Prof. Dr. Rudolf Mathar, Dr. Arash Behboodi, Markus Rothe

Tutorial 6
- Proposed Solution -

Friday, November 23, 2018

Solution of Problem 1

a) The dominant eigenvalue λ_{dom} is visible when the ratio $\gamma_2 = \frac{p}{n_2}$ is less than β_{dom}^2. With $\beta_{\text{dom}} = \beta_2 = 0.5$ we obtain $n_{\min} = n_2 = \frac{p}{\beta_2^2} = 2000$. For this number of samples, the dominant eigenvalue of the sample covariance $\mathbf{S}_n$ tends to $(1+\sqrt{\gamma_2})^2 = (1+0.5)^2 = 2.25 \gg 1.5$. The distance $\langle \boldsymbol{v}_2, \boldsymbol{v}_{\text{dom}} \rangle = \frac{1-\gamma_1/\beta_1^2}{1-\gamma_1/\beta_1}$ is equal to zero. Figure 1 shows eigenvalue distributions for this choice.

b) To see both eigenvalues the ratio $\gamma_1 = \frac{p}{n_1}$ must be less than β_1^2. With $\beta_1 = 0.2$ we obtain $n_1 = \frac{p}{\beta_1^2} = 12500$. For this number of samples, the dominant eigenvalue λ_{dom} of the sample covariance $\mathbf{S}_n$ tends to $(1+\beta_2)(1+\frac{\gamma_1}{\beta_2}) = 1.5 \cdot 1.08 = 1.62 \approx 1.5 = 1+\beta_2$. The distance $\langle \boldsymbol{v}_2, \boldsymbol{v}_{\text{dom}} \rangle = \frac{1-\gamma_1/\beta_2^2}{1-\gamma_1/\beta_2}$ is equal to $0.913 \approx 1$ which shows that $\boldsymbol{v}_2$ is nearly a unit norm vector parallel to the dominant eigenvector $\boldsymbol{v}_{\text{dom}}$. Figure 2 shows eigenvalue distributions for this choice.

By enlarging n to 50000 both eigenvalues β_1 and β_2 become visible in the Marchenko-Pastur density as shown in Figure 3.

Solution of Problem 2

a)

$$\mathbf{E}_k \mathbf{x}^{(j)} = \left(\mathbf{I}_k - \frac{1}{k}\mathbf{1}_k\mathbf{1}_k^{\mathrm{T}}\right)\mathbf{x}^{(j)} = \mathbf{x}^{(j)} - \frac{1}{k}\mathbf{1}_k\mathbf{1}_k^{\mathrm{T}}\mathbf{x}^{(j)} = \mathbf{x}^{(j)} - \frac{1}{k}\sum_{i=1}^{k} x_i^{(j)}\mathbf{1}_k$$
$$= \mathbf{x}^{(j)} - \overline{x}^{(j)}\mathbf{1}_k$$

b)

$$\left(\mathbf{E}_k\mathbf{X}^{\mathrm{T}}\right)_{ij} = \left[\mathbf{E}_k\mathbf{x}^{(1)}, \mathbf{E}_k\mathbf{x}^{(2)}, \ldots, \mathbf{E}_k\mathbf{x}^{(n)}\right]_{ij} = \left(\mathbf{x}^{(j)} - \overline{x}^{(j)}\mathbf{1}_k\right)_i = x_i^{(j)} - \overline{x}^{(j)}$$

c)

$$\sum_{i=1}^{k}\left(\mathbf{E}_k\mathbf{X}^{\mathrm{T}}\right)_{ij} = \sum_{i=1}^{k}\left(x_i^{(j)} - \overline{x}^{(j)}\right) = \sum_{i=1}^{k} x_i^{(j)} - \sum_{i=1}^{k}\overline{x}^{(j)} = k\overline{x}^{(j)} - k\overline{x}^{(j)} = 0$$

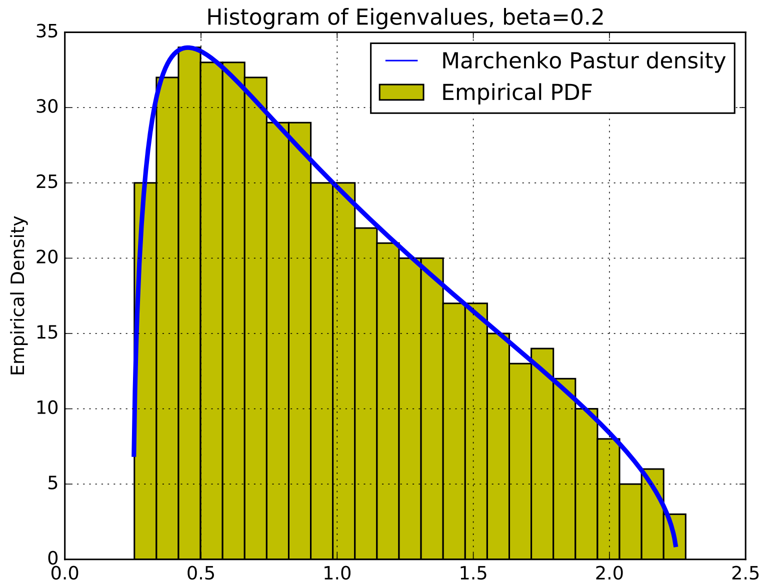

Figure 1: Eigenvalues of $\mathbf{S}_n$ for Spike model with $\beta_1 = 0.2, \beta_2 = 0.5$, $n = 2000$

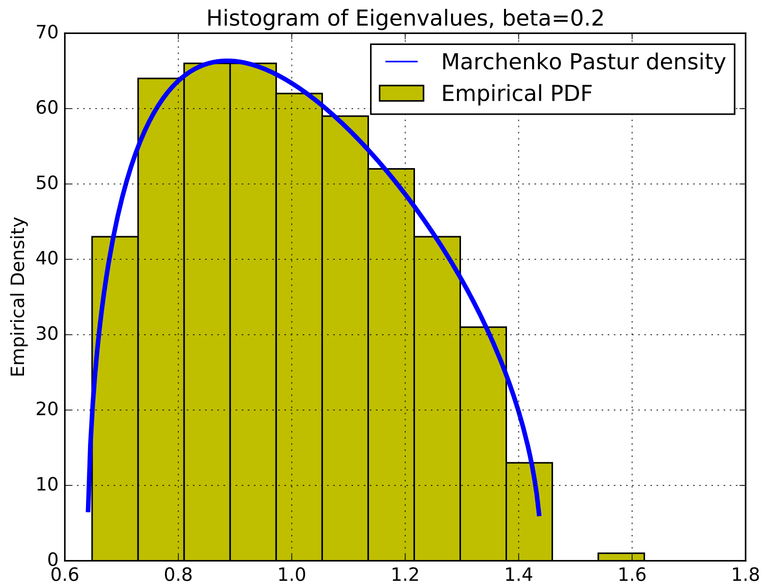

Figure 2: Eigenvalues of $\mathbf{S}_n$ for Spike model with $\beta_1 = 0.2, \beta_2 = 0.5$, $n = 12500$

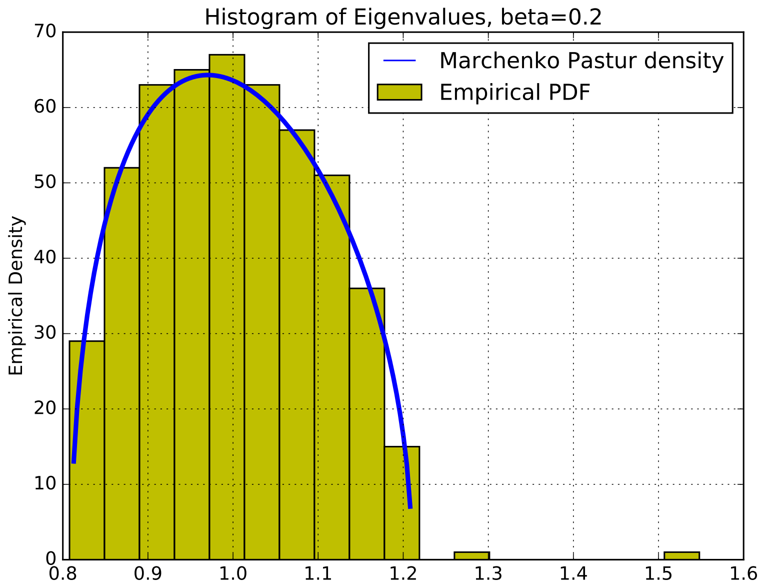

Figure 3: Eigenvalues of $\mathbf{S}_n$ for Spike model with $\beta_1 = 0.2, \beta_2 = 0.5$, $n = 50000$

Solution of Problem 3

We start by expanding the following difference

$$(1+\beta)(1+\frac{\gamma}{\beta}) - (1+\sqrt{\gamma})^2 = 1 + \frac{\gamma}{\beta} + \beta + \gamma - (1 + 2\sqrt{\gamma} + \gamma) = \frac{\gamma}{\beta} + \beta - 2\sqrt{\gamma}$$
$$= \frac{\gamma - 2\beta\sqrt{\gamma} + \beta^2}{\beta} = \frac{(\sqrt{\gamma} - \beta)^2}{\beta} \,.$$

Since $\beta > 0$ we have that

$$(1+\beta)(1+\frac{\gamma}{\beta}) - (1+\sqrt{\gamma})^2 = \frac{(\sqrt{\gamma}-\beta)^2}{\beta} > 0 \,,$$

yielding

$$(1+\beta)(1+\frac{\gamma}{\beta}) > (1+\sqrt{\gamma})^2$$

which proves the statement.