Concentration Bounds and Asymptotic Distribution for the Empirical Spectral Projectors of Sample Covariance Operators

Karim Lounici
Georgia Inst. of Technology

September 10, 2015

(Joint work with Vladimir Koltchinskii) Let $X, X_1, \ldots, X_n$ be i.i.d. Gaussian random variables in a separable Hilbert space $\mathbb{H}$ with zero mean and covariance operator $\Sigma = \mathbb{E}(X \otimes X)$, and let $\hat{\Sigma} := n^{-1} \sum_{j=1}^n (X_j \otimes X_j)$ be the sample (empirical) covariance operator based on $(X_1, \ldots, X_n)$. Denote by $P_r$ the spectral projector of $\Sigma$ corresponding to its $r$-th eigenvalue $\mu_r$ and by $\hat{P}_r$ the empirical counterpart of $P_r$. We derive tight bounds on

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P} \left\{ \frac{\| \hat{P}_r - P_r \|_2^2 - \mathbb{E}\| \hat{P}_r - P_r \|_2^2}{\text{Var}^{1/2}(\| P_r - P_r \|_2^2)} \leq x \right\} - \Phi(x) \right|,$$

where $\| \cdot \|_2$ denotes the Hilbert–Schmidt norm and $\Phi$ is the standard normal distribution function. These bounds depend on the so called effective rank of $\Sigma$ defined as $r(\Sigma) = \frac{\text{tr}(\Sigma)}{\| \Sigma \|_\infty}$, where $\text{tr}(\Sigma)$ is the trace of $\Sigma$ and $\| \Sigma \|_\infty$ is its operator norm, as well as another parameter characterizing the size of $\text{Var}(\| \hat{P}_r - P_r \|_2^2)$.

For an eigenvalue $\mu_r$ of $\Sigma$ of multiplicity 1 with associated eigenvector $\theta_r$, we derive new properties about the bias of the sample covariance eigenvector as an estimator of $\theta_r$. As a consequence, we suggest a new simple estimator of $\theta_r$ with decreased bias and we derive a concentration bound on the $l_\infty$-norm of the deviation between the estimator and $\theta_r$. This result may be of interest to perform variable selection.