STATISTICA 1, metodi matematici e statistici

Introduzione al linguaggio R

Esercitazione 8: 10-05-2004

Andrea Tancredi

Università di Roma “La Sapienza”, Rome, Italy

andrea.tancredi@uniroma1.it
http://3w.eco.uniroma1.it/utenti/tancredi

Ancora sul test ottimo di N.P.

Sia dato un campione casuale $x = (x_i \ldots, x_n)$ 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

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

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] 0.984062
```

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] 0.01593802
> 1 - pot
[1] 0.01593802
```

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] 0.1185665
```

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] 0.003909953
```

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.

Test del rapporto delle massime verosimiglianze

Consideriamo ora un sistema generale di ipotesi come

$$H_0 : \theta \in \Omega_0, \quad \theta \in \Omega$$

una zona critica intuitiva è del tipo

$$\{x : V(x) \leq \xi\}$$

dove

$$V(x) = \frac{\sup_{\theta \in \Omega_0} L(\theta)}{\sup_{\theta \in \Omega} L(\theta)}$$

è il cosiddetto rapporto delle verosimiglianze massimizzate. Il valore $\xi \in [0, 1]$ determina l'ammpiezza del test tramite la relazione

$$\sup_{\theta \in \Omega_0} P\{x : V(x) \leq \xi\} = \alpha$$

.

Supponiamo che su un campione casuale estratto da una normale con varianza σ^2 vogliamo verificare il seguente sistema di ipotesi

$$H_0 : \theta = \theta_0 \quad H_1 : \theta \neq \theta_0.$$

Allora

$$V(x) = \frac{L(\theta_0)}{L(\hat{\theta})} = (conti) = \exp\left[-\frac{n}{2\sigma^2}(\bar{x} - \theta_0)^2\right]$$

per cui la regione critica è data da

$$\left\{ x : \left(\frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}} \right)^2 \geq -2 \log \xi \right\}$$

Sotto l'ipotesi nulla la statistica

$$\left(\frac{\bar{x} - \theta_0}{\sigma/\sqrt{n}}\right)^2 \qquad (1)$$

ha distribuzione $Chi^2(1)$. Pertanto per avere un test di ampiezza α basta calcolare il quantile di livello $1 - \alpha$ da un $Chi^2(1)$ e vedere se la statistica test (1) supera o meno tale valore.

Dato

```
> set.seed(5)
> y <- rnorm(10, mean = 1, sd = 1)
> mean(y)

[1] 0.9211485
```

Verifichiamo il sistema di ipotesi

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

La statistica test (1) è pari a

```
> t <- ((mean(y) - 0)/(1/sqrt(10)))^2
> t
```

```
[1] 8.485145
```

il quantile di livello $1 - \alpha$ da un $Chi^2(1)$ è

```
> q <- qchisq(0.95, 1)
> q
```

```
[1] 3.841459
```

```
> t > q
```

```
[1] TRUE
```

Un importante aspetto applicativo è che se Ω_0 e Ω sono intervalli rispettivamente di dimensione K_0 e k, la distribuzione campionaria di $G^2 = -2 \log V(X_1, \ldots, X_n)$ sotto la condizione $\theta \in \Omega_0$ è del tipo $Chi^2(k - k_0)$.

ESERCIZIO: verifificare che per campioni provenienti da una distribuzione esponenziale con media θ pari a 1 e numerosità 10, la statistica test G^2 per la verifica dell'ipotesi nulla $H_0 : \theta = 1$ rispetto all'alternativa $H_1 : \theta \neq 1$ è asintoticamente $Chi^2(1)$.