# Principal Component Analysis: a Berry-Esseen Bound for the Spectral Projectors of the Covariance Operator

Karim Lounici
Université de Nice-Sophia

Principal Component Analysis is a popular technique to study the covariance structure of a random vector. In a recent series of papers (Koltchinskii and Lounici, 2014a,b, 2015), we proved several new results about Principal Component Analysis in an infinite dimensional setting. One result of interest is about the asymptotic distribution of the empirical spectral projectors.

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 obtain the following Berry-Esseen type bound that quantifies the accuracy of the normal approximation in term of the effective rank $\mathbf{r}(\Sigma)$ and another quantity $B_r(\Sigma)$ characterizing the order of magnitude of $\mathrm{Var}(\|\hat{P}_r - P_r\|_2^2)$. We have

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}\left\{ \frac{\|\hat{P}_r - P_r\|_2^2 - \mathbb{E}\|\hat{P}_r - P_r\|_2^2}{\mathrm{Var}^{1/2}(\|\hat{P}_r - P_r\|_2^2)} \leq x \right\} - \Phi(x) \right| \lesssim \left[ \frac{1}{B_r(\Sigma)} + \frac{\mathbf{r}(\Sigma)}{B_r(\Sigma)\sqrt{n}} \sqrt{\log\left( \frac{B_r(\Sigma)\sqrt{n}}{\mathbf{r}(\Sigma)} \bigvee 2 \right)} \right],$$

where $\Phi(x)$ denotes the distribution function of standard normal random variable.

This is a joint work with Vladimir Koltchinskii.

## References

Koltchinskii, V. and K. Lounici (2014a). Asymptotics and concentration bounds for spectral projectors of sample covariance. ArXiv:1408.4643, submitted.

Koltchinskii, V. and K. Lounici (2014b). Concentration inequalities and moment bounds for sample covariance operators. *Bernoulli*, to appear. ArXiv:1405.2468.

Koltchinskii, V. and K. Lounici (2015). Normal approximation and concentration of spectral projectors of sample covariance. ArXiv:1504.07333, submitted.