2017

Improved Simultaneous Estimation of Location and System Reliability via Shrinkage Ideas

Beidi Qiang
University of South Carolina

Follow this and additional works at: https://scholarcommons.sc.edu/etd

Part of the Statistics and Probability Commons

Recommended Citation

This Open Access Dissertation is brought to you for free and open access by Scholar Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact dillarda@mailbox.sc.edu.
IMPROVED SIMULTANEOUS ESTIMATION OF LOCATION AND SYSTEM RELIABILITY VIA SHRINKAGE IDEAS

by

Beidi Qiang

Bachelor of Science
Denison University, 2012

Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in
Statistics
College of Arts and Sciences
University of South Carolina
2017
Accepted by:
Edsel Pena, Major Professor
Lianming Wang, Committee Member
Paramita Chakraborty, Committee Member
Gabriel Terejanu, Committee Member
Cheryl L. Addy, Vice Provost and Dean of the Graduate School
ACKNOWLEDGMENTS

I acknowledge NIH Grant P30GM103336-01A1 and NIH R01CA154731. I would like to show my gratitude to Professor J. Lynch, and graduate students P. Liu, S. Shen, J. Thompson, T. Kim, and L. Wang for their insights, comments, and feedbacks.
Abstract

In decision theory, when several parameters need to be estimated simultaneously, many standard estimators can be improved, in terms of a combined loss function. The problem of finding such estimators has been well studied in the literature, but mostly under parametric settings, which is inappropriate for heavy-tailed distributions. In the first part of this dissertation, a robust simultaneous estimator of location is proposed using the shrinkage idea. A nonparametric Bayesian estimator is also discussed as an alternative. The proposed estimators do not assume a specific parametric distribution and they do not require the existence of finite moments. The performance of proposed estimators are examined in simulation studies and financial data applications. In the second part, we extend the idea of simultaneous estimation in the context of estimating system reliability when component data are observed. We propose an improved estimator of system reliability by using shrinkage estimators for each of the component reliabilities and then utilize the structure function to combine these estimators to obtain the system reliability estimator. The approach is general since the shrinkage is not on the estimated parameters of component reliability functions, but is instead on the estimated component hazard functions, and is therefore extendable to the nonparametric setting. The details in nonparametric setting are discussed in a later chapter. Simulation results are presented to examine the performances of the proposed estimator.
TABLE OF CONTENTS

ACKNOWLEDGMENTS ......................................................... ii

ABSTRACT ........................................................................ iii

LIST OF TABLES ................................................................. vi

LIST OF FIGURES ............................................................... viii

CHAPTER 1 INTRODUCTION .................................................. 1

CHAPTER 2 SOME REVIEW AND MOTIVATION .......................... 3
  2.1 The Decision Theory Framework .................................. 3
  2.2 On Simultaneous and Shrinkage Estimators ..................... 4

CHAPTER 3 ROBUST SIMULTANEOUS ESTIMATOR OF LOCATION .. 8
  3.1 Idea and Motivation of Estimator ................................. 8
  3.2 Evaluating the Estimators ......................................... 9
  3.3 Estimating the Shrinkage Coefficient $c^*$ ....................... 10
  3.4 The Nonparametric Empirical Bayesian Approach ............ 11
  3.5 Constructing Confidence Intervals ............................... 14
  3.6 Example and Numerical Illustration ............................ 15
  3.7 Simulated Comparisons of the Estimators ..................... 16
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.8</td>
<td>Application to Stock Return Data</td>
<td>17</td>
</tr>
<tr>
<td>3.9</td>
<td>Summary</td>
<td>20</td>
</tr>
<tr>
<td><strong>4</strong></td>
<td>Improved Estimation of System Reliability Through Shrinkage Idea in Parametric Settings</td>
<td>22</td>
</tr>
<tr>
<td>4.1</td>
<td>Reliability Concepts</td>
<td>22</td>
</tr>
<tr>
<td>4.2</td>
<td>Statistical Model and Data Structure</td>
<td>28</td>
</tr>
<tr>
<td>4.3</td>
<td>Improved Estimation Through Shrinkage Idea</td>
<td>30</td>
</tr>
<tr>
<td>4.4</td>
<td>Evaluating the Estimators</td>
<td>32</td>
</tr>
<tr>
<td>4.5</td>
<td>Estimating the Optimal $c^*$</td>
<td>33</td>
</tr>
<tr>
<td>4.6</td>
<td>Examples and Numerical Illustration</td>
<td>37</td>
</tr>
<tr>
<td>4.7</td>
<td>Simulated Comparisons of the Estimators</td>
<td>37</td>
</tr>
<tr>
<td>4.8</td>
<td>Summary</td>
<td>40</td>
</tr>
<tr>
<td><strong>5</strong></td>
<td>Improved Estimation of System Reliability Through Shrinkage Idea in Nonparametric Settings</td>
<td>42</td>
</tr>
<tr>
<td>5.1</td>
<td>The Nonparametric Model and Data Structure</td>
<td>42</td>
</tr>
<tr>
<td>5.2</td>
<td>Improved Estimation in Nonparametric Setting</td>
<td>43</td>
</tr>
<tr>
<td>5.3</td>
<td>Optimal $c^*$</td>
<td>44</td>
</tr>
<tr>
<td>5.4</td>
<td>Examples and Numerical Illustration</td>
<td>45</td>
</tr>
<tr>
<td>5.5</td>
<td>Simulated Comparisons of the Estimators</td>
<td>47</td>
</tr>
<tr>
<td>5.6</td>
<td>Summary</td>
<td>48</td>
</tr>
<tr>
<td><strong>6</strong></td>
<td>Conclusions and Future Research</td>
<td>49</td>
</tr>
<tr>
<td><strong>Bibliography</strong></td>
<td></td>
<td>52</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table 3.1</th>
<th>Simulated Estimators Under Normal Distributions. ( k=3 ) and ( n=5 ). Confidence intervals are indicated in parentheses.</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.2</td>
<td>Simulated Estimators Under Laplace Distributions. ( k=3 ) and ( n=5 ). Confidence intervals are indicated in parentheses.</td>
<td>15</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Simulated MSE of Estimators Under Different Distributions. These are based on 1000 simulation replications, ( k=10 ) and ( n=15 ). Standard errors are indicated in parentheses.</td>
<td>17</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Relative Efficiency (in %) of the Estimators Under Different Distributions. The relative efficiency is calculated with respect to the sample median.</td>
<td>17</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>Daily Stock Return of AAPL, AXP, BA, CAT and CSCO for the First 10 Consecutive Trading Days Starting from January 3rd.</td>
<td>19</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>Prediction MSEs of Different Location Estimators Using Daily Dow Jones Stock Return Data. Standard errors are indicated in parentheses.</td>
<td>20</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Average Losses of the Estimators in Series and Parallel Systems under Exponential Lifetime Distribution. These are based on 1000 simulation replications. The systems had 10 components, and the sample size for each replication was 10. Also indicated are the average values of the estimated shrinkage coefficient ( c ).</td>
<td>39</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Average Losses of the Estimators in Series and Parallel Systems under Weibull Lifetime Distribution. These are based on 1000 simulation replications. The systems had 10 components, and the sample size for each replication was 10. Also indicated are the average values of the estimated shrinkage coefficient ( c ).</td>
<td>40</td>
</tr>
</tbody>
</table>
Table 5.1 Average Losses of the Nonparametric Estimators in Series and Parallel Systems under Exponential Lifetime Distribution. These are based on 1000 simulation replications. The systems had 5 components, and the sample size for each replication was 15. Also indicated are the average values of the estimated shrinkage coefficient $c$. 

\[ \ldots \]
Figure 3.1  The Effect of Shrinkage v.s. Sample Size. Simulation based on 1000 Replications, p=10. The relative efficiency is calculated with respect to the sample median. ............................................. 18

Figure 3.2  Histogram and QQ plot of Apple Stock Returns for 2683 Consecutive Business Days, from January 2005 to August, 2015. The returns seem to follow a heavy-tailed distribution. .............. 20

Figure 4.1  System reliability functions for the series, parallel, series-parallel, and bridge systems when the components have common unit exponential lifetimes and there are 5 components in the series and parallel systems. ......................................................... 26

Figure 4.2  Estimated and True System Reliability Functions Over Time for a Series System with 10 components. The estimators are the ML based on component data and the improved estimator based on a sample of size 10 systems. ................................. 38

Figure 4.3  Estimated and True System Reliability Functions Over Time for a Parallel System with 10 components. The estimators are the ML based on component data and the improved estimator based on a sample of size 10. ......................................................... 39

Figure 4.4  Comparing Performances of the Component-Data Based Estimators at Different Values of $t$ for a 10-Component Parallel System. 1000 Replications we used, with each replication having a sample size of 10. ......................................................... 41

Figure 5.1  Estimated and True System Reliability Functions Over Time for a Series System with 5 components. The estimators are the nonparametric ML based on component data and the improved nonparametric estimator based on a sample of size 15 systems. . . 46
Figure 5.2  Estimated and True System Reliability Functions Over Time for a Parallel System with 5 components. The estimators are the nonparametric ML based on component data and the improved nonparametric estimator based on a sample of size 15.
Chapter 1

Introduction

In decision theory, when several parameters need to be estimated simultaneously, many standard estimators can be improved under a combined loss function, by a combined estimator. The improvement of using combined information is observed, even when using independent data to estimate those parameters. One example that uses the idea of simultaneous estimation explicitly is the James-Stein shrinkage estimator (13). As a biased estimator of the mean of Gaussian random vectors, the James-Stein estimator dominates the standard least squares estimators in terms of total mean squared error (MSE), when three or more parameters are estimated simultaneously. A lot of research has been done focusing on shrinkage estimators and the problem of simultaneous estimators which dominate the usual maximum likelihood estimators (MLEs) has been well studied.

One extension presented in this work is the idea of developing a more robust shrinkage estimator under nonparametric assumptions. In the literature, most existing James-Stein type shrinkage estimators are developed under the assumption of multivariate normal distribution. Some studies relaxed the normal assumption, but still require the existence of finite moments. However, these assumptions may not always be realistic when heavy-tailed data is observed. To extend the idea, a more robust estimator based on the sample median is proposed, which does not assume a specific parametric distribution nor requires the existence of finite moments. An alternative estimator is also developed using nonparametric Bayesian approach. The practical improvement of the proposed estimators will be demonstrated through sim-
ulation studies and empirical data analysis. The proposed estimator is expected to outperform the usual ML estimates in heavy-tailed distributions especially in "large $p$, small $n$" settings.

Another natural extension is to apply the idea of simultaneous estimation to failure time analysis, notably survival analysis and reliability, especially estimating the reliability of a coherent system using lifetime data from its components. When component-level data are available, the traditional way of estimating system reliability is to find component level MLEs and then utilize these MLEs according to the system structure to obtain an estimator of the system reliability. We propose an improved estimator of system reliability under an invariant global loss function. Shrinkage type estimators for each of the components reliability functions are obtained first and then these estimators are combined according to the system structure function to obtain an improved estimator of the system reliability. The approach is general since the shrinkage is not on the estimated component reliability function parameters but is instead on the estimated components hazard functions, and are therefore extendable to the setting where the components reliability functions are specified non-parametrically. The performances of different estimators will be compared though simulation studies. The proposed estimator is expected to perform better under the given loss function, as compared to the MLEs.

The rest of the dissertation is organized as follow. In Chapter 2, the James-Stein estimator and its extensions and applications are reviewed. The ideas and results of extending the simultaneous estimation idea to nonparametric setting is discussed in Chapter 3. In Chapters 4 and 5, the shrinkage estimation method is applied to the estimation of system reliability. Some properties of the estimator and simulation analysis are presented. Ideas for future studies, improvements, and extensions are also discussed.
Chapter 2

Some Review and Motivation

2.1 The Decision Theory Framework

A decision problem has the following elements:

$$(\Theta, A, \mathcal{X}, \mathcal{F} = \{F(\cdot|\theta), \theta \in \Theta\}, L, \mathcal{D})$$

Here $\Theta$ is the parameter space containing the possible values of some parameters $\theta$, which could be finite- or infinite-dimensional; $A$ is the action space consisting of all possible actions that the decision maker could take; $L : \Theta \times A \rightarrow \mathbb{R}$ is the loss function, with $L(\theta, a)$ denoting the loss incurred by choosing action $a$ when the parameter is $\theta$. The observable data $X$ takes values in the sample space $\mathcal{X}$, with $X$, given $\theta$, having distribution $F(\cdot|\theta)$, which belongs to the family of distribution functions $\mathcal{F}$. Non-randomized decision functions are (measurable) mappings $\delta : \mathcal{X} \rightarrow A$, and the totality of such decision functions is the decision function space $\mathcal{D}$. To assess the quality of a decision function $\delta \in \mathcal{D}$, we utilize the risk function given by

$$R(\theta, \delta) = E[L(\theta, \delta(X)|\theta)],$$

which is the expected loss incurred by using decision function $\delta$ when the parameter is $\theta$. Good decision functions are those with small risks whatever the value of $\theta$. In particular, a decision function $\delta_1$ is said to dominate a decision function $\delta_2$ if for all $\theta \in \Theta$, $R(\theta, \delta_1) \leq R(\theta, \delta_2)$ with strict inequality for some $\theta \in \Theta$. In such a case the decision function $\delta_2$ is inadmissible. The statistical inference problem of parameter point estimation falls into this decision-theoretic framework with the decision functions being called estimators.
Simultaneous Estimation

This decision-theoretic framework carries over to simultaneous decision-making, in particular, simultaneous estimation. Consider, for instance, the situation where \( \theta = (\mu_1, \mu_2, \ldots, \mu_K) \in \Theta = \mathbb{R}_K \), the action space is \( A = \mathbb{R}_K \), and the data observable is \( X = (X_1, X_2, \ldots, X_K) \in \mathcal{X} = \mathbb{R}_K \) where the \( X_j \)'s are a normal distribution with mean \( \mu_j \) and variance \( \sigma^2 \), assumed known. \( X_j \)'s are assumed to be independent, given \( \mu_j \)'s. For the simultaneous estimation problem, we could use the loss function given by

\[
L(\theta, a) = ||\theta - a||^2 = \sum_{i=1}^{K} (\mu_i - a_i)^2, (\theta, a) \in \Theta \times A.
\]

This loss function is referred to as quadratic loss function. The maximum likelihood (ML) estimator of \( \theta \) is \( \delta_{ML}(X) = X \), whose risk function is given by

\[
R(\theta, \delta_{ML}) = E \left[ ||\theta - X||^2 | \theta \right] = K\sigma^2.
\]

Incidentally, this ML estimator is also the least-squares (LS) estimator of \( \theta \). When \( K = 1 \) or \( K = 2 \), \( \delta_{ML} \) is the best (risk-wise) estimator of \( \theta \).

James-Stein Shrinkage Estimation

James and Stein (13) demonstrated that there is a better estimator of \( \theta \) than \( \delta_{ML}(X) = X \), when 3 or more parameters are estimated simultaneously. An estimator that dominates the ML estimator is their so-called shrinkage estimator of \( \theta \), given by

\[
\hat{\theta}_{JS}(X) = \hat{\theta}_{JS} = \left[ 1 - \frac{(K-2)\sigma^2}{n\|X\|^2} \right] X.
\]

More generally, if \( X = (X_1, \ldots, X_n) \) are IID \( 1 \times p \) multivariate normal vectors with mean vector \( \theta \), and common covariance matrix \( \sigma^2 I_K \), then the James-Stein estimator of \( \theta \) is given by

\[
\hat{\theta}_{JS}(X) = \hat{\theta}_{JS} = \left[ 1 - \frac{(K-2)\sigma^2}{n\|X\|^2} \right] X,
\]
where $\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$ denotes the vector of sample means.

The James-Stein shrinkage estimator is one which utilizes the combined data for estimating each component parameter, even though the component variables are independent, to improve the simultaneous estimation of the components of \( \theta \). It shows that optimizing (i.e., minimizing) a global loss is not the same as optimizing individually the loss of each component estimator. In essence, there is an advantage in the borrowing of information from each of the component data, demonstrating that when dealing with a combined or global loss function, it may be beneficial to borrow information in order to improve the estimation process. Observe that the James-Stein type of shrinkage estimator is of the form

$$\delta_c(X) = \hat{\theta}_c = c \bar{X},$$

for some $c > 0$, which in this case is data-dependent.

**An Empirical Bayes Approach**

The James-Stein Shrinkage estimator can be developed using the empirical Bayes method (9). Suppose $X = (X_1, \ldots, X_n)$ are IID $1 \times p$ multivariate normal vectors with mean vector $\theta$, and common covariance matrix $\sigma^2 I_k$. If we place a $N(0, \tau^2 I_K)$ prior on the vector $\theta$, the posterior distribution of $\theta$ is

$$\theta | X_1, \ldots, X_n \sim N \left( \frac{\tau^2}{\tau^2 + \sigma^2/n} \bar{X}, \left( \frac{1}{\tau^2 + \sigma^2/n} \right)^{-1} \right).$$

The Bayes estimator of $\theta$ is the posterior mean, which is given by

$$\hat{\theta}_{Bayes} = \frac{\tau^2}{\tau^2 + \sigma^2/n} \bar{X}. \quad (2.1)$$

The hyper parameter $\tau^2$ is then estimated from the data, which makes the approach empirical Bayes. Note that marginally $\bar{X} \sim N(0, (\sigma^2/n + \tau^2)I)$, and thus, $(\sigma^2/n + \tau^2)/||X||^2$ follows an inverse chi-square distribution with K degrees of free-
dom. When \( K \geq 3 \),
\[
E \left[ \frac{\sigma^2/n + \tau^2}{\|X\|^2} \right] = \frac{1}{K - 2},
\]
and \( \tau^2 \) is estimated by
\[
\hat{\tau}^2 = \frac{\|X\|^2}{K - 2} - \frac{\sigma^2}{n}.
\]
The James-Stein Estimator is then recovered when replacing \( \tau^2 \) by \( \hat{\tau}^2 \) in (2.1).

**Motivation**

In the literature, some generalizations of James-Stein type shrinkage estimators were developed, but most existing methods use the assumption of normal distribution. However, in 'large \( p \), small \( n \)' problems, the sample covariance matrix is usually singular, and thus traditional shrinkage methods cannot be applied. Wang et al. (28) proposed a shrinkage estimator for population mean under quadratic loss with unknown covariance matrices. Although no specific distribution is assumed, their non-parametric shrinkage estimator is based on the sample mean and does require the distribution to have finite moments. This assumption can be unrealistic when the data follow some heavy-tailed distribution, such as the Cauchy. Therefore, more robust shrinkage estimators based on nonparametric assumptions are developed through the shrinkage idea and the Bayesian approach in Chapter 2.

The application of shrinkage estimation in survival analysis has also been partly addressed in the literature. In (27) three versions of shrinkage estimators under exponential lifetimes were examined. Later, in (24) shrinkage estimators under the LINEX loss function in the situation with exponential lifetime censored data were presented. The shrinkage estimation of the reliability function for other lifetime distributions were also studied in (6) and (23). The extension to the estimation of the system reliability has also been discussed under certain system structures and lifetime distributions, see, for instance, (21; 22). Instead of developing estimators under a specific system structure and lifetime distribution, in Chapter 3, the method of simultaneous
estimation is extended to general system structures and lifetime models. Improved estimators of component reliability functions are developed under an invariant global loss function. The improved estimators are developed under a general framework, which can be further extended to nonparametric lifetime models.
Chapter 3

Robust Simultaneous Estimator of Location

3.1 Idea and Motivation of Estimator

To relax the normality assumption and provide a more robust estimator in simultaneous estimation, we extend the shrinkage idea into nonparametric settings, where heavy-tailed distributions are allowed. The nonparametric shrinkage estimator is then applied to stock return data, since such data is believed to have heavy-tailed properties based on empirical evidence.

Consider i.i.d $1 \times K$ vectors $X_1, \ldots, X_n$ that satisfy

$$X_i = \mu + \epsilon_i, X_j \in \mathbb{R}_K, \epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{iK}),$$

where $\epsilon_{ij}$'s are identically and independently distributed and follow some symmetric distribution $F_j$ that is centered at 0, for $i = 1, \ldots, n$, $j = 1, \ldots, K$. $F_j$'s are unknown. Consider the simultaneous estimation problem of estimating of the location parameters $\mu_j$'s, with respect to the quadratic loss function

$$L(\mu, \hat{\mu}) = ||\hat{\mu} - \mu||^2,$$

where $\mu = (\mu_1, \mu_2, \ldots, \mu_K)$.

A common nonparametric approach is to estimate the location parameters by the sample medians. But using the idea of simultaneous estimation, some improvement can be achieved, in terms of the combined MSE. Consider estimating $\mu_j$ by a shrinkage of the sample median, $\tilde{X}_j$, i.e.

$$\hat{\mu}_{j,c} = c \cdot \tilde{X}_j.$$
When sample size \( n \) is odd, \( \tilde{X}_j \equiv X_{(\frac{n+1}{2})j} \). The goal is to find the optimal \( c \) in terms of the combined expected loss function. However, this optimal coefficient \( c^* \) may depend on the unknown parameters. Thus an estimator of \( c^* \) will be used in practice.

The James-Stein shrinkage estimator can be viewed as an empirical Bayes estimator when putting a normal prior on the parameters \( \mu \). So we also seek a Bayes approach to develop a robust shrinkage estimator. A Dirichlet process prior is used given the nonparametric setting.

### 3.2 Evaluating the Estimators

To find the optimal amount of shrinkage, the risk or expected loss \( \mathbb{E}[L(\mu)] = \mathbb{E}[||\mu - \hat{\mu}||^2] \) is minimized with respect to \( c \).

\[
R(\mu, \hat{\mu}(c)) = \mathbb{E}[L(\mu, \hat{\mu}(c))] = \sum_{j=1}^{K} \mathbb{E}[c\tilde{X}_j - \mu_j]^2 = c^2 \sum_{j=1}^{K} \text{var}(\tilde{X}_j) + (c - 1)^2 ||\mu||^2,
\]

where \( \text{var}(\tilde{X}_j) \) denotes the variance of the sample median in sample \( j \). Thus the optimal shrinkage coefficient \( c \) is given by

\[
c^* = 1 - \frac{\sum_{j=1}^{K} \text{var}(\tilde{X}_j)}{||\mu||^2 + \sum_{j=1}^{K} \text{var}(\tilde{X}_j)} < 1.
\]

The corresponding risk of \( \hat{\mu}^* \) is then

\[
\mathbb{E}(L(\mu, \hat{\mu}^*)) = (c^*)^2 \sum_{j=1}^{p} \text{var}(\tilde{X}_j) + (1 - c^*)^2 ||\mu||^2.
\]

Notice the sample median, \( \tilde{X}_j \), has an asymptotic normal distribution with mean \( \theta_j \) and variance \( [4nf_j(\theta_j)^2]^{-1} \). Thus, \( \text{var}(\tilde{X}_j) \) goes to 0 as sample size increases, and thus the shrinkage coefficient \( c^* \) converges to 1 when sample size converges to infinity, i.e. the shrinkage effect is only significant when sample size is small.
3.3 Estimating the Shrinkage Coefficient $c^*$

In practice, in the expression of $c^*$, the true location parameters and the variances of the sample medians are unknown since the $F_j$’s are unknown and need to be estimated based on the observed data. In the nonparametric setting, the sample median of $X_j$ is used as a robust estimator of $\mu_j$. Thus $||\mu||^2$ is estimated by

$$||\mu||^2 \approx \bar{X}_1^2 + \bar{X}_2^2 + \cdots + \bar{X}_K^2$$

To estimate the variance of the sample median nonparametrically, we adopt the method proposed by Maritz and Jarrett (19). Consider the case when the sample size is odd, i.e. $n = 2m + 1$. Some modifications are needed when the sample size is even, but follows the same idea. The moments of the sample median are given by the following equation,

$$E[\tilde{X}^k] = \frac{(2m + 1)!}{(m!)^2} \int_{-\infty}^{\infty} x^k [F'(x)(1 - F(x))]^m f(x) dx.$$

Substitute $y = F(x)$ to get

$$E[\tilde{X}^k] = \frac{(2m + 1)!}{(m!)^2} \int_0^1 [F^{-1}(y)]^k [y(1 - y)]^m dy.$$

The inverse of the distribution function, $F^{-1}(x)$, can be estimated piece by piece by the observed order statistics. Thus,

$$E[\tilde{X}^k] \approx \sum_{i=1}^{n} [X_{(i)}]^k W_i$$

where $X_{(i)}$ is the $i$th observed order statistics, and

$$W_i = \frac{(2m + 1)!}{(m!)^2} \int_{(i-1)/n}^{i/n} [y(1 - y)]^m dy.$$

Combining the above results, the estimated variance of the sample median is then given by

$$\hat{\text{var}}(\tilde{X}_j) \equiv V_j = \sum_{i=1}^{n} \left[X_{(i)}\right]^2 W_i - \left[ \sum_{i=1}^{n} X_{(i)} W_i \right]^2.$$
where $X_{(i)j}$ is the $i$th order statistic in sample $j$.

As proved by Maritz and Jarrett (19), $V_j$ is a consistent estimator of $\text{var}(\tilde{X}_j)$ when $\text{var}(\tilde{X}_j)$ is finite. However, $E(V_j)$ may not exist when considering long-tailed distribution, such as Cauchy. Under Cauchy distributions, the second-order moments of the four extreme order statistics $X_{(1)}$, $X_{(2)}$, $X_{(n-1)}$, and $X_{(n)}$ are not finite, and thus the expectation of $V$ does not exist. In order to obtain a finite $E(V)$ it is necessary to use Winsorized estimate which does not involve the four extreme order statistics. Winsorized estimate is obtained by replacing the smallest $l$ order statistics by the $(l+1)$th order statistic and replace the largest $l$ order statistics by the $(n-l)$th order statistic.

Finally, the estimated $c^*$ is in the form

$$
\hat{c^*} = 1 - \frac{\sum_{j=1}^{K} V_j}{\sum_{j=1}^{p} \tilde{X}_j^2 + \sum_{j=1}^{p} V_j} < 1.
$$

### 3.4 The Nonparametric Empirical Bayesian Approach

Since the James-Stein shrinkage estimator can be developed through empirical Bayes approach, we seek an analog in the nonparametric setting by assuming a Dirichlet process prior on the distribution of $X_j$, $j = 1, \cdots, K$. The definitions of Dirichlet distribution and Dirichlet process are given below and more details are provided in (11).

**Definition 3.1.** Assume $Z_1, \cdots, Z_k$ are independent random variables, with $Z_j \sim \text{Gamma}(\alpha_j, 1)$. Let $Y_j = Z_j / \sum_{i=1}^{k} Z_i$. Then $(Y_1, \cdots, Y_k)$ follows a Dirichlet distribution with parameters $(\alpha_1, \cdots, \alpha_k)$, denoted by $(Y_1, \cdots, Y_k) \sim \mathcal{D}(\alpha_1, \cdots, \alpha_k)$.

Given the definition of a Dirichlet distribution, one can show that if $(Y_1, \cdots, Y_k) \sim \mathcal{D}(\alpha_1, \cdots, \alpha_k)$, then $E(Y_j) = \alpha_j / \sum_{i=1}^{k} \alpha_i$. As a main property of the Dirichlet distribution (11), if the distribution of $(Y_1, \cdots, Y_k)$ is $\mathcal{D}(\alpha_1, \cdots, \alpha_k)$ and $X$ is a random
variable with \( Pr(X = j | Y_1, \ldots, Y_k) = Y_j \), then the distribution of \( (Y_1, \ldots, Y_k) | X = j \) is also Dirichlet, with parameters \( (\alpha_1, \ldots, \alpha_j + 1, \ldots, \alpha_k) \).

**Definition 3.2.** Let \( \alpha \) be a nonnegative, finitely additive, finite measure on \( (\mathcal{X}, \mathcal{A}) \). \( P \) is a Dirichlet Process on \( (\mathcal{X}, \mathcal{A}) \) with parameter \( \alpha \), if for all \( k = 1, 2, \ldots \) and for all measurable partitions \( (B_1, \ldots, B_k) \) of \( \mathcal{X} \),

\[
(P(B_1), \ldots, P(B_k)) \sim \mathcal{D}(\alpha(B_1), \ldots, \alpha(B_k)).
\]

The following theorem provides a way of updating prior beliefs in response to observed data using Bayes' rule.

**Theorem 3.3.** If \( P \) is a Dirichlet Process on \( (\mathcal{X}, \mathcal{A}) \) with parameter \( \alpha \) and \( X_1, \ldots, X_n \) is a sample from \( P \), then the posterior \( P | (X_1, \ldots, X_n) \) is a Dirichlet Process on \( (\mathcal{X}, \mathcal{A}) \) with parameter \( \alpha + \sum_{i=1}^{n} \delta_{X_i} \), where \( \delta_{X_i}(A) = I(X_i \in A) \) \((11)\).

Consider i.i.d \( 1 \times K \) vectors \( X_1, \ldots, X_n \) that satisfy \( X_{ij} \sim F_j \), \( j = 1, 2, \ldots, k \), \( i = 1, 2, \ldots, n \). We put i.i.d Dirichlet Process prior on \( F_j \), i.e. \( F_j \sim \text{Dir}(\alpha) \). Then according to Theorem 3.3, the posterior \( F_j | (X_1, \ldots, X_n) \), is a Dirichlet process with parameter \( \alpha + \sum_{i=1}^{n} \delta_{X_{ij}} \). By definition of Dirichlet process, for a measurable set \( B \),

\[
(F_j(B), F_j(B^c)) | (X_1, \ldots, X_n) \sim \mathcal{D}(\alpha(B) + \sum_{i=1}^{n} I(X_{ij} \in B), \alpha(B^c) + \sum_{i=1}^{n} I(X_{ij} \in B^c)).
\]

Therefore, we obtain an estimate of \( \hat{F}_j(B) \), given by

\[
\hat{F}_j(B) = E[F_j(B) | X_1, \ldots, X_n] = \frac{\alpha(B) + \sum_{i=1}^{n} I(X_{ij} \in B)}{\alpha(\mathcal{R}) + n} = \frac{\alpha(\mathcal{R}) \cdot \alpha(B) + \sum_{i=1}^{n} I(X_{ij} \in B)}{\alpha(\mathcal{R}) + n \cdot \frac{n}{n} \sum_{i=1}^{n} I(X_{ij} \in B)}
\]

In the expression of \( \hat{F}_j(B) \), \( \alpha \) is unknown and need to be estimated using the observed
data. To estimate $\alpha$ empirically, observe that

$$E[I(X_{ij} \leq t)] = E\{E[I(X_{ij} \leq t)|F_j]\} = E(F_j(t)) = \alpha((-\infty, t])/\alpha(\Re).$$

As mentioned in (11), $\alpha(\Re)$ is a measure of faith in the prior and if $\alpha(\Re)$ is small compared to sample size, more weight is given to the observations. Therefore, we take $\alpha(\Re) = \sqrt{nk}$ as a reasonable choice suggested in (25). Then $\alpha((-\infty, t])$ can be estimated empirically by

$$\frac{1}{\sqrt{nk}} \sum_{j=1}^{k} \sum_{i=1}^{n} I(X_{ij} \leq t).$$

The empirical Bayes estimator of $F_j$ is then given by

$$\hat{F}_j(t) = \frac{\sqrt{nk}}{\sqrt{nk} + n} \cdot \frac{1}{nk} \sum_{j=1}^{k} \sum_{i=1}^{n} I(X_{ij} \leq t) + \frac{n}{\sqrt{nk} + n} \cdot \frac{1}{n} \sum_{i=1}^{n} I(X_{ij} \leq t).$$

Notice that $\hat{F}_j(t)$ is a weighted sum of the empirical distribution based on the data from sample $j$ and the empirical distribution based on the data from all samples. Thus, $\hat{F}_j(t)$ can be viewed as a shrinkage of the empirical distribution of sample $j$ towards the average empirical distribution of all samples.

The Nonparametric Bayesian Shrinkage Estimator (NBSE) of the location parameter is then the median of the estimated distribution, which can be found by solving the following equation for $t$.

$$\hat{F}_j(t) = \frac{\sqrt{nk}}{\sqrt{nk} + n} \cdot \frac{1}{nk} \sum_{j=1}^{k} \sum_{i=1}^{n} I(X_{ij} \leq t) + \frac{n}{\sqrt{nk} + n} \cdot \frac{1}{n} \sum_{i=1}^{n} I(X_{ij} \leq t) = \frac{1}{2}.$$ 

Since the $\hat{F}_j(t)$ is a step function with discontinuities at the $X_{ij}$’s, numerical solution will be used but exact equality may not be obtained. We take $t^* = \arg\inf[F_j(t) > 0.5]$ as the solution.
3.5 Constructing Confidence Intervals

In this section, we present one approach to constructing corresponding confidence intervals. Given the nonparametric setting and the form of the estimators, a bootstrap approach is applicable. The steps are as follows.

1. At iteration $m$, randomly resample $n$ times from the index set $\{1, 2, \cdots n\}$ with replacement, where $n$ is the size of the original data. This gives indices $I_{m1}, \cdots, I_{mn}$.

2. From each original sample $j$, $j = 1, \cdots K$, take $X_{mj} = (X_{I_{m1}j}, \cdots, X_{I_{mn}j})$. The $X_{mj}$'s forms a new data with sample size equals to the original sample size.

3. Calculate the estimator of location using methods from the previous sections based on the bootstrap data. This gives an estimate $\hat{\mu}_m = (\hat{\mu}_{m1}, \cdots, \hat{\mu}_{mK})$ of the location vector.

4. Repeat step 1 to 3 $M$ times, save $\hat{\mu}_1, \cdots, \hat{\mu}_M$.

5. At a confidence level of $100(1 - \alpha)\%$, the lower limit of the confidence interval for $\mu_j$, $j = 1, \cdots, K$ will be the $100(\alpha/2)\%$ percentile of the $M$ estimates $\hat{\mu}_{1j}, \cdots, \hat{\mu}_{Mj}$, and the upper limit will be the $100(1 - \alpha/2)\%$ percentile of the $M$ estimates.

We note that the resulting confidence intervals are with respect to the individual parameters. However, simultaneous confidence intervals may be more appropriate in the situation since we are estimating several location parameters simultaneously. The simultaneous confidence intervals may be obtained using the Bonferroni Method or family-wise error rate approaches.
3.6 Example and Numerical Illustration

To illustrate the performance of the estimators, we provide some numerical examples under different distributions. We generated data based on normal and Laplace distributions. Under each distribution, a sample of $n = 5$ and $K = 3$ was generated. Tables 3.1 and 3.2 present the estimates and confidence intervals for one replication, along with the true location parameters.

Table 3.1: Simulated Estimators Under Normal Distributions. $k=3$ and $n=5$. Confidence intervals are indicated in parentheses.

<table>
<thead>
<tr>
<th>True Parameter</th>
<th>Sample Mean</th>
<th>Shrinkage Median</th>
<th>Np</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>0.220</td>
<td>0.548</td>
<td></td>
<td>0.592</td>
</tr>
<tr>
<td></td>
<td>(-0.435,0.876)</td>
<td>(-0.846,0.837)</td>
<td></td>
<td>(-0.145,0.845)</td>
</tr>
<tr>
<td>1.0</td>
<td>0.586</td>
<td>0.781</td>
<td></td>
<td>0.844</td>
</tr>
<tr>
<td></td>
<td>(-0.066,1.240)</td>
<td>(-0.242,1.260)</td>
<td></td>
<td>(-0.556, 1.253)</td>
</tr>
<tr>
<td>1.5</td>
<td>1.235</td>
<td>1.279</td>
<td></td>
<td>1.253</td>
</tr>
<tr>
<td></td>
<td>(0.301,2.169)</td>
<td>(-0.226,2.052)</td>
<td></td>
<td>(-0.555, 2.040)</td>
</tr>
</tbody>
</table>

Table 3.2: Simulated Estimators Under Laplace Distributions. $k=3$ and $n=5$. Confidence intervals are indicated in parentheses.

<table>
<thead>
<tr>
<th>True Parameter</th>
<th>Sample Mean</th>
<th>Shrinkage Median</th>
<th>Np</th>
<th>Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>-0.028</td>
<td>0.464</td>
<td></td>
<td>1.101</td>
</tr>
<tr>
<td></td>
<td>(-1.385,1.328)</td>
<td>(-1.970,1.289)</td>
<td></td>
<td>(-1.033,1.296)</td>
</tr>
<tr>
<td>1.0</td>
<td>0.993</td>
<td>1.003</td>
<td></td>
<td>1.135</td>
</tr>
<tr>
<td></td>
<td>(0.124,1.863)</td>
<td>(-0.374,1.260)</td>
<td></td>
<td>(-0.380, 2.184)</td>
</tr>
<tr>
<td>1.5</td>
<td>2.689</td>
<td>2.398</td>
<td></td>
<td>2.358</td>
</tr>
<tr>
<td></td>
<td>(1.767,3.611)</td>
<td>(0.928,3.618)</td>
<td></td>
<td>(1.101, 3.618)</td>
</tr>
</tbody>
</table>

Since the results are only based on one replication, we cannot make definitive comparisons of the performance of the different estimators. However, from the tables, one could see that the shrinkage estimates and their confidence intervals estimate the true parameter reasonably well under both normal and heavy tailed distribution. In the
next section we present the results of simulation studies to compare the performance of the different estimators in terms of average squared errors.

3.7 Simulated Comparisons of the Estimators

The performance of our method was examined when the true distribution is Normal, Logistic, Laplace, and Cauchy. At each iteration, different true location parameters were generated randomly from a standard uniform. The mean squared errors (MSEs) are presented in Table 3.3, when \( K = 10, n = 11 \). We note that the presented MSEs are not an estimator of the risk, since the data were generated under different true parameters in each iteration.

To examine the effect of shrinkage, the relative efficiency was calculated with respect to the sample median via, for example,

\[
\text{RelEff} = \frac{\text{MSE}(\text{Shrinkage median})}{\text{MSE}(\text{Sample median})} \times 100%.
\]

The relative efficiency presents the ratio of efficiencies of the two methods, where the efficiency here represents the performance of an estimator in terms of the estimated risk. The relative efficiencies are presented in Table 3.4. Based on the simulation results, the shrinkage median estimator dominates the sample median in all cases, in terms of mean squared error. The shrinkage median estimator is also the best when error distribution is Laplace, and the nonparametric Bayes estimator is the best under Logistic and Cauchy distribution. The improvement of using robust estimators over JS estimator is most significant under the Cauchy distribution.

As mentioned earlier, the shrinkage coefficient \( c \) goes to 1 as sample size increases. Therefore, the improvement of using shrinkage is more significant in 'large \( p \), small \( n \)' cases. Figure 3.1 captures the effect of shrinkage as sample size increases, under different distributions.
Table 3.3: Simulated MSE of Estimators Under Different Distributions. These are based on 1000 simulation replications, $k=10$ and $n=15$. Standard errors are indicated in parentheses.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Sample Mean</th>
<th>JS Mean</th>
<th>Sample Median</th>
<th>Shrinkage Median</th>
<th>Np Bayesian</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>0.639</td>
<td>0.566</td>
<td>0.987</td>
<td>0.802</td>
<td>0.900</td>
<td>0.776</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.008)</td>
<td>(0.014)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.0021)</td>
</tr>
<tr>
<td>Logistic</td>
<td>2.165</td>
<td>1.801</td>
<td>2.695</td>
<td>1.624</td>
<td>1.592</td>
<td>0.643</td>
</tr>
<tr>
<td></td>
<td>(0.030)</td>
<td>(0.027)</td>
<td>(0.039)</td>
<td>(0.022)</td>
<td>(0.021)</td>
<td>(0.0032)</td>
</tr>
<tr>
<td>Laplace</td>
<td>1.383</td>
<td>1.102</td>
<td>1.006</td>
<td>0.756</td>
<td>0.761</td>
<td>0.722</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.012)</td>
<td>(0.010)</td>
<td>(0.0025)</td>
</tr>
<tr>
<td>Cauchy</td>
<td>2.532e+04</td>
<td>2.532e+04</td>
<td>2.099</td>
<td>1.300</td>
<td>1.262</td>
<td>0.612</td>
</tr>
<tr>
<td></td>
<td>(8.89e+03)</td>
<td>(8.89e+03)</td>
<td>(0.039)</td>
<td>(0.020)</td>
<td>(0.018)</td>
<td>(0.0035)</td>
</tr>
</tbody>
</table>

Table 3.4: Relative Efficiency (in %) of the Estimators Under Different Distributions. The relative efficiency is calculated with respect to the sample median.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Sample Mean</th>
<th>JS Mean</th>
<th>Sample Median</th>
<th>Shrinkage Median</th>
<th>Np Bayesian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>154.46</td>
<td>174.38</td>
<td>100</td>
<td>123.07</td>
<td>109.67</td>
</tr>
<tr>
<td>Logistic</td>
<td>124.48</td>
<td>149.64</td>
<td>100</td>
<td>165.95</td>
<td>169.28</td>
</tr>
<tr>
<td>Laplace</td>
<td>72.74</td>
<td>91.29</td>
<td>100</td>
<td>133.07</td>
<td>132.19</td>
</tr>
<tr>
<td>Cauchy</td>
<td>0.0079</td>
<td>0.0079</td>
<td>100</td>
<td>161.46</td>
<td>166.32</td>
</tr>
</tbody>
</table>

3.8 Application to Stock Return Data

In this section, we apply our non-parametric shrinkage method to stock return data. Daily trading prices of 29 Dow Jones companies (we excluded Visa Inc. since its initial public offering was in 2008) were collected in 2683 consecutive business days, from January 3rd, 2005 to August 28th, 2015. In the context of risk management and portfolio allocation, it is usually important to first understand the distribution of the returns of financial assets. Thus, the variable of interest here is the daily stock return, which is calculated as the percentage increase in the closing price from the previous day’s closing price, i.e. the return of day $i$ is calculated by

$$\text{return}_i = \frac{\text{Closing}_i - \text{Closing}_{i-1}}{\text{Closing}_{i-1}}.$$
Figure 3.1 The Effect of Shrinkage v.s. Sample Size. Simulation based on 1000 Replications, p=10. The relative efficiency is calculated with respect to the sample median.

A portion of the data is presented in Table 3.5.

The goal is to make predictions about future returns of all stocks simultaneously based on the return data from a previous time window. Many financial models are developed under the assumption of normal returns. Therefore, an intuitive way is to look at the moving averages for each stock individually. Using the idea presented in previous sections, we expect to improve this prediction by shrinking the averages.
Table 3.5: Daily Stock Return of AAPL, AXP, BA, CAT and CSCO for the First 10 Consecutive Trading Days Starting from January 3rd.

<table>
<thead>
<tr>
<th></th>
<th>AAPL</th>
<th>AXP</th>
<th>BA</th>
<th>CAT</th>
<th>CSCO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.02705806</td>
<td>-1.5024064</td>
<td>-1.9423217</td>
<td>-1.1044527</td>
<td>-3.93374692</td>
</tr>
<tr>
<td>2</td>
<td>0.87582135</td>
<td>-0.9281133</td>
<td>1.6606681</td>
<td>-1.9144796</td>
<td>0.05388367</td>
</tr>
<tr>
<td>3</td>
<td>0.07748127</td>
<td>-0.5143170</td>
<td>-0.6494827</td>
<td>1.4747350</td>
<td>1.50780962</td>
</tr>
<tr>
<td>4</td>
<td>7.28131155</td>
<td>-0.7385547</td>
<td>-0.3367623</td>
<td>-0.2137263</td>
<td>-0.68966341</td>
</tr>
<tr>
<td>5</td>
<td>-0.41888197</td>
<td>0.2046050</td>
<td>1.3317404</td>
<td>-0.9209692</td>
<td>0.00000000</td>
</tr>
<tr>
<td>6</td>
<td>-6.38041456</td>
<td>-0.9281558</td>
<td>-0.3138500</td>
<td>-1.3834832</td>
<td>-0.42734912</td>
</tr>
<tr>
<td>7</td>
<td>1.39394534</td>
<td>-0.2248587</td>
<td>2.2038544</td>
<td>1.2384977</td>
<td>1.55579974</td>
</tr>
<tr>
<td>8</td>
<td>6.63000674</td>
<td>-0.6948341</td>
<td>-2.5221342</td>
<td>0.5304718</td>
<td>-0.79239203</td>
</tr>
<tr>
<td>9</td>
<td>0.57321328</td>
<td>-0.3025500</td>
<td>0.5530293</td>
<td>0.8938217</td>
<td>0.53247671</td>
</tr>
<tr>
<td>10</td>
<td>0.64098695</td>
<td>0.8156360</td>
<td>1.9053231</td>
<td>-0.8040340</td>
<td>-0.74152610</td>
</tr>
</tbody>
</table>

accordingly using the James-Stein estimator. Moreover, empirical evidences have led to attention on the tail behavior of stock returns and it is now considered that the returns possess heavy-tailed distributions. The Q-Q plot of the stock returns of AAPL is presented in Figure 3.2 as an example. In heavy tailed distributions, we expect the nonparametric shrinkage method to work better than the James-Stein shrinkage estimator. To make prediction about stock returns on the next business day, returns from previous 5 or 15 trading days were used as observed data. Table 3.6 presents the prediction result using different methods. The squared errors were calculated combining all 29 stocks, and the mean and standard deviations of squared errors were calculated, for the different methods. The shrinkage estimators perform better in general, and the shrinkage median estimator has the smallest prediction mean square error, followed by the nonparametric Bayes estimator. Although the improvement is only around 1% to 2% compared with the James-Stein shrinkage estimator, it might still be of great financial interest, given the heavy cash flow in the stock market.
Figure 3.2  Histogram and QQ plot of Apple Stock Returns for 2683 Consecutive Business Days, from January 2005 to August, 2015. The returns seem to follow a heavy-tailed distribution.

Table 3.6: Prediction MSEs of Different Location Estimators Using Daily Dow Jones Stock Return Data. Standard errors are indicated in parentheses.

<table>
<thead>
<tr>
<th>Lag(days)</th>
<th>Sample JS Mean</th>
<th>Sample JS Median</th>
<th>Sample Shrinkage Mean</th>
<th>Sample Shrinkage Median</th>
<th>Np Bayesian Mean</th>
<th>Np Bayesian Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>3.795</td>
<td>3.561</td>
<td>3.982</td>
<td>3.486</td>
<td>3.488</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.208)</td>
<td>(0.206)</td>
<td>(0.211)</td>
<td>(0.193)</td>
<td>(0.198)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.170)</td>
<td>(0.169)</td>
<td>(0.177)</td>
<td>(0.166)</td>
<td>(0.173)</td>
<td></td>
</tr>
</tbody>
</table>

3.9 Summary

In this chapter, we constructed and compared the nonparametric estimators of location using the idea of shrinkage. Confidence intervals of the location parameter were also provided using bootstrap approach. Based on simulation results, the shrinkage estimator dominates the standard ones in all cases. The shrinkage median estimator and the nonparametric Bayesian estimator are robust under unknown distributions as compared to the James-Stein estimator. The shrinkage effect is more significant when sample size is small, and when the true location parameters are close to each
other. We did see improvements of using the shrinkage median in the stock return application. Though the improvements are small, it might still be of economic interest.

The ultimate goal of predicting asset returns is to allocate assets in a way that maximize the expected portfolio return as well as control the risk (20). Since the shrinkage method provides more accurate predictions of asset returns, it might be useful to incorporate the predictions in the portfolio allocation to achieve a higher portfolio return.

The efficient-market hypothesis states that asset prices fully reflect all available information, which makes it impossible for investors to outperform the market through market timing or portfolio selection. However, the application in section 3.8 shows different prediction results when different prediction methods were used. The shrinkage methods performs better in prediction when past stock return data are used, which indicates the possibility of exploiting information from the past to get a better sense of current stock price and to come up with a better trading strategy.
Chapter 4

Improved Estimation of System Reliability

Through Shrinkage Idea in Parametric Settings


4.1 Reliability Concepts

Detailed discussions of concepts and ideas presented below are available in (3) and (17). Consider a component with a lifetime, denoted by \( T \), measured in some unit of ‘time’. Usually, the lifetime will be measured in literal time, but it need not always be the case. Such a \( T \) is a non-negative random variable. The reliability function of this component is defined via

\[
R(t) = 1 - F(t) = \Pr\{T > t\},
\]

where \( F(\cdot) \) is the corresponding distribution function. We assume that lifetime variables are continuous. For \( T \), its probability density function (pdf) is

\[
f(t) = \frac{dF(t)}{dt} = -\frac{dR(t)}{dt}.
\]

Its hazard rate function \( \lambda(t) \) is defined as

\[
\lambda(t) \equiv \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \Pr\{t \leq T < t + \Delta t | T \geq t\} = \frac{f(t)}{R(t)}.
\]

which can be interpreted as the rate of failure of the component at time \( t \), given that the component is still working just before time \( t \). Given the hazard rate function, the
cumulative hazard function is
\[ \Lambda(t) = \int_0^t \lambda(v)dv. \]

For a continuous lifetime \( T \), we have the following relationships:
\[ f(t) = \lambda(t) \exp\{-\Lambda(t)\} \quad \text{and} \quad R(t) = \exp\{-\Lambda(t)\} \quad (4.1) \]

For a component with lifetime \( T \), its associated state process is \( \{X(t) : t \geq 0\} \), where \( X(t) = I\{T > t\} \) is a binary variable taking values of 1 or 0 depending on whether the component is still working (1) or failed (0) at time \( t \). The function \( I(\cdot) \) denotes indicator function.

Consider a system composed of \( K \) components, where this system is either in a working (1) or failed (0) state. The functionality of a system is characterized by its structure function
\[ \phi : \{0, 1\}^K \rightarrow \{0, 1\}, \]
with \( \phi(x_1, x_2, \cdots, x_K) \) denoting the state of the system when the states of the components are \( \mathbf{x} = (x_1, x_2, \cdots, x_K) \in \{0, 1\}^K \). The vector \( \mathbf{x} \) is called the component state vector. Such a system is said to be coherent if each component is relevant and the structure function \( \phi \) is nondecreasing in each argument. The \( i \)th component is relevant if there exists a state vector \( \mathbf{x} \in \{0, 1\}^K \) such that \( \phi(\mathbf{x}, 0_i) = 0 \leq 1 = \phi(\mathbf{x}, 1_i) \), with the notation that \( (\mathbf{x}, a_i) = (x_1, \ldots, x_{i-1}, a_i, x_{i+1}, \ldots, x_n) \). We will only consider coherent systems in this chapter. Four simple examples of coherent systems are the (i) series; (ii) parallel; (iii) three-component series-parallel; and (iv) five-component bridge systems, whose respective structure functions are given by
\[ \phi_{\text{ser}}(x_1, \ldots, x_K) = \prod_{i=1}^K x_i; \quad (4.2) \]
\[ \phi_{\text{par}}(x_1, \ldots, x_K) = \prod_{i=1}^K x_i \equiv 1 - \prod_{i=1}^K (1 - x_i); \quad (4.3) \]
\[ \phi_{\text{serpar}}(x_1, x_2, x_3) = x_1(x_2 \lor x_3); \quad (4.4) \]
\[ \phi_{\text{br}}(x_1, x_2, x_3, x_4, x_5) = (x_1x_3x_5) \lor (x_2x_3x_4) \lor (x_1x_4) \lor (x_2x_5). \quad (4.5) \]
The binary operator ‘∨’ means taking the maximum, i.e. \( a_1 \lor a_2 = \max(a_1, a_2) = 1 - (1 - a_1)(1 - a_2) \) for \( a_i \in \{0, 1\}, i = 1, 2 \).

Let \( X_i, i = 1, \ldots, K, \) be the state (at a given point in time) random variables for the \( K \) components, and assume that they are independent. Denote by \( p_i = \Pr\{X_i = 1\}, i = 1, \ldots, K, \) and let \( p = (p_1, p_2, \ldots, p_K) \in [0, 1]^K \) be the components reliability vector (at a given point in time). Associated with the coherent structure function \( \phi \) is the reliability function defined via

\[
h_\phi(p) = E[\phi(X)] = \Pr\{\phi(X) = 1\}.
\]

This reliability function provides the probability that the system is functioning, at the given point in time, when the component reliabilities at this time are \( p_i \)'s. For the first three concrete systems given above, these reliability functions are, respectively:

\[
h_{\text{ser}}(p_1, \ldots, p_K) = \prod_{i=1}^{K} p_i; \quad (4.6)
\]

\[
h_{\text{par}}(p_1, \ldots, p_K) = \prod_{i=1}^{K} p_i \equiv 1 - \prod_{i=1}^{K} (1 - p_i); \quad (4.7)
\]

\[
h_{\text{ser-par}}(p_1, p_2, p_3) = p_1[1 - (1 - p_2)(1 - p_3)]; \quad (4.8)
\]

For the bridge structure, its reliability function at a given point in time, obtained first by simplifying the structure function, is given by

\[
h_{\text{br}}(p_1, p_2, p_3, p_4, p_5) = (p_1 p_4 + p_2 p_5 + p_2 p_3 p_4 + p_1 p_3 p_5 + 2 p_1 p_2 p_3 p_4 p_5)
\]

\[-(p_1 p_2 p_3 p_4 + p_2 p_3 p_4 p_5 + p_1 p_3 p_4 p_5 + p_1 p_2 p_3 p_5 + p_1 p_2 p_4 p_5). \quad (4.9)
\]

Of more interest, however, is viewing the system reliability function as a function of time \( t \). Denoting by \( S \) the lifetime of the system, we are interested in the function

\[
R_S(t) = \Pr\{S > t\}
\]

which is the probability that the system does not fail in \([0, t]\). Let \( T = (T_1, \cdots, T_K) \) be the vector of lifetimes of the \( K \) components. The vector of component state
processes is \( \{X(t) = (X_1(t) \cdots , X_K(t)) : t \geq 0\} \). The system lifetime is then
\[
S = \sup \{t \geq 0 : \phi[X_1(t), \cdots , X_K(t)] = 1\}.
\]
The component reliability functions are \( R_i(t) = E[X_i(t)] = \Pr\{T_i > t\}, i = 1, \ldots , K. \) If the component lifetimes are independent, then the system reliability function becomes
\[
R_S(t) = E[\phi(X_1(t), \ldots , X_K(t))] = h_\phi(R_1(t), \ldots , R_K(t)). \tag{4.10}
\]
That is, under independent component lifetimes, to obtain the system reliability function, we simply replace the \( p_i \)'s in the reliability function \( h_\phi(p_1, \ldots , p_K) \) by \( R_i(t) \)'s.

For the concrete examples of coherent systems given in (4.2–4.5), we therefore obtain:
\[
R_{\text{ser}}(t) = \prod_{i=1}^{K} R_i(t); \tag{4.11}
\]
\[
R_{\text{par}}(t) = 1 - \prod_{i=1}^{K} (1 - R_i(t)); \tag{4.12}
\]
\[
R_{\text{serpar}}(t) = R_1(t)[1 - (1 - R_2(t))(1 - R_3(t))]. \tag{4.13}
\]
For the bridge structure, in (4.9), we replace each \( p_i \) by \( R_i(t) \) to obtain its system reliability function. As an illustration of these system reliability functions for the four examples of coherent systems, with \( R_i(t) = \exp(-\lambda t), i = 1, \ldots , K, \) the system reliability functions are
\[
R_{\text{ser}}(t; \lambda, K) = \exp(-K\lambda t);
\]
\[
R_{\text{par}}(t; \lambda, K) = 1 - [1 - \exp(-\lambda t)]^K;
\]
\[
R_{\text{serpar}}(t; \lambda) = \exp(-\lambda t)[1 - (1 - \exp(-\lambda t))^2];
\]
\[
R_{\text{br}}(t; \lambda) = 2 \exp\{-2\lambda t\} + 2 \exp\{-3\lambda t\} + 2 \exp\{-5\lambda t\} - 5 \exp\{-4\lambda t\}.
\]
When \( \lambda = 1 \) and \( K = 5 \), these system reliability functions are plotted in Figure 4.1.

An important concept in reliability is measuring the relative importance of each of the components in the system. There are several possible measures of component
importance (cf., (3)). We focus on the so-called reliability importance measure as this will play an important role in the improved estimation of the system reliability. The reliability importance of component $j$ in a $K$-component system with reliability function $h_\phi(\cdot)$ is

$$I_\phi(j; p) = \frac{\partial h_\phi(p_1, \cdots, p_{j-1}, p_j, p_{j+1}, \cdots, p_K)}{\partial p_j} = h_\phi(p, 1_j) - h_\phi(p, 0_j).$$  (4.14)

This measures how much system reliability changes when the reliability of component $j$ changes, with the reliabilities of the other components remaining the same. For a coherent system, the reliability importance of a component is positive. As examples, the reliability importance of the $j$th component in a series system is

$$I_{ser}(j; p) = \frac{h_{ser}(p)}{p_j}, j = 1, \ldots, K,$$

showing that in a series system the weakest (least reliable) component is the most important ("the system is as good as its weakest link"). For a parallel system, the

Figure 4.1 System reliability functions for the series, parallel, series-parallel, and bridge systems when the components have common unit exponential lifetimes and there are 5 components in the series and parallel systems.
reliability importance of the $j$th component is

$$I_{\text{par}}(j; p) = \frac{1 - h_{\text{par}}(p)}{1 - p_j}, j = 1, \ldots, K,$$

indicating that the most reliable component is the most important component in a parallel system. For the 3-component series-parallel system, the reliability importance of the three components are

$$I_{\text{serpar}}(1; p) = 1 - (1 - p_2)(1 - p_3);$$

$$I_{\text{serpar}}(2; p) = p_1(1 - p_3);$$

$$I_{\text{serpar}}(3; p) = p_1(1 - p_2).$$

Evaluated at $p = (p, p, p)$, they become $I_{\text{serpar}}(1; p) = p(2 - p)$ and $I_{\text{serpar}}(2; p) = I_{\text{serpar}}(3; p) = p(1 - p)$, which confirms the intuitive result that when the components are equally reliable, the component in series (component 1) is the most important component. In general, however, component 1 is not always the most important. For instance, if components 2 and 3 are equally reliable with reliability $p_2$, then the reliability importance of components 1, 2 and 3 become

$$I_{\text{serpar}}(1; (p_1, p_2, p_2)) = p_2(2 - p_2);$$

$$I_{\text{serpar}}(2; (p_1, p_2, p_2)) = I_{\text{serpar}}(3; (p_1, p_2, p_2)) = p_1(1 - p_2).$$

In this case, component 2 (and 3) is more important than component 1 whenever

$$p_1(1 - p_2) > p_2(2 - p_2),$$

or equivalently,

$$p_1 > \frac{p_2(2 - p_2)}{1 - p_2}.$$
4.2 Statistical Model and Data Structure

We now consider the problem of estimating the system reliability function on the basis of observed data from the system or its components. We suppose that the system has $K$ components and the observed data will be the $K$ components time-to-failures, but which could be right-censored by the system lifetime or the end of monitoring period.

We let $T_j$ denote the time-to-failure of component $j$, and we assume that the $T_j$’s are independent of each other. We denote by $R_j(\cdot; \theta_j)$ the reliability function of $T_j$, where $\theta_j \in \Theta_j$ with the parameter space $\Theta_j$ an open subset of $\mathbb{R}_{m_j}$. Furthermore, note that it is possible that there could be common parameters among the $(\theta_j, j = 1, 2, \ldots, K)$.

We also assume that the system structure function $\phi(\cdot)$ is known, hence we also know the reliability function $h_\phi(p)$. The system reliability function $R_S(\cdot)$ can therefore be expressed via

$$R_S(t) = R_S(t; \theta_1, \theta_2, \ldots, \theta_K)$$
$$= h_\phi[R_1(t; \theta_1), R_2(t; \theta_2), \ldots, R_K(t; \theta_K)]. \quad (4.15)$$

Suppose there are $n$ identical systems, so that the observable component lifetimes are

$$\{T_{ij} : i = 1, 2, \ldots, n; j = 1, 2, \ldots, K\}.$$

We assume that the $T_{ij}$’s are independent, and that for each $j$, $(T_{ij}, i = 1, 2, \ldots, n)$ are identically-distributed with reliability function $R_j(\cdot; \theta_j)$. However, in practice, the exact values of the $T_{ij}$’s are not all observable. Rather, they could be right-censored by either the system life or by the monitoring period. The observable right-censored data is

$$D = \{(Z_{ij}, \delta_{ij}) : i = 1, 2, \ldots, n; j = 1, 2, \ldots, K\} \quad (4.16)$$

where $\delta_{ij} = 1$ means that $T_{ij} = Z_{ij}$, whereas $\delta_{ij} = 0$ means that $T_{ij} > Z_{ij}$. We shall suppose that for each $i \in \{1, 2, \ldots, n\}$, we have a random variable $C_i$ (e.g., the upper
bound of monitoring time for the \( i \)th system) and also the system life \( S_i \), and

\[
Z_{ij} = \min\{T_{ij}, \min(C_i, S_i)\} \quad \text{and} \quad \delta_{ij} = 1\{T_{ij} \leq \min(C_i, S_i)\}.
\]

On the basis of this observable data \( D \), it is of interest to estimate the system reliability function given in (4.15). A simple estimator of the system reliability function is to utilize only the observed system lifetimes, the \( S_i \)'s. However, these system lives may be right-censored by the \( C_i \)'s, so that we may only be able to observe \( Z_i = \min(S_i, C_i) \) and \( \delta_i = 1\{S_i \leq C_i\} \) for \( i = 1, 2, \ldots, n \). If the component lifetime distributions are governed by just one parameter vector, then a parametric approach to estimating this parameter may be possible, for example when all the component lifetime distributions are all exponential with rate parameter \( \lambda \). When component level data is available, we could improve the estimation of system reliability by utilizing the information on the internal structure. The lifetime of components are independent with reliability function

\[
R_j(t; \theta_j) = 1 - F_j(t; \theta_j), \quad j = 1 \cdots K.
\]

Classically, the estimator of the system reliability function, \( R_S(t) \), is given by

\[
\hat{R}_S(t) = h_\phi[\hat{R}_1(t), \cdots, \hat{R}_K(t)],
\]

where \( \hat{R}_j(\cdot) \) is the MLE of component reliability based on \( (Z_j, \delta_j) = \{(Z_{ij}, \delta_{ij}) : i = 1, 2, \ldots, n\} \). In the parametric setting, \( \hat{R}_j(t) = R_j(t; \hat{\theta}_j) \), where \( \hat{\theta}_j \) is the MLE of \( \theta_j \).

For component \( j \), to find \( \hat{\theta}_j \), denote the density associated with component \( j \) by \( f_j(t; \theta_j) \), so the likelihood function based on the completely observed lifetimes of component \( j \) is given by

\[
L(\theta_j | t_{1j}, \ldots, t_{nj}) = \prod_{i=1}^{n} f_j(t_{ij} | \theta_j).
\]

In the presence of right-censoring, the likelihood function based on the observed censored data for component \( j \) becomes

\[
L_j(\theta_j; (Z_{ij}, \delta_{ij})) = \prod_{i=1}^{n} f(z_{ij} | \theta_j)^{\delta_{ij}} R(z_{ij} | \theta_j)^{1-\delta_{ij}}.
\]
This likelihood could be maximized with respect to $\theta_j$ to obtain the ML estimate $\hat{\theta}_j$, which will be a function of the $(z_{ij}, \delta_{ij}), i = 1, 2, \ldots, n$. The resulting system reliability function estimator is

$$\tilde{R}_S(t) = h_\phi[R_1(t; \hat{\theta}_1), R_2(t; \hat{\theta}_2), \ldots, R_K(t; \hat{\theta}_K)].$$  \hspace{1cm} (4.17)$$

From the theory of ML estimators for right-censored data (cf., (2)) under parametric models, as $n \to \infty$, we have that

$$\hat{\theta}_1, \ldots, \hat{\theta}_K \sim AN\left((\theta_1, \ldots, \theta_K), \frac{1}{n}BD\left[I_1^{-1}, \ldots, I_K^{-1}\right]\right)$$  \hspace{1cm} (4.18)$$

where $I_j^{-1}$ is the inverse of the Fisher information matrix for the $j$th component, and $BD$ means ‘block diagonal’. Assuming no common parameters among the $K$ components, by using the Delta-Method, we find that, as $n \to \infty$,

$$\tilde{R}_S(t) \sim AN \left[R_S(t), \frac{1}{n} \sum_{j=1}^K I_\phi(j; t) \cdot \hat{R}_j(t) I_j^{-1} \cdot \hat{R}_j(t) I_\phi(j; t)\right]$$

with

$$\hat{R}_j(t) = \frac{\partial}{\partial \theta_j} R(t; \theta_j).$$

In principle, this asymptotic variance could be estimated, though the difficulty may depend on the structure function and/or distributional form of the $R_j$’s. Later we will instead perform comparisons through numerical simulations.

### 4.3 Improved Estimation Through Shrinkage Idea

Suppose that the component level data are available. The estimation of the parameters, $(\theta_j, j = 1, 2, \ldots, K)$, becomes a problem of simultaneous estimation. In this context, the problem is to estimate simultaneously the reliability of all components in the system. Thus, we propose an improved estimator following the idea of James and Stein(13). Consider estimators of component reliabilities of the form

$$\tilde{R}_j(t) \equiv \tilde{R}_j(t; c) = [\hat{R}_j(t)]^c, j = 1, \ldots, K, c > 0,$$
where \( \hat{R}_j(t) = R_i(t, \hat{\theta}_j) \) is the ML estimator of \( R_j(t) \) based on \((Z_j, \delta_j)\) under the assumed parametric model. The system reliability estimator then becomes

\[
\hat{R}_S(t; c) = h_{\phi}[[R_1(t; \hat{\theta}_1)]^c, \ldots, [R_K(t; \hat{\theta}_K)]^c].
\]

Notice that when \( c = 1 \), we obtain the standard ML estimator discussed in the preceding subsection. If \( \hat{\Lambda}_j(t) \) denotes the estimator of the cumulative hazard function for component \( j \), according to the relationship mentioned in (4.1), we have

\[
[\hat{R}_j(t)]^c = \exp[-c\hat{\Lambda}_j(t)].
\]

Thus, we are essentially putting a shrinkage coefficient on the ML estimators of the cumulative hazard functions. We remark at this point that it is not always the case that the optimal \( c^* \) is less than 1, so that in some cases, instead of shrinking, we are expanding the estimators!

The goal is to find the optimal \( c \) in terms of a global risk function for the system reliability estimator. If the optimal shrinkage coefficient is \( c^* \), the improved estimator of system reliability becomes

\[
\hat{R}_S(t; c^*) = h_{\phi}[[R_1(t; \hat{\theta}_1)]^{c^*}, \ldots, [R_K(t; \hat{\theta}_K)]^{c^*}].
\]

However, the optimal \( c^* \) may depend on the unknown parameters \((\theta_1, \cdots, \theta_K)\). Therefore, we also need to find an estimator of \( c^* \), denoted by \( \hat{c}^* \). An intuitive and simple way to obtain an estimator of \( c^* \) is to simply replace the unknown \( \theta_j \)'s in the \( c^* \) expression with their corresponding MLEs. The final estimator becomes

\[
\tilde{R}_S(t) = h_{\phi}[[R_1(t; \hat{\theta}_1)]^{\hat{c}^*}, \cdots, [R_K(t; \hat{\theta}_K)]^{\hat{c}^*}].
\] (4.19)
4.4 Evaluating the Estimators

From a decision-theoretic viewpoint, the performance of an estimator \( a(\cdot) \) of \( R(\cdot) = R_S(\cdot) \) will be evaluated under the following class of loss functions,

\[
L(R, a) = - \int |R(t) - a(t)|^k h(R(t))dR(t).
\] (4.20)

where \( h(\cdot) \) is a positive function. Note that the negative sign is because the differential element \( dR(t) \) is negative since \( R(\cdot) \) is a non-increasing functions. The notation \(-dR(t) = dF(t)\) will be used interchangeably, where \( F \) is the associated distribution function. The decision problem of estimating \( R_S \), when using this class of loss function, becomes invariant with respect to the group of monotone increasing transformations on the lifetimes. A special member of this class of loss functions is the weighted Cramer-von Mises loss function given by

\[
L(R, a) = - \int \frac{[R(t) - a(t)]^2}{R(t)(1 - R(t))}dR(t).
\]

This loss function is a global loss function and can be viewed as the weighted squared losses aggregated (integrated) over time.

The expected loss or risk function is then given by

\[
Risk(R, a) = EL(R, a) = \mathbb{E} \left[ - \int \frac{[R(t) - a(t)]^2}{R(t)(1 - R(t))}dR(t) \right],
\] (4.21)

with the expectation taken with respect to the random elements in the estimator \( a(t) \). To find the optimal coefficient \( c \), \( \hat{R}_S(t; c) \) is plugged into the risk function and then the risk is minimized with respect to \( c \). To simplify notation, \( Risk(c) \) is used to denote the risk with respect to \( c \).
To demonstrate the existence of an optimal $c$, notice that as $c \to \infty$,

$$
Risk(c) = E \left[ \int \frac{h \phi([R_1(t; \hat{\theta}_1)]^c, \ldots, [R_K(t; \hat{\theta}_K)]^c) - R(t)^2}{R(t)(1 - R(t))} dF(t) \right]
$$

$$
\to E \left[ \int \frac{[0 - R(t)]^2}{R(t)(1 - R(t))} dF(t) \right]
$$

$$
= \int \frac{R(t)}{1 - R(t)} dF(t)
$$

$$
= \infty.
$$

Similarly, when $c \to 0$,

$$
Risk(c) \to \int \frac{1 - R(t)}{R(t)} dF(t) = \infty.
$$

Therefore, since for every finite $c > 0$, $Risk(c)$ is finite, and since $Risk(c)$ is a continuous function of $c$, there exists a value of $c$, denoted by $c^* \in [0, \infty)$, that minimizes the risk function $Risk(c)$.

Using properties of maximum likelihood estimators, as $n \to \infty$, $\hat{\theta}_j$ converges in probability to $\theta_j$, $j = 1, \cdots, K$. Thus, when $n \to \infty$, the optimal $c^*$ that minimizes the expected loss converges in probability to $c^* = 1$. Therefore, for large $n$, the improved estimator becomes probabilistically close to the ML estimator.

### 4.5 Estimating the Optimal $c^*$

Given the form of the expected loss function, to find the optimal shrinkage coefficient $c$, some approximations are used based on the asymptotic properties of maximum likelihood estimators and Taylor expansion. A first-order Taylor expansion on the estimated system reliability function $\hat{R}_S(t; c)$ at $(\hat{R}_1(t) = R_1(t), \cdots, \hat{R}_K(t) = R_K(t))$ gives

$$
\hat{R}_S(t; c) \approx h \phi[R_1(t), \cdots, R_k(t)] + \sum_{j=1}^{K} I \phi[j, R_1(t), \cdots, R_k(t)][\hat{R}_j(t) - R_j(t)]
$$

where $\hat{R}_j(t)$’s are independent given $c$, and $I \phi(j, R_1(t), \cdots, R_k(t))$ denotes the reliability importance of component $j$ under system structure $\phi$, as defined in (4.14). The
loss function can now be written as

\[
L(R, \tilde{R}) = \int_0^\infty \frac{[R(t) - \tilde{R}(t)]^2}{R(t)(1 - R(t))} dF(t)
\]

\[
\approx \int_0^\infty \frac{\sum_{j=1}^K I_\phi(j, t)[\tilde{R}_j(t) - R_j(t)]^2}{R(t)(1 - R(t))} dF(t)
\]

\[
= \int_0^\infty \sum_{j=1}^K \frac{I_\phi^2(j, t)[\tilde{R}_j(t) - R_j(t)]}{R(t)(1 - R(t))} dF(t)
\]

\[
+ \int_0^\infty \sum_{j \neq k} I_\phi(j, t)I_\phi(k, t)[\tilde{R}_j(t) - R_j(t)][\tilde{R}_k(t) - R_k(t)] dF(t).
\]

Take expectation with respect to \( \tilde{R}_j(t) \)'s to evaluate the expected loss. First examine the following expectations:

\[
E[\tilde{R}_j(t) - R_j(t)]^2 = \text{Var}[\tilde{R}_j(t)^c] + [R_j(t)^c - R_j(t)]^2
\]

Employing the Delta-Method and the asymptotic properties of MLEs, \( \text{Var}[\tilde{R}_j(t)^c] \) can be asymptotically approximated by \( c^2 R_j^{2c - 2}(t) \text{Var}[\tilde{R}_j(t)] \). Thus,

\[
E[\tilde{R}_j(t) - R_j(t)]^2 \approx c^2 R_j^{2c - 2}(t) V_j(t) + [R_j(t)^c - R_j(t)]^2,
\]

where \( V_j(t) = \text{Var}[\tilde{R}_j(t)] \). And since the components are independent,

\[
E[\tilde{R}_j(t) - R_j(t)][\tilde{R}_k(t) - R_k(t)]
\]

\[
= \text{Cov}[\tilde{R}_j(t)^c, \tilde{R}_k(t)^c] + [R_j(t)^c - R_j(t)][R_k(t)^c - R_k(t)]
\]

\[
= [R_j(t)^c - R_j(t)][R_k(t)^c - R_k(t)].
\]

Re-arranging terms, the expected loss can be expressed as

\[
\text{Risk}(c) \approx \int_0^\infty \sum_{j=1}^K \frac{I_\phi^2(j, t)c^2 R_j^{2c - 2}(t)V_j(t) + [R_j(t)^c - R_j(t)]^2}{R(t)(1 - R(t))} dF(t)
\]

\[
+ \int_0^\infty \sum_{j \neq k} I_\phi(j, t)I_\phi(k, t)c^2 R_j^{2c - 2}(t)V_j(t) \frac{R(t)(1 - R(t))}{R(t)(1 - R(t))} dF(t)
\]

\[
= \int_0^\infty \sum_{j=1}^K \frac{I_\phi^2(j, t)c^2 R_j^{2c - 2}(t)V_j(t)}{R(t)(1 - R(t))} dF(t)
\]

\[
+ \int_0^\infty \frac{\sum_j I_\phi(j, t)[R_j(t)^c - R_j(t)]^2}{R(t)(1 - R(t))} dF(t)
\]
To minimize this approximate risk function with respect to $c$, note the approximations

$$R^c_j(t) \approx R_j(t) + (c - 1)R_j(t) \log R_j(t);$$

$$R_{2c}^j(t) = [R^2_j(t)]^c(t) \approx [R_j(t)]^2 + 2(c - 1)[R_j(t)]^2 \log R_j(t).$$

Then the approximate risk function becomes a polynomial in $c$, given by

$$Risk(c) \approx A \cdot c^2 + 2B \cdot c^2(c - 1) + D \cdot (c - 1)^2; \quad (4.22)$$

where

$$A = \int_0^\infty \sum_{j=1}^K \frac{I^2(j, t)V_j(t)}{R(t)[1 - R(t)]} dF(t) > 0;$$

$$B = \int_0^\infty \sum_{j=1}^K \frac{I^2(j, t)V_j(t) \log R_j(t)}{R(t)[1 - R(t)]} dF(t) < 0;$$

$$D = \int_0^\infty \frac{\sum_{j=1}^K I(j, t)R_j(t) \log R_j(t)^2}{R(t)[1 - R(t)]} dF(t) > 0.$$

The optimal shrinkage coefficient $c^*$ based on this approximation is the minimizer of the polynomial (4.22), which is of the form

$$c^*[R_1(t), \cdots, R_K(t)] = \frac{2B - A - D + \sqrt{(2B - A - D)^2 + 12BD}}{6B}. \quad (4.23)$$

As mentioned before, in practice, the true component reliability functions, $R_j(t)$’s, and also the $V_j(t)$’s are unknown. When estimating $c^*$ empirically, the occurrences of $R_j(t)$’s in the expressions (4.23) are replaced by their MLEs, $\hat{R}_j(t) = R_j(t, \hat{\theta}_j)$, and $V_j(t)$’s are replaced by the Var[$\hat{R}_j(t)$], for $j = 1, \cdots, K$.

Notice that at $c = 1$, $Risk(1) \approx A > 0$ and the derivative $Risk'(1) \approx 2(A + B)$. If $A + B > 0$, or equivalently

$$\int_0^\infty \sum_{j=1}^K \frac{I^2(j, t)V_j(t)[1 + \log R_j(t)]}{R(t)[1 - R(t)]} dR(t) < 0, \quad (4.24)$$

the approximated risk is increasing at $c = 1$, which means the optimal $c^*$ is below 1. On the other hand, if $A + B < 0$, the optimal $c^*$ is above 1. We also notice that we
are shrinking the reliability of each component when \( c^* > 1 \), and we are increasing the reliability of each component when \( c^* < 1 \).

From the simulation results in later sections, we see that the average value of \( c^* \) is below 1 when system structure is series, and the average value of \( c^* \) is above 1 when system structure is parallel. Here we provide some intuitive explanation of this result. For a series system, the importance function for a component \( j \) can be expressed by

\[
I_\phi(j, t) = \frac{R(t)}{R_j(t)}.
\]

Thus, the left side of 4.24 becomes

\[
\int_0^\infty R(t) \frac{1}{1-R(t)} \frac{K}{\sum_{j=1}^{K} \frac{1 + \log R_j(t)}{R_j^2(t)} V_j(t)dR(t)}.
\]

When \( 1 + \log R_j(t) < 0 \), \( R_j(t) \) is close to 0, and when \( 1 + \log R_j(t) > 0 \) \( R_j(t) \) is close to 1. And because of the denominator \( R_j^2(t) \), \( [1 + \log R_j(t)]/R_j^2(t) \) is going to have larger absolute values when it is negative. This provides some insights of why the left side (4.24) becomes negative under series system structure. On the other hand, if the system structure is parallel, the importance function for component \( j \) can be expressed by

\[
I_\phi(j, t) = \frac{1 - R(t)}{1 - R_j(t)}.
\]

Thus, the left side of 4.24 becomes

\[
\int_0^\infty \frac{1 - R(t)}{R(t)} \frac{K}{\sum_{j=1}^{K} \frac{1 + \log R_j(t)}{[1 - R_j(t)]^2} V_j(t)dR(t)}.
\]

In this case, because of the denominator \( [1 - R_j(t)]^2 \), \( [1 + \log R_j(t)]/[1 - R_j(t)]^2 \) is going to have larger absolute values when it is positive. And therefore, the left side (4.24) becomes positive under parallel system structure.

We also notice that as \( n \to \infty \), \( \hat{V}_j(t) = \text{Var}[\hat{R}_j(t)] \) converges to 0 according to the asymptotic property of MLEs. Thus, when \( n \to \infty \), both \( A \) and \( B \) converge to 0. As a consequence, for large \( n \),

\[
\text{Risk}(c) \approx D \cdot (c - 1)^2.
\]
This again demonstrates that as \( n \to \infty \), the optimal \( c \) converges to 1, i.e. the effect of \( c \) will be more significant when sample size is small.

### 4.6 Examples and Numerical Illustration

To illustrate the performance of the different estimators, we provide some numerical examples under different system structures. We generated system and component lifetime data based on series and parallel structures with 10 components. A sample of 10 systems was created, and component lifetimes were generated according to exponential distributions with different means. Figure 4.2 and Figure 4.3 present the estimates for one replication, along with the true system reliability function for the series and the parallel structures, respectively.

Since this is only based on one replication, we obviously cannot make definitive comparisons of the performance of the different estimators. However, from these plots, one could see that for the series system the estimates do track the true system reliability function well. However, for the parallel system, there appears to be a big discrepancy between the true system reliability function and the three estimates. In the next section we present the results of simulation studies to compare the performance of the different estimators of the system reliability function.

### 4.7 Simulated Comparisons of the Estimators

The advantage of the improved estimator over the standard MLE is demonstrated using simulated data. Table 4.1 and 4.2 present comparisons of the performance of the ML estimator, and the improved estimator in terms of the average loss based on 1000 replications. The simulated data were for \( K = 10 \) components each with \( n = 10 \) complete or randomly right-censored observations. Exponential and Weibull lifetime models were considered under series and parallel systems. The means of the estimated shrinkage coefficients \( c \)'s are also presented.
Figure 4.2 Estimated and True System Reliability Functions Over Time for a Series System with 10 components. The estimators are the ML based on component data and the improved estimator based on a sample of size 10 systems.

The improved estimator dominates the maximum likelihood estimator in the context of the estimated global risk. The improvement is more significant when sample size $n$ is small; in the parallel system; and when there is censoring. Observe that for the series system, the average of the $c$-values is less than 1, but for the parallel system, the average value of $c$ is greater than 1.

Figure 4.4 presents the bias, variance, and mean-squared error of the estimators of the system reliability at each time point. The estimators are based on a parallel system with 10 components that follow Weibull lifetimes. The systems were randomly right censored. Although there were some under-performance in the tail region, the improved estimator dominated the ML estimator in general. The relative efficiency
Figure 4.3 Estimated and True System Reliability Functions Over Time for a Parallel System with 10 components. The estimators are the ML based on component data and the improved estimator based on a sample of size 10.

Table 4.1: Average Losses of the Estimators in Series and Parallel Systems under Exponential Lifetime Distribution. These are based on 1000 simulation replications. The systems had 10 components, and the sample size for each replication was 10. Also indicated are the average values of the estimated shrinkage coefficient $c$.

<table>
<thead>
<tr>
<th></th>
<th>Series System</th>
<th>Parallel System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non censored</td>
<td>Censored</td>
</tr>
<tr>
<td>ML</td>
<td>0.0076</td>
<td>0.0103</td>
</tr>
<tr>
<td>Improved</td>
<td>0.0069</td>
<td>0.0096</td>
</tr>
<tr>
<td>$c$</td>
<td>0.9906</td>
<td>0.9856</td>
</tr>
</tbody>
</table>

was calculated via

$$
\text{RelEff} = \left[ \frac{\text{MSE(Shrink)}}{\text{MSE(ML)}} \right] \times 100.
$$
Table 4.2: Average Losses of the Estimators in Series and Parallel Systems under Weibull Lifetime Distribution. These are based on 1000 simulation replications. The systems had 10 components, and the sample size for each replication was 10. Also indicated are the average values of the estimated shrinkage coefficient $c$.

<table>
<thead>
<tr>
<th></th>
<th>Series System</th>
<th>Parallel System</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non censored</td>
<td>Censored</td>
</tr>
<tr>
<td>ML</td>
<td>0.0098</td>
<td>0.0182</td>
</tr>
<tr>
<td>Improved</td>
<td>0.0168</td>
<td>0.0096</td>
</tr>
<tr>
<td></td>
<td>Non censored</td>
<td>Censored</td>
</tr>
<tr>
<td>ML</td>
<td>0.0300</td>
<td>0.0430</td>
</tr>
<tr>
<td>Improved</td>
<td>0.0278</td>
<td>0.0386</td>
</tr>
<tr>
<td>$c$</td>
<td>0.9905</td>
<td>0.9847</td>
</tr>
<tr>
<td></td>
<td>1.0234</td>
<td>1.0263</td>
</tr>
</tbody>
</table>

4.8 Summary

This chapter discusses improved estimation of the system reliability function, when component level data are available. Results show that if a global assessment is used to determine the performance of estimators, it is beneficial to incorporate the idea of simultaneous estimation instead of treating each component individually. Although some approximations, based on asymptotic properties, were used in developing the improved simultaneous estimators of the component reliabilities, the resulting system reliability estimator indeed dominated the usual ML estimator, at least for the series and parallel structures, and for the component lifetime distributions (exponential and Weibull distributions) considered in the simulation studies. The resulting estimator can be further examined under other types of system structure (e.g. bridge system, "K out of N" system), and for other lifetime distributions (e.g. lognormal). It is rather surprising that in some cases the estimated coefficient $c$ is above 1. We provided some intuitive explanations of this phenomenon, but close examination may be needed in order to provide a mathematical proof.
Figure 4.4 Comparing Performances of the Component-Data Based Estimators at Different Values of $t$ for a 10-Component Parallel System. 1000 Replications were used, with each replication having a sample size of 10.
Chapter 5

Improved Estimation of System Reliability
Through Shrinkage Idea in Nonparametric Settings

5.1 The Nonparametric Model and Data Structure

Consider a system of \( K \) independent components with a lifetime variable \( S \). The observed data will be the \( K \) components time-to-failure, but which could be right-censored by the system lifetime or the end of monitoring period. Suppose there are \( n \) identical systems, so that the observable component lifetimes are

\[
\{T_{ij} : i = 1, 2, \ldots, n; j = 1, 2, \ldots, K\}.
\]

A simple estimator of the system reliability function is to utilize only the observed system lifetimes, in the form

\[
D = \{(Z_i, \delta_i) : i = 1, 2, \ldots, n\}.
\]

where \( Z_{ij} \) is the (possibly censored) observed system lifetime and \( \delta_{ij} \) indicates whether the observed system lifetime is complete (\( \delta_i = 1 \)) or censored (\( \delta_i = 0 \)). In nonparametric setting, the system reliability function can be estimated by the product limit (PL) estimator \((15)\). The PL estimator of \( R_S \) is given by

\[
\hat{R}_S(t) = \prod_{\{m: t_m \leq t\}} \left[ 1 - \frac{d_m}{n_m} \right],
\]

where \( t_{(1)} < \cdots < t_{(M)} \) are the ordered \( M \) distinct uncensored observations among the \( Z_i \)'s, \( d_m = \sum_i I\{Z_i = t_{(m)}, \delta_i = 1\} \) is the number of the \( n \) systems that failed.
At \( t(m) \), and \( n_m = \sum_i I\{Z_i \geq t(m)\} \) is the number of the \( n \) systems that are at-risk at \( t(m) \). Under some regularity conditions, when \( n \to \infty \), \( \hat{R}_S(t) \) is asymptotically normally-distributed with asymptotic mean \( R_S(t) \) and an estimate of its asymptotic variance, called Greenwood’s formula (18), given by

\[
\text{Avar}[\sqrt{n} \hat{R}_S(t)] = [\hat{R}_S(t)]^2 \sum_{\{m: t(m) \leq t\}} \frac{d_m}{n_m(n_m - d_m)}. 
\]

When component level data are available, we could improve the estimation of system reliability by utilizing the information on the system structure. Assume the lifetime of components, \( T_{ij} \)'s, are independent and that for each \( j \), \( (T_{ij}, i = 1, 2, \ldots, n) \) are identically-distributed with reliability function \( R_j(\cdot) \). In practice, the exact values of the \( T_{ij} \)'s are not all observable. Rather, they could be right-censored by either the system life or by the end of monitoring period. The observable right-censored data is

\[
D = \{(Z_{ij}, \delta_{ij}) : i = 1, 2, \ldots, n; j = 1, 2, \ldots, K\}. 
\]

A component-data-based estimator of the system reliability function is given by

\[
\hat{R}_S(t) = h_{\phi}[^{\hat{R}_1(t)}, \cdots, ^{\hat{R}_K(t)}],
\]

where \( \hat{R}_j(t) \) is the PL estimator of \( R_j(t) \) using the \( j \)th component lifetime data, \( (Z_j; \delta_j) \). The important aspect to note at this point is that the estimation problem of the \( K \) components reliability functions is intrinsically a problem of simultaneous estimation. Thus, the idea of shrinkage can be adopted in this situation to potentially obtain a more efficient estimator of the system reliability function.

### 5.2 Improved Estimation in Nonparametric Setting

Consider estimators of component reliabilities of the form

\[
\tilde{R}_j(t; c) = [\hat{R}_j(t)]^c, j = 1, \cdots, K;
\]
with $c > 0$ and where $\hat{R}_j$ is the PL estimator of $R_j$ based on the $j$th-component data $(Z_j, \delta_j)$ as described in previous section. A candidate class of system reliability function estimators of $R_S$ is, for $c > 0$,

$$\hat{R}_S(t; c) = h_0[\hat{R}_1(t; c), \ldots, \hat{R}_K(t; c)]$$

$$= h_0\{[R_1(t; \hat{\theta}_1)]^c, \ldots, [R_K(t; \hat{\theta}_K)]^c\}.$$  

Notice that when $c = 1$, $\hat{R}_S(t; 1)$ is the nonparametric MLE. If $\hat{\Lambda}_j(t)$ denotes an estimator of the cumulative hazard function $\Lambda_j(t)$ for component $j$, then through a plug-in-type of estimator, we could have the relationship

$$[\hat{R}_j(t)]^c = \exp[-c\hat{\Lambda}_j(t)].$$

Thus, essentially starting with a shrinkage coefficient on the PL estimators is tantamount to shrinking the cumulative hazard functions estimators.

To obtain a specific estimator of the system reliability function in this class of system reliability function estimators, it remains to decide on how to determine, hopefully optimally, the shrinking coefficient $c$ using the observed data.

From a decision-theoretic viewpoint, the performance of an estimator $\hat{R}_S(\cdot)$ of $R_S(\cdot)$ will be evaluated under the weighted Cramer-von Mises loss function in (4.20), and the corresponding expected loss or risk function is then given by (4.21).

### 5.3 Optimal $c^*$

To find the optimal coefficient $c$, $\hat{R}_S(t; c)$ is plugged into the risk function and then this risk is minimized with respect to $c$. To simplify notation, $\text{Risk}(c)$ is used to denote the risk with respect to $c$. Some approximations are used similar to those in the previous section based on the asymptotic properties of PL estimators and via Taylor expansion. The approximate risk function is again a polynomial in $c$ in the form of (4.22). The optimal shrinkage coefficient $c^*$ based on this approximation is the minimizer of the polynomial as given in (4.23).
Again, in practice, the true component reliability functions, $R_j(t)$’s, and also the $V_j$’s in the coefficients $A$, $B$ and $D$ are unknown. In the nonparametric setting, when estimating $c^*$ empirically, the occurrences of $R_j(t)$’s are replaced by the PL estimators $\hat{R}_j$’s for $j = 1, \cdots, K$ in the expressions, while $V_j = \text{Var}[\hat{R}_j(t)]$ is estimated using Greenwood’s formula for the PLE variance. This results in a data-determined value of $c^*$ given by $\hat{c}^*$. The resulting proposed estimator of the system reliability function based on the component-level data is

$$\hat{R}_S(t) = h_\phi[\hat{R}_1(t)^{\hat{c}^*}, \ldots, \hat{R}_K(t)^{\hat{c}^*}].$$

Notice that, as $n \to \infty$, $V_j$ converges to 0 according to Greenwood’s formula. Thus, when $n \to \infty$, both $A$ and $B$ converge to 0. As a consequence, for large $n$,

$$\text{Risk}(c) \approx D \cdot (c - 1)^2.$$ 

This demonstrates that, as $n \to \infty$, the optimal $c$ converges to 1. That is, in essence, the effect of the shrinkage will be more pronounced when the sample size or the number of systems is small.

### 5.4 Examples and Numerical Illustration

To illustrate the performance of the different estimators, we provide some numerical examples under different system structures. We generated system and component lifetime data based on series and parallel structures with 5 components. A sample of 15 systems is generated, and component lifetimes are generated according to exponential distributions with different means. The component and system lifetimes are not censored. Figure 5.1 and Figure 5.2 present the estimates for one replication, along with the true system reliability function for the series and the parallel structures, respectively.

For this numerical example, the $\hat{c}^*$ for the series system is 0.9203, which means that we are increasing the component reliabilities by imposing $\hat{c}^*$. While the $\hat{c}^*$ for
the parallel system is 1.0082, which means that we are shrinking the component reliabilities. Since this is only based on one replication, we obviously cannot make definitive comparisons of the performance of the different estimators. However, based on the results from one replication, we see the estimates do track the true system reliability function well. In the next section we present the results of simulation studies to compare the performance of the different nonparametric estimators of the system reliability function.
Figure 5.2 Estimated and True System Reliability Functions Over Time for a Parallel System with 5 components. The estimators are the nonparametric ML based on component data and the improved nonparametric estimator based on a sample of size 15.

5.5 Simulated Comparisons of the Estimators

The advantage of the improved nonparametric estimator over the nonparametric MLE is demonstrated using simulated data. Table 5.1 presents comparisons of the performance of the nonparametric ML estimator, and the improved nonparametric estimator in terms of the average loss based on 1000 replications. The simulated data were for $K = 5$ components each with $n = 15$ complete observations. Component lifetimes were not censored by the system life in the simulations. Exponential lifetime models were used under series and parallel systems.
Table 5.1: Average Losses of the Nonparametric Estimators in Series and Parallel Systems under Exponential Lifetime Distribution. These are based on 1000 simulation replications. The systems had 5 components, and the sample size for each replication was 15. Also indicated are the average values of the estimated shrinkage coefficient $c$.

<table>
<thead>
<tr>
<th></th>
<th>Series System</th>
<th>Parallel System</th>
</tr>
</thead>
<tbody>
<tr>
<td>ML</td>
<td>0.02596</td>
<td>0.03486</td>
</tr>
<tr>
<td>Improved</td>
<td>0.02485</td>
<td>0.03434</td>
</tr>
<tr>
<td>$c$</td>
<td>0.9139</td>
<td>1.0135</td>
</tr>
</tbody>
</table>

The improved nonparametric estimator performs better than the nonparametric MLE in the context of the average global losses. Again we observe that for the series system, the average of the optimal $c$-values is less than 1, so that on average there is shrinkage that is going on. However, for the parallel system, the average value of $\hat{c}^*$ is greater than 1. Some intuitive explanations were provided in Chapter 4, section 4.5.

5.6 Summary

In this chapter, we constructed and improved estimation of system reliability function using component lifetimes under the nonparametric setting. By putting a shrinkage coefficient on the estimated component hazard functions, we were able to extend the result in Chapter 4. Simulation results yield similar conclusions as in the parametric setting. Simultaneous estimation performs better than individual ones if a global loss function is used, although some approximations, based on asymptotic properties of the product-limit estimator, were used. The resulting estimator can be further examined under other types of system structure and when component lifetimes are censored.
Chapter 6

Conclusions and Future Research

This dissertation discusses improved estimation when several parameters or distributions need to be estimated simultaneously. Results from simulation studies show that if a global assessment is used to determine the performance of estimators, it is beneficial to incorporate the idea of simultaneous estimation instead of treating each estimation problems individually. In Chapter 3, we focused on improving the estimation of location parameters in nonparametric settings. We developed shrinkage estimators from a decision theoretic viewpoint and from a Bayesian approach. Both estimators utilize combined information, and thus outperform the vector of sample medians in terms of mean squared loss, especially when the number of parameters, $K$, is large compared to sample size. Notice that by imposing the shrinkage coefficient $c < 1$, the sample medians are shrunk towards 0. However, using nonparametric Bayesian approach, the resulting estimators of individual distributions are rather shrinking toward the empirical distribution based on all samples.

In Chapters 4 and 5, we developed improved estimators of system reliability under parametric and nonparametric settings. Based on the simulation results, the improved estimators that incorporate the idea of simultaneous estimation dominates the standard ML estimators in terms of the average losses, although some approximations were used. We also notice that under the two extreme types of system structure, the estimated coefficient $c$ behaves differently. On average, the estimated coefficient $c$ is below 1 for the series system, indicating an expanding of component reliabilities towards 1. While for the parallel system, the mean of the estimated coefficient $c$ is
above 1, indicating a shrinking of component reliabilities towards 0. We provided some intuitive explanations based on system structure function and the reliability importance functions, but further examinations are needed in order to fully explain the phenomenon.

To improve the estimation of system reliability in the nonparametric setting, an alternative way is to use a nonparametric Bayesian approach similar to the empirical Bayes approach discussed in Chapter 3. A possible way is to put gamma process priors with common parameters on the component cumulative hazard functions, and then estimate the parameters empirically using observed data. Some adjustments may be needed when the component lifetimes are censored. When using the empirical Bayesian approach, we expect the resulting estimator to shrink towards the average hazard of all components.

As a direction for future studies, we consider a more general setting for the simultaneous estimation problem. Assume we observe $K$ samples of data from distributions $F_1, \cdots, F_K$. Consider the problem of estimating distributions $F_1, \cdots, F_K$, when a global loss function is defined on a mapping $h$ of $F_1, \cdots, F_K$, i.e. $h(F_1, \cdots, F_K)$. Notice that when estimating the location parameters in Chapter 3, we are actually taking $h$ to be the vector of median function of $F_1, \cdots, F_K$. And, in Chapters 4 and 5, we are taking $h$ to be the structure function of a coherent system. Those can be viewed as special cases of $h$, but the question remains if we can find an improved estimator of $h(F_1, \cdots, F_K)$, when $h$ is of a general form. A possible approach is to write $F_j = 1 - e^{-\Lambda_j}$, and then

$$h(F_1, \cdots, F_K) = h^*(\Lambda_1, \cdots, \Lambda_K),$$

where

$$h^*(u_1, \cdots, u_K) = h(1 - e^{u_1}, \cdots, 1 - e^{u_K}).$$

To estimate $F_j$’s simultaneously, we try to estimate $\Lambda_j$’s. We consider writing $\Lambda_j$ in
terms of basis functions, i.e.

\[ \Lambda_j(t) = \Lambda_0(t) \exp[c \cdot \theta_j^T \phi(t)], \]

where \( \phi(t) = (\phi_0(t), \cdots, \phi_p(t)) \) is a vector of basis functions, which is assumed to be known. \( \Lambda_0, \theta_1, \cdots, \theta_K \) are unknown, and need to be estimated using observed data. The goal is to find an optimal shrinkage coefficient \( c \) according to the defined loss function. The resulting estimator of \( \Lambda_j(t) \) will have the form

\[ \hat{\Lambda}_j(t) = \hat{\Lambda}_0(t) \exp[\hat{c}^* \cdot \hat{\theta}_j^T \phi(t)]. \]

A Bayesian approach may provide similar results in this setting, if we put gamma process prior on \( \Lambda_0 \), and multivariate normal priors on \( \theta_j \)'s. These approaches are still under investigation. We anticipate possible complications when the observed data is censored and we may also need to impose some restrictions on the mapping \( h \).
Bibliography


