ESERCIZI DI STATISTICA

Soluzioni degli esercizi sugli intervalli di confidenza.

A cura di Nazareno Maroni

Soluzione dell'esercizio 1. La funzione di densità di Y è data da

$$f_Y(y) = \frac{1}{\theta} e^{-\frac{y}{\theta}}, \quad y \geqslant 0.$$

Possiamo vedere che $U = \frac{Y}{\theta}$ ha distribuzione esponenziale di parametro 1 e non dipende, quindi, da θ.

$$f_U(u) = f_Y(\theta u) \left| \frac{\mathrm{d}}{\mathrm{d}u}(\theta u) \right| = \frac{1}{\theta} e^{-\frac{\theta u}{\theta}} \theta = e^{-u}, \quad u \geqslant 0.$$

Usiamo U come quantità pivotale. Dobbiamo trovare a e b tali che $\mathbb{P}(a \leqslant U \leqslant b) = 0,90$. Un modo può essere scegliere a come il quantile di livello $0,05$ (il punto t.c. $\mathbb{P}(U \leqslant a) = 0,05$) e b come il quantile di livello $0,95$. Abbiamo che

$$\mathbb{P}(U \leqslant u) = \int_0^u e^{-x} \, \mathrm{d}x = -e^{-x} \Big|_0^u = 1 - e^{-u},$$

e quindi poniamo $1 - e^{-a} = 0,05$ e $e^{-b} = 0,05$, ovvero $a = 0,051$ e $b = 2,996$. Quindi

$$0,90 = \mathbb{P}(0,051 \leqslant U \leqslant 2,996) = \mathbb{P}\left(0,051 \leqslant \frac{Y}{\theta} \leqslant 2,996\right) =$$
$$= \mathbb{P}\left(\frac{0,051}{Y} \leqslant \frac{1}{\theta} \leqslant \frac{2,996}{Y}\right) = \mathbb{P}\left(\frac{Y}{2,996} \leqslant \theta \leqslant \frac{Y}{0,051}\right).$$

Notate che $\mathbb{P}(Y > 0) = 1$ e quindi possiamo dividere per Y senza dover cambiare il verso delle disuguaglianze.

Soluzione dell'esercizio 2. Sappiamo che

$$f_Y(y) = \frac{\beta^2}{\Gamma(2)} y e^{-\beta y},$$

calcoliamo, quindi, la densità di Z tramite il cambio di variabili:

$$f_Z(z) = f_Y\left(\frac{z}{2\beta}\right) \left| \frac{\mathrm{d}}{\mathrm{d}z} \frac{z}{2\beta} \right| = \beta^2 \frac{z}{2\beta} e^{-\beta \frac{z}{2\beta}} \frac{1}{2\beta} = \frac{z}{4} e^{-\frac{z}{2}} = \frac{1}{\Gamma\left(\frac{4}{2}\right) 2^{\frac{4}{2}}} z^{\frac{4}{2}-1} e^{-\frac{z}{2}} = f_{\chi^2_4}(z).$$

Calcoliamo la funzione generatrice dei momenti di $Z = 2\beta Y$:

$$\mathbb{E}\left(e^{t2\beta Y}\right) = \int_0^{+\infty} \beta^2 y e^{y(2\beta t - \beta)} \, \mathrm{d}y = \frac{\beta^2 y}{2\beta t - \beta} e^{y(2\beta t - \beta)} \Big|_0^{+\infty} +$$
$$- \int_0^{+\infty} \frac{\beta^2}{2\beta t - \beta} e^{y(2\beta t - \beta)} \, \mathrm{d}y = -\frac{\beta^2}{(2\beta t - \beta)^2} e^{y(2\beta t - \beta)} \Big|_0^{+\infty} =$$
$$= \frac{\beta^2}{(2\beta t - \beta)^2} = \frac{1}{(1 - 2t)^2} = \left(\frac{1}{1 - 2t}\right)^{\frac{4}{2}} \qquad t < \frac{1}{2}$$

e questa è la f.g.m. di una χ^2_4.
Troviamo l'intervallo di confidenza di livello $0,90$:

$$0,90 = \mathbb{P}(\chi^2_{4,0,05} \leqslant Z \leqslant \chi^2_{4,0,95}) = \mathbb{P}(0,711 \leqslant 2\beta Y \leqslant 9,49) = \mathbb{P}\left(\frac{0,711}{2Y} \leqslant \beta \leqslant \frac{9,49}{2Y}\right).$$

Soluzione dell'esercizio 3. Si deve calcolare la distribuzione del massimo di n variabili uniformi su $(0,\theta)$ ed è

$$F_{Y_{(n)}}(y) = \left(\frac{y}{\theta}\right)^n, \qquad f_{Y_{(n)}}(y) = \frac{n}{\theta^n} y^{n-1}.$$

a) Calcoliamo prima la densità di U e poi la integriamo per ottenere la funzione di distribuzione.
$$f_U(u) = \frac{n}{\theta^n}(\theta u)^{n-1}\theta = nu^{n-1}$$
Ovviamente è $0 \leqslant U \leqslant 1$, poiché $0 \leqslant Y_i \leqslant \theta \ \forall i$, quindi $0 \leqslant Y_{(n)} \leqslant \theta$ e quindi $0 \leqslant U = \frac{Y_{(n)}}{\theta} \leqslant 1$. Integrando la densità si ha:
$$F_U(u) = \int_0^u nx^{n-1}\,\mathrm{d}x = x^n\Big|_0^u = u^n, \quad 0 \leqslant u \leqslant 1.$$

b) Sappiamo che
$$\mathbb{P}(U \leqslant a) = a^n = 0,95 \quad \Leftrightarrow \quad a = \sqrt[n]{0,95}.$$
Quindi
$$0,95 = \mathbb{P}(U \leqslant \sqrt[n]{0,95}) = \mathbb{P}\left(\frac{Y_{(n)}}{\theta} \leqslant \sqrt[n]{0,95}\right) = \mathbb{P}\left(\frac{Y_{(n)}}{\sqrt[n]{0,95}} \leqslant \theta\right).$$

Soluzione dell'esercizio 4. Stimiamo la differenza $p_1 - p_2$ tramite la differenza $\hat{p}_1 - \hat{p}_2$ delle proporzioni di fallimento nei campioni. Scriviamo

$$\hat{p}_1 = \frac{1}{n_1}\sum_{i=1}^{n_1} Y_i^{(1)} \qquad \text{e} \qquad \hat{p}_2 = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i^{(2)}$$

dove $Y_i^{(1)} \sim Bernoulli(p_1, q_1 = 1 - p_1)$ e $Y_i^{(2)} \sim Bernoulli(p_2, q_2 = 1 - p_2)$. Quindi

$$\mathbb{E}(\hat{p}_1) = p_1, \qquad \mathbb{E}(\hat{p}_2) = p_2 \quad \text{e} \qquad \mathbb{E}(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$$
$$Var(\hat{p}_1) = \frac{p_1 q_1}{n_1}, \qquad Var(\hat{p}_2) = \frac{p_2 q_2}{n_2}$$

e

$$Var(\hat{p}_1 - \hat{p}_2) = Var(\hat{p}_1) + Var(\hat{p}_2) = \frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}$$

Sappiamo, per il teorema del limite centrale, che $\dfrac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{Var(\hat{p}_1 - \hat{p}_2)}}$ può essere approssimata con una $N(0,1)$ per campioni abbastanza grandi e che l'approssimazione resta valida se usiamo l'approssimazione per la varianza che si ottiene sostituendo $p_1 \leftrightarrow \hat{p}_1, q_1 \leftrightarrow \hat{q}_1, p_2 \leftrightarrow \hat{p}_2, q_2 \leftrightarrow \hat{q}_2$.

$$\begin{aligned}
0,98 &= \mathbb{P}\left(z_{0,01} \leqslant \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}} \leqslant -z_{0,01} \right) = \\
&= \mathbb{P}\left((\hat{p}_1 - \hat{p}_2) + z_{0,01}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \leqslant p_1 - p_2 \leqslant (\hat{p}_1 - \hat{p}_2) - z_{0,01}\sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}} \right) = \\
&= \mathbb{P}(-0,1452 \leqslant p_1 - p_2 \leqslant 0,2252),
\end{aligned}$$

dove si sono sostituiti i valori di $\hat{p}_1, \hat{q}_1, \hat{p}_2, \hat{q}_2, n_1, n_2$ e $z_{0,01}$ è il quantile di livello $0,01$ di una normale standard.