Profundidades en Espacios Finitos e Infinitos Dimensionales: Un Poco de Historia y Aplicaciones.

Lucas Fernandez Piana

Un conjunto de datos en la recta real cuenta con un orden usual que da una noción de ranking entre las observaciones, al pasar a dimensiones mayores esto no se puede hacer de manera natural. Las medidas de profundidad para datos multivariados surgieron con el objetivo de extender la noción de los estadísticos de orden. Los mismos son sumamente útiles al emplear técnicas no paramétricas, sobre todo en casos donde la distribución de los datos no es gaussiana. Además, es importante contar con técnicas de "ordenación" que tengan en cuenta la geometría subyacente. Lo que no ocurre, por ejemplo si se busca definir la observación central mediante la media o la mediana coordenada a coordenanda. A grandes rasgos, para una distribución P en $\mathbb{R}^d$, una medida de profundidad es una función D(x, P) que provee un orden del "centro" hacia afuera para puntos $x \in \mathbb{R}^d$ basado en P, Zuo y Selfling (2000) [16].

Por otra parte, juegan un rol importante a la hora de describir características de la distribución subyacente a las observaciones. Es decir, no sólo proveen un funcional robusto multivariado para el problema de localización a través del punto más profundo, sino que también revelan información sobre la dispersión, forma y simetría de la distribución cuando se contruyen las regiones de profundidad como describe Serfling (2004) [15]. Otras posibilidades que ofrecen se pueden encontrar en los llamados métodos basados en profundidades que permiten abordar distintos problemas de inferencia. Por ejemplo, tests de posición y diferencias de escala basados en los DD-plots, que fueron introducidos en primera instancia como una herramienta gráfica para exploración de datos, Liu et al. (1990) [13], Li y Liu (2004) [12]; diagnóstico de no normalidad (Liu et al. (1990) [13]) y detección de outliers (Chen et al. 2009 [4]). Recientemente, las profundidades se han utilizado extensivamente en el contexto de clasificación como se ve reflejado, entre muchos otros autores, en los artículos de Ghosh y Chaudhuri (2005) [9], Li et al. (2012) [11], Cuevas y Fraiman (2009) [6], Dutta y Ghosh (2012) [7], Paindaveine and Van Bever (2012) [14], Cuesta-Albertos et al. (2017) [6].

Las nuevas tecnologías nos permiten recopilar datos con exactitud y alta frecuencia en una variable temporal o espacial generando grandes volúmenes de información. El aumento en la velocidad de procesamiento de las computadoras hizo que el análisis de esta clase de datos sea viable no sólo en grandes centros de cómputos sino también en computadoras personales. Esta democratización hizo que los datos funcionales se conviertan en un tema de investigación estadística muy activo. Además cabe mencionar que se utilizan para modelar fenómenos en la ciencia, ingeniería, medicina, economía y cualquier tópico donde tenga sentido registrar una curva. La extensión de las profundidades al contexto de espacios infinito dimensionales no es inmediata. El foco del problema se encuentra en que el análogo natural en muchos casos resulta en medidas nulas en un conjunto de probabilidad uno: Dutta et al. (2011) [8], Chakraborty y Chaudhuri (2014)

[3] y Kuelbs y Zinn (2013) [10].

Una debilidad de las profundidades se encuentra en su incapacidad para detectar multimodalidad. Sin embargo, la existencia de múltiples centros, su localización e importancia relativa son de interés tanto desde la perspectiva teórica como aplicada. Es el caso de las distribuciones mixtas y en análisis de clusters o conglomerados. En los últimos años, con el espiríritu de resolver este problema, se han introducido definiciones de profundidades locales. Por el momento, en la literatura hay muy pocos trabajos sobre este nuevo concepto: Agostinelli y Romanazzi (2011) [2], Pandeveine y Van Bever (2013) [14] y Agostinelli (2016) [1]; aunque todavía es un campo poco explorado.

Referencias

[1] Agostinelli, C. "Local half-region depth for functional data". (2018). *Journal of Multivariate Analysis*, 163, 6779.

[2] Agostinelli, C. and Romanazzi, M. (2011). "Local Depth." *Journal of Statistical Planning and Inference,* **141**, 817-830.

[3] Chakraborty, B. y Chaudhuri, P. "On data depth in infinite dimensional spaces". (2014). *Annals of the Institute of Statistical Mathematics*, 66, 303324.

[4] Chen, Y., Dang, X., Peng, H., y Bart, H., L., J. "Outlier detection with the kernelized spatial depth function". *IEEE Transactions in Pattern Analysis and Machine Intelligence*, 31, 288305.

[5] Cuesta-Albertos, J.A., Febrero-Bande, M. y Oviedo de la Fuente, M. (2017). "The DDG-classifier in the functional setting". *Test*, 26 (1), 119142.

[6] Cuevas A. y Fraiman R. (2009). "On Depth Measure and Dual Statistics. A Methodology for Dealing With General Data". *Journal of Multivariate Analysis*, 100, 753-766.

[7] Dutta, S. y Ghosh, A, K. "On robust classification using projection depth". (2012) *Annals of the Institute of Statistical Mathematics*, 64, 657676.

[8] Dutta, S., Ghosh, A, K. y Chaudhuri, P. "Some intriguing properties of Tukeys halfspace depth". (2011). *Bernoulli*, 17, 14201434.

[9] Ghosh, A. K. y Chaudhuri, P. "On maximum depth and related classifiers". (2005). *Scandinavian Journal of Statistics*, 32 (2), 327-350.

[10] Kuelbs, J. y Zinn, J. "Concerns with functional depth". (2013). *ALEA*, 10, 831855.

[11] Li, J., Cuesta-Albertos, J. A., y Liu, R. Y. (2012) "DDclassifier: Nonparametric classification procedure based on DDplot. *Journal of the American Statistical Association*, 107 (498), 737753.

[12] Li, J. y Liu, R. Y. (2004). "New nonparametric tests of multivariate locations and scales using data depth". *Statistical Science*, 19 (4), 686696.

[13] Liu, R. Y., Parelius, J. M. y Singh, K. (1990). "Multivariate analysis by data depth: Descriptive statistics, graphics and inference, (with discussion and a rejoinder by Liu and Singh)". *The Annals of Statistics*, 27 (3), 783858.

[14] Paindavaine, D. y Van vever, G. (2013). "From Depth to Local Depth: A Focus on Centrality". *Journal of the American Statistical Association*, **108**, 503, 1105-1119.

[15] Serfling, R. "Nonparametric multivariate descriptive measures based on spatial quantile". (2004). *Journal of Statistical Planning Inference*, 123, 259278.

[16] Zuo, Y. y Serfling R. (2000). "General Notion of Statistical Depth Function". *The Annals of Statistics*, **28**, 461-482.