

ESTIMACIÓN PARA EL MODELO DE REGRESIÓN ZIP PARCIALMENTE LINEAL: UNA PROPUESTA ROBUSTA

María José Llop

Facultad de Ingeniería Química, Universidad Nacional del Litoral, Argentina
llopmariajose@gmail.com

En diversas áreas del conocimiento surgen datos de conteo que pueden ser modelados mediante distribuciones discretas como la Poisson o la Binomial Negativa, sin embargo, en determinadas situaciones es frecuente que los datos exhiban una gran proporción de ceros y, por lo tanto, no permitan suponer que la distribución subyacente sea alguna de las mencionadas. El modelo de regresión de Poisson inflado con ceros (ZIP, por sus siglas en inglés) es un caso particular de los modelos lineales generalizados (MLG). El mismo utiliza la distribución binomial para modelar el hecho de que una observación provenga del proceso de ceros estructurales (con probabilidad π), o bien, que provenga de una distribución de Poisson de parámetro λ (con probabilidad $1 - \pi$). La estimación de los parámetros de este modelo se puede realizar mediante el algoritmo EM, incluyendo variables auxiliares como si fueran observables y escribiendo la función de verosimilitud como la suma de componentes que se pueden optimizar por separado. Una desventaja de los estimadores basados en verosimilitud es que su función de influencia no es acotada y por consiguiente valores extremos tanto en la respuesta como en las covariables pueden afectar considerablemente a los estimadores. En ese contexto, estimadores robustos han sido desarrollados para el modelo de regresión ZIP, utilizando, por ejemplo, funciones de pérdida acotadas en lugar de la función de verosimilitud.

Una forma natural de dotar de mayor flexibilidad a los MLG es incorporar algunas variables predictoras de manera no paramétrica. Esto da lugar a los modelos parcialmente lineales generalizados (MPLG). En el campo de la estadística robusta, se han realizado propuestas para MPLG derivando estimadores robustos tanto para la componente lineal como para la no paramétrica. Estos estimadores involucran esencialmente funciones de pérdida acotadas, con ciertos pesos que permiten controlar el efecto de las variables predictoras sobre el estimador resultante.

En este trabajo se obtienen estimadores para el modelo de regresión ZIP parcialmente lineal combinando el algoritmo EM con una adaptación del procedimiento de tres pasos propuesto por [2] que permite estimar tanto la componente lineal como la componente no paramétrica. Este procedimiento se implementa utilizando la función de verosimilitud, así como funciones de pérdida robustas. En particular, para la estimación del parámetro de regresión y la componente no paramétrica asociados al proceso de Poisson se utiliza la pérdida bicuadrada de Tukey así como las que fueron propuestas por [3] y [4]. Además para el parámetro de regresión asociado a la distribución binomial se utiliza la pérdida propuesta por [1]. Finalmente, se compara el comportamiento y desempeño de los estimadores en diferentes escenarios de contaminación mediante estudios de simulación.

Trabajo en conjunto con María José Llop, Andrea Bergesio y Anne-Françoise Yao.

Referencias

- [1] Bianco, A.M. and Yohai, V.J. (1996). Robust Estimation in the Logistic Regression Model. In: Rieder, H. (eds) Robust Statistics, Data Analysis, and Computer Intensive Methods. Lecture Notes in Statistics, vol 109. Springer, New York, NY.
- [2] Boente, G. and Rodriguez, D. (2010). Robust inference in generalized partially linear models. Computational Statistics and Data Analysis, 54(12):2942–2966.
- [3] Muller, N. and Yohai, V. (2002). Robust estimates for arch processes. Journal of Time Series Analysis, 23(3):341–375.
- [4] Valdora, V. and Yohai, V. (2014). Robust estimators for generalized linear models. Journal of Statistical Planning and Inference, 146:31–48.