Stat-581: Probability and Statistical Inference for Data Science
Quiz - 5
James Stein estimator

Prakruti Joshi (phj15)

Problem:
Compare the risk of the James stein estimator vs the MLE for k =1 to 100 for random thetai, plot it, use 1000 samples at each k to estimate risk., use N(thetai, 1) for the samples.

Aim:
We want to compare the estimates for $\boldsymbol{\theta}$, which is a k-dimensional vector. The two estimators that are compared here are the James-Stein estimator and the Maximum Likelihood estimator (MLE).
We compare the risk (estimated mean square errors) of both the estimators as the dimension of $\boldsymbol{\theta}$ ranges from k=1 to 100. The k-dimensional vector essentially is the mean vector of the k-variate normal random variable. In our case, we take the variance (σ^2) as 1.
We first generate a K-dimensional vector of mean $\boldsymbol{\theta}$ as zero. We generate 1000 samples for each random variable ~ N($\boldsymbol{\theta}$k, 1).
We thus estimate the MLE estimator, James-stein estimator and the corresponding risks from these observations.

James-Stein estimator:
The intuition behind considering an estimator like James-stein estimator is that it is optimizing for the mean-squared error of a *combined* estimator. It is not the same as optimizing for the errors of separate estimators of the individual parameters.
This is based on **Stein's paradox**. It is basically the phenomenon that when three or more parameters are estimated simultaneously, there exist combined estimators more accurate on average (that is, having lower expected mean squared error) than any method that handles the parameters separately.

If σ^2 is known, the James-Stein estimator is given by:

$$\widehat{\boldsymbol{\theta}}_{JS} = \left(1 - \frac{(m-2)\sigma^2}{\|\mathbf{y}\|^2}\right)\mathbf{y}.$$

Where m = k dimensions, y = single observation

If more than one vector observations are available,

$$\widehat{\boldsymbol{\theta}}_{JS} = \left(1 - \frac{(m-2)\frac{\sigma^2}{n}}{\|\overline{\mathbf{y}}\|^2}\right)\overline{\mathbf{y}},$$

where $\overline{\mathbf{y}}$ is the m-length average of the n observations.

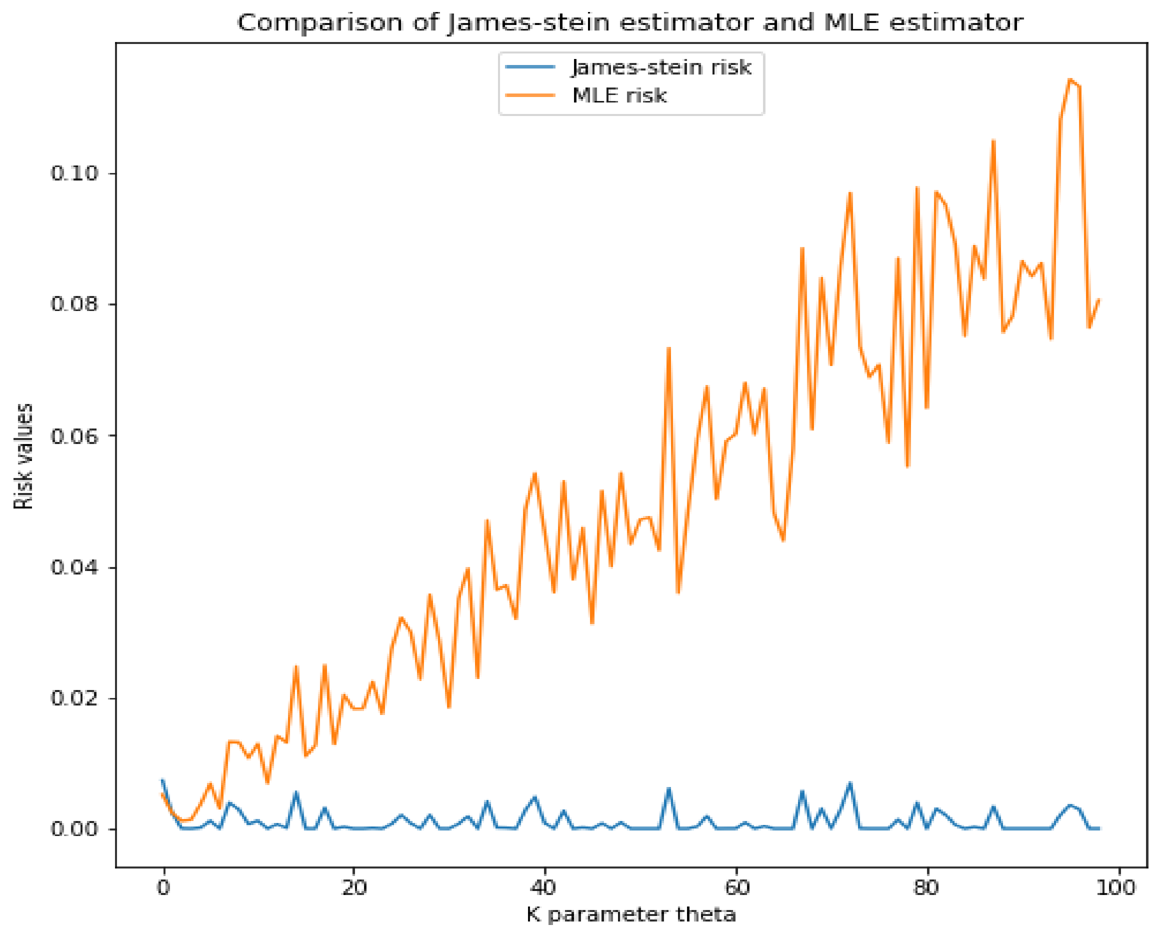
<u>Maximum Likelihood estimator (MLE) :</u>
The maximum likelihood estimate for K-dimensional $\boldsymbol{\Theta}$ mean is the sample mean from the observations sampled from a multivariate Gaussian distribution.

<u>Risk of an estimator:</u>

The quality of such an estimator is measured by its risk function. A commonly used risk function is the mean squared error of the original parameter and the estimated parameter, defined as

$$\mathrm{E}\left[\left\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\right\|^2\right].$$

<u>Analysis:</u>



For k > 2, the James-Stein estimator dominates the Maximum likelihood estimator as the risk of James-stein estimator is lower than the risk of MLE.
An estimator X is said to dominate another estimator Y, if, for all values of $\boldsymbol{\theta}$, the risk of X is lower than, or equal to, the risk of Y and if the inequality is strict for some $\boldsymbol{\theta}$. An estimator is said to be admissible if no other estimator dominates it, otherwise it is *inadmissible*.
From our observations, we can see that the least squares estimators are inadmissible when k>=3. The ordinary decision rule for estimating the mean of a multivariate Gaussian distribution is inadmissible under mean squared error risk. Thus, when three or more unrelated parameters are measured, their total MSE can be reduced by using a combined estimator such as the James–Stein estimator; whereas when each parameter is estimated separately, the least squares (LS) estimator is admissible.

References:

1. https://en.wikipedia.org/wiki/James%E2%80%93Stein_estimator
2. https://bookdown.org/content/922/james-stein.html
3. https://en.wikipedia.org/wiki/Stein%27s_example