiane si potrebbero applicare (tenendo conto dell'errore di approssimazione).
- Una soluzione semplice e spesso efficace è l'**approssimazione di Laplace**.

Si considera lo sviluppo del logaritmo della densità

$$x \mapsto \log(p(X = x)),$$

in un punto di massimo x_{\max} (anche un massimo locale). Si trova

$$\begin{aligned} & \log(p(X = x)) \\ &= \log(p(X = x_{\max})) + \frac{1}{2}(x - x_{\max}) \cdot H(x_{\max})(x - x_{\max}) + O(|x - x_{\max}|^3), \end{aligned}$$

dove

$$H(x) = \left(\frac{\partial^2 \log(p(X = x))}{\partial x_i \partial x_j} \right)_{i,j=1}^d$$

la matrice delle derivate seconde (detta anche hessiana).

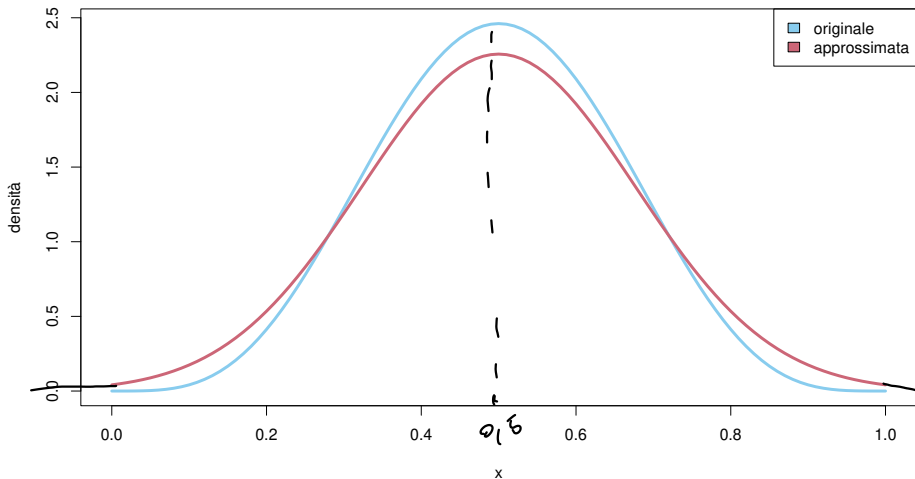
- $H(x_{\max})$ è una matrice (semi-)definita negativa.

- L'approssimazione di Laplace è data dalla densità gaussiana di valor medio $\underline{x_{\max}}$ e matrice delle covarianze $\underline{-(H(x_{\max}))^{-1}}$.

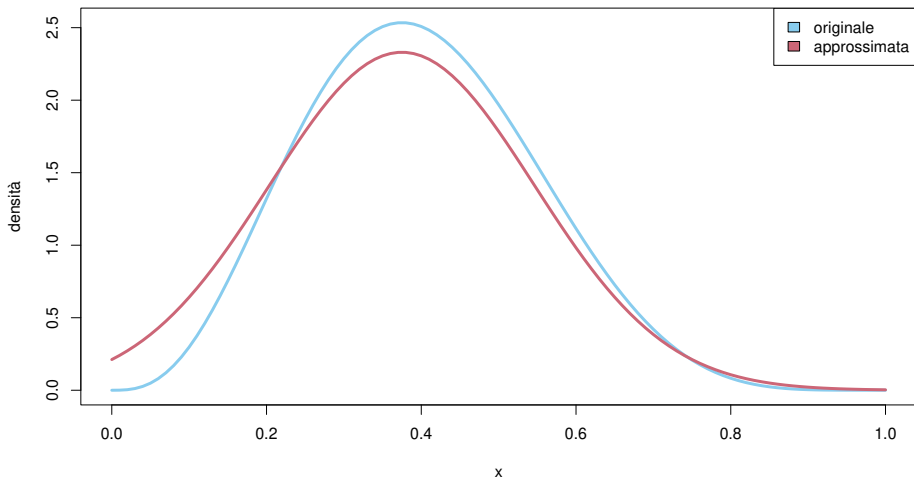
- L'approssimazione di Laplace è data dalla densità gaussiana di valor medio x_{\max} e matrice delle covarianze $-(H(x_{\max}))^{-1}$.
- Spesso si calcola x_{\max} tramite metodi numerici che restituiscono anche la matrice hessiana.

- L'approssimazione di Laplace è data dalla densità gaussiana di valor medio x_{\max} e matrice delle covarianze $-(H(x_{\max}))^{-1}$.
- Spesso si calcola x_{\max} tramite metodi numerici che restituiscono anche la matrice hessiana.
- Non è garantita in generale che l'approssimazione sia vicina alla densità di X , in particolare per valori x lontani da x_{\max} . Vediamo degli esempi.

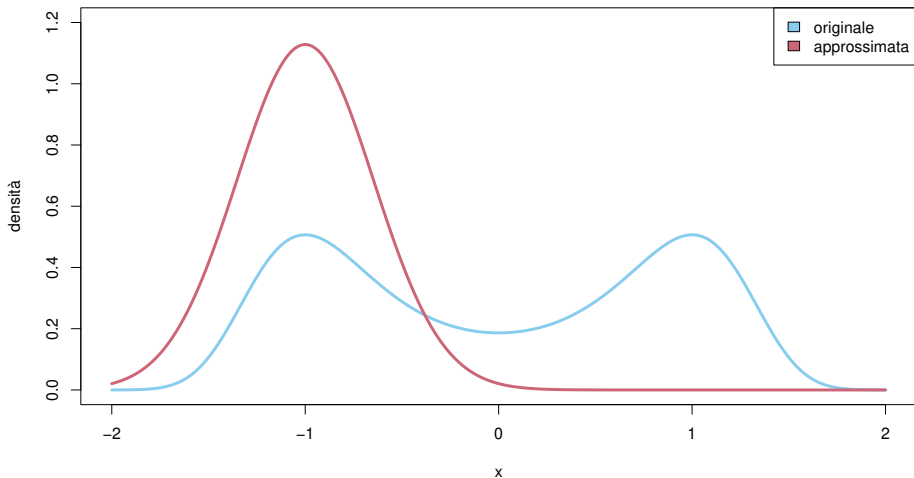
$p(X = x) = x^4(1 - x)^4 \cdot c$ per $x \in [0, 1]$, nulla altrimenti.



$$p(X = x) = x^3(1 - x)^5 \quad \text{per } x \in [0, 1], \text{ nulla altrimenti.}$$



$p(X = x) = \exp(-(1 - x)^2(1 + x)^2)$ per $x \in [-2, 2]$, nulla altrimenti.



Vediamo un esempio nel caso vettoriale.

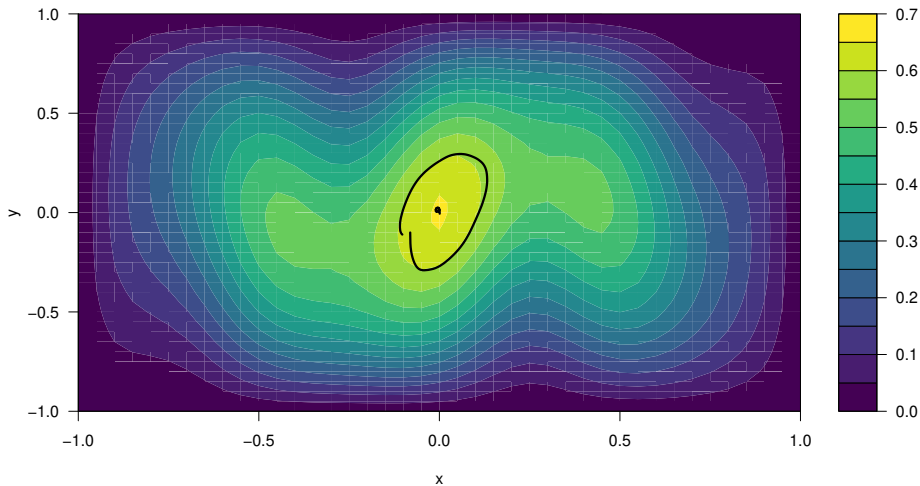


Figure 1: heatmap della densità originale

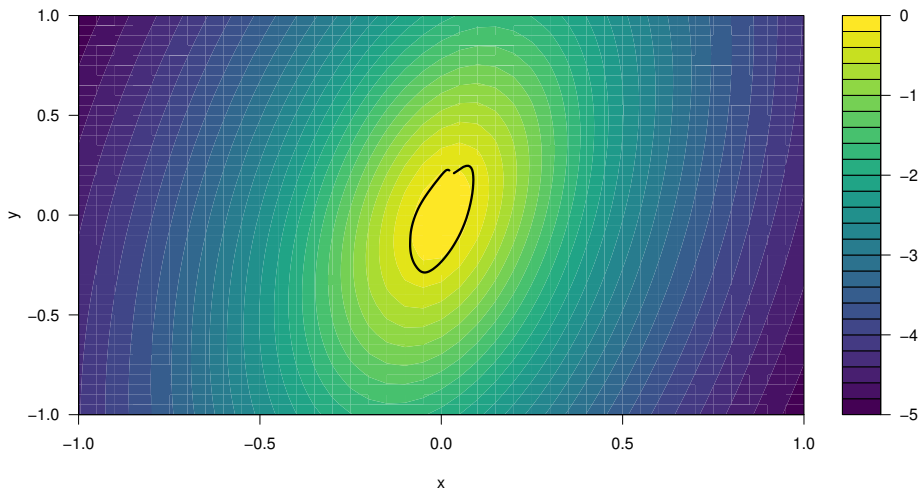


Figure 2: heatmap dell'approssimazione di Laplace.

Section 4

Esercizi

Eventi e variabili aleatorie discrete

Un nostro amico appassionato di dadi ci propone il seguente gioco: egli ha con sé un dado a 4 facce (tetraedro) numerate da 1 a 4 e uno a sei facce, numerate da 1 a 6. Sceglie uno dei due dadi (senza mostrarci quale) ed esegue con esso 2 lanci consecutivi, ottenendo come esiti due numeri X_1, X_2 (a noi non noti).

- 1 Calcolare media e varianza della variabile X_1 .

Eventi e variabili aleatorie discrete

Un nostro amico appassionato di dadi ci propone il seguente gioco: egli ha con sé un dado a 4 facce (tetraedro) numerate da 1 a 4 e uno a sei facce, numerate da 1 a 6. Sceglie uno dei due dadi (senza mostrarci quale) ed esegue con esso 2 lanci consecutivi, ottenendo come esiti due numeri X_1 , X_2 (a noi non noti).

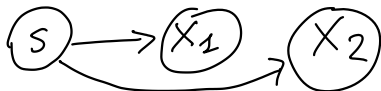
- 1 Calcolare media e varianza della variabile X_1 .
- 2 Supponendo che l'esito del primo lancio sia 3, determinare la densità di X_2 .

Eventi e variabili aleatorie discrete

Un nostro amico appassionato di dadi ci propone il seguente gioco: egli ha con sé un dado a 4 facce (tetraedro) numerate da 1 a 4 e uno a sei facce, numerate da 1 a 6. Sceglie uno dei due dadi (senza mostrarci quale) ed esegue con esso 2 lanci consecutivi, ottenendo come esiti due numeri X_1, X_2 (a noi non noti).

- 1 Calcolare media e varianza della variabile X_1 .
- 2 Supponendo che l'esito del primo lancio sia 3, determinare la densità di X_2 .
- 3 Le variabili X_1, X_2 sono indipendenti? sono correlate?

$$X_1, X_2 \in \{1, 2, 3, 4, 5, 6\} \quad S \in \{4, 6\}$$



Rispetto a $S=4 \Rightarrow X_1, X_2$ indep. uniformi su $\{1, 2, 3, 4\}$

$S=6 \Rightarrow X_1, X_2$ indep. e unif. $\{1, \dots, 6\}$

$$\begin{aligned} \mathbb{E}[X_1] &= \mathbb{E}[X_1 | S=4] P(S=4) + \mathbb{E}[X_1 | S=6] P(S=6) \\ &= \frac{1+2+3+4}{4} \cdot P(S=4) + \frac{1+2+3+4+5+6}{6} P(S=6) \end{aligned}$$

$$P(S=4) = P(S=6) = \frac{1}{2} \quad = 2,5 \cdot \frac{1}{2} + 3,5 \cdot \frac{1}{2} = 3$$

$$\begin{aligned} \mathbb{E}[X_1^2] &= \mathbb{E}[X_1^2 | S=4] \cdot \frac{1}{2} + \mathbb{E}[X_1^2 | S=6] \cdot \frac{1}{2} \\ &= \frac{1^2 + 2^2 + 3^2 + 4^2}{4} \cdot \frac{1}{2} + \frac{1^2 + 2^2 + \dots + 6^2}{6} \cdot \frac{1}{2} \end{aligned}$$

$$\Rightarrow \text{Var}(X_1) = \mathbb{E}[X_1^2] - (\mathbb{E}[X_1])^2$$

Esercizio confrontare $\text{Var}(X_1)$ con $\text{Var}(U)$ $U \in \{1, 2, \dots, 5\}$

$$\underline{X_1=3} \Rightarrow P(X_2=k | X_1=3) = ?$$

$$P(X_2=k | \cancel{X_1=3}, S=4) P(S=4 | X_1=3) + P(X_2=k | \cancel{X_1=3}, S=6) \cdot P(S=6 | X_1=3)$$

$\left\{ \begin{array}{l} \frac{1}{4} \text{ se } k \in \{1, 2, 3, 4\} \\ 0 \text{ se } k = 5, 6 \end{array} \right.$

$\frac{1}{6}$

$$\begin{aligned}
 P(S=4 | X_1=3) &= \frac{P(S=4) P(X_1=3 | S=4)}{P(X_1=3)} \\
 &= \frac{\frac{1}{2} \cdot \frac{1}{4}}{P(X_1=3)}
 \end{aligned}$$

$$\begin{aligned}
 P(X_1=3) &= P(X_1=3 | S=4) P(S=4) + P(X_1=3 | S=6) P(S=6) \\
 &= \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{6} \cdot \frac{1}{2}
 \end{aligned}$$

Le v.v. X_1, X_2 Non sono indep.

In fatti $P(X_2 = k | X_1 = h) \underset{\approx}{=} P(X_2 = k)$

Se $\frac{h=6}{k=6}$ $P(X_2=6 | X_1=6) \neq P(X_2=6)$

$$P(X_2=6 | X_1=6) = P(X_2=6 | X_1=6, S=4) P(S=4 | X_1=6) +$$

$$+ P(X_2=6 | X_1=6, S=6) P(S=6 | X_1=6)$$

$$P(X_2=6) = P(X_2=6 | \cancel{S=4}) P(S=4) \frac{1}{2} +$$

$$+ P(X_2=6 | \cancel{S=6}) P(S=6) \frac{1}{2}$$

$$\text{Cov}(X_1, X_2) = \mathbb{E}[X_1 X_2] - \overbrace{\mathbb{E}[X_1] \mathbb{E}[X_2]}$$

$$\begin{aligned} \mathbb{E}[X_1 X_2] &= \mathbb{E}[X_1 X_2 | S=4] P(S=4) + \mathbb{E}[X_1 X_2 | S=6] \cdot \frac{1}{2} \\ &= \mathbb{E}[X_1 | S=4] \mathbb{E}[X_2 | S=4] \cdot \frac{1}{2} + \\ &\quad + \mathbb{E}[X_1 | S=6] \mathbb{E}[X_2 | S=6] \cdot \frac{1}{2} \\ &= \left[(2, 5)^2 + (3, 5)^2 \right] \cdot \frac{1}{2} \end{aligned}$$

$$\text{Cov}(X_1, X_2) = \left((2, 5)^2 + (3, 5)^2 \right) \cdot \frac{1}{2} - 3^2 \quad \underline{\geq 0}$$

Due amici, Andrea e Bruno, vanno a pesca. La probabilità che Andrea prenda almeno un pesce in un dato giorno è $1/2$ e anche la probabilità che Bruno ne prenda almeno uno è $1/2$. Tuttavia Bruno sospetta che se Andrea pesca qualcosa in un dato giorno le sue possibilità diminuiscano. Poniamo $p \in (0, 1)$ la probabilità che Bruno prenda almeno un pesce, sapendo che Andrea ne ha pescato almeno uno.

- 1 Supponendo p noto, e sapendo che uno solo tra Andrea e Bruno è riuscito a prendere qualcosa in un dato giorno, calcolare la probabilità che si tratti di Andrea.

Due amici, Andrea e Bruno, vanno a pesca. La probabilità che Andrea prenda almeno un pesce in un dato giorno è $1/2$ e anche la probabilità che Bruno ne prenda almeno uno è $1/2$. Tuttavia Bruno sospetta che se Andrea pesca qualcosa in un dato giorno le sue possibilità diminuiscano. Poniamo $p \in (0, 1)$ la probabilità che Bruno prenda almeno un pesce, sapendo che Andrea ne ha pescato almeno uno.

- 1 Supponendo p noto, e sapendo che uno solo tra Andrea e Bruno è riuscito a prendere qualcosa in un dato giorno, calcolare la probabilità che si tratti di Andrea.
- 2 Gli amici hanno riportato in tabella i pesci pescati in 5 giorni di attività.

Giorno	1	2	3	4	5
Andrea	1	0	2	1	4
Bruno	0	0	2	0	0

Supponendo che gli eventi relativi ai diversi giorni siano indipendenti determinare la stima di massima verosimiglianza per p .

Variabili aleatorie continue

Un supermercato ha acquistato un nuovo frigorifero per l'esposizione dei surgelati in vendita. La durata di funzionamento dell'apparecchio è descritta da una variabile aleatoria T con distribuzione esponenziale di media 120 mesi.

- 1 Si calcoli la funzione di sopravvivenza del tempo di funzionamento.

Variabili aleatorie continue

Un supermercato ha acquistato un nuovo frigorifero per l'esposizione dei surgelati in vendita. La durata di funzionamento dell'apparecchio è descritta da una variabile aleatoria T con distribuzione esponenziale di media 120 mesi.

- 1 Si calcoli la funzione di sopravvivenza del tempo di funzionamento.
- 2 Sapendo che sono passati 60 mesi di funzionamento e il frigorifero funziona correttamente, si calcoli la densità del tempo di funzionamento rimanente (ossia $T - 60$).

Variabili aleatorie continue

Un supermercato ha acquistato un nuovo frigorifero per l'esposizione dei surgelati in vendita. La durata di funzionamento dell'apparecchio è descritta da una variabile aleatoria T con distribuzione esponenziale di media 120 mesi.

- 1 Si calcoli la funzione di sopravvivenza del tempo di funzionamento.
- 2 Sapendo che sono passati 60 mesi di funzionamento e il frigorifero funziona correttamente, si calcoli la densità del tempo di funzionamento rimanente (ossia $T - 60$).
- 3 Mentre i primi 5 anni di funzionamento sono coperti da garanzia, il costo della riparazione del frigorifero è per contratto $\exp(T/10)$ se una rottura avviene in un tempo T tra i 60 e i 120 mesi dall'installazione. Confrontare il valore atteso del costo di una riparazione nei due casi: appena il frigorifero è installato; sapendo che non si è rotto nei primi 5 anni di funzionamento.

Il prezzo in una data futura delle azioni di un'azienda quotata in borsa è rappresentabile, secondo alcuni studiosi dei mercati, da una variabile della forma

$$C = \exp(X)$$

dove X è una variabile aleatoria reale con densità $\mathcal{N}(m, \sigma^2)$.

- 1 Calcolate la densità di C , il valore atteso e la varianza.

Il prezzo in una data futura delle azioni di un'azienda quotata in borsa è rappresentabile, secondo alcuni studiosi dei mercati, da una variabile della forma

$$\underline{C = \exp(X)}$$

dove X è una variabile aleatoria reale con densità $\mathcal{N}(m, \sigma^2)$.

- ① Calcolate la densità di C , il valore atteso e la varianza.
- ② Gli studiosi hanno stimato che $\sigma^2 = 1$, mentre sono indecisi sul parametro m . Supponendo di osservare nella data futura che $C = 3$, determinare la stima di massima verosimiglianza per m .

$$C = 3 \Rightarrow \exp(X) = 3 \Rightarrow \underline{\underline{X = \ln 3}}$$

$$\Rightarrow \boxed{m = \ln 3}$$

<!-

Section 5

Processi a stati discreti

Presentazione

- Introduciamo il linguaggio di base della teoria dei processi con alcune definizioni generali.

Presentazione

- Introduciamo il linguaggio di base della teoria dei processi con alcune definizioni generali.
- Vediamo la teoria di base delle catene di Markov

Presentazione

- Introduciamo il linguaggio di base della teoria dei processi con alcune definizioni generali.
- Vediamo la teoria di base delle catene di Markov
- e poi dei processi di Markov a salti.

Presentazione

- Introduciamo il linguaggio di base della teoria dei processi con alcune definizioni generali.
- Vediamo la teoria di base delle catene di Markov
- e poi dei processi di Markov a salti.
- Studiamo esistenza e unicità delle distribuzioni invarianti associate ad un processo di Markov

Presentazione

- Introduciamo il linguaggio di base della teoria dei processi con alcune definizioni generali.
- Vediamo la teoria di base delle catene di Markov
- e poi dei processi di Markov a salti.
- Studiamo esistenza e unicità delle distribuzioni invarianti associate ad un processo di Markov
- Accenniamo al problema della stima dei parametri di un processo a partire dalle osservazioni.

Presentazione

- Introduciamo il linguaggio di base della teoria dei processi con alcune definizioni generali.
- Vediamo la teoria di base delle catene di Markov
- e poi dei processi di Markov a salti.
- Studiamo esistenza e unicità delle distribuzioni invarianti associate ad un processo di Markov
- Accenniamo al problema della stima dei parametri di un processo a partire dalle osservazioni.
- Concludiamo con degli esempi fondamentali dalla teoria delle code.

Definizioni generali

Un **processo stocastico** è una collezione di variabili aleatorie $(X_t)_{t \in \mathcal{T}}$,

- tutte a valori nello stesso insieme E , detto insieme degli **stati** del processo,

Definizioni generali

Un **processo stocastico** è una collezione di variabili aleatorie $(X_t)_{t \in \mathcal{T}}$,

- tutte a valori nello stesso insieme E , detto insieme degli **stati** del processo,
- indicizzate da un insieme $\mathcal{T} \subseteq \mathbb{R}$ detto insieme dei **tempi** del processo.

Il calcolo delle probabilità fornisce strumenti utili per affrontare problemi circa

- il *futuro* di un processo (questo è il problema della *previsione*)

Il calcolo delle probabilità fornisce strumenti utili per affrontare problemi circa

- il *futuro* di un processo (questo è il problema della *previsione*)
- il *passato*,

Il calcolo delle probabilità fornisce strumenti utili per affrontare problemi circa

- il *futuro* di un processo (questo è il problema della *previsione*)
- il *passato*,
- oppure anche il *presente* (se non è esattamente osservato, è il *filtraggio*).

Classificazione generale

Analogamente alle singole variabili aleatorie, si classificano i processi stocastici in base all'insieme degli stati:

- a **stati discreti** se E discreto (quindi finito oppure infinito numerabile, ad esempio $E = \mathbb{Z}$ oppure \mathbb{N})

In base all'insieme \mathcal{T} dei tempi:

Classificazione generale

Analogamente alle singole variabili aleatorie, si classificano i processi stocastici in base all'insieme degli stati:

- a **stati discreti** se E discreto (quindi finito oppure infinito numerabile, ad esempio $E = \mathbb{Z}$ oppure \mathbb{N})
- a **stati continui** se E è infinito continuo, $E = \mathbb{R}$, $E = \mathbb{R}^k$ (e di solito ciascuna X_t ammetta densità continua)

In base all'insieme \mathcal{T} dei tempi:

Classificazione generale

Analogamente alle singole variabili aleatorie, si classificano i processi stocastici in base all'insieme degli stati:

- a **stati discreti** se E discreto (quindi finito oppure infinito numerabile, ad esempio $E = \mathbb{Z}$ oppure \mathbb{N})
- a **stati continui** se E è infinito continuo, $E = \mathbb{R}$, $E = \mathbb{R}^k$ (e di solito ciascuna X_t ammetta densità continua)

In base all'insieme \mathcal{T} dei tempi:

- a **tempi discreti** se \mathcal{T} è discreto (ad esempio finito, oppure $\mathcal{T} = \mathbb{N}$),

Classificazione generale

Analogamente alle singole variabili aleatorie, si classificano i processi stocastici in base all'insieme degli stati:

- a **stati discreti** se E discreto (quindi finito oppure infinito numerabile, ad esempio $E = \mathbb{Z}$ oppure \mathbb{N})
- a **stati continui** se E è infinito continuo, $E = \mathbb{R}$, $E = \mathbb{R}^k$ (e di solito ciascuna X_t ammetta densità continua)

In base all'insieme \mathcal{T} dei tempi:

- a **tempi discreti** se \mathcal{T} è discreto (ad esempio finito, oppure $\mathcal{T} = \mathbb{N}$),
- a **tempi continui** se $\mathcal{T} = [0, T]$ è un intervallo (anche illimitato, ad esempio $\mathcal{T} = [0, \infty)$).

Traiettorie e marginali

È utile pensare a $(X_t)_{t \in \mathcal{T}}$ come ad una variabile aleatoria vettoriale a valori in uno spazio di **traiettorie**,

$$E^{\mathcal{T}} = \{(x_t)_{t \in \mathcal{T}}\}.$$

- Ad esempio, se $\mathcal{T} = \{1, \dots, d\}$, allora un processo $(X_i)_{i=1}^d$ può essere pensato come una variabile aleatoria congiunta X , a valori in E^d .

Traiettorie e marginali

È utile pensare a $(X_t)_{t \in \mathcal{T}}$ come ad una variabile aleatoria vettoriale a valori in uno spazio di **traiettorie**,

$$E^{\mathcal{T}} = \{(x_t)_{t \in \mathcal{T}}\}.$$

- Ad esempio, se $\mathcal{T} = \{1, \dots, d\}$, allora un processo $(X_i)_{i=1}^d$ può essere pensato come una variabile aleatoria congiunta X , a valori in E^d .
- Ricordiamo la differenza tra la legge delle marginali

$$P(X_t \in U | I),$$

al variare di $U \subseteq E$ e $t \in \mathcal{T}$, e la legge congiunta, in questo caso detta semplicemente **legge del processo** $(X_t)_{t \in \mathcal{T}}$, che è definita come tutte le probabilità del tipo

$$P(X_{t_1} \in U_1, X_{t_2} \in U_2, \dots, X_{t_k} \in U_k | I),$$

Nel caso di processi a stati discreti, per ogni $t \in \mathcal{T}$ la densità discreta della marginale al tempo t è

$$P(X_t = x|I).$$

- La densità discreta del processo è la collezione di tutte le probabilità

$$P(X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_k} = x_k|I),$$

Nel caso di processi a stati discreti, per ogni $t \in \mathcal{T}$ la densità discreta della marginale al tempo t è

$$P(X_t = x|I).$$

- La densità discreta del processo è la collezione di tutte le probabilità

$$P(X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_k} = x_k|I),$$

- Nel caso di processi a stati continui (con densità continua), basta sostituire la “ P ” di probabilità con “ p ” della densità di probabilità.

Come parametrizzare un processo?

Determinare la legge di un processo tramite pochi parametri è un problema difficile, soprattutto se l'insieme dei tempi diventa grande.

- Se $E = \{0, 1\}$, la densità discreta di un processo su $\mathcal{T} = \{1, \dots, d\}$ è praticamente una qualsiasi funzione da $\{0, 1\}^d$ a valori in $[0, 1] \rightarrow$ circa 2^d “parametri”.

Come parametrizzare un processo?

Determinare la legge di un processo tramite pochi parametri è un problema difficile, soprattutto se l'insieme dei tempi diventa grande.

- Se $E = \{0, 1\}$, la densità discreta di un processo su $\mathcal{T} = \{1, \dots, d\}$ è praticamente una qualsiasi funzione da $\{0, 1\}^d$ a valori in $[0, 1] \rightarrow$ circa 2^d “parametri”.
- Le d densità marginali si ottengono descrivendo solo d “parametri” (la probabilità $P(X_t = 1|I)$), anche meno se le leggi sono tutte uguali.

Come parametrizzare un processo?

Determinare la legge di un processo tramite pochi parametri è un problema difficile, soprattutto se l'insieme dei tempi diventa grande.

- Se $E = \{0, 1\}$, la densità discreta di un processo su $\mathcal{T} = \{1, \dots, d\}$ è praticamente una qualsiasi funzione da $\{0, 1\}^d$ a valori in $[0, 1] \rightarrow$ circa 2^d “parametri”.
- Le d densità marginali si ottengono descrivendo solo d “parametri” (la probabilità $P(X_t = 1|I)$), anche meno se le leggi sono tutte uguali.
- Non si può ricostruire la densità del processo a partire dalle densità marginali, senza ulteriori ipotesi.

Proprietà di Markov

La proprietà afferma che *il futuro e il passato sono condizionatamente indipendenti, noto esattamente il presente.*

- Un processo $(X_t)_{t \in \mathcal{T}}$ è *di Markov* (o markoviano) se, per ogni $x \in E$, $t \in \mathcal{T}$, le due variabili congiunte relative ai tempi “passati” $(X_s)_{s < t}$ e “futuri” $(X_r)_{r > t}$ sono indipendenti, rispetto all'informazione in cui si conosce esattamente il presente, ossia $\{X_t = x\}$.

Proprietà di Markov

La proprietà afferma che *il futuro e il passato sono condizionatamente indipendenti, noto esattamente il presente.*

- Un processo $(X_t)_{t \in \mathcal{T}}$ è *di Markov* (o markoviano) se, per ogni $x \in E$, $t \in \mathcal{T}$, le due variabili congiunte relative ai tempi “passati” $(X_s)_{s < t}$ e “futuri” $(X_r)_{r > t}$ sono indipendenti, rispetto all'informazione in cui si conosce esattamente il presente, ossia $\{X_t = x\}$.
- Se A è una affermazione che riguarda solo le variabili $(X_s)_{s < t}$, e B è una riguarda solamente le variabili $(X_r)_{r > t}$, allora A , B sono indipendenti rispetto all'informazione $\{X_t = x\}$:

$$P(A, B | X_t = x) = P(A | X_t = x)P(B | X_t = x),$$

oppure

$$P(A | X_t = x, B) = P(A | X_t = x),$$

o anche

$$P(B | X_t = x, A) = P(B | X_t = x).$$

In termini grafici, la proprietà di Markov si traduce in una rete bayesiana associata al processo $(X_t)_{t \in \mathcal{T}}$ del seguente tipo:

Densità di un processo di Markov

Processi omogenei

Per procedere ulteriormente e sviluppare una teoria semplice:

- 1 consideriamo come insiemi di tempi \mathcal{T} intervalli discreti $\mathcal{T} = \{0, 1, 2, \dots, n\}$ o continui $\mathcal{T} = [0, T]$.

Processi omogenei

Per procedere ulteriormente e sviluppare una teoria semplice:

- 1 consideriamo come insiemi di tempi \mathcal{T} intervalli discreti $\mathcal{T} = \{0, 1, 2, \dots, n\}$ o continui $\mathcal{T} = [0, T]$.
- 2 consideriamo processi di Markov **omogenei**, ossia tali che le probabilità di transizione dal tempo s al tempo t dipendano solamente dalla differenza dei tempi $t - s$, o equivalentemente, per ogni $\Delta t \geq 0$ si abbia

$$P(X_t = y | X_s = x) = P(X_{t+\Delta t} = y | X_{s+\Delta t} = x)$$

per stati $x, y \in E$.

Processi stazionari

Un processo $(X_t)_{t \in \mathcal{T}}$ si dice **stazionario** se, per ogni $\Delta t \geq 0$, la legge (congiunta) del processo coincide con quella del “traslato” $(X_{t+\Delta t})_{t \in \mathcal{T}}$ (purché i tempi $t + \Delta t$ appartengano a \mathcal{T}).

- Più precisamente, per ogni $k \geq 1$ e $t_1, t_2, \dots, t_k \in \mathcal{T}$ e $\Delta t \geq 0$, la legge congiunta di $(X_{t_1}, \dots, X_{t_k})$ coincide con quella di $(X_{t_1+\Delta t}, \dots, X_{t_k+\Delta t})$, purché i tempi $t_i + \Delta t$ appartengano a \mathcal{T} .

Processi stazionari

Un processo $(X_t)_{t \in \mathcal{T}}$ si dice **stazionario** se, per ogni $\Delta t \geq 0$, la legge (congiunta) del processo coincide con quella del “traslato” $(X_{t+\Delta t})_{t \in \mathcal{T}}$ (purché i tempi $t + \Delta t$ appartengano a \mathcal{T}).

- Più precisamente, per ogni $k \geq 1$ e $t_1, t_2, \dots, t_k \in \mathcal{T}$ e $\Delta t \geq 0$, la legge congiunta di $(X_{t_1}, \dots, X_{t_k})$ coincide con quella di $(X_{t_1+\Delta t}, \dots, X_{t_k+\Delta t})$, purché i tempi $t_i + \Delta t$ appartengano a \mathcal{T} .
- In particolare, nel caso di stati discreti, vale

$$P(X_{t_1} = x_1, \dots, X_{t_k} = x_k) = P(X_{t_1+\Delta t} = x_1, \dots, X_{t_k+\Delta t} = x_k),$$

per qualsiasi scelta di stati $x_1, \dots, x_k \in E$.

Processi stazionari

Un processo $(X_t)_{t \in \mathcal{T}}$ si dice **stazionario** se, per ogni $\Delta t \geq 0$, la legge (congiunta) del processo coincide con quella del “traslato” $(X_{t+\Delta t})_{t \in \mathcal{T}}$ (purché i tempi $t + \Delta t$ appartengano a \mathcal{T}).

- Più precisamente, per ogni $k \geq 1$ e $t_1, t_2, \dots, t_k \in \mathcal{T}$ e $\Delta t \geq 0$, la legge congiunta di $(X_{t_1}, \dots, X_{t_k})$ coincide con quella di $(X_{t_1+\Delta t}, \dots, X_{t_k+\Delta t})$, purché i tempi $t_i + \Delta t$ appartengano a \mathcal{T} .
- In particolare, nel caso di stati discreti, vale

$$P(X_{t_1} = x_1, \dots, X_{t_k} = x_k) = P(X_{t_1+\Delta t} = x_1, \dots, X_{t_k+\Delta t} = x_k),$$

per qualsiasi scelta di stati $x_1, \dots, x_k \in E$.

- Nel caso continuo l'identità sopra vale per le densità continue (scrivendo la densità p al posto della probabilità P).

Processi stazionari

Un processo $(X_t)_{t \in \mathcal{T}}$ si dice **stazionario** se, per ogni $\Delta t \geq 0$, la legge (congiunta) del processo coincide con quella del “traslato” $(X_{t+\Delta t})_{t \in \mathcal{T}}$ (purché i tempi $t + \Delta t$ appartengano a \mathcal{T}).

- Più precisamente, per ogni $k \geq 1$ e $t_1, t_2, \dots, t_k \in \mathcal{T}$ e $\Delta t \geq 0$, la legge congiunta di $(X_{t_1}, \dots, X_{t_k})$ coincide con quella di $(X_{t_1+\Delta t}, \dots, X_{t_k+\Delta t})$, purché i tempi $t_i + \Delta t$ appartengano a \mathcal{T} .
- In particolare, nel caso di stati discreti, vale

$$P(X_{t_1} = x_1, \dots, X_{t_k} = x_k) = P(X_{t_1+\Delta t} = x_1, \dots, X_{t_k+\Delta t} = x_k),$$

per qualsiasi scelta di stati $x_1, \dots, x_k \in E$.

- Nel caso continuo l'identità sopra vale per le densità continue (scrivendo la densità p al posto della probabilità P).
- Se un processo X è stazionario, necessariamente tutte le leggi delle marginali X_t coincidono: basta usare $k = 1$ nella definizione sopra.