High-dimensionality effects in the Markowitz problem and quadratic programs with linear constraints

Noureddine El Karoui
Department of Statistics
University of California, Berkeley
nkaroui@stat.berkeley.edu

Abstract

It is often the case in statistics and various branches of applied mathematics that one wishes to solve optimization problems involving parameters that are estimated from data. It is therefore natural to try to characterize the relationship between the solution of the optimization problem involving estimated parameters (the sample version) and the solution we would get if we knew the actual value of the parameters (the population version). An example of particular interest to some is the classical Markowitz portfolio optimization problem in finance, which is an instance of a quadratic program with linear equality constraints.

I discussed some of these questions in the large dimensional setting when the optimization is performed over vectors of size p, and p is comparable to n, the number of observations we use to get our estimates. From a practical standpoint, this asymptotic setting (p and n go to infinity while p/n does not go to zero) tries to capture the difficulties arising from the fact that we have limited amount of data to estimate the parameters appearing in the problem.

I presented results showing that the high-dimensionality of the data (i.e p/n not small) implies significant and quantifiable risk underestimation, in both the case of quadratic programs with linear equality constraints and linear inequality constraints, when the number of constraints is fixed in the asymptotics. I also considered the question of robustness of the conclusions to various distributional assumptions, focusing on understanding the sensitivity of the results to heavy-tails and time correlation. Finally, I discussed the impact of working with non-independent observations and a significant non-classical failure of the bootstrap. A possible robust correction of the problems was also proposed.

The analysis is based on random matrix theory.