20 Dicembre 2021

Esercitazione n. 4
Metodologie di Analisi Dati

Esercizio 1

Questo esercizio riguarda il fit di maximum likelihood (ML) e si basa sul programma di minimizzazione `MINUIT`, utilizzando la sua implementazione in `ROOT/C++ TMinuit`.

Scaricare il software necessario a questo link:

https://cernbox.cern.ch/s/QjGkDneHGvfytgy

Il programma fornito genera un campione di dati di $n = 200$ valori da una pdf che è data dalla somma di una funzione esponenziale e di una gaussiana:

$$f(x;\theta,\xi) = \theta \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2} + (1-\theta)\frac{1}{\xi}e^{-x/\xi} \qquad (x \geq 0)$$

La pdf viene modificata in modo da essere troncata nell'intervallo $0 \leq x \leq x_{max}$.

(a) Per impostazione predefinita, il programma `mlFit.cc` fissa i parametri μ e σ e tratta solo i parametri θ e ξ come liberi. Eseguendo il programma, produci i seguenti grafici:

- pdf risultante dal fit con i dati;
- un grafico dello "scan" di $-\log L$ verso θ;
- un "contour" plot di $\log L = \log L_{max} - 1/2$ nel piano (θ, ξ).

Dal grafico di $-\log L$ verso θ, mostra che la deviazione standard di $\hat{\theta}$ è uguale al valore stampato dal programma. Dal grafico di $\log L = \log L_{max} - 1/2$, mostra che le distanze degli stimatori di ML alle linee tangenti al contorno danno le stesse deviazioni standard $\sigma_{\hat{\theta}}$ e $\sigma_{\hat{\xi}}$ stampate dal programma.

(b) Ricordiamo che l'inversa della varianza della matrice di covarianza degli stimatori di ML $V_{ij} = \text{cov}[\hat{\theta}_i, \hat{\theta}_j]$ può essere approssimata nel limite di grande campione di dati da

$$V_{ij}^{-1} = -E\left[\frac{\partial^2 \log L}{\partial\theta_i \partial\theta_j}\right] = -\int \frac{\partial \log P(\boldsymbol{x}\,|\,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j} P(\boldsymbol{x}\,|\,\boldsymbol{\theta}) d\boldsymbol{x}$$

dove qui $\boldsymbol{\theta}$ rappresenta il vettore di tutti i parametri. Mostra che V^{-1} è proporzionale alla dimensione del campione n e quindi mostra che le deviazioni standard dei MLE di tutti i parametri diminuiscono di $1/\sqrt{n}$. (Suggerimento: scrivi la forma generale della probabilità per un campione indipendente e identicamente distribuito: $L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta})$. Non è necessario utilizzare la specifica $f(x; \boldsymbol{\theta})$ per questo problema.)

(c) Modificando la riga

```
numVal = 200
```

riesegui il programma per diverse dimensioni del campione di dati con $n = 100,\ 400,\ 800$ eventi e trova in ogni caso la deviazione standard di $\hat{\theta}$. Fai un plot di $\sigma_{\hat{\theta}}$ verso n per $n = 100,\ 400,\ 800$ e commenta come questo si pone in relazione a ciò che ti aspetti.

(d) Utilizzando le routine `TMinuit` `FixParameter` e `Release`, trova $\hat{\theta}$ e la sua deviazione standard $\sigma_{\hat{\theta}}$ nei seguenti quattro casi:

- θ libero e $\mu,\ \sigma,\ \xi$ fissati;
- θ e ξ liberi e μ, σ fissati;
- θ, ξ, μ liberi e σ fissati;
- θ, ξ, μ e σ tutti liberi.

Commentare come la deviazione standard $\sigma_{\hat{\theta}}$ dipende dal numero di parametri liberi nel fit.