Universitat Autònoma de Barcelona
Facultat de Ciències
Estadística
M. Saucedo

Seminari 6: Contrasts d'hipòtesis

Aquesta pràctica té com a objectiu implementar en R els tests de quocient de versemblances (LRT) que hem estat estudiant, i altres tests asimptòtics (Wald).

1 Tests asimptòtics (LRT, Wald Test)

LRT és més o menys universal, i es pot aplicar en la majoria dels casos, almenys quan el nombre de paràmetres és finit: Per provar la hipòtesi nul·la $H_0 : \theta \in \Theta_0$ contra l'alternativa $H_1 : \theta \in \Theta_1 = \Theta \backslash \Theta_0$, el LRT rebutja H_0 per a valors petits de

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(x;\theta)}{\sup_{\theta \in \Theta} L(x;\theta)}.$$

Observem que si

$$\Lambda(\mathbf{x}) = -2 \log \lambda(\mathbf{x}) = 2 \left[l(\widehat{\theta}; \mathbf{x}) - l(\theta_0; \mathbf{x}) \right]$$

la regió crítica del LRT es pot escriure com

$$C_1 = \{\mathbf{x} : \Lambda(\mathbf{x}) \geq c\}$$

Sota les mateixes condicions de regularitat que necessitem per garantir que el EMV és assimptòticament normal, tenim

$$\Lambda(\mathbf{X}) \xrightarrow{D} \chi^2_p \qquad \text{si} \quad p = dim(\Theta) - dim(\Theta_0) \geq 1$$

Wald va proposar utilitzar com a estadístic per provar $H_0 : \theta = \theta_0$:

$$W = (\hat{\theta} - \theta_0)^2 I(\hat{\theta}) \sim \chi^2_1 \quad \text{assimptòticament, sota}\, H_0,$$

o en la seva versió vectorial

$$W = (\hat{\theta} - \theta_0)^{tr} I(\hat{\theta})(\hat{\theta} - \theta_0) \sim \chi^2_p \quad \text{assimptòticament, sota}\, H_0, \text{ on}\, p = dim(\Theta).$$

Observem que W és el quadrat de la distància entre θ_0 i el EMV de θ, ponderada per una estimació consistent de la informació continguda en la mostra.

- el LRT requereix el càlcul del EMV i el EMV restringit a Θ_0.
- el test de Wald només requereix el càlcul del EMV.

El LRT per tant usa més informació i hi ha estudis empírics que suggereixen que per a mostres de grandària moderada resulta més confiable, encara que asintóticamente són equivalents.

Problemes

1. L'ecologista I.C. Pielou, va estudiar el patró d'arbres sans i malalts (infectats amb *Armillaria*) en una plantació d'avets. Va registrar les longituds de 109 successions d'arbres malalts.

Longitud de las sucesiones de árboles enfermos

Longitud	1	2	3	4	5	6
Número de sucesiones	71	28	5	2	2	1

Basant-se en consideracions d'índole biològic, Pielou va proposar un model geomètric per a la distribució de les longituds.

Nota: Si $X \sim$Geo(p), $\mathbf{P}(X = h) = (1-p)^{h-1}p$, para $h = 1, 2, \ldots$, és a dir, X compta el número d'assajos fins al primer èxit en una successió d'assajos independents de Bernoulli amb probabilitat d'èxit p.

   (a) Un grup de defensors dels boscos afirmen que 3/5 dels arbres estan malalts. Pots rebutjar aquesta afirmació amb una significació de 0.05 fent servir el LRT?

   (b) Repeteix l'apartat anterior fent servir el test de Wald.

   (c) Construeix un interval de confiança del 95% (aproximat) per a p a partir de Wald?

2. Utilitzeu les dades del nombre de gols de les lligues de futbol europees corresponents a les temporades des de 1993-1994 fins a la 2003-2004 disponibles a:

`http://mat.uab.cat/~mbarcelona/est/goles.csv`

   (a) Suposeu que les dades tenen una distribució de Poisson amb paràmetre θ. Feu un LRT amb hipòtesis nul·la que la mitjana de gols a la lliga espanyola és 3 i $\alpha = 0.05$.

   (b) Feu un test del quocient de versemblances per a la hipòtesis nul·la que la mitjana del nombre de gols per partit en cadascuna de les lligues és la mateixa. Suposeu que les dades de cada lliga tenen una distribució de Poisson amb paràmetres respectius $\lambda_i, i = 1, \ldots, 5$.

3. El nombre de fractures d'extremitats tractades durant un cert període de temps en les emergències d'un hospital és

	dret	esquerra
braç	46	49
cama	22	32

Volem saber si podem concloure que les fractures ocorren amb més probabilitat al costat esquerra amb confiança del 95%:

Considerem variables aleatòries i.i.d. discretes $X_1, \ldots X_n$ amb

$$\mathbf{P}(X_i = j) = p_j \qquad j = 1, 2, 3, 4.$$

El paràmetre d'interès és $\theta = (p_1, p_2, p_3, p_4)$, i $\sum_{j=1}^{4} p_j = 1$.

Es vol fer un test per a

$$H_0 : p_1 = p_2, p_3 = p_4.$$

   (a) Escriure l'estimador de màxima versemblança (global) de θ.

   (b) Escriure l'estimador de màxima versemblança de θ sota la hipòtesi nul·la.

   (c) Escriure la regió crítica del test de raó de versemblances per a aquest problema.

   (d) Quina és la conclusió del test?

Del examen de pràctiques 2016

4. Les següents dades corresponen a la mitjana mensual de la velocitat del vent (km/h) a Castelldefells entre els anys 2006 i 2012[1].

```
wind=c(5.86,6.64,8.8,7.81,7.78,7.63,7.51,6.95,5.13,5.2,4.79,5.3)
```

Els meteoròlegs pensen que un bon model per a les velocitats mitjanes del vent és la distribució de Weibull, amb funció de distribució

$$F(t) = 1 - \mathrm{e}^{-(t/\beta)^{\alpha}} \qquad x \geq 0$$

amb paràmetre d'escala $\beta > 0$ i paràmetre de forma $\alpha > 0$ desconeguts.

(a) Calcular (numèricament) els estimadors de màxima versemblança amb `nlm` escrivint una funció que tingui $-\log L$ on L és la versemblança de la mostra.

HINT: Si es vol fer més fàcilment, convé usar que, si $Y \sim$Weibull(α, β), llavors $X = -\log(Y) \sim$Gumbel$(-\log(\beta), \alpha^{-1})$. Una variable aleatoria $X \sim$ Gumbel(ξ, θ) té funció de distribució[2]

$$F_X(x) = \mathrm{e}^{-\mathrm{e}^{-(x-\xi)/\theta}} \quad \text{i densitat} \quad f_X(x) = \frac{1}{\theta}\mathrm{e}^{-(z+\mathrm{e}^{-z})} \quad \text{amb} \quad z = \frac{x-\xi}{\theta}$$

$$\mathbf{E}X = \xi + \theta\gamma \quad \mathbf{Var}X = \frac{\pi^2}{6}\theta^2 \quad \text{ón} \quad \gamma \approx 0.5772 \quad \text{és la constant d' Euler.}$$

(b) Fes un *qq*-plot per verificar gràficament si el model Weibull és raonable.

(c) Quins són els estimadors de α i β?

(d) Verifica els resultats obtinguts fent servir

```
install.packages("fitdistrplus");library(fitdistrplus)
fitdist(wind, "weibull")
```

[1] http://www.castelldefels.org/es/doc_generica.asp?dogid=2015.

[2] https://en.wikipedia.org/wiki/Gumbel_distribution