# Lasso robusto para la selección de covarianza en modelos gráficos Gaussianos

Expositor: Marcelo Ruiz (Universidad Nacional de Río Cuarto, ivanmarce@gmail.com)
Autor/es: Marcelo Ruiz (Universidad Nacional de Río Cuarto, ivanmarce@gmail.com); Rubén Zamar (Universidad de British Columbia, ruben@stat.ubc.ca); Ginette Lafit (Universidad de Leuven, ginettelafit19@gmail.com); Francisco Nogales (Universidad Carlos III de Madrid, fco-javier.nogales@uc3m.es)

Los modelos gráficos Gaussianos (MGG) de alta dimensión son utilizados para representar la dependencia lineal entre variables dada por las correlaciones parciales de cada par de variables condicional a las restantes. Esta estructura de dependencia queda caracterizada por las entradas no nulas - fuera de la diagonal- de la inversa de la matriz de covarianza. La selección de covarianza (SC) consiste en, basada en una muestra, determinar cuáles son esas entradas significativamente no nulas.

Sea $\mathbf{X} = (X_1, \ldots, X_p)^\top \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma)$ y $\Omega = \widehat{\Sigma}^{-1}$ la matriz de precisión. Si $\mathbf{x}_1, \ldots, \mathbf{x}_n$ es una muestra de $\mathbf{X}$ y $n > p$ entonces la matriz de covarianza muestral $\widehat{\Sigma}$ es un buen estimador de $\Sigma$ y puede ser utilizado para estimar $\Omega$ definiendo $\widehat{\Omega} = \widehat{\Sigma}^{-1}$. Pero, si $p > n$ la matriz de covarianza muestral no es invertible y el estimador de máxima verosimilitud de $\boldsymbol{\Sigma}$ no existe.

Para tratar con este problema se han desarrollado alternativas de SC asumiendo que $\Omega$ es rala, en particular la del tipo lasso que define

$$\widehat{\Omega}_L = \mathrm{argmin}_{\Omega \succ 0} \{ \mathrm{tr}(\Omega \widehat{\Sigma}) - \mathrm{logdet}(\Omega) + \lambda \parallel \Omega \parallel_{1,\mathrm{off}} \} \tag{1}$$

donde

$$\parallel \Omega \parallel_{1,\mathrm{off}} := \sum_{i \neq j} |\omega_{lj}| \quad \text{for } i, j = 1, \ldots, p,$$

con $\lambda > 0$ una constante de regularización.

La matriz de covarianza y de correlación muestrales son muy sensibles a la presencia de outliers multidimensionales provocando una pobre recuperación del MGG y una estimación sesgada de $\Omega$ y, peor aún es el resultado, si la contaminación obedece al modelo de contaminación independiente.

Teniendo en cuenta (1), en una estrategia de tipo plug-in, se puede alcanzar un estimador de $\Omega$ resistente a contaminación utilizando un estimador robusto de la matriz de covarianza, $\widehat{\Sigma}_R$. Nosotros proponemos como estimador robusto de $\Omega$ a aquel basado en la propuesta de Khan (2006) de estimación bivariada Winsorizada y ajustada de $\Sigma_{ij}$. En esta presentación mostramos que el desempeño de nuestro estimador es superior a los existentes en la literatura tales como los de Tarr, Müller y Weber (2016) y Öllerer y Croux (2015).