
STATISTICA 1, metodi matematici e statistici

Introduzione al linguaggio R

Esercitazione 5: 22-04-2005

Luca Monno

Università degli studi di Pavia

`luca.monno@unipv.it`

`http://www.lucamonno.it`

Richiami di teoria

Disuguaglianza di Cramèr-Rao

In un problema di stima regolare sia $T(y)$ uno stimatore tale che esistono

$$a(\theta) = E[T(Y); \theta]$$

e la sua derivata

$$a'(\theta) = \frac{d}{d\theta} \int_{\mathcal{Y}} T(y)p(y; \theta)dy = \int_{\mathcal{Y}} T(y) \frac{d}{d\theta} p(y; \theta)dy$$

Allora, indicando con $I(\theta)$ la matrice di informazione di Fisher, se $0 < I(\theta) < \infty$

$$\text{var}[T(Y); \theta] \geq \{a'(\theta)\}^2 / I(\theta)$$

Distribuzione asintotica della SMV

Consideriamo un campione y_1, \dots, y_n *i.i.d.* con densità (o d.d.p.) $p(y; \theta)$ che soddisfa le condizioni dei problemi regolari di stima; sia $i(\theta)$ l'informazione attesa per una singola osservazione e sia $\hat{\theta}$ consistente e θ_* il vero valore del parametro

$$\sqrt{n}(\hat{\theta} - \theta_*) \xrightarrow{d} \mathcal{N}(0, i(\theta_*)^{-1})$$

ovvero

$$\hat{\theta} \xrightarrow{d} \mathcal{N}\left(\theta_*, \frac{i(\theta_*)^{-1}}{n}\right)$$

La stima di massima verosimiglianza è quindi asintoticamente corretta ed efficiente (infatti raggiunge il limite inferiore di Cramer-Rao)

Stima della media in una lognormale

Sia (Y_1, \dots, Y_n) un campione casuale estratto da una distribuzione lognormale

$$f(y; \mu, \sigma^2) = \frac{1}{y} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2}.$$

Ricordiamo che $Z = \log Y$ è normale con media μ e varianza σ^2 .

Consideriamo $\mu = 0$ e $\sigma^2 = 1$ e calcoliamo numericamente

$$\int_0^\infty f(y; \mu, \sigma^2) dy$$

```
> f <- function(x, mu, sigma) (1/sqrt(2 * pi * sigma^2 * x^2))  
+   exp(-0.5 * (log(x) - mu)^2/sigma^2)  
> integrate(f, lower = 0, upper = Inf, mu = 0, sigma = 1)
```

1 with absolute error < 2.5e-07

Si verifica analiticamente (provare per esercizio) che

$$\psi = E(Y) = \int_0^\infty y f(y; \mu, \sigma^2) dy = e^{\mu + \sigma^2/2}.$$

Scriviamo una funzione che restituisce la media di una v.a. lognormale attraverso l'integrazione numerica.

```
> media.f <- function(mu, sigma) {  
+   f1 <- function(x, mu, sigma) x * f(x, mu, sigma)  
+   integrate(f1, lower = 0, upper = Inf, mu = mu, sigma = sigma)  
+ }  
> media.f(1, 2)
```

```
[1] 20.08549
```

```
> exp(1 + 2^2/2)
```

```
[1] 20.08554
```

L'integrazione numerica sembra funzionare!!!. La utilizziamo anche per calcolare la varianza.

```

> varianza.f <- function(mu, sigma) {
+   f2 <- function(x, mu, sigma) (x - media.f(mu, sigma))^2 *
+     f(x, mu, sigma)
+   integrate(f2, lower = 0, upper = Inf, mu = mu, sigma = sigma)
+ }

```

Supponiamo che $\mu = 0$ e che vogliamo stimare la media $\psi = e^{\sigma^2/2}$. Uno stimatore non distorto per ψ è $T = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ e la sua varianza è pari alla varianza di Y diviso n .

Il primo obiettivo è confrontare la varianza di T con il limite inferiore di Cramer-Rao che è pari a (calcolare per esercizio) $\frac{\sigma^4 (e^{\sigma^2/2})^2}{2n}$. Scriviamo allora una funzione che restituisca il limite inferiore di Cramer-Rao per ψ in funzione di σ e n

```

> cr.lb <- function(n, sigma) {
+   sigma^4 * (exp(sigma^2/2))^2 / (2 * n)
+ }

```

Scriviamo anche una funzione che restituisca la varianza di T in funzione di σ e n

```
> var.T <- function(n, sigma) {  
+   varianza.f(mu = 0, sigma)/n  
+ }
```

Possiamo infine confrontare la varianza di T con il limite di Cramer-Rao. Consideriamo ad esempio

```
> n <- c(5, 10, 20, 40, 80, 160, 320)  
> round(cr.lb(n = n, sigma = 1.5), 2)  
  
[1] 4.80 2.40 1.20 0.60 0.30 0.15 0.08
```

```
> round(var.T(n = n, sigma = 1.5), 2)  
  
[1] 16.11  8.05  4.03  2.01  1.01  0.50  0.25
```

Consideriamo ora lo stimatore di massima verosimiglianza di ψ

$$\hat{\psi} = \exp \left\{ \sum_{i=1}^n (\log Y_i)^2 / 2n \right\}$$

E' immediato osservare che $\sum_{i=1}^n (\log Y_i)^2 \stackrel{D}{=} \sigma^2 V$ dove V ha distribuzione chi quadrato con n gradi di libertà. Per cui il momento r -esimo di $\hat{\psi}$ è dato da

$$E \left(\hat{\psi}^r \right) = E \left\{ \exp(r\sigma^2 V/2n) \right\} = \int_0^\infty \frac{\exp(r\sigma^2 v/2n)}{(2^{n/2} \Gamma(n/2))} v^{n/2-1} e^{-v/2} dv$$

Anche in questo caso l'integrale si risolve facilmente ed è pari a $(1 - r\sigma^2/n)^{-n/2}$, ma noi scriviamo delle funzione che lo calcoli utilizzando l'integrazione numerica.

```
> mr.psi <- function(r, n, sigma) {  
+   integrate(function(x) exp(r * sigma^2 * x/(2 * n)) * dchisq(x,  
+     df = n), lower = 0, upper = 1000)$val  
+ }
```

Effettivamente abbiamo che per la media si ha

```
> (1 - 1.5^2/10)^(-10/2) - mr.psi(1, 10, 1.5)  
  
[1] -4.440892e-16
```

La varianza sarà quindi data da

```
> varianza.psi <- function(n, sigma) {  
+   mr.psi(2, n, sigma) - mr.psi(1, n, sigma)^2  
+ }
```

Possiamo anche scrivere una funzione che dia la distorsione di $\hat{\psi}$ in funzione di n e σ

```
> bias.psi <- function(n, sigma) {  
+   mr.psi(r = 1, n = n, sigma = sigma) - exp(sigma^2/2)  
+ }
```

e l'errore quadratico medio

```
> eqm.psi <- function(n, sigma) {  
+   varianza.psi(n, sigma) + bias.psi(n, sigma)^2  
+ }
```

Costruiamo adesso una tabella dove nella prima riga abbiamo il limite inferiore di Cramer-Rao per gli stimatori non distorti quando $\sigma = 1.5$ e $n=5,10,20,40,80,160,320$; nella seconda riga abbiamo le varianze di T ; nella terza le varianze di $\hat{\psi}$; nella quarta la distorsione di $\hat{\psi}$ e nella quinta l'errore quadratico medio di $\hat{\psi}$

```
tabella<-rbind(  
+ tapply(n,INDEX=n,FUN=cr.lb,sigma=1.5),  
+ tapply(n,INDEX=n,FUN=var.T,sigma=1.5),  
+ tapply(n,INDEX=n,FUN=varianza.psi,sigma=1.5),  
+ tapply(n,INDEX=n,FUN=bias.psi,sigma=1.5),  
+ tapply(n,INDEX=n,FUN=eqm.psi,sigma=1.5))  
tabella<-round(tabella,2)  
quantita<-c("limite di Cramer Rao","varianza di T",  
+ "varianza della SMV", "distorsione della SMV","MSE  
+ della SMV")
```

```
> data.frame(quantita, tabella, check.names = F)
```

```
      quantita      5      10      20      40      80     160     320
1 limite di Cramer Rao  4.80  2.40  1.20  0.60  0.30  0.15  0.08
2      varianza di T  16.11  8.05  4.03  2.01  1.01  0.50  0.25
3  varianza della SMV 296.36  7.08  1.91  0.75  0.33  0.16  0.08
4 distorsione della SMV  1.38  0.50  0.22  0.10  0.05  0.02  0.01
5      MSE della SMV 298.26  7.32  1.96  0.76  0.34  0.16  0.08
```

Per $n \geq 80$ il limite inferiore di Cramer-Rao viene raggiunto da $\hat{\psi}$ che, ricordiamo, è asintoticamente non distorto e asintoticamente efficiente;

La distorsione di $\hat{\psi}$ è sempre molto piccola rispetto alla sua varianza e quindi il contributo della distorsione all'errore quadratico medio è trascurabile;

Lo stimatore non distorto T è più efficiente solo quando $n = 5$, altrimenti è sempre battuto da $\hat{\psi}$;

Per la normalità asintotica della stima di massima verosimiglianza deve essere $\hat{\psi} \sim \mathcal{N}(\psi, I(\psi)^{-1})$. Per verificare che tale risultato è vero relativamente al nostro esempio dobbiamo trovare l'informazione di Fisher per ψ .

Ricordiamo che (calcoli alla lavagna) se $\psi(\theta)$ è una parametrizzazione di θ allora vale il risultato

$$I_{\psi}(\psi) = I_{\theta}(\theta = \theta(\psi)) \left(\frac{d\theta(\psi)}{d\psi} \right)^2$$

. Nel nostro caso $\psi = \exp(\sigma^2/2)$ e ricordando che $I_{\sigma^2}(\sigma^2) = n/(2(\sigma^2)^2)$ abbiamo che $I_{\psi} = (2 \log \psi)^{-2} (2/\psi)^2 (n/2)$

Deve quindi risultare che $\hat{\psi}$ si distribuisce per n elevato come una normale con media ψ e varianza $[(2 \log \psi)^{-2} (2/\psi)^2 (n/2)]^{-1}$

```
> n <- c(n, 640)
> par(mfrow = c(2, 4))
> for (i in 1:length(n)) {
+   s <- n[i]
+   m <- 1000
+   simulazioni <- rlnorm(s * m, 0, 1.5)
+   campioni <- matrix(simulazioni, nrow = m)
+   psi.hat <- exp(apply((log(campioni))^2, FUN = mean, MAR =
+   psi.vero <- exp(1.5^2/2)
+   i.f <- (2/psi.vero)^2 * (s/2) * (1/(2 * log(psi.vero))^2)
+   hist(psi.hat, main = s, prob = T)
+   curve(dnorm(x, mean = psi.vero, sd = sqrt(i.f^(-1))), add
+     col = 4)
+ }
```