Normality in high-dimensional statistics

Andrea Montanari

Classical statistical theory delineates precise conditions under which statistical estimators are asymptotically normal. The basic mechanism is relatively simple. Consider, for instance, maximum likelihood. The estimator converges to its population value, and its distribution is determined by the distribution of the score at the population value, which is a sum of i.i.d. terms (for i.i.d. samples). The central limit theorem applies. It has long been known that this theory breaks down when the number of parameters to be estimated is comparable with the number of samples. In these settings, statistical theory has mostly focused on deriving estimation rates, i.e. upper and lower bounds on the estimation error that hold up to constant factors. However, over the last few years, it has become possible to establish normality results, of an entirely different type, in high-dimensional settings. In these problems, the distance of the estimator from its population value is not small, and its distribution needs to be determined in a non-perturbative way. I will discuss a few examples of this analysis, and its implications for statistical inference.