Math 3: Introduction to Probability and Statistics

Bayesian Shrinkage, Admissibility, and Empirical Bayesian Methods
Bayesian and Empirical Bayesian Prediction

I. Bayesian Analysis of Means in the Normal Model

II. Admissibility in Predicting Expectations
   A. Mean Square Error and Estimator Admissibility
   B. Stein’s Shrinkage Estimator for the Sample Mean

III. Learning Expectations from Other Expectations

IV. Bayesian Shrinkage Estimators for Covariance Matrices
Bayesian Expected Expectations

- Key: Treat Expectations for a random variable as, themselves, being random variables
- Posterior Expectations for a random variable are a weighted average of prior expectations and expectations from data

\[ E_{post}\{\mu\} = \alpha E_{prior}\{\mu\} + (1 - \alpha) E_{data}\{\mu\} \]

- Posterior beliefs “shrink” sample estimates toward prior
- Analytical Challenge: What is the right value for $\alpha$?
  - Depends on degree of belief in prior
  - Depends on amount of information in data
Example: Updating Historical Averages

• We first observe data from 1950-1989 (Sample I), during which the average return on an asset was 10%.

• We then observe additional data from 1990-1999 (Sample II) with an average return on the asset of 20%.

• What is the updated historical average?
  – Total Sample Period: $40 + 10 = 50$ years
  – Total Sample Return: $40 \times 10\% + 10 \times 20\% = 600\%$
  – Total Sample Average: $600\%/50$ years $= 12\%/year$

• “Bayesian” Updated Average
  – Additional data is $\frac{1}{4}$ as informative as prior sample
  – Posterior average is $0.8 \times 10\% + 0.2 \times 20\% = 12\%/year$
Posterior Expectations: The Normal Model

- Let the data point $y$ have a normal distribution with unknown mean $\mu$ and known standard deviation $\sigma_d$
- Suppose prior beliefs treat $\mu$ as a Normally distributed random variable with mean $\mu$ and std deviation $\sigma$
- Now suppose you observe a realization from $y$ of $\mu_d$

- Fact: The **Posterior Expectation** of $\mu$ is

  \[
  E_{\text{post}}[\mu] = \frac{1}{\frac{1}{\sigma^2} + \frac{1}{\sigma_d^2}} \left( \frac{\mu}{\sigma^2} + \frac{\mu_d}{\sigma_d^2} \right) = \frac{\sigma_d^2 \mu + \sigma^2 \mu_d}{\sigma_d^2 + \sigma^2}
  \]

- Key Intuition: This is a weighted average of the prior and observed means, with weights proportional to the **precision**, or the inverse of the variance of the signals
Special Features of Conjugate Priors

• The Normal model with known variance is a special case where the prior is **conjugate** to the likelihood

• Conjugate priors have the same distributional structure as the observed data generating process itself
  – Can think of the prior as representing the estimation results from a prior sample
  – Prior equivalent to “fake data” written up by the experimenter

• Important: The sample of “fake data” in your prior must be completely independent of the observed sample of real data
Posterior Mean for Two Samples

- Suppose returns are normally distributed with variance $\sigma$
- Sample $I$ corresponds to prior beliefs with mean 10 and std deviation

$$\sigma^2_I = \frac{\sigma^2}{40}$$

- Sample $II$ corresponds to a sample of data with mean 20 and std deviation

$$\sigma^2_{II} = \frac{\sigma^2}{10}$$

- Then the Posterior Expectation is given by

$$E_{post} [\mu] = \frac{\sigma^2}{10} \frac{10}{10} + \frac{\sigma^2}{40} \frac{20}{40} = \frac{1 + 0.5}{0.1 + 0.025} = \frac{1.5}{0.125} = 12$$
Bayesian and Empirical Bayesian Prediction

I. Bayesian Analysis of Means in the Normal Model

II. Admissibility in Predicting Expectations
   A. Mean Square Error and Estimator Admissibility
   B. Stein Estimator for the Sample Mean

III. Learning Expectations from Other Expectations

IV. Bayesian Shrinkage Estimators for Covariance Matrices
Estimation Errors: Bias and Variance

• Suppose you are trying to estimate a scalar parameter \( \mu \), whose true value is \( \mu^* \) with estimator \( \hat{\mu} \)

\[
\hat{\mu} : \{\text{Data}\} \rightarrow \mathbb{R}
\]

• The mean square error of the estimator indicates its expected error rate before seeing the data

\[
E \left[ (\hat{\mu} - \mu^*)^2 \right] = E \left[ (\hat{\mu} - E[\hat{\mu}] + E[\hat{\mu}] - \mu^*)^2 \right]
\]

\[
= E \left[ (\hat{\mu} - E[\hat{\mu}])^2 + (E[\hat{\mu}] - \mu^*)^2 + 2(\hat{\mu} - E[\hat{\mu}]) (E[\hat{\mu}] - \mu^*) \right]
\]

\[
= E \left[ (\hat{\mu} - E[\hat{\mu}])^2 \right] + (E[\hat{\mu}] - \mu^*)^2 + 2E \left[ (\hat{\mu} - E[\hat{\mu}]) \right] (E[\hat{\mu}] - \mu^*)
\]

\[
= \text{Var}(\hat{\mu}) + \text{Bias}(\hat{\mu})^2 + 2 \times 0 \times (E[\hat{\mu}] - \mu^*)
\]
Estimation Errors: Bias and Variance

• Suppose you are trying to estimate a scalar parameter $\mu$, whose true value is $\mu_0$ with estimator $\hat{\mu}$

$$\hat{\mu} : \{\text{Data}\} \rightarrow \mathbb{R}$$

• The mean square error of the estimator indicates its expected error rate before seeing the data

$$E\left[ (\hat{\mu} - \mu^*)^2 \right] = \text{Var}(\hat{\mu}) + \text{Bias}(\hat{\mu})^2$$

• This decomposition suggests two aspects of an estimator we want to control

  – Bias: $E[\hat{\mu} - \mu^*]$

  – Variance: $\text{Var}[\hat{\mu}]$
Example: MSE for Two Estimators of a Mean

• Suppose you don’t know the mean for a random variable

\[ y \sim N(\mu^*, 1) \]

• Consider the sample average estimator for a mean

\[ \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} y_t \]

• This estimator is unbiased, so its MSE is just its variance

\[ MSE[\hat{\mu}] = Var[\hat{\mu}] = \frac{1}{T} \]

• Consider a naïve estimator for a mean

\[ \tilde{\mu} = 42 \]

• This estimator has zero variance, so its MSE is just its bias

\[ MSE[\tilde{\mu}] = Bias[\tilde{\mu}]^2 = (\mu^* - 42)^2 \]
Best Linear Unbiased Estimators

• Continuing: Consider a sample of independent normal random variables with unit variance and common mean

\[ y_t \sim iid \ N(\mu^*, 1), \ t = 1, \ldots, T \]

• Define the class of Linear Unbiased Estimators as the set of all weighted averages for the observed \( y \)'s

\[ \hat{M} = \left\{ \hat{\mu} : \frac{1}{\sum_{t=1}^{T} \omega_t} \sum_{t=1}^{T} \omega_t y_t \to \mathbb{R} \mid E[\hat{\mu}] = \mu^* \right\} \]

• The sample average is the **Best LUE**, meaning it has the minimum variance of any unbiased linear estimator

\[ \bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t = \text{arg min}_{\hat{\mu} \in \hat{M}} \text{Var}(\hat{\mu}) \]
Admissibility

• An estimator is **admissible** in a class of estimators if it minimizes the mean square error of the estimate

• Example: A BLUE is admissible in the class of Linear Unbiased Estimators

\[
\hat{M} = \left\{ \hat{\mu} : \frac{1}{\sum_{t=1}^{n} \omega_t} \sum_{t=1}^{n} \omega_t y_t \rightarrow \mathbb{R} \mid E[\hat{\mu}] = \mu^* \right\}
\]

- Note that \( \hat{\mu} \in \hat{M} \Rightarrow Bias(\hat{\mu}) = 0 \Rightarrow MSE(\hat{\mu}) = Var(\hat{\mu}) \)

- Then, since

\[
\bar{y} = \frac{1}{n} \sum_{t=1}^{n} y_t = \arg \min_{\hat{\mu} \in \hat{M}} Var(\hat{\mu}) = \arg \min_{\hat{\mu} \in \hat{M}} MSE(\hat{\mu})
\]

\( \bar{y} \) is admissible in \( \hat{M} \)
Bayesian and Empirical Bayesian Prediction

I. Bayesian Analysis of Means in the Normal Model

II. Admissibility in Predicting Expectations
   A. Mean Square Error and Estimator Admissibility
   B. Stein’s Shrinkage Estimator for the Sample Mean

III. Learning Expectations from Other Expectations

IV. Bayesian Shrinkage Estimators for Covariance Matrices
Bayesian Shrinkage and Admissibility

- Motivation for Shrinkage: When making predictions, we are often willing to accept some bias to get lower MSE.

- Clearly, adding estimators to our set of candidates can only reduce the MSE of the admissible estimator.
  - Bayesian estimators, while biased, are typically consistent and have relatively low sampling variances.
  - Prior information biases an estimator unless you already know the true value for the parameter.

- Empirical Bayesian Goal: Balance Bias and Variance to minimize MSE, yielding admissible estimators within the class of Bayesian estimators.
Example: MSE for a Shrinkage Estimator

• Continuing to look at the mean for a random variable

\[ y \sim N(\mu^*, 1) \]

• Recall the sample average estimator for the mean

\[ \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} y_t \]
\[ MSE[\hat{\mu}] = Var[\hat{\mu}] = \frac{1}{T} \]

• And our naïve estimator for the mean

\[ \tilde{\mu} = 42 \]
\[ MSE[\tilde{\mu}] = Bias[\tilde{\mu}]^2 = (\mu^* - 42)^2 \]

• A “shrinkage” estimator combines these two

\[ \bar{\mu}(\alpha) = \alpha \hat{\mu} + (1 - \alpha) \tilde{\mu} \]

• The MSE for this shrinkage estimator is:

\[ MSE[\bar{\mu}] = Bias[\bar{\mu}]^2 + Var[\bar{\mu}] = (1 - \alpha)^2 (\mu^* - 42)^2 + \alpha^2 \frac{1}{T} \]
Minimizing MSE for Optimal Shrinkage

• Continuing to look at the mean for a random variable

\[ y \sim N(\mu^*, 1) \]

• Now consider a shrinkage estimator for the mean:

\[ \bar{\mu}(\alpha) = \alpha \hat{\mu} + (1 - \alpha) \bar{\mu} \]

• The MSE for the shrinkage estimator is given by:

\[ MSE[\bar{\mu}(\alpha)] = (1 - \alpha)^2 (\mu^* - 42)^2 + \alpha^2 \frac{1}{T} \]

• We can now choose \( \alpha \) to minimize the MSE:

\[ \frac{\delta}{\delta \alpha} MSE[\bar{\mu}(\alpha)] = -2(1 - \alpha^*)(\mu^* - 42)^2 + 2\alpha^* \frac{1}{T} = 0 \]

\[ \Rightarrow \alpha^* = \frac{(\mu^* - 42)^2}{(\mu^* - 42)^2 + 1/T} = \frac{MSE[\tilde{\mu}]}{MSE[\tilde{\mu}] + MSE[\hat{\mu}]} \]
Stein’s Optimal Shrinkage Intensity

• Continuing to look at the mean for a random variable

\[ y \sim N(\mu^*, 1) \]

• Now consider a shrinkage estimator for the mean:

\[ \bar{\mu}(\alpha) = \alpha \hat{\mu} + (1 - \alpha) \tilde{\mu} \]

• The MSE-Optimal shrinkage intensity is:

\[ \alpha^* = \frac{T(\mu^* - 42)^2}{T(\mu^* - 42)^2 + 1} = \frac{MSE[\tilde{\mu}]}{MSE[\tilde{\mu}] + MSE[\hat{\mu}]} \]

  – Increases as \( T \) grows (and sample mean variance falls)
  – Increases as \( \mu \) moves away from 42 (bias gets worse)

• Key Intuition: Optimally combining signals to minimize MSE weights each signal by the inverse of their MSE
Stein’s Optimal Shrinkage Intensity

• Continuing to look at the mean for a random variable

\[ y \sim N(\mu^*, 1) \]

• Now consider a shrinkage estimator for the mean:

\[ \bar{\mu}(\alpha) = \alpha \hat{\mu} + (1 - \alpha) \tilde{\mu} \]

• The MSE-Optimal shrinkage intensity is:

\[ \alpha^* = \frac{T(\mu^* - 42)^2}{T(\mu^* - 42)^2 + 1} = \frac{\text{MSE}[\tilde{\mu}]}{\text{MSE}[\hat{\mu}] + \text{MSE}[\tilde{\mu}]} \]

• **Result:** The Infeasible Stein Estimator:

\[ \bar{\mu}_S = \bar{\mu}(\alpha^*) = \alpha^* \hat{\mu} + (1 - \alpha^*) \tilde{\mu} \]

is admissible in the class of estimators \( \bar{\mathcal{M}} = \{\bar{\mu}(\alpha)\} \)
Feasible Stein-Optimal Shrinkage

• Result: The Infeasible Stein Shrinkage Estimator:

\[ \bar{\mu}_S = \bar{\mu}(\alpha^*) = \alpha^* \hat{\mu} + (1 - \alpha^*) \tilde{\mu} \]

is admissible in the class of estimators \( \bar{\mu}(\alpha) \)

• Infeasibility: The MSE-Optimal shrinkage intensity requires knowing \( \mu^* \) to calculate \( \alpha^* \)

• Solution: Estimate

\[ \hat{\alpha}^* = \frac{T(\hat{\mu} - 42)^2}{T(\hat{\mu} - 42)^2 + 1} \]

with an unbiased and consistent estimator, \( \hat{\mu} \rightarrow_p \mu^* \)

• Then the Feasible Stein Shrinkage Estimator is:

\[ \hat{\mu}_S = \bar{\mu}(\hat{\alpha}^*) \]
“Bayesian” Interpretation of Stein Shrinkage

• Back to the normal model:  \( y \sim N(\mu^*, 1) \)
• Then we know the sample mean is also normal:
  \[
  \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} y_t \sim N\left( \mu, \frac{1}{T} \right)
  \]
• Suppose prior beliefs treat \( \mu \) as a Normally distributed random variable with
  \[
  \mu = 42 \quad \sigma^2 = T^2 \left( \hat{\mu} - 42 \right)^2
  \]
• Then the posterior expectation is:  \( E_{\text{post}}[\mu] = \bar{\mu}(\hat{\alpha}^*) = \hat{\mu}_S \)
• Empirical Bayesian Estimators aren’t truly Bayesian
  – Prior beliefs for \( \sigma^2 = T^2 \left( \hat{\mu} - 42 \right)^2 \) depend on data
  – Prior is not independent of data and Bayes’ rule is no longer being used correctly
Empirical Bayesian Estimation Strategy

• Given prior beliefs, Bayesian updating provides a mechanism for combining these beliefs with data to construct posterior expectations or forecasts.

• An Empirical Bayesian estimator calibrates these prior beliefs using data to obtain feasibly admissible estimators, but this calibration exercise confounds applying Bayes’ rule.

• Feasible Empirical Bayesian estimators are not admissible in finite samples, but can be “asymptotically admissible”:
  – Calibrated prior doesn’t reach the true optimum.
  – Intuitive condition: Calibrating parameters need to converge to true values.
Bayesian and Empirical Bayesian Prediction

I. Bayesian Analysis of Means in the Normal Model

II. Admissibility in Predicting Expectations
   A. Mean Square Error and Estimator Admissibility
   B. Stein’s Shrinkage Estimator for the Sample Mean

III. Learning Expectations from Other Expectations

IV. Bayesian Shrinkage Estimators for Covariance Matrices
Bayesian Updating for Multivariate Normal Means

• Instead of estimating one mean, consider analyzing $N$ means for $N$ independent normal random variables with unit variance

$$y_{(n \times 1)} \sim MVN \left( \mu^*, I_n \right)$$

• The sample average then has a normal distribution

$$\bar{y} = \frac{1}{T} \sum_{t=1}^{T} y_t \sim MVN \left( \mu^*, T^{-1}I_n \right)$$

• Consider shrinking this sample average towards a common mean:

$$\bar{\mu}(\alpha, \underline{\mu}) = (1 - \alpha) \underline{\mu}_{(N \times 1)} + \alpha \bar{y}$$
Total MSE for Estimating Many Means

• The shrinkage estimator for $\mu$ is:
  \[ \bar{\mu}_n (\mu, \alpha) = (1 - \alpha) \mu + \alpha \bar{y}_n \]

• The sample average is not biased, but the posterior expectation inherits some bias from the prior

• For the $n^{th}$ expectation
  \[ \text{Bias} \left[ \bar{\mu}_n (\mu, \alpha) \right] = (1 - \alpha) (\mu - \mu_n) = (1 - \alpha) \text{Bias}_n (\mu) \]

• The variance of the estimator the $n^{th}$ expectation is
  \[ \text{Var} \left[ \bar{\mu}_n (\mu, \alpha) \right] = (1 - \alpha)^2 \text{Var} (\mu) + \alpha^2 \text{Var} (\bar{y}_n) \]
  \[ = (1 - \alpha)^2 \text{Var} (\mu) + \alpha^2 \frac{1}{T} \]
Total MSE for Estimating Many Means

• The shrinkage estimate for the $n^{th}$ mean $\mu_n$ is:
  \[ \bar{\mu}_n (\bar{\mu}, \alpha) = (1 - \alpha) \bar{\mu} + \alpha \bar{y}_n \]

• This estimator’s bias and variance are:
  \[ \text{Bias} \left[ \bar{\mu}_n (\bar{\mu}, \alpha) \right] = (1 - \alpha) (\bar{\mu} - \mu_n) \]
  \[ \text{Var} \left[ \bar{\mu}_n (\bar{\mu}, \alpha) \right] = (1 - \alpha)^2 \text{Var} (\bar{\mu}) + \alpha^2 \frac{1}{T} \]

• So the total MSE of the posterior expectation is:
  \[ \sum_{n=1}^{N} \text{MSE} [\bar{\mu}_n] = \sum_{n=1}^{N} \text{Bias} [\bar{\mu}_n]^2 + \sum_{n=1}^{N} \text{Var} (\bar{\mu}_n) \]
Total MSE for Estimating Many Means

• The shrinkage estimate for the $n^{th}$ mean $\mu_n$ is:

$$\bar{\mu}_n(\mu, \alpha) = (1-\alpha)\mu + \alpha \bar{y}_n$$

• This estimator’s bias and variance are:

$$\text{Bias} \left[ \bar{\mu}_n(\mu, \alpha) \right] = (1-\alpha)(\mu - \mu_n)$$

$$\text{Var} \left[ \bar{\mu}_n(\mu, \alpha) \right] = (1-\alpha)^2 \text{Var}(\mu) + \alpha^2 \frac{1}{T}$$

• So the total MSE of the posterior expectation is:

$$\sum_{n=1}^{N} \text{MSE}[\bar{\mu}_n] = \sum_{n=1}^{N} (1-\alpha)^2 (\mu - \mu_n)^2 + \sum_{n=1}^{N} \left\{ (1-\alpha)^2 \text{Var}(\mu) + \alpha^2 \frac{1}{T} \right\}$$

$$= (1-\alpha)^2 \left[ \sum_{n=1}^{N} (\mu - \mu_n)^2 + NV \text{Var}(\mu) \right] + \frac{N}{T} \alpha^2$$

$$= (1-\alpha)^2 \text{MSE}(\mu) + \alpha^2 \text{MSE}(\bar{y})$$
Total MSE for Estimating Many Means

- The shrinkage estimate for the \( n^{th} \) mean \( \mu_n \) is:
  \[
  \bar{\mu}_n(\mu, \alpha) = (1 - \alpha) \bar{\mu} + \alpha \bar{y}_n
  \]

- This estimator’s bias and variance are:
  \[
  \text{Bias}\left[\bar{\mu}_n(\mu, \alpha)\right] = (1 - \alpha)(\bar{\mu} - \mu_n)
  \]
  \[
  \text{Var}\left[\bar{\mu}_n(\mu, \alpha)\right] = (1 - \alpha)^2 \text{Var}(\bar{\mu}) + \alpha^2 \frac{1}{T}
  \]

- So the total MSE of the posterior expectation is:
  \[
  \sum_{n=1}^{N} \text{MSE}[\bar{\mu}_n] = (1 - \alpha)^2 \left[ \sum_{n=1}^{N} (\mu - \mu_n)^2 + N\text{Var}(\bar{\mu}) \right] + \frac{N}{T} \alpha^2
  \]
  \[
  = (1 - \alpha)^2 \text{MSE}(\bar{\mu}) + \alpha^2 \text{MSE}(\bar{y})
  \]
Choosing the Common Shrinkage Target

- **Goal:** Minimize the posterior MSE of the shrinkage estimator

\[
\sum_{n=1}^{N} MSE\left(\bar{\mu}_n(\mu, \alpha)\right) = (1-\alpha)^2 \sum_{n=1}^{N} (\mu - \mu_n)^2 + N(1-\alpha)^2 Var(\mu) + \frac{N}{T} \alpha^2
\]

- **Taking First Order Conditions with respect to \( \mu \)**

\[
\frac{\delta}{\delta \mu} \sum_{n=1}^{N} MSE\left(\bar{\mu}_n(\mu, \alpha)\right) = 2(1-\alpha)^2 \sum_{n=1}^{N} (\mu_S - \mu_n) = 0
\]

\[
\Rightarrow N\mu_S = \sum_{n=1}^{N} \mu_n
\]

\[
\Rightarrow \mu_S = \frac{1}{N} \sum_{n=1}^{N} \mu_n
\]

- **This gives the Infeasible Stein Optimal prior expectation**
Choosing the Shrinkage Intensity

• Goal: Minimize the posterior MSE of the shrinkage estimator

\[
\sum_{n=1}^{N} \text{MSE}\left[ \bar{\mu}_n\left( \mu_S, \alpha \right) \right] = (1 - \alpha)^2 \sum_{n=1}^{N} \text{MSE}\left[ \mu_{S,n} \right] + \alpha^2 \sum_{n=1}^{N} \text{MSE}\left[ \bar{y}_n \right]
\]

• Taking First Order Conditions with respect to \( \alpha \)

\[
\frac{\delta}{\delta \alpha} \sum_{n=1}^{N} \text{MSE}\left[ \bar{\mu}_n\left( \mu_S, \alpha \right) \right] = -2(1 - \alpha_s) \sum_{n=1}^{N} \text{MSE}\left[ \mu_{S,n} \right] + 2\alpha_s \sum_{n=1}^{N} \text{MSE}\left[ \bar{y}_n \right] = 0
\]

\[
\Rightarrow (1 - \alpha_s) \sum_{n=1}^{N} \text{MSE}\left[ \mu_{S,n} \right] = \alpha_s \sum_{n=1}^{N} \text{MSE}\left[ \bar{y}_n \right]
\]

\[
\Rightarrow \alpha_s = \frac{\sum_{n=1}^{N} \text{MSE}\left[ \mu_{S,n} \right]}{\sum_{n=1}^{N} \text{MSE}\left[ \bar{y}_n \right] + \sum_{n=1}^{N} \text{MSE}\left[ \mu_{S,n} \right]}
\]
Stein Shrinkage Estimator Admissibility

- Given the shrinkage estimator’s MSE
  \[
  \sum_{n=1}^{N} \text{MSE} \left[ \bar{\mu}_n \left( \mu_S, \alpha \right) \right] = (1 - \alpha)^2 \sum_{n=1}^{N} \text{MSE} \left[ \mu_{S,n} \right] + \alpha^2 \sum_{n=1}^{N} \text{MSE} \left[ \bar{y}_n \right]
  \]

- Minimizing the Mean Square error yields optimal priors:
  \[
  \bar{\mu}_S = \frac{1}{N} \sum_{n=1}^{N} \mu_n \\
  \alpha_s = \frac{\text{MSE} \left[ \mu_S \right]}{\text{MSE} \left[ \bar{y} \right] + \text{MSE} \left[ \mu_S \right]}
  \]

- Result: The Stein Shrinkage estimator: \( \bar{\mu} \left( \mu_S, \alpha_s \right) \)
  is admissible for \( \mu \) in the class of shrinkage estimators with common prior means
Feasible Stein Estimator for Many Means

- Implementing the Stein Estimator requires knowing:
  \[
  \mu_S = \frac{1}{N} \sum_{n=1}^{N} \mu_n
  \]

- If we estimate:
  \[
  \hat{\mu}_S = \frac{1}{N} \sum_{n=1}^{N} \bar{y}_n
  \]

- Then:
  \[
  \text{Var}(\hat{\mu}_S) = \text{Var}\left(\frac{1}{N} \sum_{n=1}^{N} \bar{y}_n\right) = \frac{1}{N^2} \sum_{n=1}^{N} \text{Var}(\bar{y}_n) = \frac{1}{N^2} \sum_{n=1}^{N} \frac{1}{T} = \frac{1}{NT}
  \]
Feasible Stein Estimator for Many Means

• Implementing the Stein Estimator requires knowing:

\[ \mu_s = \frac{1}{N} \sum_{n=1}^{N} \mu_n \]

\[ \alpha_s = \frac{MSE[\mu_s]}{MSE[\bar{y}] + MSE[\mu_s]} \]

• If we estimate: \[ \hat{\mu}_s = \frac{1}{N} \sum_{n=1}^{N} \bar{y}_n \]

Then:

\[ Var(\hat{\mu}_s) = \frac{1}{NT} \]

• We can also estimate the Bias for the \( n^{th} \) expectation using the sample average for that expectation

\[ \widehat{Bias}_n(\hat{\mu}_s) = (\bar{y}_n - \hat{\mu}_s) \]

• So:

\[ \widehat{MSE}[\hat{\mu}_s] = \sum_{n=1}^{N} (\bar{y}_n - \hat{\mu}_s)^2 + \sum_{n=1}^{N} \frac{1}{NT} = \sum_{n=1}^{N} (\bar{y}_n - \hat{\mu}_s)^2 + \frac{1}{T} \]
Feasible Stein Estimator for Many Means

- Implementing the Stein Estimator requires knowing:

\[ \mu_S = \frac{1}{N} \sum_{n=1}^{N} \mu_n \]
\[ \alpha_S = \frac{MSE[\mu_S]}{MSE[\bar{y}] + MSE[\mu_S]} \]

- If we estimate: \( \hat{\mu}_S = \frac{1}{N} \sum_{n=1}^{N} \bar{y}_n \) Then: \( Var(\hat{\mu}_S) = \frac{1}{NT} \)

- Estimating the Bias by: \( \text{Bias}_n(\hat{\mu}_S) = (\bar{y}_n - \hat{\mu}_S) \)

- Gives an estimated \( \widehat{MSE}[\hat{\mu}_S] = \sum_{n=1}^{N} (\bar{y}_n - \hat{\mu}_S)^2 + \frac{1}{T} \)

- And the Feasible Stein Estimator is \( \bar{\mu}_n(\hat{\mu}_S, \alpha_S) \) with

\[ \alpha_S = \frac{\widehat{MSE}[\hat{\mu}_S]}{MSE[\bar{y}] + \widehat{MSE}[\hat{\mu}_S]} \]
Bayesian and Empirical Bayesian Prediction

I. Bayesian Analysis of Means in the Normal Model

II. Admissibility in Predicting Expectations
   A. Mean Square Error and Estimator Admissibility
   B. Stein’s Shrinkage Estimator for the Sample Mean

III. Learning Expectations from Other Expectations

IV. Bayesian Shrinkage Estimators for Covariance Matrices
Shrinkage Estimators for Covariance Matrices

• The sample covariance matrix \( S \) is poorly specified
  – Free Parameters: \( N(N+1)/2 \)
  – Too little structure leads to very high variance in estimates

• The shrinkage approach mixes the sample covariance matrix with the Identity matrix
  – Automated procedures select the amount of structure (shrinkage intensity) based on the data
  – Other models can use richer specifications for structure
Covariance Matrix Shrinkage: Identity Matrix

- Estimate Covariance Matrix as a combination of the sample with the identity matrix:
  \[ \Sigma^* = \alpha S + (1 - \alpha) I \]

- Optimizes Shrinkage Intensity (\( \alpha \)) to minimize expected loss under Frobenius Norm (fancy way of saying Matrix MSE):
  \[ R = E \left[ \| \alpha I + (1 - \alpha) S - \Sigma \|^2 \right] = \sum_{i=1}^{N} \sum_{j=1}^{N} E \left[ \left( \alpha I_{ij} + (1 - \alpha) s_{ij} - \sigma_{ij} \right)^2 \right] \]
    - Analytical results give solutions for optimal (\( \alpha \)) in terms of underlying parameters
    - Estimating parameters yield feasible estimates of the optimal shrinkage intensity
Alternative Approaches to Feasibility

• Bootstrap Approach
  – Resample the data to estimate variances & covariances in loss function
  – Use resampled estimates to solve for optimal shrinkage intensities

• Cross Validation Approach
  – Estimate Target Covariance Matrices using pre-sample data
  – Compute Shrinkage Intensity to optimize portfolio performance in Training Sample
  – Evaluate performance of portfolio out of sample
Bayesian and Empirical Bayesian Prediction

I. Bayesian Analysis of Means in the Normal Model

II. Admissibility in Predicting Expectations
   A. Mean Square Error and Estimator Admissibility
   B. Stein Estimator for the Sample Mean

III. Learning Expectations from Other Expectations

IV. Bayesian Shrinkage Estimators for Covariance Matrices