

# Analisi Statistica dei Dati: foglio di lavoro su modelli mistura e verosimiglianza profilo

Luca La Rocca

4 dicembre 2018

## Example 5.10

1. Create un vettore numerico `countData` contenente i dati  $z^\bullet$  presentati nella sezione 1.1.5 del testo di riferimento.
2. Scrivete una funzione `logLik` che, prendendo in input un vettore `ab` di lunghezza 2 e un vettore `z` di lunghezza  $N = 6$  e somma  $n = 196$ , restituisca in output la log-verosimiglianza del modello beta-binomiale troncato:

$$\ell(\alpha, \beta; z) \triangleq \sum_{k=1}^N z_k \log \frac{B(\alpha + k, \beta + N - k)}{B(\alpha, \beta)} - n \log \left\{ 1 - \frac{B(\alpha, \beta + N)}{B(\alpha, \beta)} \right\},$$

dove  $B(a, b) = \int_0^1 \pi^{a-1} (1 - \pi)^{b-1} d\pi$ ,  $a, b \in \mathbb{R}_+^*$ , è la funzione beta di Eulero ed  $\alpha, \beta$  sono rispettivamente il primo e secondo elemento di `ab`.

3. Scrivete una funzione `mr2ab` che, prendendo in input un vettore `mr` con componenti  $\mu = \alpha/(\alpha + \beta)$  e  $\rho = 1/(\alpha + \beta + 1)$ , restituisca in output un vettore con componenti  $\alpha = \mu(1 - \rho)/\rho$  e  $\beta = (1 - \mu)(1 - \rho)/\rho$ .
4. Scrivete una funzione `xi2mu` che, prendendo in input tre scalari `xi`, `psi` ed `N`, restituisca in output la soluzione  $\mu$  dell'equazione

$$\frac{B(\mu\psi, (1 - \mu)\psi + N)}{B(\mu\psi, (1 - \mu)\psi)} = \xi,$$

dove  $\psi = (1 - \rho)/\rho$  e  $\xi$  è la probabilità che  $N$  test falliscano su un malato.

5. Scrivete una funzione `logLikRep` che calcoli la log-verosimiglianza al punto 2 a partire da un vettore `xr` con componenti  $\xi$  e  $\rho$ .
6. Massimizzate numericamente la log-verosimiglianza al punto 5, per i dati al punto 1, usando la funzione `optim` e salvando l'output in un vettore `xrHat`.
7. Stimate il conteggio mancante come  $\hat{z}_0^\bullet = n\hat{\xi}^\bullet/(1 - \hat{\xi}^\bullet)$  e la sensibilità media come  $\hat{\mu}^\bullet$  ottenuta da  $\hat{\xi}^\bullet$  (oltre che  $\hat{\rho}^\bullet$  ed  $N$ ) tramite la funzione al punto 4.
8. Visualizzate i risultati ottenuto al punto 7 nella console di R.

9. Scrivete una funzione `logLikRhoXi` che calcoli la log-verosimiglianza al punto 5 prendendo un vettore `rho` come primo argomento, un vettore `xi` come secondo argomento e il vettore `z` dei dati come terzo argomento.
10. Scrivete una funzione `rhoHatCon` che, prendendo in input un vettore `xi` e il vettore dei dati `z`, restituisca in output la stima di massima verosimiglianza vincolata  $\hat{\rho}(\xi, z)$  necessaria per calcolare la log-verosimiglianza profilo di  $\xi$ .
11. Scrivete una funzione `logLikProXi` che, prendendo in input un vettore `xi` e il vettore dei dati `z`, restituisca in output la log-verosimiglianza profilo del parametro  $\xi$ .
12. Confrontate graficamente la log-verosimiglianza profilo e quella stimata di  $\xi$ , verificando come la log-verosimiglianza stimata, con la finzione che  $\rho$  sia esattamente pari a  $\hat{\rho}^\bullet$ , sottovaluti l'incertezza residua su  $\xi$ .

#### Testo di riferimento

L. Held & D. Sabanés Bové. *Applied Statistical Inference*. Springer, 2014.