MAT-266: Transformaciones estabilizadoras de varianza

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Para introducir ideas, considere

$$\mathsf{E}(Y) = \mu, \qquad \mathsf{var}(Y) = \sigma^2 h(Y),$$

y suponga la transformación $z = g(y)$ tal que la varianza de Z es aproximadamente independiente de μ.

Suponga una expansión de primer orden de $g(y)$ en torno de μ, esto es

$$g(y) \approx g(\mu) + g'(\mu)(y - \mu).$$

De este modo,

$$\begin{aligned} \mathsf{E}(Z) &\approx g(\mu) \\ \mathsf{var}(Z) &\approx \{g'(\mu)\}^2 \,\mathsf{var}(Y - \mu) = \{g'(\mu)\}^2 \sigma^2 h(y) \end{aligned}$$

Transformaciones estabilizadoras de varianza

Así, para determinar una transformación tal que var$(Z) = \sigma^2$ necesitamos que

$$g'(\mu) = \frac{1}{\sqrt{h(y)}},$$

o de forma análoga,

$$g(\mu) = \int \frac{\mathsf{d}\,\mu}{\sqrt{h(\mu)}}.$$

En particular, se podría considerar la clase de transformaciones en que $h(y)$ es una potencia de μ. Algunos ejemplos son los siguientes:

$h(\mu)$	z	Descripción
μ^4	y^{-1}	recíproco
μ^2	$\log y$	logarítmico
μ	$\sqrt{y}$	raíz cuadrada
$\mu(1-\mu)$	$\sin^{-1}(\sqrt{y})$	seno inverso
$(1-\mu^2)^2$	$\log(\frac{1+y}{1-y})$	correlación

Box y Cox (1964)[1] sugirieron llevar a cabo la estimación ML para una clase general de transformaciones. El supuesto fundamental es que $Y > 0$ y que para alguna potencia de Y tal que su varianza es aproximadamente constante y satisface el supuesto de normalidad

En concreto, ellos consideraron la familia de transformaciones

$$Y(\lambda) = \begin{cases} \dfrac{Y^{\lambda} - 1}{\lambda}, & \lambda \neq 0, \\ \log(Y), & \lambda = 0, \end{cases}$$

el objetivo es obtener una estimación de λ desde los datos observados.

[1] Journal of the Royal Statistical Society, Series B **26**, 211-252.

Sea $U_i = Y_i(\lambda)$ para $i = 1, \dots, n$, y asuma que $\boldsymbol{U} \sim \mathsf{N}_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I})$ para alguna elección de λ. El Jacobiano de la transformación es dado por

$$J = \prod_{i=1}^{n} \frac{\partial Y_i(\lambda)}{\partial Y_i} = \prod_{i=1}^{n} Y_i^{\lambda - 1} = \Big(\prod_{i=1}^{n} Y_i^{1/n} \Big)^{n(\lambda-1)} = G^{n(\lambda-1)},$$

donde G es la media geométrica de las observaciones.

De este modo, la log-verosimilitud requerida adopta la forma

$$\ell(\boldsymbol{\theta}, \lambda) = -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} Q_\lambda(\boldsymbol{\beta}) + \log J,$$

donde

$$Q_\lambda(\boldsymbol{\beta}) = \|\boldsymbol{Y}(\lambda) - \boldsymbol{X\beta}\|^2.$$

Para λ fijado, tenemos

$$\widehat{\boldsymbol{\beta}}(\lambda) = (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top \boldsymbol{Y}(\lambda),$$
$$\widehat{\sigma}^2(\lambda) = \frac{1}{n}\|\boldsymbol{Y}(\lambda) - \boldsymbol{X}\widehat{\boldsymbol{\beta}}(\lambda)\|^2.$$

Esto lleva a la log-verosimilitud perfilada, dada por

$$\ell_*(\lambda) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log(\mathsf{RSS}(\lambda)/n) - \frac{n}{2} + \log J,$$

donde

$$\mathsf{RSS}(\lambda) = \boldsymbol{Y}^\top(\lambda)(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}(\lambda).$$

Considere por conveniencia,

$$\boldsymbol{Z}(\lambda) = \frac{\boldsymbol{Y}(\lambda)}{G^{\lambda-1}},$$

de este modo podemos escribir

$$\ell_*(\lambda) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log(\mathsf{RSS}_Z(\lambda)/n) - \frac{n}{2},$$

con

$$\mathsf{RSS}_Z(\lambda) = \boldsymbol{Z}^\top(\lambda)(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Z}(\lambda),$$

luego la maximización de $\ell_*(\lambda)$ es equivalente a la minimización de $\mathsf{RSS}_Z(\lambda)$.

Típicamente se realiza la transformación $Z(\lambda)$ para un rango de valores de λ, se realiza el ajuste del modelo lineal y se examina aquél valor de λ que corresponde al valor más pequeño de $\mathsf{RSS}_Z(\lambda)$[2]

Para los datos transformados la función Box-Cox asume la forma

$$Z(\lambda) = \begin{cases} \dfrac{Y^{\lambda} - 1}{\lambda\, G^{\lambda-1}}, & \lambda \neq 0, \\ G \log(Y), & \lambda = 0, \end{cases}$$

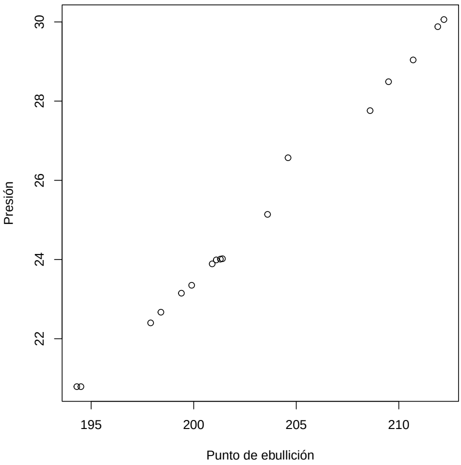
obteniendo el valor $\widehat{\lambda}$ se lleva a cabo el ajuste $\mathsf{E}(\boldsymbol{Z}) = \boldsymbol{X}\boldsymbol{\beta}$ con $\boldsymbol{Z} = \boldsymbol{Z}(\widehat{\lambda})$.

Observación:

Note que $\lambda = 1$ implica que el modelo no debe ser transformado.

[2] Usualmente basta considerar $-2 \leq \lambda \leq 2$ con incrementos no muy pequeños.

Datos de Forbes



Datos de Forbes

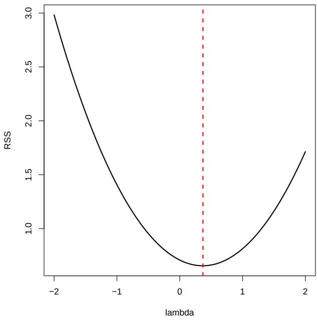
```r
boxcox.lm <- function(x, y, lambda) {
  boxcox <- function(y, lambda) {
    n <- length(y)
    lambda <- rep(lambda, n)
    z <- ifelse(lambda != 0., (y^lambda - 1.) / lambda, log(y))
    z
  }
  n <- nrow(x)
  p <- ncol(x)
  k <- length(lambda)
  RSS <- rep(0, k)
  logLik <- rep(0, k)
  for (i in 1:k) {
    geom <- geomean(y)
    z <- boxcox(y, lambda = lambda[i])
    z <- z / geom^(lambda[i] - 1.)
    fm <- ols.fit(x, z, method = "sweep")
    RSS[i] <- fm$RSS
    logLik[i] <- -.5 * n * log(2 * pi) - .5 * n * log(RSS[i] / n) - .5 * n
  }
  idx <- order(RSS)[1]
  opt <- lambda[idx]
  obj <- list(lambda = lambda, RSS = RSS, logLik = logLik, opt = opt)
  obj
}
```

Datos de Forbes

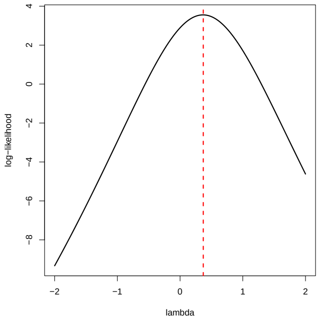
```
# carga datos desde MASS
> library(MASS)
> data(forbes)

# gráfico de los datos
> plot(pres ~ bp, data = forbes, xlab = "Punto de ebullición",
+       ylab = "Presión")

# ajuste preliminar
library(fastmatrix)
fm <- ols(pres ~ bp, data = forbes, x = TRUE, y = TRUE)

# extrae vector de respuestas y matriz de diseno
x <- fm$x
y <- fm$y

# interpreta script, crea 'grilla' e invoca método de ajuste
source("boxcox.lm.R")
lambda <- seq(-2, 2, by = 0.01)
z <- boxcox.lm(x, y, lambda)
```

```
# gráficos de RSS y log-likelihood
> plot(lambda, z$RSS, type = "l", ylab = "RSS", lwd = 2)
> abline(v = z$opt, col = "red", lwd = 2, lty = 2)
> plot(lambda, z$logLik, type = "l", ylab = "log-likelihood", lwd = 2)
> abline(v = z$opt, col = "red", lwd = 2, lty = 2)

# removiendo dato 12
> y12 <- y[-12]
> x12 <- x[-12,]
> z <- boxcox.lm(x12, y12, lambda)

# ajustando diversos modelos
f0 <- ols(pres ~ bp, data = forbes)
f1 <- ols(log(pres) ~ bp, data = forbes)
f2 <- ols((pres^.37 - 1) / .37 ~ bp, data = forbes)
f3 <- ols((pres^.10 - 1) / .10 ~ bp, data = forbes)
```

Datos de Forbes



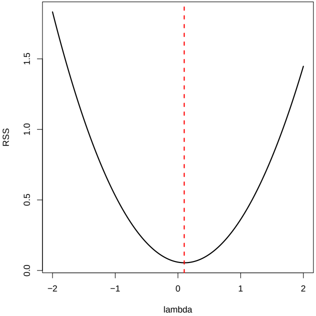
(a)



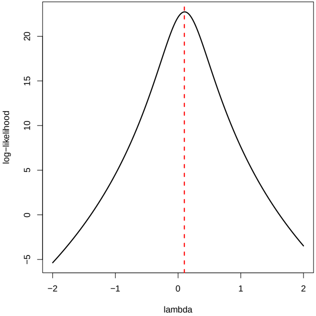
(b)

Todos los datos, $\lambda_{\text{opt}} = 0.37$.

Datos de Forbes



(a)



(b)

Dato 12 removido, $\lambda_{\text{opt}} = 0.10$.

Resumen de estimación para los datos de Forbes:

Parámetro	λ			
	—	0.00	0.10[3]	0.37
β_0	-81.0637	-0.9709	-1.9885	-7.6462
β_1	0.5229	0.0206	0.0285	0.0681
σ^2	0.0542	0.0001	0.0001	0.0008
RSS	0.8131	0.0011	0.0021	0.0114
$\ell(\widehat{\boldsymbol{\theta}})$	1.7186	57.5378	52.3791	37.9802

[3] Observación 12 removida