# Math 690: Topics in Data Analysis and Computation

Lecture notes for Fall 2017

Scribed by Yixin Lin, Shen Yan

# September 12

Last time we talked about Marchenko-Pastur Law, which is the null case, i.e. $y_i = z_i \sim N(0, \sigma^2 I)$. Today we will talk have a few remarks on M-P law, and move on to introduce spiking law and possible better estimator than $S$.
Remark:

1. if we replace Gaussian distribution with other distributions, say $x_{ij} \sim Bernoulli, P(x_{ij} = 1) = P(x_{ij} = 1) = 1/2$, we still get the semi-circle law. Generally speaking, if $[X]_{p\times n} = [x_1, \cdots, x_n]$, $S = \frac{1}{n} XX^T$, then if $x_{ij} \sim p(x)$ i.i.d. , and has finite first 4 moment, then the distribution of $S$'s eigenvalues, eigenvectors satisfy M-P law.

2. "edge law": the largest eigenvalue of $S$, $\hat{\lambda}_1(S) \to^{a.s.} b(\gamma)$. "Tracy-widon" law: $\frac{\hat{\lambda}_1(S) - b}{n^{-3/2}} \to^{distribution}$ Tracy-Widon density.
please add: … in the limit of both p and n…
and "same for smallest eigenvalue (when gamma \neq 1)".

3. Recall that $\mathbb{E}\|S - \Sigma\|_F^2 \sim \frac{cp^2}{n}$, if the true covariance matrix is identity matrix, $\Sigma = I_p$, then we do have curse of dimensionality, i.e. the bound is tight and $p$ matters. To see this, we can compute
\sim (this is the limit)
$$\|S - I\|_F^2 = \sum_{i=1}^{p} (\lambda_i - 1)^2 = p \int_{\mathbb{R}} (t-1)^2 p_{MP}(t) dt = const(\gamma) p \sim c\frac{p}{n} p$$
The second equality comes from the definition of weakly convergence.
in next line in text:
"while the previous upper bound gives that cp^2/n which is c*p when p \sim \gamma n."

## Spiking Model

Moving past the null case, we know consider $y_i = x_i + z_i, x_i \sim p_x = N(0, \Sigma_x), z_i \sim N(0, \sigma^2 I)$, also $\Sigma_x = uu^T, \|u\| = 1$. Now we want to find the properties of the eigenvalues, eigenvectors of $S_y = \frac{1}{n}\sum y_i y_i^T$ under the conditions that $p, n \to \infty, p/n \to \gamma$.
We define $R = \frac{1}{\sigma^2}$, this can be seen as the signal-to-noise ratio. We rescale $y_i \to y_i\sqrt{R}$, correspondingly $z_i \sim N(0, I), x_i \sim N(0, Ruu^T)$. Let $\hat{\lambda}_1$ be the largest eigenvalue of $S_y$, $\hat{v_1}$ be the corresponding eigenvector, then the theorem states:
When $p, n \to \infty, p/n \to \gamma$,

$$\hat{\lambda}_1 \to^{a.s.} (\lambda_1)_\infty = \begin{cases} b(\gamma) = (1+\sqrt{\gamma})^2 & R \le \sqrt{\gamma} \\ (1+R)(1+\frac{\gamma}{R}) & R > \sqrt{\gamma} \end{cases}$$

In terms of eigenvectors, we have

$$|\hat{v}_1^T u|^2 \to^{a.s.} c_\infty = \begin{cases} 0 & R \le \sqrt{\gamma} \\ \frac{1-\frac{\gamma}{R^2}}{1+\frac{\gamma}{R}} & R > \sqrt{\gamma} \end{cases}$$

If we plot how $(\lambda_1)_{infty}$ and $c_\infty$ changes with $R$(skip the plot :) ), we can see the phase transition at $R = \sqrt{\gamma}$, called the BBP transition.

## Better estimator than $S$?

One way of improving the estimation is spectral shrinkage applied to the sample covariance matrix.

Suppose $S_y$ is the sample covariance matrix of $\{y_i\}_i$, and we have the eigendecomposition

$$S_y = \hat{V}\hat{\Lambda}\hat{V}^T.$$

The estimator has the form of

\hat{\Sigma}_x

$$S_x = \hat{V} f(\hat{\Lambda})\hat{V}^T = \sum_{k=1}^{p} f(\hat{\lambda}_k)\hat{v}_k\hat{v}_k^T,$$

where $f : \mathbb{R} \to \mathbb{R}$ is a function. Due to the relation between $R$ and $(\lambda_1)_\infty$ in the phase transition theorem above, we would like $f(\lambda_1)$ to recover $R$, and this leads to $f$ being a threshold function which vanishes whenever $\lambda_1 < \sqrt{\gamma}$ (Ex. verify this).

Stein's Phenomenon:

Can we find estimators with lower risks than MLE?

For example, $y \sim N(\mu, \Sigma)$ and we want to estimate $\mu$, lower risk means having lower $\mathbb{E}\|\hat{\mu} - \mu\|^2$. Consider 2 dimension (1 dimension is similar). Given $n = 1$, the MLE is sample mean which is $y$ itself, namely $\hat{\mu}_1(y) = y$. We can let $\hat{\mu}_2 = \rho y, \rho < 1$. For $\hat{\mu}_1$ the risk is $\mathbb{E}\|y - \mu\|^2 = 2$. For $\hat{\mu}_2$, the risk is $\mathbb{E}\|\rho y - \mu\|^2 = (\rho - 1)^2\|u\|^2 + 2\rho^2$, we can certainly choose $\rho$ such that it is smaller than 2. More generally, Charles Stein had the following result

$$\text{If } p \geq 3, y \sim N(\mu, \sigma^2 I), \exists \tilde{\mu}(y) \text{ s.t. } risk(\tilde{\mu}) < risk(\hat{\mu}^{MLE}), \tilde{\mu}(y) = (1 - \frac{\sigma^2}{\|\hat{\mu}^{MLE}\|})\hat{\mu}^{MLE}$$

it does not depend on p? James-Stein estimator has a (p-2) factor. check again.