STATISTICA 1, metodi matematici e statistici

Introduzione al linguaggio R

Esercitazione 6: 06-05-2005

Luca Monno

Università degli studi di Pavia

luca.monno@unipv.it
http://www.lucamonno.it

Test di Neyman e Pearson

Dato il seguente campione

```
> y <- c(0.4, 1.76, 0.65, 0.52, 0.27, 0.53, 1.22, 1.25, 0.17, 1
```

estratto da una v.a. esponenziale verificare l'ipotesi $H_0: \theta = 1$ contro l'ipotesi alternativa $H_1: \theta = 2$ attraverso il test di Neyman e Pearson imponendo un errore di prima specie pari a 0.05 e calcolare la potenza del test.

Per prima cosa ci costruiamo una funzione che restituisca il rapporto delle verosimiglianze

```
> rapp.ver <- function(data, theta0, theta1) {
+     theta1^(length(data)) * exp(-theta1 * sum(data))/theta0^(
+         exp(-theta0 * sum(data))
+ }
```

e ci calcoliamo il valore di tale rapporto per il nostro campione

```
> t <- rapp.ver(data = y, theta0 = 1, theta1 = 2)
```

Per vedere se `t` cade nella zona di accettazione o in quella di rifiuto ci dobbiamo calcolare il quantile di livello 0.95 della distribuzione del test del rapporto delle verosimiglianze sotto l'hp nulla (ricordiamoci che valori elevati del rapp.ver portano a rifiutare H_0)

```
> n <- 10
> m <- 10000
> theta.vero = 1
> a <- rexp(n * m, theta.vero)
> campioni <- matrix(a, nrow = m)
> lrt <- apply(campioni, FUN = rapp.ver, MAR = 1, theta0 = 1, t
> lim <- quantile(lrt, 0.95)
```

Accettiamo l'ipotesi se il nostro valore `t` è più piccolo di `lim`

```
> if (t < lim) print(c("accettiamo H_0"))

[1] "accettiamo H_0"
```

Per calcolarci la potenza del test dobbiamo calcolare la probabilità di avere valori maggiori di `lim` sotto l'ipotesi alternativa

```
> n <- 10
> m <- 10000
> theta.vero = 2
> a <- rexp(n * m, theta.vero)
> campioni <- matrix(a, nrow = m)
> lrt2 <- apply(campioni, FUN = rapp.ver, MAR = 1, theta0 = 1,
+     theta1 = 2)
> pot <- sum(lrt2 > lim)/m
```

Il seguente grafico ci aiuta a capire meglio la situazione

```
> plot(density(log(lrt)), col = 2, xlim = c(-70, 70), ylim = c(
+     0.1))
> lines(density(log(lrt2)), col = 3)
> abline(v = log(lim), lty = 2)
> abline(v = log(t), lty = 3)
```

Ancora sul test ottimo di N.P.

Sia dato un campione casuale $x = (x_i \dots, x_n)$ da una $\mathcal{N}(\theta, \sigma^2)$ con σ noto e consideriamo le ipotesi

$$H_0 : \theta = \theta_0 \quad H_1 : \theta = \theta_1 \quad \theta_0 < \theta_1$$

Prefissando un livello di ampiezza pari ad α, il test ottimo nel senso del Lemma di N.P. ha una zona critica del tipo

$$\{x : p(x; \theta_1) \geq kp(x; \theta_0)\}\,.$$

Supponiamo di volere un ampiezza del test pari ad α. Dobbiamo quindi trovare k tale che l'evento

$$\frac{p(X_1, \dots, X_n; \theta_1)}{p(X_1, \dots, X_n; \theta_0)} > k$$

ha probabilità α sotto θ_0

Dopo un pò di conti (provare a farli) si ottiene che l'evento sopra indicato è equivalente a

$$\exp\left\{\frac{n(\theta_1-\theta_0)}{\sigma^2}(\bar{x}-\bar{\theta})\right\} > k$$

dove $\bar{\theta}=(\theta_0+\theta_1)/2$. La zona di rifiuto è quindi formata da quei campioni tali che

$$\bar{x} \geq c$$

dove c va determinato in modo tale che

$$Prob\left\{\bar{X} > c|H_0\right\} = \alpha.$$

Ricordando che sotto H_0 si ha $\bar{X} \sim \mathcal{N}(\theta_0, \sigma^2/n)$ abbiamo che c è il quantile di livello $1-\alpha$ di una $\mathcal{N}(\theta_0, \sigma^2/n)$

Consideriamo allora il seguente campione

```
> set.seed(5)
> x <- rnorm(10, 1, 1)
```

e assumendo (come in realtà è) che il campione proviene da una normale con varianza unitaria, supponiamo di voler verificare il sistema di ipotesi

$$H_0 : \theta = 0 \quad H_1 : \theta = 1.$$

con un test di ampiezza $\alpha = 0.05$.

Ci calcoliamo il quantile di livello 0.95 da una normale con media $\theta_0 = 0$ e varianza $\sigma^2/n = 1/n$

```
> c <- qnorm(0.95, 0, sd = sqrt(1/10))
> c

[1] 0.5201484
```

Nel nostro caso la media del campione è

```
> media <- mean(x)
> media
```

```
[1] 0.9211485
```

abbiamo quindi che

```
> media > c
```

```
[1] TRUE
```

e quindi rifiutiamo l'ipotesi H_0. La potenza del test è pari a $Prob(\text{accettare}\, H_1|H_1)$ e quindi nel nostro caso è $Prob(\bar{X} > c|\theta_1)$. Ricordando che sotto θ_1 si ha $\bar{X} \sim \mathcal{N}(\theta_1, \sigma^2/n)$ la potenza del test è quindi

```
> pot <- pnorm(c, 1, sd = sqrt(1/n), lower.tail = F)
> pot

[1] 0.9354202
```

Osserviamo dall'help di `pnorm` che la sintassi è

```
pnorm(q, mean=0, sd=1, lower.tail = TRUE, log.p = FALSE)
```

dove

```
lower.tail: logical; if TRUE (default), probabilities are
P[X <= x], otherwise, P[X > x]
```

L'errore di II specie, ovvero la probabilità di rifiutare H_1 quando è vera è quindi

```
> beta <- pnorm(c, 1, sd = sqrt(1/n))
```

```
> beta
[1] 0.06457983
> 1 - pot
[1] 0.06457983
```

Vediamo ora cosa succede all'errore di II specie se diminuiamo α. Consideriamo $\alpha = 0.01$

```
> c.2 <- qnorm(0.99, 0, sd = sqrt(1/10))
> pot.2 <- pnorm(c.2, 1, sd = sqrt(1/n), lower.tail = F)
> 1 - pot.2
[1] 0.2015972
```

L'errore di seconda specie è quindi aumentato

Vediamo ora cosa succede all'errore di II specie se aumentiamo α.
Consideriamo $\alpha = 0.1$

```
> c.3 <- qnorm(0.9, 0, sd = sqrt(1/10))
> pot.3 <- pnorm(c.3, 1, sd = sqrt(1/n), lower.tail = F)
> 1 - pot.3

[1] 0.03000459
```

L'errore di II specie è quindi diminuito.

Vediamo infine che succede all'errore di II specie se fissato α facciamo aumentare la numerosità campionaria

```
> c.4 <- qnorm(0.95, 0, sd = sqrt(1/20))
> pot.4 <- pnorm(c.4, 1, sd = sqrt(1/20), lower.tail = F)
> 1 - pot.4

[1] 0.002347246
```

Rispetto al test con ampiezza α ed $n = 10$ l'errore di II specie è quindi diminuito.