PCA in high dimensions

- \( X_n = \{x_i\}_{i \leq n} \subset \mathbb{R}^p \)
- \( X_n = [x_1 \cdots x_n] \in \mathbb{R}^{p \times n} \)
- \( x_i \sim \mathcal{N}(0, I) \)
- \( S_n = (1/n)X_nX_n^T \)
- Examine spectral properties of \( S_n \) when \( p, n \to \infty \) and \( p/n = \gamma \leq 1 \)
- In other words, how does PCA perform in high dimensions?
PCA in high dimensions

\[ p = 500, \ n = 1000 \]
Eigenvalues of \( S_n = (1/n)X_nX_n^T \)
PCA in high dimensions

$p = 500$, $n = 1000$

Histogram of eigenvalues of $S_n = (1/n)X_nX_n^T$
### Spike model

- \( X_n = \{x_i\}_{i \leq n} \subset \mathbb{R}^p \)
- \( X_n = [x_1 \cdots x_n] \in \mathbb{R}^{p \times n} \)
- \( x_i \sim \mathcal{N}(0, I + \beta vv^T) \), where \( \beta \geq 0 \) and \( v \in \mathbb{R}^p \)
- \( S_n = (1/n)X_nX_n^T \)
- Examine spectral properties of \( S_n \) when \( p, n \to \infty \) and \( p/n = \gamma \leq 1 \)
- Can PCA find the 1-dimensional structure of the data with variance \( 1 + \beta \) and along the direction \( v \)?
Spike model

\[ p = 500, \ n = 1000, \ \beta = 2 \]

Histogram of eigenvalues of \( S_n = \frac{1}{n}X_nX_n^T \)
Spike model

$p = 500$, $n = 1000$, $\beta = 1/2$

Histogram of eigenvalues of $S_n = (1/n)X_nX_n^T$