

Sub-Gaussian estimators of the mean of a random matrix with entries possessing only two moments

Stanislav Minsker, University of Southern California

Estimation of the covariance matrix has attracted a lot of attention of the statistical research community over the years, partially due to important applications such as Principal Component Analysis.

However, frequently used empirical covariance estimator (and its modifications) is very sensitive to outliers in the data. As P. Huber wrote in 1964, "...This raises a question which could have been asked already by Gauss, but which was, as far as I know, only raised a few years ago (notably by Tukey): what happens if the true distribution deviates slightly from the assumed normal one? As is now well known, the sample mean then may have a catastrophically bad performance..."

Motivated by this question, we develop a new estimator of the (element-wise) mean of a random matrix, which includes covariance estimation problem as a special case. Assuming that the entries of a matrix possess only finite second moment, this new estimator admits sub-Gaussian or sub-exponential concentration around the unknown mean in the operator norm. We will explain key ideas behind our construction, as well as applications to covariance estimation and matrix completion problems.