Esercizi - Inferenza 1

E9.1 Siano date le coppie $(x_1, y_1) = (5, -2)$, $(x_2, y_2) = (-3, -2)$, $(x_3, y_3) = (-2, -1)$, $(x_4, y_4) = (0, 1)$ ottenute da un modello lineare con rumore Gaussiano (con media $\mu = 0$ e varianza σ^2) indipendente su ciascun campione. Utilizzando il principio di massima verosimiglianza, calcolare una funzione lineare $y = a_* x + b_*$ che descriva tali punti. Si calcoli, inoltre, il valore dello stimatore nei punti $x = 19$, $x = -19$ e $x = 38$.

Soluzione:
L'equazione che governa le coppie (x_i, y_i) è la seguente

$$y_i = ax_i + b + \epsilon_i,$$

che può essere riscritta come

$$\epsilon_i = y_i - ax_i - b.$$

L'ipotesi di rumore gaussian implica che la funzione di verosimiglianza prenda la seguente forma

$$L(a, b, \sigma) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\epsilon_i^2/2\sigma^2}.$$

Applicando il logaritmo si ottiene

$$\log L(a, b, \sigma) = \sum_i [-2\log(2\pi) - 2\log(\sigma^2) - \frac{1}{2\sigma^2}\epsilon_i^2].$$

Le condizioni $\frac{\partial}{\partial a} \log L(a, b, \sigma) = 0$ e $\frac{\partial}{\partial b} \log L(a, b, \sigma) = 0$ sono equivalenti a

$$\frac{\partial}{\partial a}[\epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + \epsilon_4^2] = 0,$$

e

$$\frac{\partial}{\partial b}[\epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2 + \epsilon_4^2] = 0.$$

La prima diventa

$$(-2 - 5a - b)^2 + (-2 + 3a - b)^2 + (-1 + 2a - b)^2 + (1 - b)^2 = 0$$

da cui si ottiene $a^* = -1/19$. Risolvendo la seconda rispetto a b si ottiene $b^* = -1$. La legge lineare diventa quindi

$$f(x) = -\frac{1}{19}x - 1,$$

e può essere utilizzata per calcolare $f(\pm 19)$ e $f(38)$.

E9.2 Calcolare lo stimatore di massima verosimiglianza della media di n campioni i.i.d. da una distribuzione esponenziale.

Soluzione:
Data la pdf

$$f_\lambda = \lambda e^{-\lambda x},$$

il logaritmo della funzione di verosimiglianza rispetto a n campioni i.i.d. è

$$\log L(\lambda) = \log \prod_i f_\lambda(x_i) = \sum_i (\log \lambda - \lambda x_i).$$

Ponendo la derivata rispetto a λ uguale a zero si ottiene

$$\frac{n}{\lambda} - \sum_i x_i = 0,$$

e quindi $1/\lambda^* = n^{-1} \sum_i x_i$, ovvero la media campionaria.

E9.3 Sia data una sequenza di campioni indipendenti $(x_1, \ldots, x_n)$ campionati dalla stessa distribuzione X con media $E[X] = \mu$ e varianza $Var[X] = \sigma^2$. Dato lo stimatore $\hat{\sigma}^2(x_1, \ldots, x_n) = \frac{1}{n}\left(\sum_{i=1}^n x_i\right)^2 - x_2 x_3$, calcolarne la distorsione. Se risulta distorto, modificarlo in modo da ottenere uno stimatore corretto.

Soluzione:

$$\begin{aligned}
E\hat{\sigma}^2 &= E\left[\frac{1}{n}\left(\sum_i x_i\right)^2 - x_2 x_3\right] = \frac{1}{n}E\left[\left(\sum_i x_i\right)^2\right] - E\left[x_2 x_3\right] \\
&= \frac{1}{n}Var\left(\sum_i x_i\right) + \frac{1}{n}\left(E\left[\sum_i x_i\right]\right)^2 - E\left[x_2 x_3\right] \\
&= \frac{1}{n}\sum_i Var(x_i) + \frac{1}{n}\left(\sum_i E[x_i]\right)^2 - E[x_2]E[x_3] \\
&= \frac{1}{n}\sum_i \sigma^2 + \frac{1}{n}\left(\sum_i \mu\right)^2 - \mu^2 \\
&= \frac{1}{n}n\sigma^2 + \frac{1}{n}n^2\mu^2 - \mu^2 \\
&= \sigma^2 + (n-1)\mu^2.
\end{aligned}$$

Lo stimatore è distorto. Con la seguente modifica $x_2x_3 \to nx_2x_3$, il secondo termine non contribuisce e lo stimatore diventa corretto.