

Miguel Andrés Marcos
 FIQ, CONICET - UNL, Argentina
 mmarcos@santafe-conicet.gov.ar

El método de PLS fue introducido en 1983 [7] como un método de estimación en un modelo lineal en el caso de que el número de predictores es más grande que el tamaño de muestra, sin una contraparte poblacional correspondiente. Es uno de los métodos de predicción más importantes de regresión lineal en el área de la quimiometría, sin embargo su naturaleza algorítmica dificultó el estudio de sus propiedades estadísticas, con la consecuencia de que PLS fue considerada durante mucho tiempo como una técnica que es útil, pero cuyas propiedades estadísticas son elusivas.

En 1990 Helland [6] mostró que en un contexto de regresión lineal, el algoritmo PLS en realidad obtiene una expresión cerrada para la estimación de parámetros, obteniendo también una correspondiente fórmula para la versión poblacional. Finalmente en un trabajo de Cook, Helland y Su [5] se describe un modelo de reducción conocido como envelope regression y se prueba una conexión entre este y PLS.

En 2010, Chun y Keles [2] probaron que el estimador PLS era consistente cuando $p/n \rightarrow 0$ pero inconsistente en otros casos, bajo ciertas condiciones (“sparsity”). Por esta razón se limitó la discusión al estudio de propiedades del estimador muestral para el caso $n > p$ con p fijo y $n \rightarrow \infty$, mientras que el caso $p > n$ quedó abierto. Más adelante, Cook y Forzani [3,4] estudiaron este modelo en un contexto de alta dimensión y obtuvieron resultados sobre la consistencia n, p -asintótica bajo condiciones distintas de las de Chun y Keles (“abundance”).

Basándose en los trabajos [5,3], en [1] se estudiaron condiciones para obtener la distribución n, p -asintótica del estimador para el caso PLS de una componente. En esta charla presentaremos la distribución n, p -asintótica del estimador PLS en el caso general, a partir de los resultados de consistencia de [4].

Trabajo en conjunto con R. Dennis Cook (University of Minnesota) y Liliana Forzani (FIQ, CONICET - UNL, Argentina).

Referencias

- [1] J. Basa, D. Cook, L. Forzani, M. Marcos. Asymptotic distribution of one-component partial least squares regression estimators in high dimensions. *Can J Statistics*, 52: 118–130 (2024).
- [2] H. Chun, S. Keles. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society Series B*, 72(1): 3–25 (2010).
- [3] R.D. Cook, L. Forzani. Big data and partial least-squares prediction. *Canadian Journal of Statistics*, 46(1): 62–78 (2018).
- [4] R.D. Cook, L. Forzani. Partial least squares prediction in high-dimensional regression. *Ann. Statist.* 47 (2) 884–908 (2019).
- [5] R.D. Cook, I.S. Helland, Z. Su. Envelopes and partial least squares regression. *Journal of the Royal Statistical Society Series B* 75: 851–877 (2013).
- [6] I. S. Helland. Partial least squares regression and statistical models. *Scandinavian Journal of Statistics*, 17(2): 97–114 (1990).
- [7] S. Wold, H. Martens, H. Wold. The multivariate calibration problem in chemistry solved by the pls method. In Kågström, B. and Ruhe, A., editors, *Matrix Pencils*, 286–293, Berlin, Heidelberg (1983). Springer Berlin Heidelberg.