Estimadores de tipo MM para el modelo lineal multivariado

Nadia L. Kudraszow (Universidad Nacional de La Plata - CONICET)
Ricardo A. Maronna (Universidad Nacional de La Plata - CIC)

Resumen

Sean $\mathbf{y}_i = (y_{i1}, \ldots, y_{iq})'$ y $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})'$ con $1 \leq i \leq n$ los vectores de las respuestas y de los predictores, respectivamente, que cumplen el modelo lineal multivariado

$$\mathbf{y}_i = \mathbf{B_0}'\mathbf{x}_i + \mathbf{u}_i,$$

donde $\mathbf{B_0}$ es una matriz de $p \times q$ y los $\mathbf{u}_1, \ldots, \mathbf{u}_n$ son vectores aleatorios de dimensión q independientes e idénticamente distribuidos (i.i.d.). Los $\mathbf{x}_i$ se suponen aleatorios, i.i.d. e independientes de los $\mathbf{u}_i$.

Si la distribución de los $\mathbf{u}_i$ es normal multivariada $N_q(\mathbf{0}, \mathbf{\Sigma}_0)$, el estimador clásico para este modelo es el estimador de máxima verosimilitud (EMV) de $\mathbf{B}_0$ que se obtiene aplicando mínimos cuadrados a cada una de las respuestas separadamente, y el EMV de $\mathbf{\Sigma}_0$ es la matriz de covarianzas muestral de los residuos. Estos estimadores son equivariantes por regresión y por transformaciones lineales de $\mathbf{x}_i$ y de $\mathbf{y}_i$. Sin embargo se sabe que este estimador es extremadamente sensible a observaciones atípicas. Por eso, se han investigado varias alternativas robustas, aunque sólo unos pocos de los estimadores propuestos conservan las propiedades de equivarianza y tienen una alta eficiencia bajo el modelo normal. Para cubrir ambas necesidades presentamos a los MM-estimadores para regresión lineal multivariada.

Los MM-estimadores para regresión lineal univariada fueron introducidos por Yohai (1987) para combinar robustez y eficiencia en el caso univariado. La extensión al caso multivariado se realizó de la siguiente manera. Sea $(\tilde{\mathbf{B}}_n, \tilde{\mathbf{\Sigma}}_n)$ (con $|\tilde{\mathbf{\Sigma}}_n| = 1$) un estimador inicial de $(\mathbf{B}_0, \mathbf{\Sigma}_0)$ con alto punto de ruptura aunque posiblemente ineficiente. Definimos las distancias $d_i^2(\mathbf{B}, \mathbf{\Sigma}) = (\mathbf{y}_i - \mathbf{B}'\mathbf{x}_i)' \, \mathbf{\Sigma}^{-1} \, (\mathbf{y}_i - \mathbf{B}'\mathbf{x}_i)$. Sea $\hat{\sigma}_n$ un M-estimador de escala de $\{d_i(\tilde{\mathbf{B}}_n, \tilde{\mathbf{\Sigma}}_n), i = 1, .., n\}$ basado en una "función-ρ" ρ_0. Sea ρ_1 otra función-ρ tal que $\rho_1 \leq \rho_0$. Entonces el MM-estimador de regresión multivariado $(\hat{\mathbf{B}}_n, \hat{\mathbf{\Sigma}}_n)$ está dado por

$$(\hat{\mathbf{B}}_n, \hat{\mathbf{\Gamma}}_n) = \arg\text{mín} \left\{ \sum_{i=1}^{n} \rho_1 \left(\frac{d_i(\mathbf{B}, \mathbf{\Gamma})}{\hat{\sigma}_n} \right) : (\mathbf{B}, \mathbf{\Gamma}) \in \mathbb{R}^{p \times q} \times \mathcal{S}_q \text{ y } |\mathbf{\Gamma}| = 1 \right\}$$

$$\hat{\mathbf{\Sigma}}_n = \hat{\sigma}_n^2 \hat{\mathbf{\Gamma}}_n$$

siendo $\mathcal{S}_q$ el conjunto de las matrices simétricas y definidas positivas de dimensión q.

Probamos que estos estimadores tienen punto de ruptura asintoticamente 0.5, eficiencia asintótica bajo errores normales y son asintóticamente normales.

Referencias

[1] Yohai V. (1987). High Breakdown-point and high efficiency estimates for regression, *The Annals of Statistics*, **15**, 642-656