James-Stein Estimation and Ridge Regression

MA677 Final Project

Fengyuan (Vincent) Shen

Department of Mathematics and Statistics
Boston University

May 6, 2024

Abstract

This paper examines the advancements in statistical estimation through the integration of shrinkage techniques, specifically the James-Stein Estimator and Ridge Regression, within an empirical Bayes framework. These methods demonstrate significant improvements in estimation accuracy and reliability, particularly in scenarios involving small sample sizes or high-dimensional data. The exploration begins with a deep dive into the mathematical foundations of these estimators, providing proofs of their superiority over traditional methods and linking their theoretical underpinnings. Practical computational methods are then applied to illustrate their effectiveness through simulations and comparative analyses. Finally, the implications of these techniques are discussed, emphasizing their role in enhancing the dependability of statistical predictions and addressing challenges such as overfitting and multicollinearity. This study highlights the practical and conceptual benefits of shrinkage methods in statistical practice, showcasing their potential to revolutionize traditional estimation approaches.

Keywords: *James-Stein Estimator, Empirical Bayes, Ridge Regression*

Contents

Introduction

The field of statistical estimation has seen substantial advancements with the introduction of shrinkage techniques. This project is inspired by the empirical Bayes framework and aims to delve deeper into its application through the James-Stein Estimator (JSE) and Ridge Regression—two pivotal methods that exemplify the integration of shrinkage to improve estimation accuracy and reliability.

My interest in this topic stems from a fascination with how Bayesian principles can be pragmatically employed within frequentist settings to address real-world problems, a synergy often overlooked in traditional methods. This approach is particularly appealing due to its potential to provide more accurate estimates in situations where classical methods falter, such as small sample sizes or high-dimensional data contexts.

The paper is organized as follows:

- **Section 2: Essential Insights from the Chapter** - This section presents the principal concepts and critical insights obtained from the study of the James-Stein Estimator and Ridge Regression, alongside pertinent questions and ideas that emerged.
- **Section 3: Mathematical Foundations** - This section delves into the mathematical underpinnings of the James-Stein Estimator, providing a detailed derivation and a proof of its superiority over traditional estimators. It also links these concepts to Ridge Regression, highlighting their similarities and differences.
- **Section 4: Computational Methods** - Practical computational approaches for implementing these estimators are explored through simulations and comparative analyses, demonstrating their effectiveness.
- **Section 5: Conclusions** - Summarizes the findings and discusses the implications of these methods for statistical practice, particularly emphasizing their role in improving the accuracy and reliability of statistical estimations.

Essential Insights from the Chapter

2.1 Main Points

- The James-Stein Estimator introduces shrinkage to improve mean squared error estimates compared to traditional Maximum Likelihood Estimation (MLE).
- Unlike the MLE, the James-Stein Estimator adjusts each estimate using all observed data, shrinking estimates towards the mean to reduce estimation errors.
- An illustrative application using baseball data showcases the effectiveness of JSE, which consistently outperforms MLE in terms of mean squared error, particularly for moderate sample sizes.
- Ridge Regression introduces a penalty term to the least squares objective, shrinking coefficients and effectively addressing issues of multicollinearity in high-dimensional data.
- Ridge Regression and the James-Stein Estimator both embody shrinkage principles, differing in their contexts but sharing the objective of improving statistical estimation.

2.2 Key Insights and Implications

- **Shrinkage Methods**: The demonstrated superiority of JSE over MLE underlines the practical importance of shrinkage methods in modern statistical analysis. Ridge Regression, similarly, is crucial in high-dimensional data scenarios to mitigate multicollinearity.
- **Choosing Estimators**: The findings suggest practitioners need to be mindful of the bias-variance trade-off when choosing estimators. Introducing bias deliberately, as with JSE or Ridge Regression, often results in better overall performance, especially with high-dimensional data or multiple estimates.
- **Empirical Bayes Approach**: The application of JSE shows the value of empirical Bayes approaches in shrinking estimates towards a central value based on the distribution of observed data. This highlights the potential to combine frequentist and Bayesian perspectives in practical statistical problems.
- **High-Dimensional Data**: Ridge Regression exemplifies a crucial approach to manage the complexity of high-dimensional datasets, which are increasingly common in data science. By shrinking coefficient estimates, Ridge Regression can help identify significant predictors while reducing noise.

2.3 Questions for Further Exploration

- How does the choice of the shrinkage factor in the James–Stein estimator affect its performance in practical scenarios?
- How might the strengths of the James-Stein Estimator and Ridge Regression be combined or extended for more robust estimation in complex data situations?
- The section suggests that both methods improve estimation by introducing bias. Could this be a potential drawback in certain practical applications where unbiasedness is critical?

Mathematical Foundations

3.1 The James-Stein Estimator: Principles and Insights

The James-Stein estimator represents a significant advancement in statistical estimation, serving as a shrinkage estimator designed to improve the mean squared error of estimates. This powerful estimator is introduced through a compelling application to baseball batting averages.

3.1.1 What is the James-Stein Estimator?

Consider a scenario in statistical inference:

$$z_i|\mu_i \sim N(\mu_i, 1) \quad \text{for } i = 1, 2, \ldots, N \tag{1}$$

Here, z_i represents the observed value, and μ_i denotes the unknown parameter to be estimated. Traditionally, the maximum likelihood estimator (MLE) for μ_i is straightforward:

$$\hat{\mu}_i^{(\text{MLE})} = z_i \tag{2}$$

However, the James-Stein estimator introduces a more refined approach [1, 2]:

$$\hat{\mu}_i^{(\text{JSE})} = \left(1 - \frac{N-2}{\|z\|^2}\right) z_i \tag{3}$$

Unlike the MLE, which uses only the individual observation z_i for estimating μ_i, the James-Stein estimator utilizes all observed values to formulate each estimate. This collective use of data points typically results in estimates that are shrunk towards the overall mean, thereby potentially reducing the total estimation error under squared loss.

3.1.2 Illustrative Example Using Baseball Data

Table 1 showcases the batting averages of 18 Major League Baseball players from the 1970 season. The `MLE` column shows each player's average over the first 90 at bats, which serves as an early-season snapshot of their performance. Conversely, the `TRUTH` column reveals the players' averages over the remainder of the season, based on an additional 370 at bats on average, providing a more comprehensive assessment of their true skill.

This dataset offers a real-world application of the James-Stein theorem, suggesting that enhancements from this estimator can be significant under certain conditions. The `MLE` values, treated as binomial proportions, are refined using the James-Stein estimator to potentially reduce prediction error and offer a more accurate forecast of the `TRUTH` values.

Player	MLE	JS	TRUTH
1	.345	.283	.298
2	.333	.279	.346
3	.322	.276	.222
4	.311	.272	.276
5	.289	.265	.263
6	.289	.264	.273
7	.278	.261	.303
8	.255	.253	.270
9	.244	.249	.230
10	.233	.245	.264
11	.233	.245	.264
12	.222	.242	.210
13	.222	.241	.256
14	.222	.241	.269
15	.211	.238	.316
16	.211	.238	.226
17	.200	.234	.285
18	.145	.212	.200

Table 1: Comparison of MLE and James-Stein Estimator (JS) Predictions with True Values

The James-Stein theorem states that for $N \geq 3$ [1, 3]:

$$E(\|\mu^{(\text{JSE})} - \mu\|^2) < E(\|\mu^{(\text{MLE})} - \mu\|^2) \tag{4}$$

This implies that without any additional assumptions or conditions, the James-Stein estimator consistently outperforms the MLE in terms of the expected squared error, making it a superior choice in the presence of squared loss.

Why is the James-Stein estimator considered a pivotal empirical Bayesian method? The forthcoming sections on the derivation of the James-Stein estimator and the proof of the James-Stein theorem elucidate its mathematical foundation and justify its empirical effectiveness.

3.1.3 Derivation of the James-Stein Estimator

The derivation of the James-Stein estimator begins with the assumption of normality in our observations. Consider the model [3]:

$$z_i | \mu_i \sim N(\mu_i, 1) \tag{5}$$

Where μ_i is the true but unknown parameter, and z_i is the observed value.

This implies the likelihood for z_i given μ_i is known. Let ε_i, the error term, be normally distributed:

$$z_i = \mu_i + \varepsilon_i, \quad \text{where } \varepsilon_i \sim N(0,1). \tag{6}$$

Assuming a prior distribution $\mu_i \sim N(0, \sigma^2)$ for the parameters, the marginal distribution of z_i can be derived as follows:

$$z_i \sim N(0, 1+\sigma^2) \tag{7}$$

By applying the Bayesian theorem and leveraging the inherent properties of Gaussian distributions—namely marginal, conditional, and joint distributions—we significantly streamline our inference process. This approach is grounded in the principle that Gaussian characteristics inherently extend through all aspects of these distributions. As a result, we derive the posterior distribution of μ_i given z_i as follows:

$$\mu_i | z_i \sim N\left(\left(1 - \frac{1}{\sigma^2+1}\right) z_i, \frac{\sigma^2}{1+\sigma^2}\right) \tag{8}$$

The expected value of μ_i given z_i becomes the James-Stein estimator:

$$E(\mu_i|z_i) = \left(1 - \frac{1}{\sigma^2+1}\right) z_i = \hat{\mu}_i^{(\text{JSE})} \tag{9}$$

To fully implement this estimator, we need an estimate of σ^2. Estimating σ^2 involves using the observed data:

$$z_i \sim N(0, 1+\sigma^2) \tag{10}$$

Upon standardizing, we get a χ^2 distribution with N degrees of freedom for the sum of squared standardized observations:

$$Q = \sum_{i=1}^{N} \frac{z_i^2}{1+\sigma^2} \sim \chi_N^2 \tag{11}$$

Given that Q follows a χ^2 distribution, $1/Q$ follows an inverse χ^2 distribution, which leads to:

$$E\left(\frac{1}{Q}\right) = \frac{1}{N-2} \Rightarrow E\left(\frac{1}{\sum_{i=1}^{N} \frac{z_i^2}{1+\sigma^2}}\right) = \frac{1}{N-2} \tag{12}$$

Finally, we estimate $\frac{1}{1+\sigma^2}$ using:

$$\frac{N-2}{\sum_{i=1}^{N} z_i^2} \tag{13}$$

Substituting this into the estimator formula gives us the James-Stein estimator:

$$\hat{\mu}_i^{(JSE)} = \left(1 - \frac{N-2}{\sum_{i=1}^{N} z_i^2}\right) z_i \tag{14}$$

3.1.4 Proof of the James-Stein Theorem

The James-Stein theorem demonstrates the superiority of the James-Stein estimator over the MLE under certain conditions. This proof details the conditions under which the expected squared error of the James-Stein estimator is less than that of the MLE when estimating multiple parameters.

Starting Point We begin with the identity for the squared errors [3]:

$$\sum_{i=1}^{N} (\hat{\mu}_i - \mu_i)^2 = \sum_{i=1}^{N} [(z_i - \hat{\mu}_i)^2 - (z_i - \mu_i)^2 + 2(\hat{\mu}_i - \mu_i)(z_i - \mu_i)] \tag{15}$$

Taking expectations on both sides, and recognizing that $\hat{\mu}$ is a function of z, we have:

$$E(\|\hat{\mu} - \mu\|^2) = E(\|z - \hat{\mu}\|^2) - N + 2\sum_{i=1}^{N} \text{Cov}(\hat{\mu}_i, z_i) \tag{16}$$

where $\text{Cov}(\hat{\mu}_i, z_i) = E[(\hat{\mu}_i - \mu_i)(z_i - \mu_i)]$.

Utilizing Stein's Identity Stein's identity provides a pivotal simplification, stating:

$$\text{Cov}(\hat{\mu}_i, z_i) = E\left[\frac{\partial \hat{\mu}_i}{\partial z_i}\right] \tag{17}$$

which can be derived from integration by parts. Substituting this into our equation, we obtain:

$$E(\|\hat{\mu} - \mu\|^2) = E(\|z - \hat{\mu}\|^2) - N + 2\sum_{i=1}^{N} E\left[\frac{\partial \hat{\mu}_i}{\partial z_i}\right] \tag{18}$$

Substituting Estimators Inserting $\hat{\mu}^{(MLE)} = z$ and $\hat{\mu}^{(JSE)} = \left(1 - \frac{N}{\|z\|^2}\right) z$, we compare:

$$E(\|\mu^{(MLE)} - \mu\|^2) = N \tag{19}$$

$$E(\|\mu^{(JSE)} - \mu\|^2) = N - E\left[\frac{(N-2)^2}{\|z\|^2}\right] \tag{20}$$

Concluding the Proof From the above, it is evident that when $N > 2$,

$$E(\|\mu_i^{(JSE)} - \mu_i\|^2) < E(\|\mu_i^{(MLE)} - \mu_i\|^2) \tag{21}$$

thus proving the theorem. This result highlights the effectiveness of the James-Stein Estimator, particularly when estimating multiple parameters, as it consistently yields lower expected squared errors compared to the MLE.

3.2 Linking Ridge Regression and James-Stein Estimation

Ridge Regression and the James-Stein Estimator share a fundamental approach to addressing the shortcomings of traditional least squares in the presence of multicollinearity or when the number of predictors exceeds the number of observations. Both methods adjust the estimates by introducing a form of shrinkage, improving the reliability of predictions especially when dealing with high-dimensional data.

3.2.1 Basic Linear Model and Shrinkage

In the context of linear regression, we often start with the model:

$$y = X\beta + \epsilon \tag{22}$$

where y is the vector of responses, X is the matrix of predictors, β is the vector of unknown coefficients, and ϵ is the noise vector, typically assumed to be normally distributed, $\epsilon \sim N(0, \sigma^2 I)$.

3.2.2 Ordinary Least Squares and Its Limitations

The ordinary least squares (OLS) estimator is given by:

$$\hat{\beta}^{OLS} = (X^T X)^{-1} X^T y \tag{23}$$

However, when $X^T X$ is nearly singular or not invertible (which happens in cases of multicollinearity or when $p > n$), the OLS estimator becomes highly sensitive to random errors in the observed data, resulting in large variances for the coefficient estimates.

3.2.3 Introduction of Ridge Regression

Ridge Regression modifies the OLS approach by adding a penalty term to the loss function [4, 5, 6]:

$$\hat{\beta}^{ridge} = \arg\min_{\beta} \left\{ \|y - X\beta\|^2 + \lambda\|\beta\|^2 \right\} \tag{24}$$

This penalty term $\lambda\|\beta\|^2$ discourages the flexibility of the model by penalizing the magnitude of the coefficients, effectively shrinking them towards zero. This is particularly useful in reducing the variance of the estimates.

3.2.4 James-Stein Estimator and Its Connection to Ridge

Similarly to how Ridge introduces shrinkage, the James-Stein estimator also applies a shrinkage factor to improve the estimation of means from a multivariate normal distribution. For Ridge Regression, the shrinkage factor is controlled by λ, while for the James-Stein estimator, the shrinkage depends on the number of dimensions and the variability of the estimates. Specifically, for estimates $\hat{\beta}$ under a normal model, the James-Stein-type shrinkage can be expressed as [2]:

$$\hat{\beta}^{JS} = \left[1 - \frac{(p-2)\sigma^2}{\hat{\beta}^T S \hat{\beta}}\right] \hat{\beta} \tag{25}$$

where S is related to the covariance structure of the predictors (akin to $X^T X$ in Ridge Regression).

Computational Methods

This section focuses on practical computational approaches to implementing the James-Stein Estimator and Ridge Regression. Through simulations and comparisons, we demonstrate the effectiveness of these methods in statistical estimation.

4.1 Implementing James-Stein Estimator

The James-Stein Estimator is implemented to demonstrate its shrinkage effect on simulated data. This estimator is particularly notable for its ability to reduce estimation errors by shrinking estimates towards their overall mean.

4.1.1 Algorithm for James-Stein Estimator

Algorithm 1 James-Stein Estimator

procedure JAMESSTEINESTIMATOR(z, $sigma\_squared$)
    $n \leftarrow \text{length}(z)$
    $z\_bar \leftarrow \text{mean}(z)$
    $s\_squared \leftarrow \text{sum}((z - z\_bar)^2)/n$
    $shrinkage\_factor \leftarrow 1 - ((n-3) * sigma\_squared/s\_squared)$
    $shrinkage\_factor \leftarrow \max(0, shrinkage\_factor)$
    for $i \leftarrow 1$ to n **do**
        $z[i] \leftarrow z\_bar + shrinkage\_factor * (z[i] - z\_bar)$
    end for
    return z
end procedure

4.1.2 Simulation Results

Data are generated from a normal distribution with a mean of 5 and a standard deviation of 1. The James-Stein Estimator is then applied to these data, demonstrating significant shrinkage of estimates towards the mean. The impact of this shrinkage is clearly illustrated in a scatter plot that compares the original data with the James-Stein estimation.

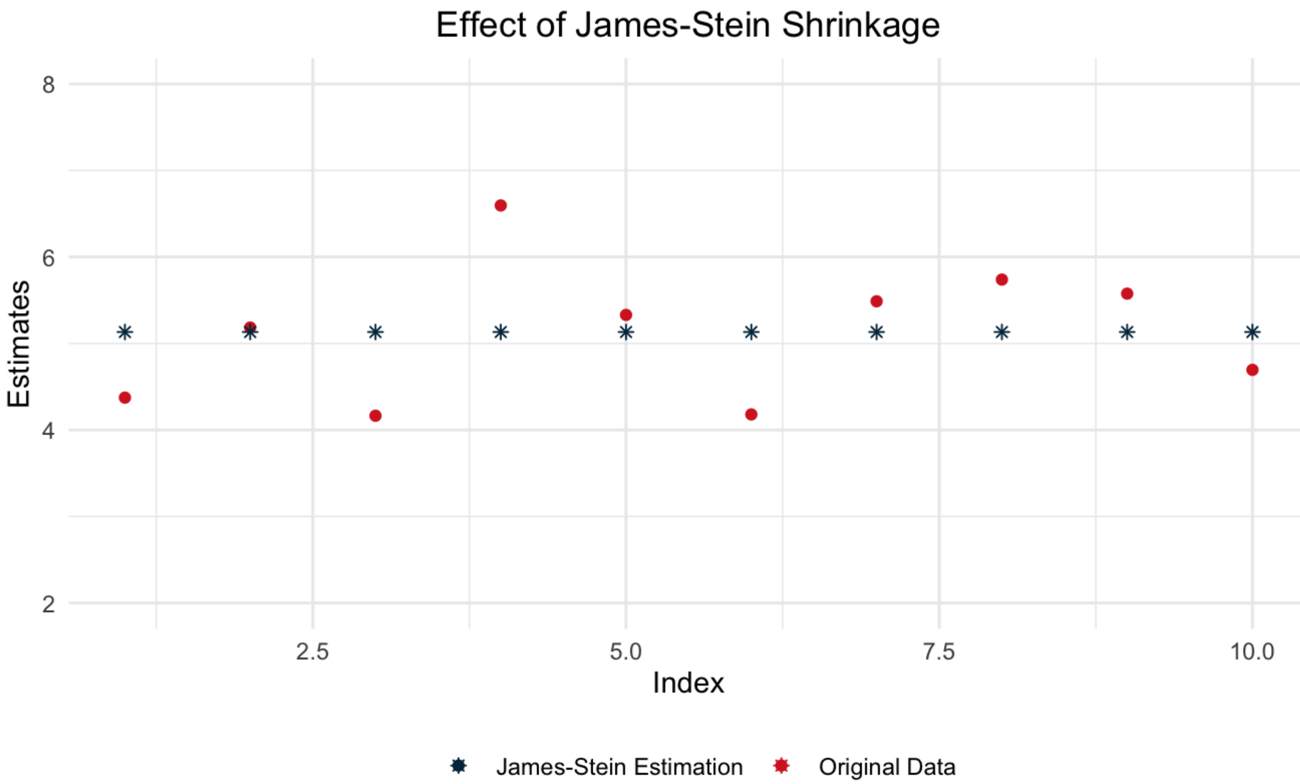

Figure 1: Effect of James-Stein shrinkage on simulated data.

4.2 Simulation of Estimator Effectiveness

We compare the James-Stein Estimator with the Maximum Likelihood Estimator across different sample sizes to illustrate the reduction in squared errors.

4.2.1 Algorithm for Comparing Estimator Effectiveness

Algorithm 2 Compare Estimators

procedure COMPAREESTIMATORS(num_sim, $sample\_sizes$, mu, $sigma$)
  for n in $sample\_sizes$ **do**
    $mle\_errors \leftarrow$ vector of size num_sim
    $jse\_errors \leftarrow$ vector of size num_sim
    for $i \leftarrow 1$ to num_sim **do**
      $z \leftarrow$ generate normal data$(n, mu, sigma)$
      $mle \leftarrow$ mean(z)
      $s\_squared \leftarrow$ sum$((z - mle)^2)/n$
      $shrinkage\_factor \leftarrow 1 - ((n - 3) * sigma^2/s\_squared)$
      $shrinkage\_factor \leftarrow$ max$(0, shrinkage_factor)$
      $jse \leftarrow mle + shrinkage\_factor * (z - mle)$
      $mle\_errors[i] \leftarrow$ sum$((mle - mu)^2)$
      $jse\_errors[i] \leftarrow$ sum$((jse - mu)^2)$
    end for
    Store mle_errors and jse_errors in data frame
  end for
  return data frame
end procedure

4.2.2 Visualizing Comparison Results

The log-squared errors of both estimators are plotted to show the James-Stein Estimator's efficiency, especially in smaller samples.

Generally, the James-Stein Estimator demonstrates lower median log squared errors compared to the MLE across all sample sizes, particularly as the sample size increases. This demonstrates the effectiveness of the James-Stein shrinkage, especially in reducing variance and improving estimation accuracy over the traditional MLE approach.

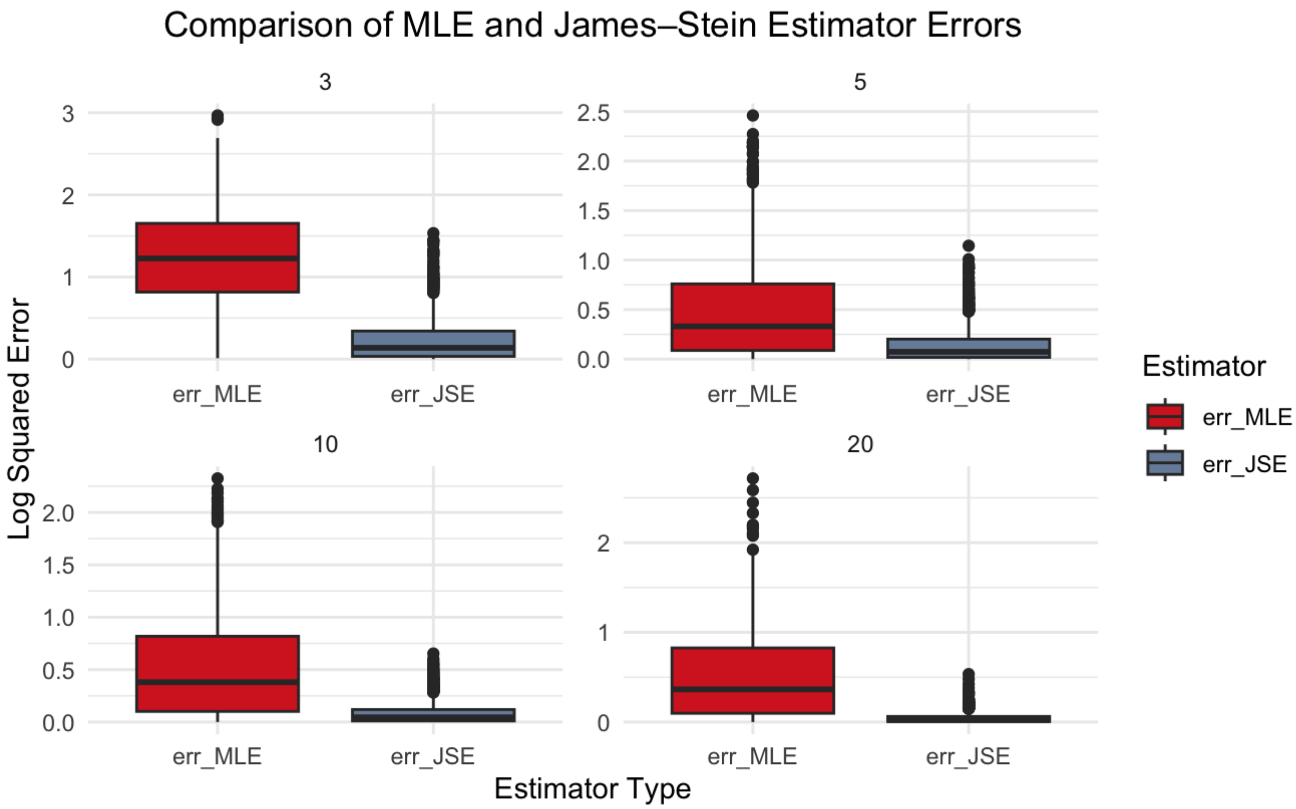

Figure 2: Log-squared error comparison of MLE and JSE across different sample sizes.

4.3 Exploring Ridge Regression

Ridge Regression is investigated as an auxiliary method that also demonstrates shrinkage, similar to the James-Stein approach, but within the framework of regression analysis. Data and coefficients are generated and Ridge Regression is implemented using the glmnet package in R.

The plot of coefficient shrinkage, as the regularization parameter increases, showcases Ridge Regression's capacity to effectively manage overfitting and multicollinearity, highlighting its utility in high-dimensional statistical modeling.

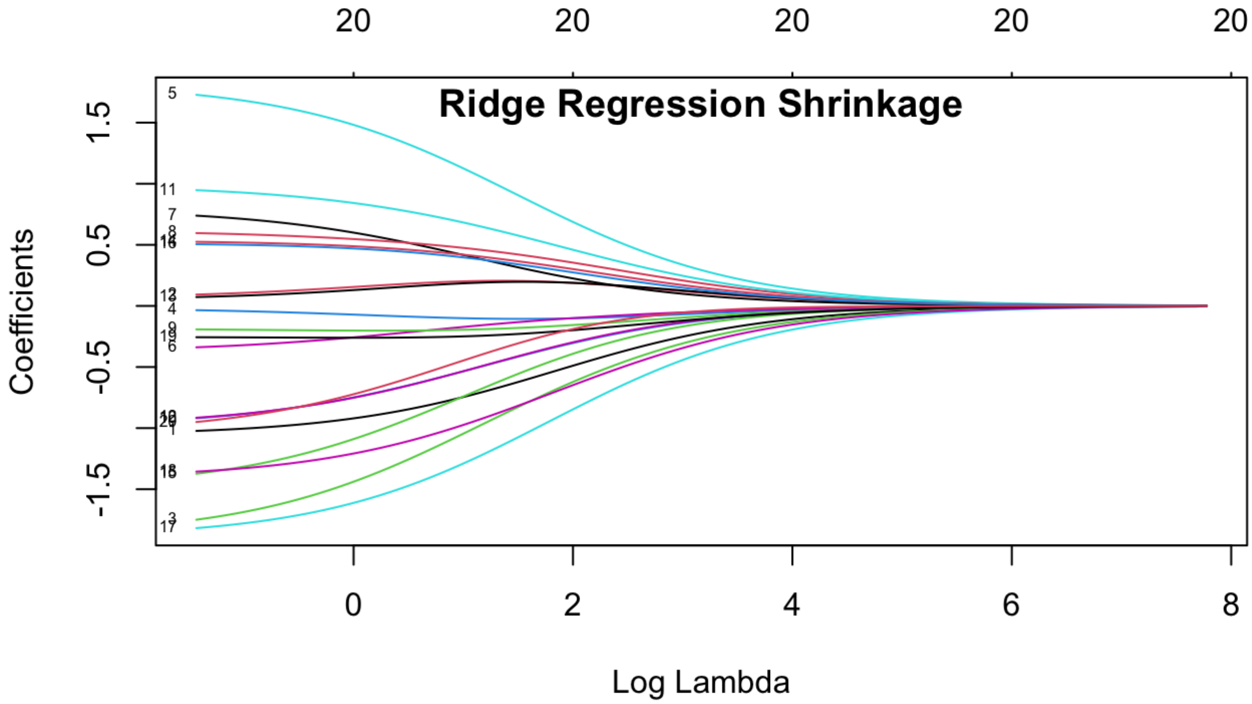

Figure 3: Coefficient shrinkage in Ridge Regression as a function of the regularization parameter.

4.4 Conclusion

This section has showcased the practical utility and advantages of the James-Stein Estimator and Ridge Regression through detailed computational demonstrations. Both methods are united by their use of shrinkage to enhance the accuracy and dependability of statistical estimates.

Conclusions

This study has explored the significant impact of the James-Stein Estimator and Ridge Regression on statistical practice, emphasizing their conceptual and practical benefits in improving estimation accuracy and reliability. These methods exemplify the power of shrinkage techniques in addressing challenges in statistical estimation, particularly in scenarios of high dimensionality or small sample sizes.

5.1 Implications for Statistical Practice

- **James-Stein Estimator**: With its robust performance improvement over traditional Maximum Likelihood Estimation, especially in the context of small sample sizes, James-Stein Estimator demonstrates a critical advancement in statistical estimation theory. It effectively reduces estimation error by leveraging the strength of shrinkage towards the mean, which is particularly beneficial in empirical Bayesian contexts. This approach not only enhances the reliability of estimates but also provides a substantial reduction in the risk of overfitting, making it an invaluable tool in statistical analysis.
- **Ridge Regression**: It complements the James-Stein Estimator by applying similar principles of shrinkage within a regression framework. This technique addresses the limitations of Ordinary Least Squares estimation in the presence of multicollinearity or when the number of predictors exceeds the number of observations. By introducing a penalty term that constrains the size of the coefficients, Ridge Regression ensures more stable and interpretable models, which is crucial for complex data analyses involving numerous predictors.

5.2 Synthesis of Key Findings

The integration of computational methods to illustrate the efficacy of these estimators underscores their practical significance in real-world applications. The James-Stein Estimator's ability to shrink estimates toward the overall mean offers a pragmatic solution for enhancing the accuracy of predictions derived from small datasets or datasets with high variance in observations. Similarly, Ridge Regression's capability to manage the complexity of models by penalizing the magnitude of coefficients helps in mitigating issues arising from overfitting and multicollinearity, promoting more reliable predictive modeling.

References

[1] William James and Charles Stein. “Estimation with quadratic loss”. In: *Breakthroughs in statistics: Foundations and basic theory*. Springer, 1992, pp. 443–460.

[2] Bradley Efron and Trevor Hastie. *Computer age statistical inference, student edition: algorithms, evidence, and data science*. Vol. 6. Cambridge University Press, 2021.

[3] Bradley Efron. *Large-scale inference: empirical Bayes methods for estimation, testing, and prediction*. Vol. 1. Cambridge University Press, 2012.

[4] Trevor Hastie. “Ridge regularization: An essential concept in data science”. In: *Technometrics* 62.4 (2020), pp. 426–433.

[5] Gareth James et al. *An introduction to statistical learning*. Vol. 112. Springer, 2013.

[6] Trevor Hastie et al. *The elements of statistical learning: data mining, inference, and prediction*. Vol. 2. Springer, 2009.

Appendix

A.1 R Code for Implementation of the James-Stein Estimator

```
# Function to apply James-Stein Estimator
james_stein_estimator <- function(z, sigma_squared) {
  n <- length(z)
  z_bar <- mean(z)
  s_squared <- sum((z - z_bar)^2) / n
  shrinkage_factor <- 1 - ((n-3) * sigma_squared / s_squared)
  shrinkage_factor <- max(0, shrinkage_factor)  # Ensure non-negative shrinkage
  return(z_bar + shrinkage_factor * (z - z_bar))
}

# Simulate data
set.seed(1)
z <- rnorm(10, mean = 5, sd = 1)
sigma_squared <- 1

# Apply the JSE
jse_results <- james_stein_estimator(z, sigma_squared)
print(jse_results)

data <- data.frame(Index = 1:length(z),
                   Original = z,
                   JamesStein = jse_results)

ggplot(data) +
  geom_point(aes(x = Index, y = Original, color = "Original Data"), pch = 19) +
  geom_point(aes(x = Index, y = JamesStein, color = "James-Stein Estimation"), pch = 8) +
  scale_color_manual(values = c("Original Data" = "#d62828",
                                "James-Stein Estimation" = "#003049")) +
  labs(x = "Index", y = "Estimates", title = "Effect of James-Stein Shrinkage") +
  theme_minimal() +
  theme(
    plot.title = element_text(hjust = 0.5),
    legend.title = element_blank(),
    legend.position = 'bottom'
    ) +
  ylim(2, 8)
```

A.2 R Code for Simulation of Estimator Effectiveness

```
# Perform simulations and calculate MLE and JSE
simulate_estimators <- function(n, mu = 5, sigma = 1, num_sim = 1000) {
  mle_errors <- numeric(num_sim)
  jse_errors <- numeric(num_sim)

  for (i in 1:num_sim) {
    z <- rnorm(n, mean = mu, sd = sigma)
    mle <- mean(z)
    s_squared <- sum((z - mle)^2) / n
    shrinkage_factor <- 1 - ((n-3) * sigma^2 / s_squared)
    shrinkage_factor <- max(0, shrinkage_factor)
    jse <- mle + shrinkage_factor * (z - mle)

    mle_errors[i] <- sum((mle - mu)^2)
    jse_errors[i] <- sum((jse - mu)^2)
  }

  data.frame(
    Error = c(jse_errors, mle_errors),
    Estimator = rep(c("err_MLE", "err_JSE"), each = num_sim),
    N = rep(n, 2 * num_sim)
  )
}

# Parameters
sample_sizes <- c(3, 5, 10, 20)
num_simulations <- 1000

# Run simulations for multiple sample sizes
results <- do.call(rbind, lapply(sample_sizes, function(n) {
  simulate_estimators(n = n, num_sim = num_simulations)
}))
results$Estimator <- factor(results$Estimator, levels = c("err_MLE", "err_JSE"))

# Plot
ggplot(results, aes(x = Estimator, y = log(Error + 1), fill = Estimator)) +
  geom_boxplot() +
  facet_wrap(~ N, scales = "free") +
  labs(title = "Comparison of MLE and James{Stein Estimator Errors",
       x = "Estimator Type",
       y = "Log Squared Error") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5)) +
```

```
  scale_fill_manual(values = c("err_MLE" = "#d62828", "err_JSE" = "#778da9"))
```

A.3 R Code for Comparing with Ridge Regression

```
# Simulate data
set.seed(1)
x <- matrix(rnorm(100*20), 100, 20)  # 100 observations, 20 predictors
beta <- runif(20, -2, 2)  # True coefficients
y <- x %*% beta + rnorm(100)  # Response variable

# Ridge regression using glmnet
fit <- glmnet(x, y, alpha = 0)

# Plot coefficient shrinkage
plot(fit, xvar = "lambda", label = TRUE)
title(main = "Ridge Regression Shrinkage", line = -1)
```