

IECD-325: Transformaciones estabilizadoras de varianza

Felipe Osorio

felipe.osorio@uv.cl

Transformaciones estabilizadoras de varianza

Para introducir ideas, considere

$$\mathbb{E}(Y) = \mu, \quad \text{var}(Y) = \sigma^2 h(Y),$$

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,

$$\mathbb{E}(Z) \approx g(\mu)$$

$$\text{var}(Z) \approx \{g'(\mu)\}^2 \text{var}(Y - \mu) = \{g'(\mu)\}^2 \sigma^2 h(y)$$

Transformaciones estabilizadoras de varianza

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

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

o de forma análoga,

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

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

Transformaciones estabilizadoras de varianza

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 su varianza sea aproximadamente constante y satisfaga el supuesto de normalidad

En concreto, ellos consideraron la familia de transformaciones

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

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

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

Transformaciones estabilizadoras de varianza

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

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

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

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

Transformaciones estabilizadoras de varianza

Para λ fijado, tenemos

$$\begin{aligned}\hat{\beta}(\lambda) &= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}(\lambda), \\ \hat{\sigma}^2(\lambda) &= \frac{1}{n} \|\mathbf{Y}(\lambda) - \mathbf{X}\hat{\beta}(\lambda)\|^2.\end{aligned}$$

Esto lleva a la log-verosimilitud perfilada, dada por

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

donde

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

Transformaciones estabilizadoras de varianza

Considere por conveniencia,

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

de este modo podemos escribir

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

con

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

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

Transformaciones estabilizadoras de varianza

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 $\text{RSS}_Z(\lambda)$ ²

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

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

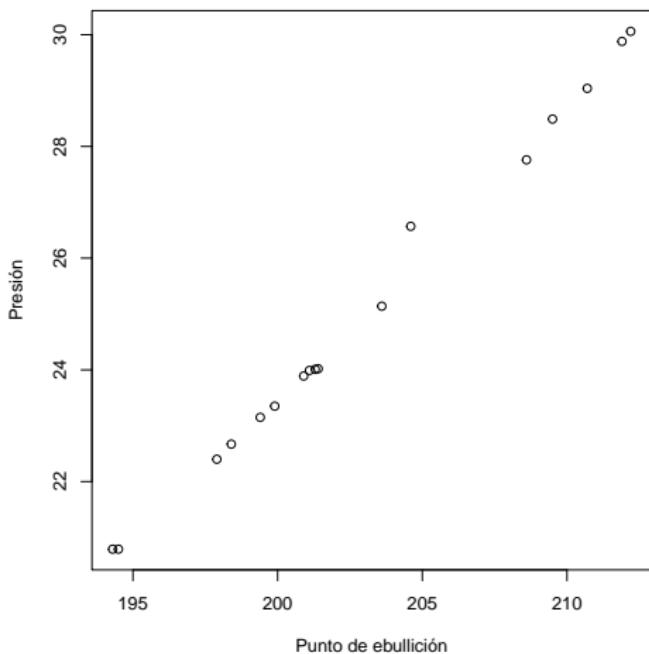
obteniendo el valor $\hat{\lambda}$ se lleva a cabo el ajuste $E(\mathbf{Z}) = \mathbf{X}\boldsymbol{\beta}$ con $\mathbf{Z} = Z(\hat{\lambda})$.

Observación:

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

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

Datos de Forbes



Datos de Forbes

```
1 boxcox.lm <- function(x, y, lambda) {
2   boxcox <- function(y, lambda) {
3     n <- length(y)
4     lambda <- rep(lambda, n)
5     z <- ifelse(lambda != 0., (y^lambda - 1.) / lambda, log(y))
6     z
7   }
8   n <- nrow(x)
9   p <- ncol(x)
10  k <- length(lambda)
11  RSS <- rep(0, k)
12  logLik <- rep(0, k)
13  for (i in 1:k) {
14    geom <- geomean(y)
15    z <- boxcox(y, lambda = lambda[i])
16    z <- z / geom^(lambda[i] - 1.)
17    fm <- ols.fit(x, z, method = "sweep")
18    RSS[i] <- fm$RSS
19    logLik[i] <- -.5 * n * log(2 * pi) - .5 * n * log(RSS[i] / n) -
20      .5 * n
21  }
22  idx <- order(RSS)[1]
23  opt <- lambda[idx]
24  obj <- list(lambda = lambda, RSS = RSS, logLik = logLik,
25            opt = opt)
26  obj
27 }
```

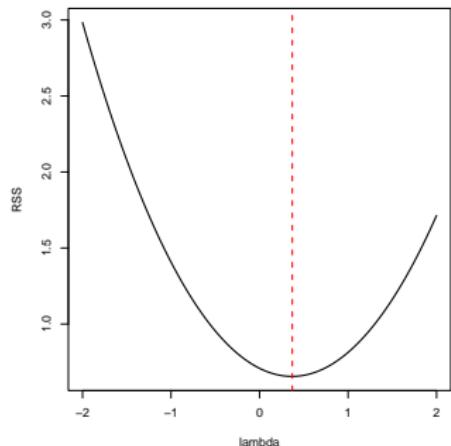
Datos de Forbes

```
1 # carga datos desde MASS
2 > library(MASS)
3 > data(forbes)
4
5 # gráfico de los datos
6 > plot(pres ~ bp, data = forbes, xlab = "Punto de ebullición",
7 +       ylab = "Presión")
8
9 # ajuste preliminar
10 library(fastmatrix)
11 fm <- ols(pres ~ bp, data = forbes, x = TRUE, y = TRUE)
12
13 # extrae vector de respuestas y matriz de diseño
14 x <- fm$x
15 y <- fm$y
16
17 # interpreta script, crea 'grilla' e invoca método de ajuste
18 source("boxcox.lm.R")
19 lambda <- seq(-2, 2, by = 0.01)
20 z <- boxcox.lm(x, y, lambda)
21
```

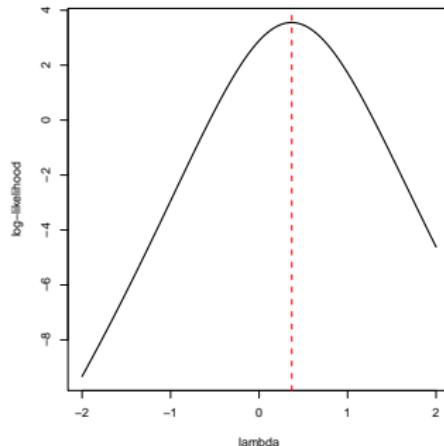
Datos de Forbes

```
1 # gráficos de RSS y log-likelihood
2 > plot(lambda, z$RSS, type = "l", ylab = "RSS", lwd = 2)
3 > abline(v = z$opt, col = "red", lwd = 2, lty = 2)
4 > plot(lambda, z$logLik, type = "l", ylab = "log-likelihood", lwd = 2)
5 > abline(v = z$opt, col = "red", lwd = 2, lty = 2)
6
7 # removiendo dato 12
8 > y12 <- y[-12]
9 > x12 <- x[-12,]
10 > z <- boxcox.lm(x12, y12, lambda)
11
12 # ajustando diversos modelos
13 f0 <- ols(pres ~ bp, data = forbes)
14 f1 <- ols(log(pres) ~ bp, data = forbes)
15 f2 <- ols((pres^.37 - 1) / .37 ~ bp, data = forbes)
16 f3 <- ols((pres^.10 - 1) / .10 ~ bp, data = forbes)
17
```

Datos de Forbes



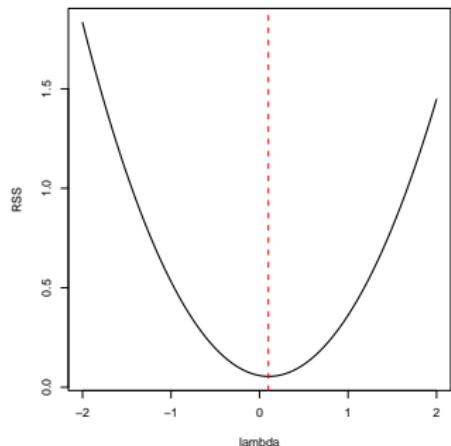
(a)



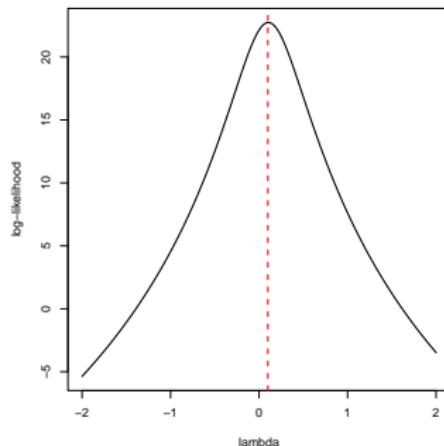
(b)

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

Datos de Forbes



(a)



(b)

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

Datos de Forbes

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
$\ell(\hat{\theta})$	1.7186	57.5378	52.3791	37.9802

³Observación 12 removida