

Statistica - Introduzione a



a cura di Antonio Iovanella

`iovanella@disp.uniroma2.it`

`http://www.disp.uniroma2.it/Users/iovanella`

Massima verosimiglianza

Un esempio

Massima verosimiglianza

ESEMPIO:

I dati provenienti da rilevazioni per il monitoraggio ambientale suscitano in genere domande alle quali non possiamo dare risposta mediante semplici metodi descrittivi.

Per tentare di studiare un problema del genere possiamo usare un dataset di R che mostra la serie storica delle concentrazioni medie di ozono rilevate a New York nel periodo maggio-settembre 1973.

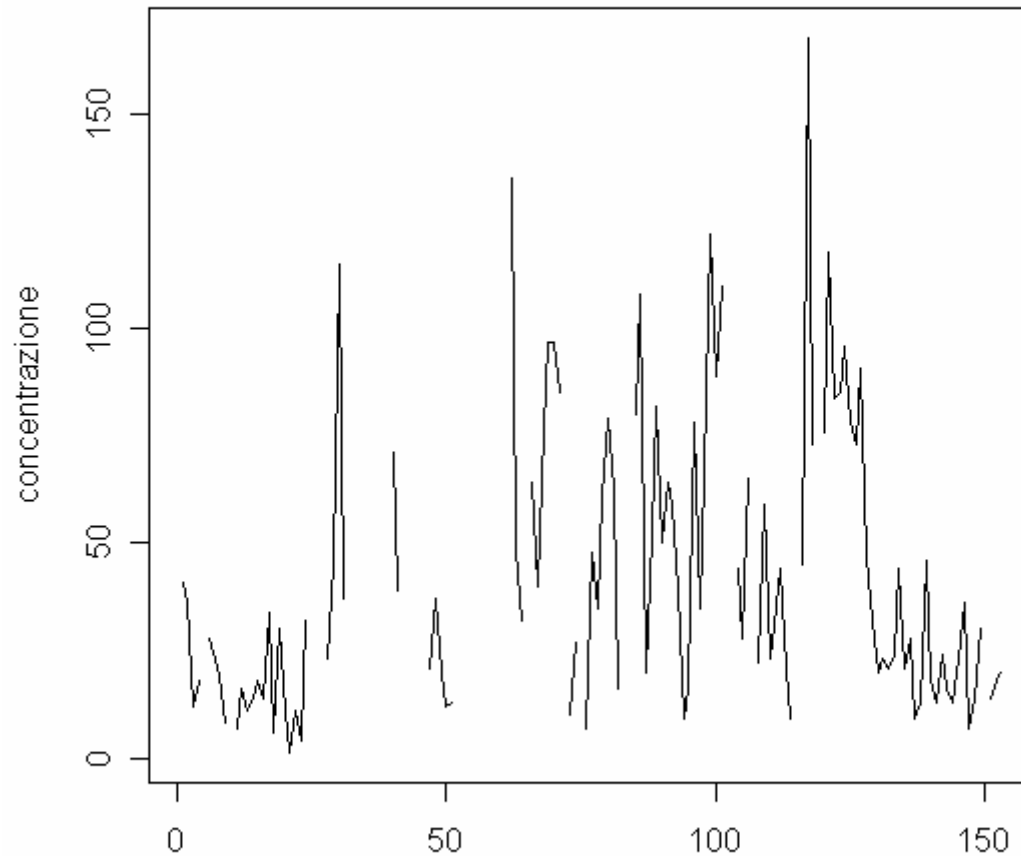
```
##CARICAMENTO DATI
```

```
data(airquality)
```

```
plot(airquality[,1], type = "title(main = "Ozono: medie  
giornaliere ", maggio-settembre 1973"))
```

Massima verosimiglianza

Ozono: medie giornaliere



New-York:
maggio-settembre 1973

Massima verosimiglianza

Domande:

- e possibile ricostruire la serie stimando i valori mancanti?
- è possibile prolungare la serie in modo da poter prevedere l'andamento dell'ozono nel futuro?
- se x_c è un valore critico oltre il quale la concentrazione di ozono può provocare danni alla salute, è possibile stimare quante volte tale soglia sarà superata nel corso di un anno?
- è sufficientemente lunga la serie riportata nella figura precedente per rispondere alle domande precedenti?

Per rispondere a domande come queste sarebbe necessaria una completa e certa conoscenza del processo generatore dei dati.

Massima verosimiglianza

Supponiamo che dalle osservazioni su 10 giorni ci sia stato un superamento della soglia ammissibile in tre giorni diversi. Per esempio, la sequenza potrebbe essere la seguente:

$$x_0 = (0; 1; 0; 1; 0; 0; 1; 0; 0; 0)$$

La probabilità di osservare ciò che è stato effettivamente osservato è data da:

$$p(x_0) = \theta^3(1 - \theta)^7$$

Come risolverlo?

Massima verosimiglianza

Supponiamo che dalle osservazioni su 10 giorni ci sia stato un superamento della soglia ammissibile in tre giorni diversi. Per esempio, la sequenza potrebbe essere la seguente:

$$x_0 = (0; 1; 0; 1; 0; 0; 1; 0; 0; 0)$$

La probabilità di osservare ciò che è stato effettivamente osservato è data da:

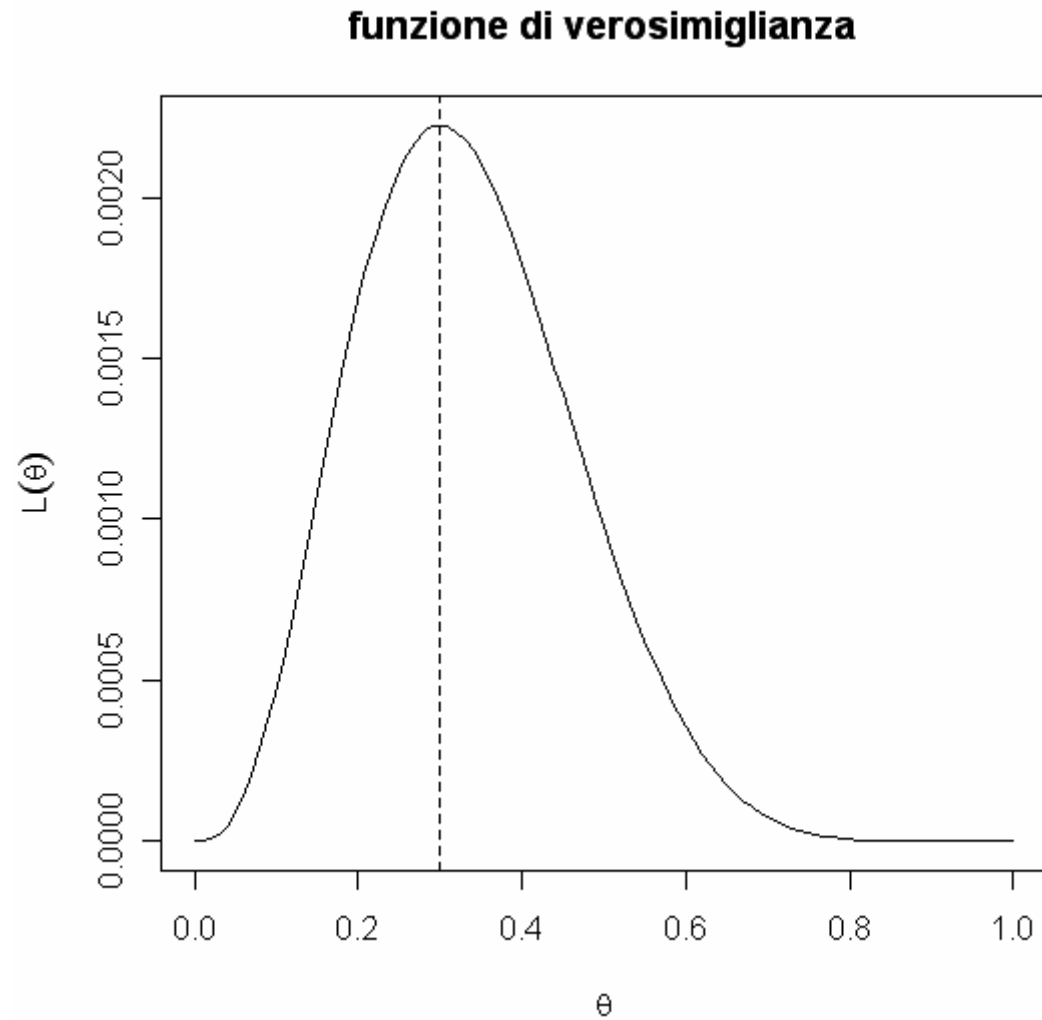
$$p(x_0) = \theta^3(1 - \theta)^7$$

```
Theta<- seq(0,1,by=0.01)

plot(theta, theta^(3)*(1-theta)^(7), xlab =
expression(theta), ylab = expression(L(theta)), type = "l")

abline(v = 0.3, lty=2, title(main = "funzione di
verosimiglianza"))
```

Massima verosimiglianza



Massima verosimiglianza

Il fatto che tale funzione non si annulli mai in $(0; 1)$ indica che sulla base della sequenza osservata, nessun valore di θ può essere a priori escluso. Tuttavia, la funzione indica che esistono valori di θ che rendono la probabilità di osservare la sequenza osservata estremamente bassa, mentre altri valori (nell'intorno di 0.3) rendono tale probabilità estremamente alta: questi ultimi sono allora valori “più verosimili” e 0.3 è il valore più verosimile tra tutti.

Massima verosimiglianza

Se x è un campione di T osservazioni, la funzione di verosimiglianza è, nel nostro caso:

$$L(\theta) = \theta^{T\bar{x}} (1 - \theta)^{T - T\bar{x}}$$

passando ai logaritmi:

$$l(\theta) = T\bar{x} \log \theta + (T - T\bar{x}) \log(1 - \theta)$$

la soluzione è:

$$\frac{\partial}{\partial \theta} l(\theta) = \frac{T\bar{x}}{\theta} - \frac{(T - T\bar{x})}{1 - \theta} = 0 \Rightarrow \hat{\theta}(x) = \bar{x}$$

Massima verosimiglianza

Nell'esempio che abbiamo considerato, il principio di massima verosimiglianza ci porta a considerare uno stimatore piuttosto intuitivo e così naturale da mettere in dubbio l'efficacia dell'impalcatura formale che abbiamo introdotto.

Tuttavia, è bene precisare subito che, soprattutto quando abbiamo a che fare con modelli parametrici più complessi, la forma dello stimatore è spesso tutt'altro che ovvia.