Laboratorio di Statistica Computazionale - 5

Matteo Grigoletto

Argomenti di oggi

- Ottimizzazione.
  - Caso monoparametrico.
  - Caso multiparametrico.

Alcuni esercizi per oggi /1

- Esercizio n. 1. Sia $(Y_1, \ldots, Y_n)$ un campione casuale semplice estratto da una v.c. di Bernoulli con probabilità di successo p ignota. Supponiamo in particolare che in $n = 10$ prove siano stati ottenuti $w = \sum_{i=1}^{10} y_i = 4$ successi.
  - **a.** Si scriva la funzione di verosimiglianza e si ottenga con una procedura di ottimizzazione numerica la stima di massima verosimiglianza di p per i dati a disposizione.
  - **b.** Si costruisca un intervallo di confidenza per p basato sia sulla distribuzione asintotica dello SMV che sulla distribuzione asintotica del rapporto di verosimiglianza.

Alcuni esercizi per oggi /2 I

- Esercizio n. 2. Sia $(Y_1, \ldots, Y_n)$ un campione casuale semplice estratto da una v.c. Y con distribuzione binomiale negativa: $Y \sim \text{Bin.neg}(k, p)$, con k e p parametri ignoti. Qui Y rappresenta il numero di insuccessi che precedono il k-mo successo, in una successione di prove indipendenti con probabilità di successo p. La funzione di probabilità di Y è data da

$$p(y; k, p) = P(Y = y) = \binom{y+k-1}{k-1} p^k (1-p)^y$$

con $y = 0, 1, \ldots$; $k = 1, 2, \ldots$; $p \in (0, 1)$. Si noti che:

$$\text{E}(Y) = \frac{k(1-p)}{p} \quad \text{e} \quad \text{Var}(Y) = \frac{k(1-p)}{p^2}$$

Sia $(y_1, \ldots, y_n)$ una realizzazione campionaria, in cui i valori $0, 1, \ldots, 13$ sono stati osservati con frequenze $(6, 6, 6, 8, 6, 4, 5, 4, 1, 2, 1, 1, 1, 1)$, rispettivamente. Se n_r è la frequenza del valore r, la funzione di log-verosimiglianza è data da:

$$l(k, p) = \sum_{r=0}^{13} n_r \, log[p(r; k, p)] \, .$$

Alcuni esercizi per oggi /2 II

a. Si scriva una funzione che calcola la log-verosimiglianza. Nel farlo, si noti che, per come abbiamo definito la funzione di probabilità, k è un parametro *discreto*. Questo, per gli algoritmi di ottimizzazione (che lavorano nel continuo) è un grosso problema. Per ovviare all'inconveniente, notiamo che

$$\binom{y+k-1}{k-1} = \frac{(y+k-1)!}{(k-1)! \cdot y!}$$

è definito solo per k intero. Tuttavia,

$$\frac{\Gamma(y+k)}{\Gamma(k) \cdot y!}$$

coincide con $\binom{y+k-1}{k-1}$ per k intero, ma è definito anche per qualsiasi k reale positivo. Nel definire la log-verosimiglianza, si usi quindi quest'ultima espressione invece dell'usuale coefficiente binomiale.

b. Si tracci un grafico di $l(k, p)$ e si commenti il suo andamento.

c. Si trovino gli stimatori di k e p basati sul metodo dei momenti.

Alcuni esercizi per oggi /2 III

d. Si effettui il calcolo numerico delle stime di massima verosimiglianza di k e p mediante la funzione `optim`, adottando un'opportuna riparametrizzazione che permetta di tenere conto dei vincoli sullo spazio parametrico[1]. Come valori di partenza si usino le stime basate sul metodo dei momenti ottenute al punto precedente.

e. Si calcolino le SMV senza riparametrizzazioni, sfruttando la capacità della funzione `nlminb` di cercare l'ottimo in uno spazio parametrico vincolato.

f. Si traccino le regioni di confidenza per (k, p) basate sulle distribuzioni asintotiche dello SMV bidimensionale e del rapporto di verosimiglianza.

g. Si calcolino gli intervalli di confidenza per k e p, basati sulle distribuzioni asintotiche degli SMV e delle verosimiglianze profilo.

---

[1] Si usi, in particolare

$$\psi = \log(k-1) \qquad \text{e} \qquad \phi = \log\left(\frac{p}{1-p}\right) ,$$

da cui segue, ovviamente

$$k = 1 + e^{\psi} \qquad \text{e} \qquad p = \frac{e^{\phi}}{1+e^{\phi}} .$$

Alcuni esercizi per oggi /3

- Esercizio n. 3. La distribuzione Weibull ha densità

$$f(y;\alpha,\beta)=\frac{\beta}{\alpha}\left(\frac{y}{\alpha}\right)^{\beta-1}e^{-(y/\alpha)^{\beta}},\quad y\geq 0,\ \alpha>0,\ \beta>0\ .$$

Dunque, la funzione quantile è

$$Q(p;\alpha,\beta)=F^{-1}(p;\alpha,\beta)=\alpha\left[-\log(1-p)\right]^{1/\beta}\ .$$

Vengongo osservati i seguenti valori:

3, 5, 5, 13, 14, 15, 22, 22, 23, 30, 36, 39, 44,
46, 50, 72, 79, 88, 97, 102, 139, 188, 197, 210.

a. Si scriva la log-verosimiglianza come funzione dei parametri α e $q=Q(0.1;\alpha,\beta)$. Si usi tale log-verosimiglianza per calcolare la SMV di (α,q). Si consideri che valori iniziali appropriati per α e β sono, rispettivamente, 63 e 1.1.

b. Si usi il metodo della verosimiglianza profilo per calcolare gli intervalli di confidenza per α e per q.

Alcuni esercizi per oggi /4

- Esercizio n. 4. Si assuma che le v.c. $Y_1, \ldots, Y_n$ siano indipendenti e che Y_i, per $i = 1, \ldots, n$, abbia distribuzione di Bernoulli con probabilità di successo

$$\pi_i = \pi_i(x_i) = \frac{\mathrm{e}^{\alpha+\beta\, x_i}}{1+\mathrm{e}^{\alpha+\beta\, x_i}} \ ,$$

dove $x_1, \ldots, x_n$ sono valori noti, mentre α e β sono parametri reali ignoti[2]. Si osservano i valori

y_i	0	1	1	1	1	0	1	0	0	0	1	1
x_i	0	0	0	0	0	0	1	1	1	1	1	1

a. Si trovino le SMV di α e β[3].

b. Si tracci una regione di confidenza per (α, β) usando la distribuzione asintotica del rapporto di verosimiglianza.

---

[2]Si tratta di un modello di regressione logistica.

[3]Si noti che, nel caso presente, i valori x dicono semplicemente che la probabilità di successo cambia per 2 gruppi di individui. In questa situazione le SMV possono essere trovate anche esplicitamente. Per un andamento più complesso dei valori x, tuttavia, non esistono soluzioni esplicite. Si noti anche che, per la regressione logistica, l'origine $(\alpha, \beta) = (0, 0)$ è di solito un buon punto di partenza per la ricerca dell'ottimo.

Alcuni esercizi per oggi /5

- Esercizio n. 5. Le v.c. $Y_1, \ldots, Y_n$ siano i.i.d. con distribuzione di Cauchy, avente funzione di densità

$$f(y; \alpha, \beta) = \frac{1}{\pi\beta\left[1 + \left(\frac{y-\alpha}{\beta}\right)^2\right]} ,$$

con $y \in \mathbb{R}, \alpha \in \mathbb{R}$, $\beta \in \mathbb{R}^+$. In particolare, è noto che $\beta = 1$. Vengono osservati i dati contenuti nel file `cauchy.dat`.

  - Si usi `nlminb` per trovare la SMV di α, partendo da questi diversi valori iniziali:

    -11, -1, 0, 1.5, 4, 4.7, 7, 8, 38.

    Si commentino i risultati trovati.