APPLICAZIONI
DI
INFERENZA STATISTICA

ESERCIZIO 1

Le ore di costante funzionamento giornaliero di cinque lampadine speciali (l_i, i = 1, …, 5) campionate indipendentemente l'una dalle altre (c.c.s.) da una distribuzione di tipo Gamma (2; λ) sono state:

l_1	l_2	l_3	l_4	l_5
7.76	19.38	14.99	2.56	10.64

1) Si determini e rappresenti graficamente la funzione di log-verosimiglianza;
2) si individui la stima di massima verosimiglianza per λ;
3) si individuino gli stimatori di $\mu(X)$ e di $\sigma^2(X)$;
4) si dimostri che lo stimatore di $\mu(X)$ è corretto.

Si determinino, inoltre:

5) la distribuzione esatta e approssimata dello stimatore media campionaria e i relativi intervalli di confidenza.

Soluzione 1)

Si ricorda che la funzione di densità di una v.a. Gamma di parametri $r = 2$ e λ è (*Probabilità e statistica metodologica*, formula IV.3.28):

$$f(X) = \begin{cases} \lambda^2 e^{-\lambda x} x & x \geq 0 \\ 0 & x < 0 \end{cases}.$$

La funzione di verosimiglianza (*Probabilità e statistica metodologica*, formula IX.2.1) risulta essere:

$$L[r;\lambda] = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{\lambda^r e^{-\lambda x_i} x^{r-1}}{\Gamma(r)}$$

$$\Downarrow$$

$$L[2;\lambda] = \prod_{i=1}^{5} \lambda^2 e^{-\lambda x_i} x_i = \lambda^{10} e^{-55,33\lambda} 61404,29$$

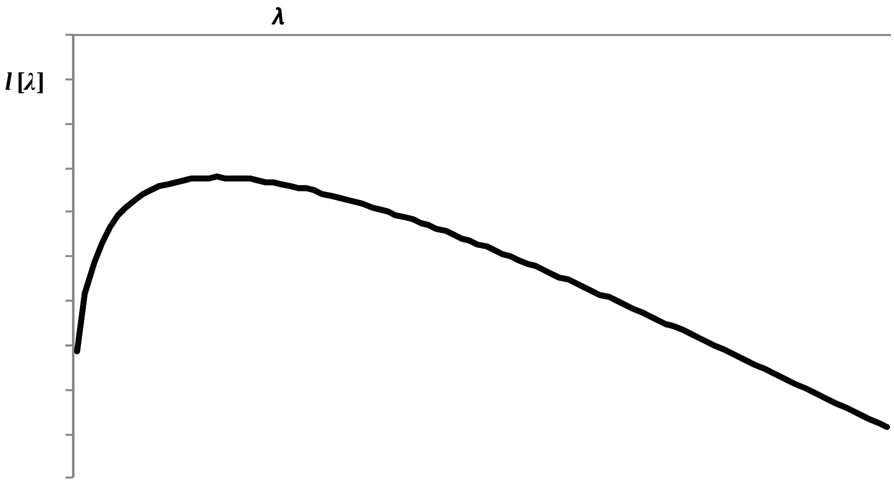
da cui la funzione di log-verosimiglianza (*Probabilità e statistica metodologica*, formula IX.2.3):

$$l[2;\lambda] = \ln\left(\prod_{i=1}^{5} \lambda^2 e^{-\lambda x_i} x_i\right) =$$

$$= \ln\left(61404,29 \cdot \lambda^{10} e^{-55,33\lambda}\right) = \ln 61404,29 + \ln \lambda^{10} - 55,33\lambda =$$

$$= 10 \ln \lambda - 55,33\lambda + 11,025\,.$$

La rappresentazione grafica è la seguente:



Soluzione 2)

La stima di massima verosimiglianza viene ricavata eguagliando a zero la derivata prima della funzione di log-verosimiglianza (*Probabilità e statistica metodologica*, formula IX.2.12):

$$l_*[\lambda] = 0$$

$$\Downarrow$$

$$\frac{10}{\lambda} - 55,33 = 0\,.$$

La soluzione dell'equazione porta a:

$$\hat{\lambda} = \frac{10}{55,33} = 0,18\,.$$

Soluzione 3)

Individuata la stima di λ e ricordando le formulazioni della media e della varianza della v.a. Gamma (*Probabilità e statistica metodologica*, formule IV.3.37-38), si possono ricavare le stime di queste ultime come segue:

$$\hat{\mu}_X = \frac{r}{\hat{\lambda}} = \frac{2}{0{,}18} = 11{,}11$$

$$\hat{\sigma}_X^2 = \frac{r}{\hat{\lambda}^2} = \frac{2}{0{,}18^2} = 61{,}73 \,.$$

Soluzione 4)

Dalle soluzioni 2) e 3) si deduce che la stima di λ è:

$$\hat{\lambda} = \frac{2n}{\sum_{i=1}^{n} x_i}$$

$$\hat{\mu}_X = \frac{2}{\hat{\lambda}} = \frac{2}{\frac{2n}{\sum_{i=1}^{n} x_i}} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\hat{\sigma}_X^2 = \frac{2}{\hat{\lambda}^2} = \frac{2}{\left(\frac{2}{\sum_{i=1}^{n} x_i}\right)^2} = \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{2n^2} = \frac{\hat{\mu}_X}{2} \quad .$$

Ne segue che affinché gli stimatori per la media e la varianza siano corretti deve risultare:

$$\mathrm{M}\left(\hat{\mu}_X\right) = \frac{2}{\lambda}$$

$$\mathrm{M}\left(\hat{\sigma}_X^2\right) = \frac{2}{\lambda^2} \;.$$

Essendo $\hat{\mu}_X$ la stima della media aritmetica nel campionamento casuale semplice, è sicuramente corretta.

Il valore atteso della stima della varianza è:

$$\mathrm{M}\left(\hat{\sigma}_X^2\right)=\mathrm{M}\left(\frac{2}{\hat{\lambda}^2}\right)=\mathrm{M}\left[\frac{\left(\sum_{i=1}^{n}x_i\right)^2}{2n^2}\right]=\frac{1}{2}\mathrm{M}\left(\frac{\sum_{i=1}^{n}x_i}{n}\right)^2.$$

Si noti che la quantità $\left(\sum_{i=1}^{n}x_i\right)^2$ è pari a:

$$\left(x_1^2+x_2^2+\ldots+2x_1x_2+\ldots+2x_1x_n+\ldots+2x_{n-1}x_n\right).$$

Conseguentemente, nel momento in cui si deve calcolare il valore atteso della suddetta somma è necessario ricavare quello dei suoi singoli addendi. Ciò implica che il calcolo di $\mathrm{M}\left(\sum_{i=1}^{n}x_i\right)^2$ richiede la determinazione di nuove variabili aleatorie. Si può ovviare a tale problema tramite simulazione computazionale; tuttavia, la numerosità campionaria non è sufficientemente alta da ottenere risultati soddisfacenti al fine di dare un giudizio sulla correttezza dello stimatore della varianza.

Soluzione 5)

La distribuzione esatta della media campionaria:

$$\hat{\mu}_X=\frac{\sum_{i=1}^{n}x_i}{n}$$

si ottiene come somma di *n* v.a. Gamma $\gamma(r+s,\lambda)$ divise per *n*. Dal momento che se $X\sim\gamma(r,\lambda)$ e $Y\sim\gamma(s,\lambda)$ $\Rightarrow Z=\left(X+Y\right)\sim\gamma(r+s,\lambda)$, la somma di cinque variabili Gamma $\gamma(r,\lambda)$ identicamente distribuite è una $\gamma(5r,\lambda)$. Pertanto, definita $\Sigma=X_1+X_2+X_3+X_4+X_5$ si ha che $\Sigma\sim\gamma\left(10,\lambda\right)$. La v.a. Media Campionaria è dunque:

$$\overline{X}=\frac{\Sigma}{5}$$

con distribuzione $\gamma(5r,5\lambda)$:

$$f(\overline{X})=\frac{\left(5\lambda\right)^{5r}e^{-5\lambda\overline{x}}\,\overline{x}^{5r-1}}{\Gamma(5r)}.$$

Pertanto l'intervallo di confidenza esatto a un livello $1-\alpha\%$ si ottiene nel seguente modo:

$$\int_0^{Inf_{esatto}} f\left(\overline{x}\right) d\ \overline{x} = \alpha/2$$
$$\int_{Sup_{esatto}}^{+\infty} f\left(\overline{x}\right) d\ \overline{x} = 1-\alpha/2 \ .$$

Ipotizzando un valore di α pari al 5%, dalle tavole della v.a. Gamma di parametri $5r = 10$ e $5\lambda = 5 \cdot 0{,}18 = 0{,}9$ si ottiene rispettivamente per l'estremo inferiore 3,345 e per l'estremo superiore 17,998.

Ai fini di individuare l'intervallo di confidenza approssimato si ipotizza per la media campionaria la distribuzione *t* di Student con 4 gradi di libertà. L'intervallo, ipotizzando $\alpha = 0{,}05$, è dunque:

$$\overline{x} - t_{g;\alpha/2} \cdot \overline{s} < \mu < \overline{x} + t_{g;\alpha/2} \cdot \overline{s}$$
$$\Downarrow$$
$$11{,}11 - 2{,}77 \cdot 7{,}856 < \mu < 11{,}11 + 2{,}77 \cdot 7{,}856$$
$$\Downarrow$$
$$-10{,}65 < \mu < 32{,}87 \, .$$

Si osservi che l'intervallo di confidenza approssimato è molto differente da quello esatto. In particolare, poichè $x > 0$ non ha senso che l'intervallo di confidenza comprenda valori negativi. Inoltre, l'intervallo esatto ha un'ampiezza molto minore di quello approssimato.

L'inesattezza dell'intervallo di confidenza approssimato è legata alla bassa numerosità campionaria ($n = 5$) in conseguenza della quale non ci si può appellare al Teorema del Limite Centrale.

ESERCIZIO 2

Le lunghezze l_i (i = 1, ..., 6) in cm di 6 tavole estratte casualmente dalla produzione di un macchinario che segue una legge Normale $N\left(120,\sigma\right)$ sono:

l_1	l_2	l_3	l_4	l_5	l_6
130.44	137.20	113.92	130.21	110.81	120.86

1) si determini e rappresenti graficamente la funzione di log-verosimiglianza $l(\sigma)$;
2) si individui la stima di massima verosimiglianza per σ;
3) si valuti la correttezza dello stimatore per σ;
4) si determini, inoltre l'intervallo di confidenza per la varianza.

Soluzione 1)

La funzione di verosimiglianza (*Probabilità e statistica metodologica*, formula IX.2.1) risulta:

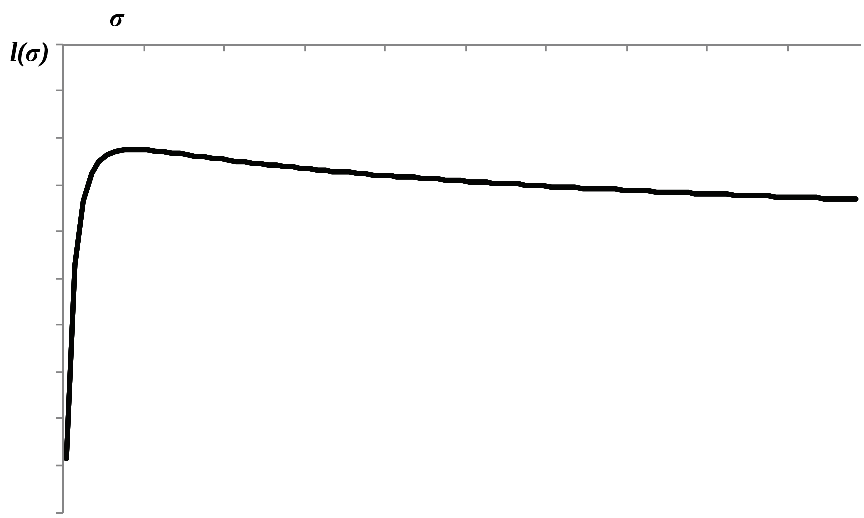
$$L(\sigma) = \prod_{i=1}^{6} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x_i-120}{\sigma}\right)^2} = \left(\frac{1}{\sqrt{2\pi}}\right)^6 \left(\frac{1}{\sigma^2}\right)^3 e^{-\frac{1}{2}\sum_{i=1}^{6}\left(\frac{x_i-120}{\sigma}\right)^2} =$$

$$= \frac{1}{8\pi^3\sigma^6} e^{-\frac{631,24}{2\sigma^2}} .$$

Ne segue la funzione di log-verosimiglianza (*Probabilità e statistica metodologica*, formula IX.2.3):

$$l(\sigma) = \ln\left(\frac{1}{8\pi^3}\frac{1}{\sigma^6} e^{-\frac{631,24}{2\sigma^2}}\right) = \ln\frac{1}{8\pi^3} - 3\ln\sigma^2 - \frac{315,62}{\sigma^2} .$$

Graficamente:



Soluzione 2)

Per individuare la stima di massima verosimiglianza per σ si pone uguale a zero la derivata prima della funzione di log-verosimiglianza (*Probabilità e statistica metodologica*, formula IX.2.12):

$$l'(\sigma) = -\frac{6}{\sigma} + \frac{631,24}{\sigma^3} = 0$$
$$\Downarrow$$
$$\hat{\sigma}\left(6\hat{\sigma}^2 - 631,24\right) = 0$$
$$\Downarrow$$
$$\hat{\sigma}^2 = 105,21$$
$$\Downarrow$$
$$\hat{\sigma} = 10,26 .$$

Soluzione 3)

Dalla soluzione 2) si osserva che:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n} .$$

Tale formulazione, come si evince dalle nozioni di inferenza statistica, non porta a una stima corretta della varianza.

Soluzione 4)

L'intervallo di confidenza risulta:

$$\frac{\hat{\sigma}^2}{\chi^2_{n-1;\alpha/2}} < \sigma^2 < \frac{\hat{\sigma}^2}{\chi^2_{n-1;1-\alpha/2}} .$$

Ipotizzando $\alpha = 0,05$ si ottiene:

$$\frac{631,24}{12,8325} < \sigma^2 < \frac{631,24}{0,8312}$$
$$\Downarrow$$
$$49,19 < \sigma^2 < 759,43 .$$