## An R package for monitoring test under density ratio model and its applications

by

### Boyi Hu

B.Sc., University of Science and Technology of China, 2013M.Sc., Lakehead University, 2016

## A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

**Master of Science** 

in

## THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Statistics)

The University of British Columbia

(Vancouver)

September 2018

© Boyi Hu, 2018

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis/dissertation entitled:

An R package for monitoring test under density ratio model and its applications

submitted by Boyi Hu in partial fulfillment of the requirements for

| the degree of         | Master of Science |  |
|-----------------------|-------------------|--|
|                       |                   |  |
| 1n                    | Statistics        |  |
| Examining Committee   | :                 |  |
|                       | Jiahua Chen       |  |
| Supervisor            |                   |  |
|                       | Natalia Nolde     |  |
| Supervisory Committee | Member            |  |
|                       |                   |  |
| Supervisory Committee | Member            |  |
|                       |                   |  |
| Additional Examiner   |                   |  |

# Abstract

Quantiles and their functions are important population characteristics in many applications. In forestry, lower quantiles of the modulus of rapture and other mechanical properties of the wood products are important quality indices. It is important to ensure that the wood products in the market over the years meet the established industrial standards. Two well-known risk measures in finance and hydrology, value at risk (VaR) and median shortfall (MS), are quantiles of their corresponding marginal distributions. Developing effective statistical inference methods and tools on quantiles of interest is an important task in both theory and applications. When samples from multiple similar natured populations are available, Chen et al. [2016] proposed to use a density ratio model (DRM) to characterize potential latent structures in these populations. The DRM enables us to fully utilized the information contained in the data from connected populations. They further proposed a composite empirical likelihood (CEL) to avoid a parametric model assumption that is subject to model-mis-specification risk and to accommodate clustered data structure. A cluster-based bootstrap procedure was also investigated for variance estimation, construction of confidence interval and test of various hypotheses.

This thesis contains complementary developments to Chen et al. [2016]. First, a user-friendly R package is developed to make their methods easy-to-use for practitioners. We also include some diagnostic tools to allow users to investigate the goodness of the fit of the density ratio model. Second, we use simulation to compare the performance DRM-CEL-based test and the famous Wilcoxin rank test for clustered data. Third, we study the performance of DRM-CEL-based inference when the data set contains observations with different cluster sizes. The simulation results show that DRM-CEL method works well in common situations.

# Lay Summary

Quantiles and their functions are important population characteristics in many applications. In this thesis, we focus on lower quantiles of lumber strength distribution, which are usually used as quality indices for wood products. It is important to ensure that the wood products in the market over the years meet the established industrial standards. Therefore, monitoring some certain lower quantile of the distribution of strength of wood products across years is of big concern. We first review some existing quantile monitoring methods. Based on one of these methods, we developed an user-friendly R package to make the method easy-to-use for practitioners. Further, with the help of this package, we investigate the performance of this monitoring method in a more general situation, which is not considered in the original paper.

# **Table of Contents**

| Ab                    | strac | ······································           |  |  |  |  |  |
|-----------------------|-------|--|--|--|--|--|--|
| Lay Summary           |       |  |  |  |  |  |  |
| Table of Contents   v |       |  |  |  |  |  |  |
| List of Tables        |       |  |  |  |  |  |  |
| List of Figures       |       |  |  |  |  |  |  |
| Acknowledgments       |       |  |  |  |  |  |  |
| Preface x             |       |  |  |  |  |  |  |
| 1                     | Intro | duction  |  |  |  |  |  |
|                       | 1.1   | The density ratio model                          |  |  |  |  |  |
|                       | 1.2   | Empirical likelihood under density ratio model 5 |  |  |  |  |  |
|                       | 1.3   | DRM-CEL for clustered data                       |  |  |  |  |  |
|                       | 1.4   | Bootstrap and the monitoring test                |  |  |  |  |  |
|                       | 1.5   | Contribution of this thesis                      |  |  |  |  |  |
| 2                     | Prob  | lem Description                                  |  |  |  |  |  |
|                       | 2.1   | Targeted statistical problems                    |  |  |  |  |  |
|                       | 2.2   | Typical data structure         14                |  |  |  |  |  |
| 3                     | Num   | erical tools for DRM-CEL Monitoring Test         |  |  |  |  |  |

|              | 3.1 | Model Fitting: drm_fit( $\cdots$ )  | 19 |  |
|--------------|-----|---|----|--|
|              | 3.2 | Quantile Estimates: $quan_est(\cdots)$  | 23 |  |
|              | 3.3 | Visualization Tool: $plot_cdf(\dots)$   | 24 |  |
|              | 3.4 | Diagnostic Tools: $plot_kde(\cdots)$ and $plot_qq(\cdots)$                        | 26 |  |
|              | 3.5 | Monitoring Test: monitor_test( $\cdots$ )   | 28 |  |
|              | 3.6 | Data Example  | 30 |  |
| 4            | Mon | itoring tests for Clustered Data  | 37 |  |
|              | 4.1 | Classical Wilcoxon Rank Sum Test  | 38 |  |
|              | 4.2 | Modified Wilcoxon Rank Sum Test   | 39 |  |
|              | 4.3 | Wilcoxon rank sum test for monitoring purpose                                     | 42 |  |
|              | 4.4 | The DRM_CEL monitoring test with unequal sized clusters                           | 45 |  |
|              | 4.5 | $Performance \ of \ DRM\_CEL \ Monitor \ Test \ when \ Model \ is \ Misspecified$ | 53 |  |
| 5            | Sum | mary  | 58 |  |
| Bibliography |     |   |    |  |

# **List of Tables**

| Table 4.1 | Rejection Rates (%) of DRM_CEL Monitoring Test and Modi-     |    |  |
|-----------|--|----|--|
|           | fied Wilcoxon Test at Nominal Level 95%                      | 45 |  |
| Table 4.2 | Coverage probabilities (%) of two-sided 95% bootstrap confi- |    |  |
|           | dence intervals under Normal models (type one)               | 49 |  |
| Table 4.3 | Coverage probabilities (%) of two-sided 95% bootstrap confi- |    |  |
|           | dence intervals under Gamma models (type one)                | 50 |  |
| Table 4.4 | Coverage probabilities (%) of two-sided 95% bootstrap confi- |    |  |
|           | dence intervals under Normal models (type two)               | 51 |  |
| Table 4.5 | Coverage probabilities (%) of two-sided 95% bootstrap confi- |    |  |
|           | dence intervals under Gamma models (type two)                | 52 |  |
| Table 4.6 | Coverage probabilities (%) of two-sided 95% bootstrap confi- |    |  |
|           | dence intervals under Normal models                          | 55 |  |
| Table 4.7 | Coverage probabilities (%) of two-sided 95% bootstrap confi- |    |  |
|           | dence intervals under Gamma models                           | 56 |  |

# **List of Figures**

| Figure 3.1 | Q-Q Plot of 2007                     | 32 |
|------------|--------------------------------------|----|
| Figure 3.2 | Q-Q Plot of 2010                     | 32 |
| Figure 3.3 | Q-Q Plot of 2011                     | 33 |
| Figure 3.4 | Kernel Density Estimate Plot of 2007 | 33 |
| Figure 3.5 | Kernel Density Estimate Plot of 2010 | 35 |
| Figure 3.6 | Kernel Density Estimate Plot of 2011 | 35 |
| Figure 3.7 | Estimated Distributions under DRM    | 36 |

# Acknowledgments

Foremost I would like to express my deepest gratitude to my supervisor, Professor Jiahua Chen. His inspiration, guidance, encouragement and insight help me through my master career at UBC. I learn a lot from him. I also want to thank Professor Natalia Nolde for being the second reviewer of my thesis.

I would like to give my sincere appreciation and gratitude to my parents for their support and encouragement. This thesis is dedicated to them.

# Preface

This dissertation is an unpublished work by the author, Boyi Hu, under the supervision of Professor Jiahua Chen.

## **Chapter 1**

# Introduction

Population quantiles and their functions are important measures in many applications. For example, in forestry, it is important to maintain the value of the fifth lower percentile of lumber strength distribution. Such engineering properties are vital to the strength of wood structures. See American Society for Testing and Materials (ASTM) Standard D1990 International [2007]. Two well-known risk measures in finance and hydrology, value at risk (VaR) Pflug [2000] and median shortfall (MS) Bertsimas et al. [2004], are quantiles of the corresponding distributions. They are generally used to evaluate the risk tolerance of financial institutions and hydrological structures. Monitoring the levels of these quantiles is of great interest in these applications.

This thesis is motivated by a research project on forestry at the University of British Columbia through funding "The NSERC CRD FPInnovations grant" and by the work of Chen et al. [2016]. The Canadian forestry industry is a major contributor to the Canadian economy. Canada has 42 percent of its land acreage covered by forests. The country contains 10 percent of the world's forested land and the forests are made up mostly of spruce, poplar and pine. To meet the society's long-term demand for forest products and the near-term economic benefit, responsible forest management is required to ensure that forests are legally harvested and managed. Data collection and statistical analysis are important parts of the effective management.

The properties of wood products may change over time due to some long term

trends such as the climate change, and some catastrophic short term impacts such as extreme forest fire series and mountain pine beetles. All of these factors might affect the mechanical strength of wood materials. Therefore, people from forestry are very serious at monitoring the change of quality of wood products. To measure the mechanical property of wood, modulus of elasticity (MOE) and modulus of rupture (MOR) Bier and Collins [1985] are two important and commonly used quality indices Barrett and Kellogg [1991], Bendtsen and Senft [2007], Boone and Chudnoff [1972], Smith et al. [1991]. MOE and MOR are measurements of lumber's elasticity and toughness respectively. Lower quantiles of the population distribution of these two indices are used to measure the reliability of wood products. The 5% quantile is usually a common choice. They are important design values for wood structures.

Imagine populations made of lumber produced by a number of mills over years. The lumber samples may be collected as follows. Each year, a number of mills are randomly chosen and then several lots of wood pieces are randomly sampled from them. From each lot sampled, several pieces of wood are selected to form a sample from this mill. They form the data set to be analyzed to decide whether the wood product meets the industrial standard and to provide other information useful to wood industry and engineers. Clearly, the observations obtained are clustered in such applications.

Many standard statistical methods are developed for data in which the observed values may be regarded independent of each other. This includes the famous Wilcoxon rank sum test, Kolmogorov goodness-of-fit test, and the method of Anderson [1979]. The data collected from wood industry and likely in other applications are, however, often clustered. This can be seen from the exploratory data analysis of real data and based on some background information Verrill et al. [2015]. Because the wood pieces sampled from the same lot/mill are likely from the same tree, or from trees grown in the same region and so on, they are more likely to have have similar strengths.

When observations are correlated but analyzed as if they are independent, the uncertainty in statistical inference is generally underestimated. This leads to undercoverage of the confidence intervals and inflated type I error of significance tests. To investigate the effect of clustered data to type I errors of some standard procedures, Verrill et al. [2015] conducted extensive simulation studies of eight statistical tests proposed by the United States Department of Agriculture. Their investigation largely confirms the damaging effect of clustered data if the clustering structure is not properly looked after. In addition, although the parameter of interest in the targeted application is lower quantiles, these methods are generally designed from other population statistics. Hence, even if they work well for their original purpose, additional developments are in great needs.

The forestry data and data from some applications have two important characteristics. The first is that we often have samples from multiple populations that share latent structures. The second is that the data can be clustered. The paper of Anderson [1979] proposes a semi-parametric density ratio model which takes the shared latent structure into consideration; it further develops an empirical likelihood approach for estimation, confidence interval and hypothesis test regarding population quantiles. For clustered data in forestry and likely in some other applications, the joint distributions are exchangeable. This leads to non-parametric cluster effect assumption in Chen et al. [2016] and the proposed composite empirical likelihood. To avoid inflated type I error in monitoring tests, Chen et al. [2016] further propose to use cluster-based bootstrapping method for variance estimation. Asymptotically valid confidence intervals and hypothesis tests are subsequently obtained.

This thesis contains complementary work to these two papers as well as new developments to the research and application area. We first give a brief review of the materials in these two papers in the next section.

#### **1.1** The density ratio model

In common problems, one may have a set of random observations from a population with distribution F. We may estimate F by its empirical distribution

$$F_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}(x_i \le x)$$

where  $\mathbb{1}(\cdot)$  is the indicator function. We have used  $x_1, \ldots, x_n$  as observed values with sample size *n*. For any  $p \in (0, 1)$ , the sample quantile is defined to be

$$\hat{\xi}_p = F_n^{-1}(p) = \inf\{x : F_n(x) \ge p\}.$$

Generally speaking, the sample quantile  $\hat{\xi}_p$  is a good nonparametric estimator Serfling [2009].

When *F* is known to be a member of parametric distribution family, say,  $F(x) = F(x; \theta)$ . One may first estimate  $\theta$  with an efficient method and subsequently estimate population quantile  $\xi_p$  by  $F^{-1}(x; \hat{\theta})$ . This is also a valid approach.

As pointed out in Anderson [1979], when we have data from multiple connected populations, the empirical quantiles fail to utilize this extra information. The efficiency of the parametric quantile estimator  $F^{-1}(x;\hat{\theta})$  cannot be further improved, but it can suffer from model mis-specification. If one wishes to avoid risky parametric model assumption and make use of the information in multiple samples, a density ratio model is an attractive choice.

Let  $G_1, G_2, \ldots, G_m$  be distribution functions of *m* connected populations from which some observations are available. In forestry, they may stand for populations from *m* calendar years. Let  $g_1, g_2, \ldots, g_m$  be their density functions with respect to some  $\sigma$ -finite measure. The DRM postulates that

$$\frac{g_k(x)}{g_1(x)} = \exp\{\boldsymbol{\theta}_k^{\tau} \mathbf{q}(x)\}$$
(1.1)

for some known function  $\mathbf{q}(x)$  of dimension d and corresponding unknown parameter vector  $\boldsymbol{\theta}_k$ , where  $\boldsymbol{\theta}_k^{\tau}$  is the transpose of  $\boldsymbol{\theta}_k$ . The first entry of  $\mathbf{q}(x)$  is generally required to be constant one so that the first element of  $\boldsymbol{\theta}_k$  is a normalization parameter. The DRM is semi-parametric and it allows users to choose or specify the basis function  $\mathbf{q}(x)$  to reflect the nature of the populations in applications. When the basis function is properly chosen, the information from multiple samples is pooled together to permit more efficient statistical inference. The pooled information may improve the efficiency of each quantile estimate. Unless otherwise specified, we will regard these densities as with respect to Lesbesgue measure.

Take two risk measures, VaR and MS, in finance and hydrology as examples.

They are quantiles of the populations under investigation. Due to the lack of observations in extreme tails of the distribution, non-parametric inferences are usually not meaningful so that some parametric assumptions are usually specified. The generalized extreme value (GEV) distribution family is often assumed for data analysis. If we assume the shape parameters  $\xi$  of the corresponding distributions are identical, we can verify that these distributions are special cases of DRM with proper basis functions. Under the assumption, it is advantageous to make inference under the semi-parametric DRM.

We now move to the review of empirical likelihood method under the density ratio model.

### **1.2 Empirical likelihood under density ratio model**

Consider the simple case where the observations in multiple samples are denoted as  $x_{k,j}$ : k = 1, 2, ..., m and  $j = 1, 2, ..., n_k$ . Suppose they are independent and the distribution of  $X_{k,j}$  is given by  $G_k$  satisfying the DRM (1.1). In the spirit of Owen [2001], Chen and Chen [2000] restrict the form of  $G_1$  to

$$G_1(x) = \sum_{k,j} p_{k,j} \mathbb{1}(x_{k,j} \le x),$$

where  $p_{k,j}$  is the likelihood contribution of observation  $x_{k,j}$ . If not specified, the range over summation or product will be all possible values of the indices. Under the DRM assumption, we may write  $G_r$  as

$$G_r(x) = \sum_{k,j} p_{k,j} \exp\{\boldsymbol{\theta}_r^{\tau} \mathbf{q}(x_{k,j})\} \mathbb{1}(x_{k,j} \le x)$$

for r = 1, 2, ..., m, with  $\boldsymbol{\theta}_1 = 0$ . Since  $G_r$ 's are distribution functions, we must have

$$\sum_{k,j} p_{k,j} \exp\{\boldsymbol{\theta}_r^{\tau} \mathbf{q}(x_{x,j})\} = 1$$
(1.2)

for all r = 1, 2, ..., m.

The empirical likelihood (EL) function is given by

$$L(G_1,\ldots,G_m)=\prod_{k,j}dG_k(x_{k,j})=\{\prod_{k,j}p_{k,j}\}\exp\{\sum_k\boldsymbol{\theta}_k^{\tau}\sum_j\mathbf{q}(x_{k,j})\}$$

with  $p_{k,i}$  and  $\boldsymbol{\theta}_k$  satisfying (1.2).

By standard method of Lagrange multipliers Nakayama et al. [1975], the fitted values of  $p_{k,j}$  are given by

$$\hat{p}_{k,j} = \frac{1}{n\sum_{r=1}^{m} \rho_r \exp\{\hat{\boldsymbol{\theta}}_r^{\tau} \mathbf{q}(x_{k,j})\}}$$

where  $\rho_r = n_r / \sum_{k=1}^m n_k$ . The fitted cumulative distribution functions are given by

$$\hat{G}_k(x) = \sum_{k,j} \hat{p}_{k,j} \exp\{\hat{\boldsymbol{\theta}}_k^{\tau} \mathbf{q}(x_{k,j})\} \mathbb{1}(x_{k,j} \le x).$$

Chen and Liu [2013] prove that the DRM-EL quantile estimators derived from the above  $\hat{G}_k(x)$  are more efficient than the empirical quantiles. They also investigate other properties such as the asymptotic normality, confidence intervals and the hypothesis test problems. We will not go over these details.

### **1.3 DRM-CEL for clustered data**

As explained earlier that the data are often clustered in applications. Ignoring the cluster structure may lead to inflated type I errors for hypothesis tests and undercoverage problem for the confidence intervals. Hence, the methods developed in Chen and Liu [2013] must be modified. For this purpose, Chen et al. [2016] further developed a DRM based composite likelihood approach. Consider the situation where each  $X_{k,j,l}$  is a *d*-dimensional random vector, where *k* is the population index, *j* is the cluster index and *l* is the data index within each cluster. Let  $F_k(y_1, \ldots, y_d)$  be the joint distribution function of  $X_{k,j,l}$ . In many applications, it is quite reasonable to assume that this distribution is exchangeable. For example, when d = 4, the exchangeability of  $F_k$  means

$$F_k(y_1, y_2, y_3, y_4) = F_k(y_1, y_3, y_2, y_4) = F_k(y_4, y_3, y_1, y_2) = \cdots$$

This implies that the marginal distributions of each entry are same. For this reason, the population distribution of any entry of  $X_{k,j,l}$  is given by

$$G_k(x) = F_k(x, \infty, \infty, \infty) = F_k(\infty, x, \infty, \infty) = F_k(\infty, \infty, x, \infty) = F_k(\infty, \infty, \infty, x).$$

If so, the DRM-EL methods seem to work, at least artificially. We need not impose additional restrictions on  $F_k$ .

We can motivate the direct use of previous DRM-EL methods through the notion of composite likelihood of Lindsay [1988].

Consider a random vector *Y*, with probability density function  $f(y; \theta)$  for some parameter vector  $\theta \in \Theta$ . Denote by  $\{A_1, ..., A_K\}$  a set of events with associated likelihoods  $L_k(\theta; y) \propto f(y \in A_k; \theta)$ . Following Lindsay [1988] a composite likelihood is the weighted product

$$L_c(\boldsymbol{\theta}; y) = \prod_{k=1}^K L_k(\boldsymbol{\theta}; y)^{w_k},$$

where  $w_k$  are nonnegative weights to be chosen. One should not assume any properties of the authentic likelihood functions for  $L_c(\boldsymbol{\theta})$ . One may choose  $w_k$  to reflect the some knowledge on the relationship between these events and to improve the inference efficiency Varin et al. [2011]. In Chen and Liu [2013], they consider the composite likelihood of each cluster. Specifically, for each cluster

$$\mathbf{x}_{k,j} = (x_{k,j,1}, x_{k,j,2}, \dots, x_{k,j,d})^{\tau},$$

Chen and Liu [2013] suggest to have  $A_i$  as the event  $\{X_{k,j,i} = x_{k,j,i}\}$  for i = 1, 2, ..., d. With  $A_1, A_2, ..., A_d$  being symmetric (the exchangeability of the joint distribution  $F_k$ ), it is natural to have  $w_1 = w_2 = \cdots = w_d$  and define a composite likelihood to be

$$L_c(G_1,...,G_m) = \prod_{k,j} \{\prod_{l=1}^d dG_k(x_{k,j,l})\}.$$

defined on the space of  $G_1, \ldots, G_m$  satisfying

$$G_r(x) = \sum_{k,j,l} p_{k,j,l} \exp\{\boldsymbol{\theta}_r^{\tau} \mathbf{q}(x_{k,j,l})\} \mathbb{1}(x_{k,j,l} \le x)$$

with  $\boldsymbol{\theta}_1 = 0$  and that for  $r = 1, 2, \dots, m$ ,

$$\sum_{k,j,l} p_{k,j,l} \exp\{\boldsymbol{\theta}_r^{\tau} \mathbf{q}(x_{k,j,l})\} = 1.$$

Following the same algebra in Section 1.2, the composite likelihood is maximized when

$$\hat{p}_{k,j,l} = \frac{1}{nd\sum_{r=1}^{m}\rho_r \exp\{\hat{\boldsymbol{\theta}}_r^{\tau} \mathbf{q}(x_{k,j,l})\}}$$

where  $\hat{\boldsymbol{\theta}}_r$  is the maximum CEL estimator. Subsequently, the maximum CEL estimator of  $G_r(x)$  is given by

$$\hat{G}_r(x) = \sum_{k,j,l} \hat{p}_{k,j,l} \exp\{\hat{\boldsymbol{\theta}}_r^{\tau} \mathbf{q}(x_{k,j,l})\} \mathbb{1}(x_{k,j,l} \leq x).$$

The DRM-CEL based quantile estimators are given by  $\hat{\xi}_{r,\alpha} = \hat{G}_r^{-1}(\alpha)$ .

Chen et al. [2016] prove that the DRM-CEL based quantile estimators remain consistent as the total sample size  $\sum_{k=1}^{m} n_k d$  increases under some conditions and are still asymptotically normal. However, their asymptotic variances are usually larger than those obtained when the data are not clustered. For this reason, the results of Chen et al. [2016] on monitoring tests cannot be directly used. To address this problem, they propose to use resampling methods to avoid variance underestimation. This is the topic of the next section.

### **1.4** Bootstrap and the monitoring test

As pointed out from the beginning, one problem of interest in forestry is to monitor the change of lower quantiles. This task may be regarded as a hypothesis test problem. We wish to know whether the population quantile of one year is significantly lower than that of the previous year, or lower than some industrial standard.

Because of this, we wish to test hypothesis in the form of

$$H_0: \xi_{1,\alpha} \leq \xi_{2,\alpha}$$

versus one-sided alternative

$$H_1:\xi_{1,\alpha}>\xi_{2,\alpha}$$

where  $\xi_{1,\alpha} \leq \xi_{2,\alpha}$  are  $\alpha$ -quantiles of the first two populations for some selected quantile level  $\alpha$ . We consider this problem when there may also be samples from  $G_3, G_4$  and so on. We also want to know whether the existing method works or not when data are clustered.

The results in Chen et al. [2016] imply that the difference of two DRM-CEL quantile estimators,  $\hat{\xi}_{k_1,\alpha} - \hat{\xi}_{k_2,\alpha}$ , where  $k_1$  and  $k_2$  are population indices, is asymptotically normal with zero mean and finite variance. Therefore, we may use Wald method for the purpose of hypothesis test if we have an easy to use estimate of the asymptotic variance. It turns out that the asymptotic variance has a complex expression and depends on specific cluster structures. Hence, a straight estimation of the asymptotic variance is not a convenient option.

To overcome this difficult, Chen et al. [2016] propose a cluster-based bootstrap method to estimate the distribution of  $\hat{\xi}_{k_1,\alpha} - \hat{\xi}_{k_2,\alpha}$ . In this scheme, the  $n_k$  bootstrap clusters are sampled from the observed clusters without replacement from the *k*th sample. A bootstrap sample is therefore formed as

$$\{\mathbf{x}_{k,j}^*; j=1,2,\ldots,n_k, k=1,2,\ldots,m\}.$$

The DRM-CEL is then applied to this bootstrap sample to obtain  $\hat{\xi}_{k_1,\alpha}^* - \hat{\xi}_{k_2,\alpha}^*$ . This process is repeated a large number *B* times. The empirical distribution formed by these *B* values of  $\hat{\xi}_{k_1,\alpha}^* - \hat{\xi}_{k_2,\alpha}^*$  is utilized to give an estimate of the distribution of  $\hat{\xi}_{k_1,\alpha} - \hat{\xi}_{k_2,\alpha}$ .

Clearly, confidence intervals for  $\hat{\xi}_{k_1,\alpha} - \hat{\xi}_{k_2,\alpha}$  can then be constructed accordingly. From the relationship between the confidence interval and the hypothesis test, we further obtain a procedure for monitoring test.

In Chen et al. [2016], the validity of this bootstrap procedure is summarized as the following theorem.

**Theorem 1** Let  $\xi_{k_1,\alpha}, \xi_{k_2,\alpha}$  be two population quantiles of  $G_{k_1}$  and  $G_{k_2}$ . Let  $\varphi(x, y)$  be a differentiable function. Let  $(\hat{\xi}_{k_1,\alpha}, \hat{\xi}_{k_2,\alpha})$  and  $\xi^*_{k_1,\alpha}, \xi^*_{k_2,\alpha}$  be DRM-CEL quantile estimators based on the original sample and bootstrap sample. Then under some

conditions and as the total sample size  $n \to \infty$ ,

$$\begin{aligned} \sup_{x} \left| \mathsf{Pr}^{*} \left( \sqrt{n} \{ \varphi(\xi_{k_{1},\alpha}^{*},\xi_{k_{2},\alpha}^{*}) - \varphi(\hat{\xi}_{k_{1},\alpha},\hat{\xi}_{k_{2},\alpha}) \} \le x \right) \\ - \mathsf{Pr} \left( \sqrt{n} \{ \varphi(\xi_{k_{1},\alpha},\xi_{k_{2},\alpha}) - \varphi(\xi_{k_{1},\alpha},\xi_{k_{2},\alpha}) \} \le x \right) \right| = o_{p}(1) \end{aligned}$$

where  $Pr^*$  denotes the conditional probability given data corresponding to bootstrapping distribution.

Note that a null hypothesis in the form of  $\xi_1 = \xi_2$  can be expressed as  $\varphi(\xi_1, \xi_2) = 0$  with  $\varphi(x, y) = x - y$ . We test the hypothesis

$$H_0: \varphi(\xi_{k_1,\alpha},\xi_{k_2,\alpha})=0$$

by observing whether an appropriate levelled confidence interval of  $\varphi(\xi_{k_1,\alpha},\xi_{k_2,\alpha})$  contains 0. We will give details of the bootstrap procedure later.

#### **1.5** Contribution of this thesis

To apply the methods developed by Chen and Liu [2013] and Chen et al. [2016] on realistic problems, a user-friendly R package is needed. Several relevant R functions have been built by the authors to carry out the simulation studies in their papers, but these R functions are not easy to use by users from industry. Based on these existing R functions, we further build a user-friendly R package called "dr-mmt" to implement these DRM based quantile estimation methods. The functions in this package are designed for application requirements and easy to access.

In this thesis, we first present the R functions in the package "drmmt". With these R functions, we use simulation studies to compare the performances of different methods for quantile estimations of clustered data, and investigate the performance of the method developed by Chen et al. [2016] in a more general case. In Chapter 2, we first present the data collection procedure for lumber example, followed by an introduction to the required data format of our R functions for any data set. Although the method developed by Chen et al. [2016] is motivated by lumber project, their method actually can be applied to many other areas as long as the corresponding model assumptions are satisfied. For our R functions, as long as the format of data set is complete, they can be applied on any data sets from any area. Then we present how to input a real data set into our R functions. In Chapter 3, we first review the statistical model and corresponding parameter estimation method Chen et al. [2016]. We present the statistical questions coming from the targeted applications and the solutions of these questions. Then we present the R functions in our package "drmmt". For each R function, we illustrate its inputs and outputs and demonstrate it by a code example. We explain the algorithms behind these function and introduce the R functions being used from other R packages. At the end of this chapter, we use a real data example to demonstrate how to apply the functions from this package to a realistic problem.

In Chapter 4, we first review the classical Wilcoxon rank sum test Lehmann [2006] and introduce a cluster-based Wilcoxon rank sum test developed by Rosner et al. [2003]. The classical Wilcoxon rank sum test is not appropriate for our targeted application because it is designed for independent dataset and it is not built to detect the change in population quantiles. Even if the cluster-based Wilcoxon rank sum test has a good performance in the clustered dataset, its hypothesis is still not testing the change in quantiles. Admittedly, under some extra assumption, the Wilcoxon rank sum tests can be used to detect the change in medians, but the median is not a useful index in the targeted applications. In some common cases, the alternative hypothesis of this cluster-based Wilcoxon rank sum has the same direction with the alternative hypothesis of the targeted applications, but they are not always identical. It may bring potential risks for the statistical inference and in the following section, we use simulation studies to illustrate this. Compared with the Wilcoxon-type hypothesis tests, the DRM-CEL monitoring test of Chen et al. [2016] can be used to test the hypothesis related to more general population quantiles, not only the median. In the paper, one of the assumptions under which the monitoring that is developed, is that the clusters of samples have a same number of observations. However, in realistic applications, a data set with equal cluster sizes cannot always be guaranteed. If this is the case, whether we still can use this method to monitor the quantiles is of our interest. For example, the cluster sizes of the real data set we use in Chapter 4 are not equal, but we still use the DRM-CEL method to analyze it. We loosen the restriction of cluster sizes of populations being equal in Chen et al. [2016] and use simulation studies to investigate the influence

of unequal cluster sizes on the DRM-CEL method. The simulation results support that as long as the maximum cluster size over all populations is finite and the number of clusters of each population is large enough, this DRM-CEL monitoring test still works. I believe that the mathematical techniques behind this conclusion are very similar to that given in Chen et al. [2016], Serfling [2009]. The theory behind this conclusion is one of my further research topics. At the end of this Chapter, we use simulation studies to investigate the performance of long term monitoring test when model is misspecified in the cases of cluster sizes being not equal. A summary is given at the end of this thesis.

## Chapter 2

# **Problem Description**

As mentioned in the introduction, the wood quality is vital to all the other related areas. This quality can be depicted by some certain lower quantile of the distribution of lumber mechanical strength. Such a quantile is called the quality index of a population in this context. Before a statistical analysis is conducted, real world problems need to be defined clearly first and then to be transfered into statistical problems. Answers to these statistical problems are sought based on data collected on representative samples from representative populations.

### **2.1** Targeted statistical problems

We identify a few typical statistical problems to be addressed as follows.

- Does the quality index of a population meet the published standard?
- Do two populations have the same quality index value?
- Has the quality index of a population decreased?

We design our numerical tools to handle data collected in the following schemes. Consider the situation where lumbers are produced in a population of mills.

• A set of mills are selected with probability proportional to their volume of productions.

• For each sampled mills, a pre-specified number of lumbers are randomly selected.

### 2.2 Typical data structure

The chosen lots from the same year make a population of representative samples of the lumber strength for that certain year. Wood pieces from the same mill have similar solid strength which has been verified by real data. Such a feature seems reasonable. Wood pieces from the same mill are very likely from the same tree, or from trees grown in the same region. They are then very likely to have similar strength. Moreover, the wood strength is also affected by the devices that they are produced from. Wood pieces produced from same devices might have some similarity and this affect their strength. Both of them implies a correlation in wood strength among the wood pieces sampled from the same mill. Therefore, wood pieces from the same mill should be treated as a cluster. The following example presents the information can be contained in a lumber strength data set.

| region | mill  | lot | piece | mor     |
|--------|-------|-----|-------|---------|
| 1      | 1     | 1   | 1     | 2.51250 |
| 1      | 1     | 1   | 2     | 7.92980 |
| 1      | 1     | 1   | 3     | 5.66030 |
| •••    | •••   | ••• | • • • | •••     |
| 1      | 1     | 2   | 11    | 3.76650 |
| 1      | 1     | 2   | 12    | 2.64550 |
| 1      | 1     | 2   | 13    | 3.68530 |
| 1      | 2     | 1   | 1     | 6.85073 |
| 1      | 2     | 1   | 2     | 7.51707 |
| 1      | 2     | 1   | 3     | 4.71568 |
| • • •  | • • • | ••• | • • • | •••     |
| 1      | 2     | 2   | 11    | 5.88808 |
| 1      | 2     | 2   | 12    | 4.91489 |
| 1      | 2     | 2   | 13    | 4.89644 |
| 1      | 2     | 2   | 14    | 6.41309 |
|        | • • • |     |       |         |

In this data set, "region" indicates the place where lumber is collected. "mill" indicates the wood manufactories. "lot" is the index of lots for each manufactory. "mor" is the abbreviation of "modulus of rupture", which is a material property, defined as the stress in a material just before it yields in a flexure test. We can observe from this example data set that 13 pieces of wood from the first mill form a cluster and 14 pieces of wood from the second mill form another cluster.

As mentioned above, to monitor the quality of lumber across years, we need to monitor the values of lumber strength distribution quantiles across years. In other words, if we regard the year when lumber strength observations measured as population index, our target is to compare some lower quantiles of different population distributions and then to figure out if the lumber quality become worse or not.

Focusing on this quantile comparison problem, Chen et al. [2016] use the density ratio model (DRM) to link the lumber strength distributions of different years and use the composite empirical likelihood method to estimate parameters. A hypothesis of the difference of two population quantiles is considered for the purpose of detecting the quantile change among different lumber strength distributions. Such a hypothesis is called monitoring test in the context because its motivation is to monitor lumber quality across years. A cluster-based bootstrap procedure is proposed to make statistical inference on this hypothesis and to construct confidence intervals of desired quantiles. In the paper, their simulation results support that such confidence intervals have satisfactory precise coverage probabilities. This monitoring test controls type I error rates tightly with good power. For convenience, let DRM-CEL denote the method from Chen et al. [2016]. More details of this long term monitoring test are presented in the next chapter.

Although this DRM-CEL-based method is motivated by the project of monitoring lumber quality across years, it is not limited to this specific area. Such a method can be applied to any data set for the purpose of monitoring the quantile change as long as the model assumptions are satisfied. Let  $x_{k,j,l}$  denote the *l*th observation in the *j*th cluster from the *k*th sample. Usually a qualified data set should contain the information of population indices, cluster indices and observed values as follows.

| population | cluster | observed values |
|------------|---------|-----------------|
| 1          | 1       | x_{1,1,1}       |
| 1          | 1       | $x_{\{1,1,2\}}$ |
|            |         |                 |
| 1          | 2       | $x_{\{1,2,1\}}$ |
| 1          | 2       | $x_{\{1,2,2\}}$ |
|            |         | •••             |
| 2          | 1       | x_{2,1,1}       |
| 2          | 1       | $x_{\{2,1,2\}}$ |
|            |         | •••             |

In Chen et al. [2016], a solid theoretical foundation has been established for their DRM-CEL method and its performance has been evaluated by the corresponding simulation study. There are two remaining tasks on this method.

The first one is to fill in the gap between the theoretical results and the corresponding applications. There are already some R program as the realization of this DRM-CEL-based method but the program is developed for the research purpose and is not easy to access for users without much statistical background. For the convenience of users from industry, a user-friendly R package that can be applied easily is needed. This package should contain the functions that can be used to obtain the desired quantile estimates, to visualize the results and to diagnose the results.

The second task is to extend their DRM-CEL quantile estimator and their cluster-based bootstrap method to a more general case. Chen et al. [2016] only consider the scenario that clusters of all populations share a same cluster size. In this thesis, we will use simulation study to investigate the performance of their method in the situation without the constraint of equal cluster size.

# **Chapter 3**

# Numerical tools for DRM-CEL Monitoring Test

Chen and Chen [2000], Chen and Liu [2013] and Chen et al. [2016] have developed an extensive set of statistical tools for data analysis regarding quantiles and quantile functions given multiple samples. They have also subscribed many numerical recipes to carry out the data analysis. At the same time, they have yet to materialize all numerical recipes. Mostly, their developments enable them to investigate the properties of proposed inference methods. In applications, users not only must carry out the data analysis suggested by the authors, but also want to do some exploratory analysis. As an important part of this thesis, we develop some numerical tools for exploratory data analysis as well as tools for more elaborate statistical inference.

As mentioned in Chapter 2, three pieces of information are required for a data set, population indices, cluster indices and observed values. We consider the situation where the samples are organized into a list with the *k*th entry being a matrix or a data frame consisting of observations from the *k*th sample. Specifically, for clustered data, each sample needs to be saved in a matrix with two columns or a data frame with two columns (one column for cluster indices and one column for observations). For the data frame format, users can specify variable names for the cluster indices and the observed values, and the order of two columns. For the matrix format, since its columns can not be named, by default, the first column is for

cluster indices and the second column is for the observed values. For independent data, each sample is just saved in a vector of observed values. Here we use the same data set in Chapter 2 as an example. The variable "mill" is treated as the population index and "lot" is treated as the cluster index.

| region | mill | lot   | piece | mor     |
|--------|------|-------|-------|---------|
| 1      | 1    | 1     | 1     | 2.51250 |
| 1      | 1    | 1     | 2     | 7.92980 |
| 1      | 1    | 1     | 3     | 5.66030 |
| •••    | •••  | •••   | •••   | • • •   |
| 1      | 1    | 2     | 11    | 3.76650 |
| 1      | 1    | 2     | 12    | 2.64550 |
| 1      | 1    | 2     | 13    | 3.68530 |
| 1      | 2    | 1     | 1     | 6.85073 |
| 1      | 2    | 1     | 2     | 7.51707 |
| 1      | 2    | 1     | 3     | 4.71568 |
| •••    |      | • • • | •••   | •••     |
| 1      | 2    | 2     | 11    | 5.88808 |
| 1      | 2    | 2     | 12    | 4.91489 |
| 1      | 2    | 2     | 13    | 4.89644 |
| 1      | 2    | 2     | 14    | 6.41309 |
| •••    | •••  | •••   | •••   |         |

If users want to use a data frame format, two possible ways of the first entry of the list are as follows.

```
lot mor

1 2.51250

1 7.92980

1 5.66030

...

2 3.76650

2 2.64550

2 3.68530
```

or

mor lot 2.51250 1 7.92980 1 5.66030 1 . . . . . . 3.76650 2 2.64550 2 3.68530 2

If users want to use a matrix format, there is only one way for the first entry of the list, which is as follows.

 1
 2.51250

 1
 7.92980

 1
 5.66030

 ...
 ...

 2
 3.76650

 2
 2.64550

 2
 3.68530

In the next section, we introduce the R functions in our package and we present their inputs, outputs with examples and the algorithms behind them.

## **3.1** Model Fitting: drm\_fit(···)

The first numerical task of the DRM-CEL approach is to find the fitted values of  $\boldsymbol{\theta}_r$ : r = 1, 2, ..., m. One must look for

$$G_r(x) = \sum_{k,j,l} p_{k,j,l} \exp\{\boldsymbol{\theta}_r^{\tau} \mathbf{q}(x_{k,j,l})\} \mathbb{1}(x_{k,j,l} \le x)$$

satisfying  $\boldsymbol{\theta}_1 = \mathbf{0}$  and for  $r = 1, 2, \dots, m$ ,

$$\sum_{k,j,l} p_{k,j,l} \exp\{\boldsymbol{\theta}_r^{\tau} \mathbf{q}(x_{k,j,l})\} = 1$$

that maximize

$$L_c(G_1,\ldots,G_m) = \prod_{k,j} \{\prod_{l=1}^d dG_k(x_{k,j,l})\}.$$

Luckily, the numerical solution to the above problem in  $\theta$  is the same as find the maximum point of the following dual likelihood function:

$$\ell_n(\boldsymbol{\theta}) = -\sum_{k,j,l} \log\left[\sum_{r=1}^m \rho_r \exp\{\boldsymbol{\theta}_r^{\tau} \mathbf{q}(x_{k,j,l})\}\right] + \sum_{k,j,l} \boldsymbol{\theta}_k^{\tau} \mathbf{q}(x_{k,j,l})$$
(3.1)

Once its maximum point  $\hat{\boldsymbol{\theta}}$  is obtained, we have

$$\hat{p}_{k,j,l} = \frac{1}{nd\sum_{r=1}^{m}\rho_r \exp\{\hat{\boldsymbol{\theta}}_r^{\tau} \mathbf{q}(x_{k,j,l})\}}.$$

Subsequently, the maximum CEL estimator of  $G_r(x)$  is given by

$$\hat{G}_r(x) = \sum_{k,j,l} \hat{p}_{k,j,l} \exp\{\hat{\boldsymbol{\theta}}_r^{\tau} \mathbf{q}(x_{k,j,l})\} \mathbb{1}(x_{k,j,l} \leq x).$$

See Chen and Liu [2013], Chen et al. [2016] Keziou and Leoni-Aubin [2008], and Qin and Zhang [1997].

It is interesting to find the dual likelihood function coincides with the likelihood function under multinomial logistic regression model Czepiel [2002]. There is a well written R-function "multinom( $\cdots$ )" from a R package called "nnet" for fitting a multinomial logistic regression model by finding the maximum likelihood estimates of parameters Ripley et al. [2016]. We will adopt this function for our purpose and provide some explanations on the connections between the dual likelihood and the likelihood under the multinomial logistic regression model.

Let there be *N* experiment units divided into *m* groups. All responses take values in categories  $\{1, 2, ..., m\}$ . For each unit, the probability that its response falls into category *r* is

$$\pi_r = \Pr\{Y = r\}.$$

Let  $\Delta_{u,r} = 1$  if the response of the *u*th unit is in category *r* and  $\Delta_{u,r} = 0$  otherwise. Let  $\Delta_u$  and  $\delta_u$  be corresponding random vector and observed response vector. We may then write

$$\mathsf{Pr}\{\Delta_u = \delta_u\} = \pi_1^{\delta_{u,1}} \pi_2^{\delta_{u,2}} \cdots \pi_m^{\delta_{u,m}}.$$

Now, assume that  $\pi_r$ 's are not the same for all units but satisfy a logistic regression model: the log-odds of a unit falls into category *r* is a linear function of some covariate **z** 

$$\log\left(\frac{\pi_r}{\pi_1}\right) = \mathbf{z}^{\tau}\boldsymbol{\beta}_r$$

for r = 2, ..., m, where  $\boldsymbol{\beta}_r$  is a regression coefficient vector. We may let  $\boldsymbol{\beta}_1 = \mathbf{0}$  for convenience.

Given data of N units, assume the independence between units and the multinomial logistic regression, we get the log likelihood function

$$\ell_N(\boldsymbol{\beta}) = \sum_{u} \left[ \sum_{r=1}^m \delta_{u,r} \{ \mathbf{z}_u^{\tau} \boldsymbol{\beta}_r \} - \log \left( \sum_{r=1}^m \exp\{ \mathbf{z}_u^{\tau} \boldsymbol{\beta}_r \} \right) \right].$$

Clearly,  $\ell_N(\beta)$  is algebraically similar to the dual empirical likelihood given in (3.1). Hence, "multinom(···)" can be used for DRM-CEL based data analysis. To use this function for DRM-CEL, we need to match the corresponding components to items in the logistic likelihood function.

Let the observations in multiple sample problem be  $\{x_{k,j,l}\}$ , where  $x_{k,j,l}$  is the *l*th observation in cluster *j* of population *k*. The total number of observations is given by  $N = d\sum_k n_k$ . Let  $\mathbf{q}(x)$  be the user specified basis function of length T + 1 and  $(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_m)$  be regression coefficient matrix of size  $(T + 1) \times m$ .

By regarding the *l*th unit in cluster *j* from population *k* as a unit whose response falls into category *k* with  $\mathbf{z}_{k,j,l} = \mathbf{q}(x_{k,j,l})$  being its covariate  $\mathbf{z}_{k,j,l}$ . Using *u* for index (k, j, l), we can see  $\delta_{u,k'} = 1$  when k = k' and  $\delta_{u,k'} = 0$  otherwise. Under the multinomial logistic regression model, we get the log likelihood function

$$\ell_N(\boldsymbol{\beta}) = \sum_{u} \left[ \sum_{r=1}^m \delta_{u,r} \{ \boldsymbol{\beta}_r^{\tau} \mathbf{q}(x_u) \} - \log \left( \sum_{r=1}^m \exp\{ \boldsymbol{\beta}_r^{\tau} \mathbf{q}(x_u) \} \right) \right].$$

Note that  $\delta_{u,r} = 0$  unless u = (r, j, l) for some j, l. The above expression, when

written in terms of (k, j, l), becomes

$$\ell_N(\boldsymbol{\beta}) = \sum_{k,j,l} \left[ \{ \boldsymbol{\beta}_k^{\tau} \mathbf{q}(x_{k,j,l}) \} - \log \left( \sum_{r=1}^m \exp\{ \boldsymbol{\beta}_r^{\tau} \mathbf{q}(x_{k,j,l}) \} \right) \right].$$

Meanwhile, the dual composite empirical likelihood would have been

$$\ell_d(\boldsymbol{\theta}) = \sum_{k,j,l} \boldsymbol{\theta}_k^{\tau} \mathbf{q}(y_{k,j,l}) - \sum_{k,j,l} \log\left(\sum_{r=1}^m \rho_r \exp\{\boldsymbol{\theta}_r^{\tau} \mathbf{q}(y_{k,j,l})\}\right)$$
(3.2)

where  $\rho_r = n_r/n$ . Let

$$\boldsymbol{\beta}_r = \boldsymbol{\theta}_r + (\log \rho_r, 0, \dots, 0)^{\tau},$$

we find

$$\ell_N(\boldsymbol{\beta}) = \ell_d(\boldsymbol{\theta}) + \sum_{k=1}^m n_k \log \rho_k.$$

Hence, if  $\hat{\boldsymbol{\theta}}$  and  $\hat{\boldsymbol{\beta}}$  are the maximum points of  $\ell_d(\boldsymbol{\theta})$  and  $\ell_N(\boldsymbol{\beta})$  respectively, we must have

$$\hat{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\beta}}_k - (\log \rho_r, 0, \dots, 0)^{\tau}.$$

This relationship enables us to use "multinom(...)" to obtain  $\hat{\beta}$  of a multinomial logistic regression and then  $\hat{\theta}$  of DRM.

With the help of "multinom( $\cdots$ )", I build an R function called "drm\_fit( $\cdots$ )" for model fitting.

#### Input:

- z: An optional character indicating the name of data values.
- cluster: An optional character indicating the name of clusters.
- data: A list of *m* samples. If data are clustered, a matrix format is required each sample. If variable names are not provided by the user, by default, the first column of the matrix is cluster identifier; the second column is for the observed values. If data are independent, each sample is saved in a numerical vector.
- basis: A vector of characters defining the basis function  $\mathbf{q}(x)$  required in DRM.

- Examples:

basis=  $c(1, 'x', 'x^2')$ 

- The basis function can be user-defined functions. Example:

 $ff = function(x) \{x + sin(x)\}$ basis=c(1, `ff(x)')

#### **Output**:

- distr\_est: The first output is a matrix with its *r*th column made of likelihood contributions of  $x_{k,j,l}$  for the *r*th distribution function  $\hat{G}_r(\cdot)$  after  $x_{k,j,l}$  is sorted. The estimated distributions  $\hat{G}_r(\cdot)$  can be obtained from this matrix.
- theta\_hat: The second output is a matrix with columns  $\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \dots, \hat{\boldsymbol{\theta}}_m$ .
- data: The third output is a matrix with the second column being the sorted observations, and the first column is its population identifier.

This output is designed to be used as input of other R functions in the package.

### **3.2** Quantile Estimates: quan\_est(···)

After DRM-CEL estimate of the population distribution  $\hat{G}_r$  is obtained, the corresponding quantile estimates are readily computed. We have written an R function quan\_est(...) for this purpose. Some numerical details are as follows.

For population distribution  $G_r$ , we obtain information on the support of  $\hat{G}_r$ , which is the original data set, and corresponding probabilities (or likelihood contributions) on the support.

For any  $\alpha \in (0,1)$ , the DRM-CEL quantile of the population *r* is defined to be

$$\hat{\xi}_r(\alpha) = \inf\{x : \hat{G}_r(x) \ge \alpha\}.$$

The above definition does not ask us to solve equation  $\hat{G}_r(x) = \alpha$  because  $\hat{G}_r(x)$  is not continuous.

To simplify the computation, I use linear interpolation to obtain a continuous distribution function  $\hat{G}'_r(x)$ . After which, I solve  $\hat{G}'_r(x) = \alpha$  and report its solution as  $\hat{\xi}_r(\alpha)$ .

I use an R function wtd.quantile(...) from R package "Hmisc" Harrell and Dupont [2008] with argument: type="i/(n+1)". This function uses linear interpolation.

My R function for computing quantiles is called  $quan\_est(\cdots)$ . Its input and output are as follows:

#### Input:

- fit: an output of  $drm_fit(\cdots)$ .
  - Example:

fit = drm\_fit(...)

• quan: A numeric vector indicating quantile levels to be estimated. It can be a single value or several values.

- Two examples:

quan = 0.05quan = c(0.05, 0.10, 0.15)

#### **Output**:

• If the input of quan is a single value,  $\alpha$ , then the output is a vector containing  $\hat{G}_r^{-1}(\alpha)$  for r = 1, ..., m.

If the input quan is a vector,  $\alpha_1, \alpha_2, ..., \alpha_t$ , then the output is a  $t \times m$  matrix with the *i*th row being a vector containing  $\hat{G}_r^{-1}(\alpha_i)$  for i = 1, ..., t.

### **3.3** Visualization Tool: plot\_cdf(···)

After DRM-CEL estimate of the population distribution  $\hat{G}_r$  is obtained, a user is likely interested to have a visual impression of fitted distributions. We have written an R function plot\_cdf(...) to make this task easy. It depicts the following function

for any r = 1, 2, ..., m:

$$\hat{G}_r(x) = \sum_{k,j,l} \hat{p}_{k,j,l} \exp\{\hat{\boldsymbol{\theta}}_r^{\tau} \mathbf{q}(x_{k,j,l})\} \mathbb{1}(x_{k,j,l} \leq x).$$

Some numerical details are as follows.

For population distribution  $G_r$ , we obtain information on the support of  $\hat{G}_r$ , which is the original data set, and (or likelihood contributions) on the support.

 $G_r(\cdot)$  under the assumption is a continuous distribution while  $\hat{G}_r(\cdot)$  is a discrete distribution function. I first compute values of

$$\hat{G}_r(x) = \sum_{k,j,l} \hat{p}_{k,j,l} \exp\{\hat{\boldsymbol{\theta}}_r^{\tau} \mathbf{q}(x_{k,j,l})\} \mathbb{1}(x_{k,j,l} \leq x).$$

at  $y = x_{k,j,l}$  for all k, j and l.

For a user-specified set of *r* values, this R function plots curves connecting the points  $(x_{k,j,l}, \hat{G}_r(x_{k,j,l}))$  over all *k*, *j* and *l* by a default R function plot(···) with argument

```
plot(\ldots, type = 'l').
```

This argument makes the function  $plot(\cdot)$  plot the points first and then connect these points with line segments.

Its inputs and outputs are given as follows:

Input:

- fit: An output of  $drm_fit(\cdots)$ .
  - Example:

fit = drm\_fit(...)

- index: A numeric vector indicating populations to be plotted. Single entry is also permitted.
- x\_name: An optional character for the name of the x-axis of the plot. If not specified, the x-axis of the plot will not have a name.

**Output**: The plot of  $\hat{G}_r(x)$  of selected populations.

### **3.4** Diagnostic Tools: $plot_kde(\cdots)$ and $plot_qq(\cdots)$

When a DRM with user specified basis function  $\mathbf{q}(x)$  is fitted, a user may wonder whether the data and the fitted model actually match each other. While formal procedures can be developed, one simpler solution is to create a density estimate based on the DRM-CEL fit and compare it with the histogram of the original data; another is to have a quantile-quantile plot of the DRM-CEL estimate  $\hat{G}_r$  against the empirical distribution from the *r*th population.

We use  $plot_kde(\dots)$  to produce a kernel density estimate and plot it with the histogram of the corresponding sample. If the model fits data well, the kernel density estimate should match the histogram well.

We use  $plot_qq(\dots)$  to generate a plot between DRM-CEL-quantiles and sample quantiles of a chosen population. If the model fits data well, these points should approximate a line with slope 1.

Here is the specifications for  $plot_kde(\cdots)$ .

#### Input:

- fit: A list that is the output of  $drm_fit(\cdots)$ .
  - Example:

fit = drm\_fit(...)

- index: identity of the population to be plotted.
- n : the number of grids required for the plot. The default value is 200.
- x\_name: An optional character indicating the x-axis name of the plot. If not specified, the x-axis will not have a name.
- main: An optional character indicating the main title of the plot. If not specified, the plot will not have a title.

**Output**: The output of plot\_kde(···) is a figure of kernel density estimation of  $\hat{G}_k(x)$  together with a histogram of the corresponding sample.
Some details are as follows: we use hist(..., breaks = 'FD', ...) with 'Freedman-Diaconis rule'. It selects the size of the bins according to

Bin size = 
$$2\frac{\text{IQR}(x)}{m^{1/3}}$$
,

where IQR(x) is the interquartile range of the data and *m* is the number of observations in the sample *x*.

We follow Chen and Liu [2013] to produce a kernel density estimate. Let  $K(\cdot)$  be a commonly used kernel function. For some bandwidth b > 0, let  $K_b(x) = (1/b)K(x/b)$ . Then a kernel estimate of  $g_r(x)$  is given by

$$g_r(x) = \int K_b(x-y) d\hat{G}_r(y).$$

In plot\_kde(...), we set K(x) to the standard normal density function. We choose the bandwidth *b* according to the rule of thumb of Silverman [2018].

We select n points uniformly in the range of data at which the kernel density estimates (KDE) are calculated. We obtain the density estimate curve by connecting each pair of consecutive points. We overlay the density estimate over the histogram.

Now we give some details for the other diagnostic tool,  $plot_qq(\cdots)$ , which can be used to produce a quantile-quantile plot stated as the above.

#### Input:

• fit: an output of  $drm_fit(\cdots)$ .

Example: fit = drm\_fit(...)

- index: identity of the population to be plotted.
- n : the number of quantiles used to plot. The default value is 200.
- main: An optional character indicating the main title of the plot. If not specified, the plot will not have a title.

**Output**: The output of plot\_qq(···) is a quantile-quantile plot between  $\hat{G}_r(x)$  and the sample quantiles of the *r*th sample. The *n* quantile levels,  $\alpha_1, ..., \alpha_n$ , are chosen evenly between 1% and 99%.

The linear interpolation is used to obtain quantiles of  $\hat{G}_r(x)$  and empirical quantiles through R function wtd.quantile(...) from the R package "Hmisc' Harrell and Dupont [2008]. We use "plot(...)" with argument "type = p", "col = blue" and "pch = 3" to plot these two vectors of quantiles of length *n*. We add a straight red line starting from the origin with slope 1.

### **3.5** Monitoring Test: monitor\_test(···)

We have a function to carry out monitoring test for the following two types of hypotheses:

$$H_0: \xi_{k_1,\alpha} \le \xi_{k_2,\alpha} \quad vs \quad H_1: \xi_{k_1,\alpha} > \xi_{k_2,\alpha}$$
$$H_0: \xi_{k_1,\alpha} = \xi_{k_2,\alpha} \quad vs \quad H_1: \xi_{k_1,\alpha} \ne \xi_{k_2,\alpha}$$

for any  $\alpha \in (0, 1)$ .

This R function, monitor\_test(...), is designed to test the above hypotheses by the bootstrap method Efron [1992], Hall [1986, 1988]. The output of this R function includes p-values of hypothesis tests of interest and one-sided/ two-sided confidence intervals of  $\xi_{k_2,\alpha} - \xi_{k_1,\alpha}$ . The following shows the inputs and outputs of this function.

#### Input:

- z: An optional character indicating the name of data values.
- cluster: An optional character indicating the name of clusters.
- data: A list containing the observations of variables. The format for data set of this function is the same as the R function drm\_fit(...).
- basis: A vector of characters indicating the basis function in DRM, that is, **q**(*y*).

```
Example: basis= c(1, 'x')
Example: ff = function(x) { x + sin(x)}; basis=c(1, 'ff(x)')
```

• quan: the levels of quantiles  $\alpha$  in the above hypotheses, which can be a single value or a vector of desired values.

**Example:** quan = c(0.05, 0.15)

- Comp: identities of two populations.
- alternative: either "two.sided", "greater" or "less".

Suppose "Comp = c(i, j)". The alternative hypotheses are

two.sided : 
$$\xi_{i,\alpha} \neq \xi_{j,\alpha}$$
;  
greater :  $\xi_{i,\alpha} \ge \xi_{j,\alpha}$ ;  
less :  $\xi_{i,\alpha} \le \xi_{j,\alpha}$ .

- B: number of bootstrap repetitions. The default value is 999.
- ci\_level: nominal level of bootstrapping confidence intervals.

#### **Output**:

- \$p\_value: p-values of desired hypothesis tests.
- \$CI\_1: confidence limits of quantiles  $\xi_{\alpha_1,i}, ..., \xi_{\alpha_L,i}$ .
- \$CI\_2: confidence limits of quantiles  $\xi_{\alpha_1,j}, ..., \xi_{\alpha_1,j}$ .
- \$CI\_diff: confidence limits of  $\xi_{\alpha_l,i} \xi_{\alpha_l,j}$  for l = 1, ..., L
  - If "alternative = 'two.sided' ", confidence limits are two-sided.
  - If "alternative = 'greater' ", then upper confidence limits.
  - If "alternative = 'less' ", then lower confidence limits.

A detailed algorithm of this R function is given as follows.

- We sample  $n_k$  clusters IIDfrom the *k*th sample for k = 1, ..., m, where  $n_k$  is the number of clusters in the *k*th sample. This forms a bootstrap sample.
- Call a function similar to quan\_est(···) to obtain {ξ<sup>\*</sup><sub>k2,α</sub> − ξ<sup>\*</sup><sub>k1,α</sub>}<sup>(b)</sup> for the bootstrap sample.

- Repeat the above to obtain,  $\{\hat{\xi}_{k_2,\alpha}^* \hat{\xi}_{k_1,\alpha}^*\}^{(b)}, b = 1, ..., B.$
- For one-sided alternative, we compute the p-value as the percentage of times when

$$\{\hat{\xi}_{k_2,\alpha}^* - \hat{\xi}_{k_1,\alpha}^*\}^{(b)} > 0$$

in B bootstrap samples.

• For two-sided alternative, we also compute the percentage of times when

$$\{\hat{\xi}_{k_2,\alpha}^* - \hat{\xi}_{k_1,\alpha}^*\}^{(b)} < 0$$

and the twice of the lower percentage is the p-value.

• The solution to the confidence interval is similar.

#### **3.6 Data Example**

In this section, I apply these R functions to a real data set. The dataset contains samples from three populations which will be referred to as 2007, 2010 and 2011. There are 98, 282 and 445 modulus of rupture (MOR) measurements in the three samples, respectively.

I save the MOR measurements from three years into "Imbdata" following the format mentioned before. The 1st, 2nd and 3rd entries of "Imbdata" are MOR data of 2007, 2010 and 2011 respectively. All of the measurements are treated as independent observations.

Several different choices for basis function q(x) are considered. We first fit the model with "drm\_fit(…)" with different basis functions q(x) and then use "plot\_qq(…)" and "plot\_kde(…)" to generate diagnostic plots. Based on these diagnostic plots, we finally use  $q(x) = (1, \log x)$  in the DRM. The choice of basis function has a significant influence on model fitting. Fokianos and Kaimi [2006] quantified the effect of choosing an incorrect linear form of q(x). In general, the point estimate is adversely affect when model is misspecified. For this reason, Fokianos [2007] consider a model selection approach for DRM. Fokianos [2007] suggest to select q(x) as a linear combination of a rich class of functions. The most appropriate q(x) is then determined by selecting a sub-vector of the current q(x). The classical model selection approaches can be applied here. With the following code, we obtain Figures 3.1, 3.2, 3.3, 3.4, 3.5 and 3.6.

```
model fit = drm fit(z = 'obs', cluster = 'id',
            data = lmbdata, basis = c(1, 'log(x)'))
plot_qq(z = 'obs', cluster = 'id', data = lmbdata,
        distr = model_fit, index = 1)
plot qq(z = 'obs', cluster = 'id', data = lmbdata,
        distr = model_fit, index = 2)
plot_qq(z = 'obs', cluster = 'id', data = lmbdata,
        distr = model_fit, index = 3)
plot_kde(z = 'obs', cluster = 'id', data = lmbdata,
         distr = model_fit, n = 1000, index = 1,
         main = 'KDE')
plot_kde(z = 'obs', cluster = 'id', data = lmbdata,
         distr = model fit, n = 1000, index = 2,
         main = 'KDE')
plot_kde(z = 'obs', cluster = 'id', data = lmbdata,
         distr = model_fit, n = 1000, index = 3,
         main = 'KDE')
```

We compute 5%, 10% and 15% quantiles of three populations with "quan\_est( $\cdots$ )".

qq = quan\_est(distr = model\_fit, quan=c(0.05, 0.1, 0.15)) print(qq)

The 5%, 10% and 15% quantile estimates of 2007 are 4.55, 4.97 and 5.26. The 5%, 10% and 15% quantile estimates of 2007 are 4.55, 4.99 and 5.28. The 5%, 10% and 15% quantile estimates of 2007 are 3.51, 4.02 and 4.48.

We now test for the following hypotheses:

$$\begin{aligned} H_0: \xi_{1,0.05} &\leq \xi_{2,0.05} \ vs \ H_1: \xi_{1,0.05} > \xi_{2,0.05}; \\ H_0': \xi_{2,0.05} &\leq \xi_{3,0.05} \ vs \ H_1': \xi_{2,0.05} > \xi_{3,0.05}. \end{aligned}$$

These can be accomplished with the following commands:



Figure 3.1: Q-Q Plot of 2007









Figure 3.3: Q-Q Plot of 2011

KDE



Figure 3.4: Kernel Density Estimate Plot of 2007

P-values of the above hypothesis tests are 0.47 and 0.01. The bootstrapping confidence intervals of  $\xi_{1,0.05} - \xi_{2,0.05}$ , and  $\xi_{2,0.05} - \xi_{3,0.05}$  are  $(-0.34,\infty)$  and  $(0.76,\infty)$ . Since zero falls in the first interval but not the second one, we do not have sufficient evident to reject  $H_0$  but we reject  $H'_0$  at the significance level 5%, which implies that the lumber quality of 2011 very probably become worse compared with that of 2010.

To visualize the estimated distributions, we use the following code to generate a plot of three estimated distributions under DRM.

In Figure 3.7, the black line is the estimated distribution of 2007, the red line is the estimated distribution of 2010 and the green line is the estimated distribution of 2011. From the plot, we notice that the lumber quality of 2010 and 2011 are very probably better than that of 2007.



Figure 3.5: Kernel Density Estimate Plot of 2010



Figure 3.6: Kernel Density Estimate Plot of 2011



Figure 3.7: Estimated Distributions under DRM

## **Chapter 4**

# Monitoring tests for Clustered Data

In Chen et al. [2016], the problem of interest is to efficiently and accurately monitor the change of population quantiles, which are often used as quality indices for wood products. In that paper, the performances of several conventional statistical tests for monitor purpose are discussed and compared. These conventional tests are often designed for assumed i.i.d. samples. When the observations are independent, the nonparametric Wilcoxon rank sum test has promised properties. However when the data are clustered, Wilcoxon rank sum test and most other tests all have inflated type I errors.

If the monitoring targets are lower quantiles, then the Wilcoxon rank sum test is not an appropriate choice in the first place. However, it has always been a familiar tool for users. If the data distribution is not too far from normal, then the change in quantile often leads to other changes for which Wilcoxon rank sum test can detect effectively. Hence, if the Wilcoxon rank sum test can be modified to work for clustered data, it may draw attention to many practitioners. For this purpose, we investigate the performance of a modified large-sample Wilcoxon rank sum test proposed by Rosner et al. [2003] and it incorporates the clustering effects. Rosner et al. [2003] have shown the type I error of this cluster-based Wilcoxon test close to the nominal level when the sample size is not very small. It is of interest to compare it with the DRM-CEL approach of Chen et al. [2016]. To distinguish the "old" Wilcoxon test and the "new" Wilcoxon test, we use "the classical Wilcoxon test" for the test designed for i.i.d. data and use "the modified Wilcoxon test" to for the test designed for clustered data proposed in Rosner et al. [2003].

The classical Wilcoxon test is often presented as an alternative to a t-test when the data are not normally distributed. Whereas a t-test is a test of population means, the Wilcoxon test is commonly regarded as a test of population medians. This is not strictly true, and treating it as such can lead to inadequate analysis of data. The Wilcoxon test is a test of both location and shape. In the case where the only distributional difference is a shift in location, this can indeed be described as a difference in medians.

In this chapter, we first review some details of the classical Wilcoxon rank sum test. In section 4.2, we introduce the modified Wilcoxon rank test for the clustered data proposed by Rosner et al. [2003].

### 4.1 Classical Wilcoxon Rank Sum Test

The Wilcoxon rank sum test is a nonparametric test for the hypothesis that it is equally likely that an observation from one population will be less than or greater than an observation from a second population. Suppose that we have two samples of sizes m and n, respectively. Wilcoxon test is conducted under the assumption that the observations from both samples are independent of each other.

Wilcoxon test works for the null hypothesis

$$H_0: \operatorname{Pr}\{X > Y\} = \operatorname{Pr}\{Y > X\}.$$

The alternative hypothesis  $H_1$  could be a two-sided:  $\Pr\{X > Y\} \neq \Pr\{Y > X\}$ ; or a one-sided such as  $\Pr\{X > Y\} < \Pr\{Y > X\}$ .

We denote the combined sample by *Z*. The sample size of *Z* will be denoted by N = m + n. Define

$$W = \sum_{i=1}^{m} \operatorname{Rank}(X_i)$$

where  $Rank(X_i)$  is the rank of *i*-th observation in the X sample among the combined

Z sample of m + n observations. The expectation and variance of W are given by

$$E(W) = m(N+1)/2$$
(4.1)

$$\mathbb{V}\mathrm{ar}(W) = [mn/\{N(N-1)\}] \sum_{i=1}^{N} (\mathrm{Rank}(Z_i) - \frac{1+N}{2})^2. \tag{4.2}$$

The Wilcoxon rank sum statistic is defined to be

$$T = \{W - E(W)\} / \{\mathbb{V}\mathrm{ar}(W)\}^{1/2},\$$

The test statistic T is asymptotically standard normal. Wilcoxon rank sum test uses the asymptotic distribution to determine the critical regions Lehmann [2006]. The expression 4.2 works well in the situation of independent data. For clustered data set, the observed values from a cluster tend to have similar ranks among the combined sample and therefore, their ranks also form a cluster as well. In this situation, 4.2 may overestimate the variance of W and this leads to a inflated type I error when using Wilcoxon rank sum test on clustered data set.

## 4.2 Modified Wilcoxon Rank Sum Test

The test statistic of the modified Wilcoxon rank sum test is derived in two steps. Consider a simple case that data are balanced so that there are equal number of observations in each cluster and the sub-units in clusters are exchangeable. Let m and n denote the numbers of clusters in X and Y respectively, and let the cluster size be g.

Let us keep *Z* for the combined data set and use  $Z_{ij}$  for *j*th sub-unit of the *i*th cluster. Let  $\delta_i = 1$  if the *i*th cluster is in sample *X* and  $\delta_i = 0$  if it is in sample *Y*. Because the observations in a cluster are assumed exchangeable, we still use *X* and *Y* for single random variables from two populations. The null hypothesis remains to be

$$H_0: \quad \mathsf{Pr}\{X > Y\} = \mathsf{Pr}\{Y > X\}.$$

The alternative hypothesis  $H_1$  again could be a two-sided:  $\Pr\{X > Y\} \neq \Pr\{Y > X\}$ ; or a one-sided such as  $\Pr\{X > Y\} < \Pr\{Y > X\}$ .

Let  $R_{ij}$  be the rank of *j*th sub-unit in the *i*th cluster among all Ng sub-units.

Define

$$W_c = \sum_{i=1}^N \delta_i R_{i+1}$$

where  $R_{i+} = \sum_{j=1}^{g} R_{ij}$ . In other words,  $W_c$  is the same as the classical Wilcoxon rank sum test statistic after regarding sub-units as independent observations from populations *X* and *Y*.

The expectation and variance of  $W_c$  are given by

$$E(W_c) = mg(Ng+1)/2$$
(4.3)

$$\mathbb{V}ar(W_c) = [mn/\{N(N-1)\}] \sum_{i=1}^{N} [R_{i+} - g(1+Ng)/2]^2.$$
(4.4)

The modified Wilcoxon rank sum test statistics  $Z_c$  is defined to be

$$T_c = \{W_c - E(W_c)\} / \{\operatorname{Var}(W_c)\}^{1/2}$$

When N is large,  $T_c$  is asymptotically standard normal.

When cluster sizes are not the same, assume the sub-units in the same cluster are exchangeable. This implies that all units from the same population have the same marginal distribution.

Let  $(m_g, n_g)$  = number of clusters of size g in population X and Y respectively. The cluster size of the *i*th cluster is denoted as  $g_i$ . Let  $g_{max} = \max_i \{g_i\}$ , then  $1 \le g_i \le g_{max}\}$ . Denote  $N_g = m_g + n_g$  and  $N = \sum_{g=1}^{g_{max}} N_g$ . Similarly, let  $R_{ij}$  be the rank of *j*th sub-unit in the *i*th cluster among all  $\sum_{g=1}^{g_{max}} gN_g$  sub-units, and  $R_{i+} = \sum_{j=1}^{g_i} R_{ij}$ . Define

$$W_c = \sum_{i=1}^N \delta_i R_{i+},$$

where  $\delta_i$  is same as the above. Let  $\Delta_{ig} = 1$  if the size of the *i*th cluster is g and

 $\Delta_{ig} = 0$  if its size is not g. The expectation and variance of  $W_c$  are given by

$$E(W_c) = \sum_{g=1}^{g_{max}} m_g [\sum_{i=1}^{N} \Delta_{ig} R_{i+i}] / N_g$$
(4.5)

$$\mathbb{V}\mathrm{ar}(W_c) = \sum_{g=1}^{g_{max}} [n_g m_g / \{N_g (N_g - 1)\}] \sum_{i=1}^{N_g} (\Delta_{ig} R_{i+} - [\sum_{i=1}^N \Delta_{ig} R_{i+}] / N_g)^2.$$
(4.6)

In this situation, they find that

$$T_c = \{W_c - E(W_c)\} / \{\mathbb{V}ar(W_c)\}^{1/2}$$

is asymptotically standard normal.

Compared with 4.2, 4.4 and 4.6 take the effect of clusters into account which improve the accuracy of normal approximation, as supported by simulations in Rosner et al. [2003]. Specifically, the simulation in Rosner et al. [2003] shows that when the sample size is above 80, the normal approximation leads to tests with accurate sizes.

#### **4.3** Wilcoxon rank sum test for monitoring purpose

Based on the simulation results, Rosner et al. [2003] find the type I error of the classical Wilcoxon rank sum test is seriously inflated and the modified Wilcoxon test has an accurate type I error and a good power. Recall that the statistical problem of interest is to test the hypothesis

$$H_0: \xi_{1,\alpha} \leq \xi_{k,\alpha}$$
 against  $H_a: \xi_{1,\alpha} > \xi_{k,\alpha}$ 

where index k refers to the kth population in the multi-sample setting.

Strictly speaking, both Wilcoxon tests are not developed for the comparison of two quantiles. Wilcoxon test works for the null hypothesis

$$H_0: \quad \mathsf{Pr}\{X > Y\} = \mathsf{Pr}\{Y > X\}.$$

The alternative hypothesis  $H_1$  could be a two-sided:  $\Pr\{X > Y\} \neq \Pr\{Y > X\}$ ; or a one-sided such as  $\Pr\{X > Y\} < \Pr\{Y > X\}$ . However, this null hypothesis is commonly regarded as a test of population medians,

$$H_0: \quad X_m = Y_m,$$

where  $X_m$  and  $Y_m$  are medians of X population and Y population, respectively. This premise is based on a misunderstanding of the null hypothesis as pointed out by Kruskal and Wallis [1952]. The practical value of this is hard to see, and thus the null hypothesis of Wilcoxon test sometimes is presented as "the two populations have equal medians". The actual null hypothesis can be expressed as the latter median hypothesis, but only under the additional assumption that the shapes of the distributions are identical in each group. In other words, the interpretation of the test as comparing the medians of the distributions requires the location-shift-only alternative to be the case. For most of cases,

$$Pr\{X > Y\} = Pr\{X < Y\} \Leftrightarrow X_m = Y_m.$$

Even if, we assume the location-shift-only alternative, the Wilcoxon-type tests

still not appropriate for the hypothesis of our interest

$$H_0: \xi_{1,\alpha} \leq \xi_{k,\alpha}$$
 against  $H_a: \xi_{1,\alpha} > \xi_{k,\alpha}$ 

What we would like to show here is that the modified Wilcoxon test can reject  $H_0$  for a wrong reason: an observation from one population has a higher probability of being larger than the observation from another population but their  $\alpha$ th quantiles are equal. We demonstrate this point with a targeted simulation experiment.

Specifically, we carry out a simulation study based on a normal random effect model. We simulate data with two samples from the following normal random effect model:

$$y_{kjl} = \mu_k + \gamma_{kj} + \varepsilon_{kjl}$$

for k = 1, 2, j = 1, ..., 40 and l = 1, ..., d, where  $\gamma_{kj}$  is the random effect and  $\varepsilon_{kjl}$  is the error term. The random effects and error terms are normally distributed independent of each other. Based on this model, we generate two samples each with 40 clusters.

In Scenario 1, we set

$$\begin{split} \mu_1 &= 0.8224, \ \gamma_{2,j} \sim N(0,0.5), \ \varepsilon_{2,j,l} \sim N(0,1.75); \\ \mu_2 &= 0, \ \gamma_{1,j} \sim N(0,0.5), \ \varepsilon_{1,j,l} \sim N(0,0.5). \end{split}$$

We put the cluster size at g = 10. The marginal distribution of  $y_{1,j,l}$  is N(0.8224, 2.25)and the marginal distribution of  $y_{2,j,l}$  is N(0,1). The 5% quantile of N(0.8224, 2.25)is equal to the 5% quantile of N(0,1), and  $Pr(y_{1,\cdot} > y_{2,\cdot}) > Pr(y_{1,\cdot} < y_{2,\cdot})$ .

In Scenario 2, we set

$$\mu_1 = 1.6449, \ \gamma_{2,j} \sim N(0,0.5), \ \varepsilon_{2,j,l} \sim N(0,3.5);$$
  
 $\mu_2 = 0, \ \gamma_{1,j} \sim N(0,0.5), \ \varepsilon_{1,j,l} \sim N(0,0.5).$ 

We put the cluster size at g = 5. The marginal distributions are N(1.6449,4) and N(0,1). The 5% quantiles of N(1.6449,2) and N(0,1) are equal, and  $\Pr(y_{1,\cdot} > y_{2,\cdot}) > \Pr(y_{1,\cdot} < y_{2,\cdot})$ .

In Scenario 3, we set

$$\mu_1 = 0, \ \gamma_{1,j} \sim N(0,0.5), \ \varepsilon_{1,j,l} \sim N(0,0.5);$$
  
 $\mu_2 = 0, \ \gamma_{2,j} \sim N(0,0.5), \ \varepsilon_{2,j,l} \sim N(0,1.75).$ 

We put the cluster size at g = 5. The marginal distributions are N(0, 1) and N(0, 2.25). Their 5% quantiles are not equal, and  $Pr(y_{1,\cdot} > y_{2,\cdot}) = Pr(y_{1,\cdot} < y_{2,\cdot})$ .

In Scenario 4, we set

$$\mu_1 = 0, \ \gamma_{1,j} \sim N(0,0.5), \ \varepsilon_{1,j,l} \sim N(0,0.5);$$
  
$$\mu_2 = 0, \ \gamma_{2,j} \sim N(0,0.5), \ \varepsilon_{2,j,l} \sim N(0,3.5).$$

We put the cluster size at g = 5. The marginal distributions are N(0,1) and N(0,4). The 5% quantiles are not equal, and  $Pr(y_{1,.} > y_{2,.}) = Pr(y_{1,.} < y_{2,.})$ .

Let two pairs of hypotheses be

$$H_0: \xi_{1,\alpha} \leq \xi_{2,\alpha}$$
 against  $H_a: \xi_{1,\alpha} > \xi_{2,\alpha}$ .

and

$$H'_{0}: \Pr(y_{1,\cdot} > y_{2,\cdot}) \leq \Pr(y_{1,\cdot} < y_{2,\cdot}) \text{ against } H'_{a}: \Pr(y_{1,\cdot} > y_{2,\cdot}) > \Pr(y_{1,\cdot} < y_{2,\cdot}).$$

In Scenarios 1 and 2,  $H_0$  is true but  $H'_0$  is not true, and in Scenarios 3 and 4,  $H'_0$  is true but  $H_0$  is not true. The simulated rejection rates are given in Table 4.1

From Table 4.1, in Scenarios 1 and 2, where  $H_0$  is true, the DRM\_CEL method has accurate type I errors when the nominal level of hypothesis test is 5%. However, in these two scenarios, the modified Wilcoxon test rejects  $H_0$  wrongly with rejection rates 94.4% and 100% among 1000 repetitions. In Scenarios 3 and 4, where  $H_0$  is not true, the DRM\_CEL monitoring test rejects  $H_0$  with rejection rates 22.7% and 60.7% among 1000 simulations. However, in these two scenarios, the modified Wilcoxon test only rejects the hypothesis with rates 6.3% and 7.2%. Therefore, if the problem of interest is to compare some quantile differences, the DRM\_CEL monitoring test is more appropriate and reasonable than the modified Wilcoxon

| Setting    | Method                       | Rejection Rate (×100) |
|------------|------------------------------|-----------------------|
| Scenario 1 | DRM_CEL<br>Modified Wilcoxon | 5.1<br>94.4           |
| Scenario 2 | DRM_CEL<br>Modified Wilcoxon | 5.7<br>100            |
| Scenario 3 | DRM_CEL<br>Modified Wilcoxon | 22.7<br>6.3           |
| Scenario 4 | DRM_CEL<br>Modified Wilcoxon | 60.7<br>7.2           |

 Table 4.1: Rejection Rates (%) of DRM\_CEL Monitoring Test and Modified

 Wilcoxon Test at Nominal Level 95%

test.

# 4.4 The DRM\_CEL monitoring test with unequal sized clusters

Chen et al. [2016] develop a DRM\_CEL monitoring test under the assumption that all clusters have the same size. From my observation, this requirement is not necessary in applications. The R-package we developed earlier can be easily applied to clustered data with unequal sized clusters. We are therefore interested to know if the tests still have accurate sizes, namely, whether their type I errors are close to nominal values. In addition, it is of interest to compare the performance of the DRM\_CEL monitoring test and the cluster-based Wilcoxon rank sum test.

We first use simulation to investigate the coverage precision of the confidence intervals constructed by the cluster-based bootstrap for quantiles and quantiles differences. The simulation study sheds light on whether unequal cluster sizes impact the performance of bootstrap-based confidence intervals. We also like to know how large the sample must be before the similar asymptotic results become applicable.

Two data generating models are used in our simulation study. They are given as follows. For convenience, we call them "normal model" and "gamma model" in the following context. • Normal Random Effect Model

$$y_{k,j,l} = \mu_k + \gamma_{k,j} + \varepsilon_{k,j,l}$$

for k = 1, 2, 3, 4,  $j = 1, ..., n_k$  and  $l = 1, ..., d_{k,j}$ , where  $n_k$  is the number of clusters of *k*th populations and  $d_{k,j}$  is the cluster size of the *j*th cluster of the *k*th population.  $\gamma_{k,j}$  is the term of random effect and  $\varepsilon_{k,j,l}$  is the error term. The random effects and error terms are normally distributed and independent of each other.

• Gamma Random Effect Model Nadarajah and Gupta [2006]

Given the population index *k* and cluster index *j*, let  $U_1,...,U_{d_{k,j}}$  be  $d_{k,j}$  i.i.d. random variables with beta distribution having shape parameters *a* and *b* (positive constant). Further let  $W_{k,j}$  be i.i.d. gamma-distributed random variables with shape parameter a + b and rate parameter  $\beta$ .

That is,

$$W_{k,j} \stackrel{i.i.d}{\sim} \operatorname{Gamma}(a+b,\beta);$$
  
 $U_1, \ldots, U_{d_{k,j}} \stackrel{i.i.d}{\sim} \operatorname{Beta}(a,b).$ 

Define

$$\mathbf{Y}_{k,i}^{\tau} = W_{k,j} \times (U_1, \ldots, U_{d_{k,i}}).$$

The distribution of  $\mathbf{Y}^{\tau}$  is the multivariate gamma with  $\operatorname{corr}(Y_i, Y_j) = a/(a + b)$ . The marginal distribution of  $Y_i$  is Gamma with shape parameter a and rate parameter  $\beta$ .

We apply the DRM\_CEL monitoring test via the R function "monitor\_test( $\cdots$ )" to datasets generated from the above models. We summarize the two-sided coverage probabilities of two chosen quantiles and their difference.

In the simulation study, we consider the problem with four populations. Two types of data generating procedures are used. Let vector  $(n_1, n_2, n_3, n_4)$  denote the numbers of clusters of the four samples. The values of  $(n_1, n_2, n_3, n_4)$  are pre-fixed before the simulations. For the first type of data generating procedure, in each repetition, we simulate data as follows.

- We first set a range for cluster sizes (l, l+1, ..., l+p);
- We randomly assign a value from (l, l+1, ..., l+p) to each cluster of each sample;
- Given the values of  $(n_1, n_2, n_3, n_4)$  and the randomized cluster sizes from (l, l+1, ..., l+p), we repeat the following simulation 1000 times:
  - In each repetition, two data sets are simulated from the above two models respectively and we apply the R function "monitor\_test(···)" on them to obtain the results of our interest.

The second type of data generating procedure is as follows.

- We first set a range for cluster sizes (l, l+1, ..., l+p);
- Based on the values of  $(n_1, n_2, n_3, n_4)$  and (l, l+1, ..., l+p), we repeat the following simulation 1000 times:
  - In each repetition, we randomly assign a value from (l, l+1,..., l+p) to each cluster of each sample;
  - Given randomized cluster sizes, two data sets are simulated from the above two models respectively;
  - We apply the R function "monitor\_test(...)" on them to obtain the results of our interest.

In the first type of data generating procedure, cluster sizes are randomly decided before the repetition is started. The same cluster sizes are used over the 1000 repetitions. While in the second type of data generating procedure, the cluster sizes are randomly assigned for every repetition. The cluster sizes are different over the 1000 repetitions. We want to investigate the influence of the two different types of unequal cluster sizes on the DRM\_CEL cluster-based bootstrap method.

For both data generating procedure, several sets of the sample sizes  $n_1$ ,  $n_2$ ,  $n_3$  and  $n_4$  are chosen in the simulation. Given  $(n_1, n_2, n_3, n_4)$ , we use different ranges of cluster sizes and parameter values. A wider range of cluster sizes means there is more difference between clusters and different parameter values in the data generating models implies different correlation strengths within clusters.

Specifically, we simulate data with two settings:  $(n_1, n_2, n_3, n_4) = (38, 45, 60, 60)$ and  $(n_1, n_2, n_3, n_4) = (63, 75, 100, 100)$ . For given  $(n_1, n_2, n_3, n_4)$ , the range of cluster sizes is chosen as (5, 6, 7, 8) to (3, ..., 10).

We calculate the ratios of the true values of target quantiles and their difference falling in the bootstrap intervals, which give the coverage probabilities. Tables 4.2 and 4.3 present coverage probabilities based on the first type of simulated data from the normal and gamma models, respectively. Tables 4.4 and 4.5 present coverage probabilities based on the second type of simulated data from two models.

Based on the simulation results, we find the coverage probabilities are closer to the nominal 95% when the sample sizes are larger under the normal model. The precision is also higher when the error variances are smaller. Although there are many occasions where the simulated coverage probability is as low as 91%, a large proportion of them are near 94% or higher, but never too much higher than 95%. Hence, we conclude that the DRM-CEL and bootstrap based confidence intervals have satisfactorily precise coverage probabilities even when data contain unequal sized clusters.

| Cluster Sizes | Target                    | $\alpha = 0.05$                             | $\alpha = 0.10$                             | $\alpha = 0.05$   | $\alpha = 0.10$                             |  |
|---------------|---------------------------|---|---|---|---|--|
|               |                           | $(n_1, n_2, n_3, n_4) = (38, 45, 60, 60)$   |   |   |   |  |
|               |                           | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $(\sigma_{\gamma,3}^2,\sigma_{\gamma,4}^2)$ | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2,\sigma_{\gamma,3}^2,\sigma_{\gamma,4}^2)$ |   |  |
|               |                           | =(2.25, 2.2                                 | 5, 1.00, 1.00)                              | =(1.44, 1.4   | 4, 1.00, 1.00)                              |  |
| 5 - 8         | $\xi_{1,\alpha}$          | 92.2  | 91.4  | 92.8  | 91.6  |  |
|               | $\xi_{2,lpha}$            | 91.0  | 93.3  | 93.6  | 93.1  |  |
|               | $\Delta \xi_{1,2,lpha}$   | 92.8  | 93.8  | 94.6  | 93.3  |  |
| 4 - 9         | $\xi_{1,\alpha}$          | 91.2  | 92.0  | 92.7  | 93.1  |  |
|               | $\xi_{2,lpha}$            | 91.7  | 92.9  | 93.5  | 94.0  |  |
|               | $\Delta \xi_{1,2,lpha}$   | 93.6  | 94.9  | 94.7  | 94.8  |  |
| 3 - 10        | $\xi_{1,\alpha}$          | 91.7  | 91.5  | 91.9  | 91.5  |  |
|               | $\xi_{2,lpha}$            | 90.7  | 92.3  | 92.0  | 92.5  |  |
|               | $\Delta \xi_{1,2,lpha}$   | 93.0  | 94.0  | 94.3  | 93.7  |  |
|               |                           | (1  | $(n_1, n_2, n_3, n_4) = (50, 60, 80, 80)$   |   |   |  |
|               |                           | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $(\sigma_{\gamma,3}^2,\sigma_{\gamma,4}^2)$ | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$   | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ |  |
|               |                           | =(2.25, 2.2                                 | 5, 1.00, 1.00)                              | =(1.44, 1.4   | 4, 1.00, 1.00)                              |  |
| 5 - 8         | $\xi_{1,\alpha}$          | 91.9  | 91.4  | 93.9  | 91.6  |  |
|               | $\xi_{2,lpha}$            | 93.4  | 93.3  | 94.6  | 92.9  |  |
|               | $\Delta \xi_{1,2,lpha}$   | 93.4  | 93.8  | 95.1  | 95.1  |  |
| 4 - 9         | $\xi_{1,\alpha}$          | 91.6  | 92.0  | 92.9  | 92.8  |  |
|               | $\xi_{2,lpha}$            | 93.2  | 92.9  | 94.1  | 93.1  |  |
|               | $\Delta \xi_{1,2,lpha}$   | 95.0  | 94.9  | 95.8  | 95.9  |  |
| 3 - 10        | $\xi_{1,\alpha}$          | 91.6  | 91.5  | 92.2  | 92.5  |  |
|               | $\xi_{2,lpha}$            | 91.4  | 92.3  | 92.7  | 92.8  |  |
|               | $\Delta \xi_{1,2,lpha}$   | 94.0  | 94.0  | 94.4  | 94.1  |  |
|               |                           | ( <i>n</i> ]                                | $(n_1, n_2, n_3, n_4) =$                    | (63, 75, 100,   | 100)  |  |
|               |                           | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$   | $(\sigma_{\gamma,3}^2,\sigma_{\gamma,4}^2)$ |  |
|               |                           | =(2.25, 2.2                                 | 5, 1.00, 1.00)                              | =(1.44, 1.4   | 4, 1.00, 1.00)                              |  |
| 5 - 8         | $\xi_{1,\alpha}$          | 93.5  | 92.4  | 94.0  | 93.3  |  |
|               | $\xi_{2,lpha}$            | 94.5  | 91.5  | 94.3  | 93.3  |  |
|               | $\Delta \xi_{1,2,lpha}$   | 94.4  | 94.5  | 94.6  | 94.5  |  |
| 4 - 9         | $\xi_{1,\alpha}$          | 91.9  | 92.5  | 93.7  | 93.0  |  |
|               | $\xi_{2,lpha}$            | 93.2  | 92.6  | 93.8  | 92.3  |  |
|               | $\Delta \xi_{1,2,lpha}$   | 95.3  | 94.9  | 94.7  | 94.6  |  |
| 3 - 10        | $\xi_{1,\alpha}$          | 92.4  | 91.6  | 93.0  | 92.9  |  |
|               | $\xi_{2,lpha}$            | 93.1  | 92.3  | 93.6  | 93.2  |  |
|               | $\Delta \xi_{1,2,\alpha}$ | 94.5  | 93.9  | 94.3  | 94.4  |  |

**Table 4.2:** Coverage probabilities (%) of two-sided 95% bootstrap confidence intervals under Normal models (type one)

| Cluster Sizes | Target                  | $\alpha = 0.05$                           | $\alpha = 0.10$     | $\alpha = 0.05$   | $\alpha = 0.10$ |  |
|---------------|-------------------------|---|---------------------|-------------------|-----------------|--|
|               |                         | $(n_1, n_2, n_3, n_4) = (38, 45, 60, 60)$ |                     |                   |                 |  |
|               |                         | b = 14                                    |                     | b = 36            |                 |  |
| 5 - 8         | $\xi_{1,\alpha}$        | 91.4                                      | 91.5                | 93.5              | 93.2            |  |
|               | $\xi_{2,lpha}$          | 90.7                                      | 92.1                | 93.2              | 94.0            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.2                                      | 93.6                | 94.1              | 93.8            |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 91.8                                      | 93.7                | 93.2              | 94.6            |  |
|               | $\xi_{2,lpha}$          | 92.5                                      | 94.5                | 93.2              | 95.2            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.0                                      | 92.9                | 94.2              | 94.6            |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 93.4                                      | 92.5                | 93.7              | 94.5            |  |
|               | $\xi_{2,lpha}$          | 93.8                                      | 93.0                | 94.5              | 94.7            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.1                                      | 93.2                | 95.0              | 95.3            |  |
|               |                         | ( <i>n</i> <sub>1</sub>                   | $(n_2, n_3, n_4) =$ | =(50, 60, 80, 80) |                 |  |
|               |                         | b =                                       | = 14                | b = 36            |                 |  |
| 5 - 8         | ξ1,α                    | 92.4                                      | 93.8                | 93.3              | 92.4            |  |
|               | $\xi_{2,lpha}$          | 93.2                                      | 93.7                | 93.2              | 93.3            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.9                                      | 94.3                | 94.4              | 94.3            |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 95.1                                      | 94.8                | 94.7              | 94.2            |  |
|               | $\xi_{2,lpha}$          | 94.6                                      | 94.3                | 94.7              | 94.9            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 95.2                                      | 94.7                | 95.0              | 95.1            |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 93.6                                      | 93.4                | 93.3              | 94.1            |  |
|               | $\xi_{2,lpha}$          | 94.1                                      | 92.4                | 94.4              | 94.5            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.8                                      | 94.0                | 94.6              | 95.2            |  |
|               |                         | $(n_1, n_2)$                              | $n_2, n_3, n_4) =$  | (63, 75, 100,     | 100)            |  |
|               |                         | <i>b</i> =                                | = 14                | <i>b</i> =        | = 36            |  |
| 5 - 8         | $\xi_{1,\alpha}$        | 92.1                                      | 92.4                | 93.0              | 92.7            |  |
|               | $\xi_{2,lpha}$          | 92.5                                      | 93.0                | 92.6              | 92.3            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.7                                      | 94.5                | 93.7              | 93.4            |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 93.2                                      | 94.6                | 94.2              | 94.4            |  |
|               | $\xi_{2,lpha}$          | 94.0                                      | 94.8                | 95.4              | 94.8            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.6                                      | 94.3                | 94.9              | 94.9            |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 93.5                                      | 92.9                | 94.5              | 93.8            |  |
|               | $\xi_{2,lpha}$          | 93.9                                      | 93.5                | 94.4              | 94.8            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.5                                      | 94.5                | 95.2              | 95.2            |  |

**Table 4.3:** Coverage probabilities (%) of two-sided 95% bootstrap confidence intervals under Gamma models (type one)

| Cluster Sizes | Target                  | $\alpha = 0.05$                             | $\alpha = 0.10$                             | $\alpha = 0.05$                             | $\alpha = 0.10$                             |  |
|---------------|-------------------------|---|---|---|---|--|
|               |                         | $(n_1, n_2, n_3, n_4) = (38, 45, 60, 60)$   |   |   |   |  |
|               |                         | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ |  |
|               |                         | =(2.25, 2.2                                 | 5, 1.00, 1.00)                              | =(1.44, 1.4                                 | 4, 1.00, 1.00)                              |  |
| 5 - 8         | $\xi_{1,lpha}$          | 93.3  | 92.7  | 93.5  | 93.0  |  |
|               | $\xi_{2,lpha}$          | 92.5  | 93.4  | 93.7  | 92.5  |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.2  | 92.9  | 94.5  | 94.2  |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 91.3  | 92.0  | 94.3  | 94.1  |  |
|               | $\xi_{2,lpha}$          | 91.5  | 92.4  | 94.9  | 93.0  |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.0  | 93.6  | 94.1  | 93.8  |  |
| 3 - 10        | $\xi_{1,lpha}$          | 92.3  | 93.2  | 93.1  | 92.7  |  |
|               | $\xi_{2,lpha}$          | 91.9  | 92.3  | 93.8  | 91.8  |  |
|               | $\Delta \xi_{1,2,lpha}$ | 95.2  | 95.2  | 94.8  | 94.4  |  |
|               |                         | ()  | $(n_1, n_2, n_3, n_4) = (50, 60, 80, 80)$   |   |   |  |
|               |                         | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ |  |
|               |                         | =(2.25, 2.2                                 | 5, 1.00, 1.00)                              | =(1.44, 1.4                                 | 4, 1.00, 1.00)                              |  |
| 5 - 8         | $\xi_{1,lpha}$          | 91.9  | 93.4  | 94.4  | 92.8  |  |
|               | $\xi_{2,lpha}$          | 92.0  | 93.3  | 94.4  | 92.7  |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.6  | 94.1  | 93.6  | 93.6  |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 91.9  | 93.4  | 93.9  | 94.3  |  |
|               | $\xi_{2,lpha}$          | 93.1  | 92.3  | 93.7  | 94.3  |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.7  | 94.3  | 94.9  | 95.4  |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 91.4  | 92.9  | 94.2  | 94.0  |  |
|               | $\xi_{2,lpha}$          | 93.8  | 93.7  | 93.9  | 93.0  |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.7  | 95.2  | 94.9  | 95.2  |  |
|               |                         | ( <i>n</i>                                  | $(n_1, n_2, n_3, n_4) =$                    | (63, 75, 100,                               | 100)  |  |
|               |                         | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ |  |
|               |                         | =(2.25, 2.2                                 | 5, 1.00, 1.00)                              | =(1.44, 1.4                                 | 4, 1.00, 1.00)                              |  |
| 5 - 8         | ξ1,α                    | 93.1  | 94.5  | 93.1  | 93.2  |  |
|               | $\xi_{2,lpha}$          | 92.8  | 93.1  | 93.6  | 92.7  |  |
|               | $\Delta \xi_{1,2,lpha}$ | 95.3  | 95.4  | 92.3  | 92.2  |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 93.3  | 94.5  | 94.0  | 93.5  |  |
|               | $\xi_{2,lpha}$          | 94.3  | 94.3  | 94.0  | 94.2  |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.9  | 94.7  | 94.4  | 93.8  |  |
| 3 - 10        | $\xi_{1,lpha}$          | 93.0  | 93.7  | 93.7  | 92.4  |  |
|               | $\xi_{2,lpha}$          | 92.4  | 92.3  | 92.8  | 93.2  |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.6  | 93.5  | 93.5  | 93.9  |  |

**Table 4.4:** Coverage probabilities (%) of two-sided 95% bootstrap confidenceintervals under Normal models (type two)

| Cluster Sizes | Target                  | $\alpha = 0.05$                           | $\alpha = 0.10$     | $\alpha = 0.05$   | $\alpha = 0.10$ |  |
|---------------|-------------------------|---|---------------------|-------------------|-----------------|--|
|               |                         | $(n_1, n_2, n_3, n_4) = (38, 45, 60, 60)$ |                     |                   |                 |  |
|               |                         | b = 14                                    |                     | b = 36            |                 |  |
| 5 - 8         | $\xi_{1,\alpha}$        | 92.4                                      | 92.0                | 91.7              | 92.9            |  |
|               | $\xi_{2,lpha}$          | 92.1                                      | 92.5                | 91.7              | 92.7            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.6                                      | 94.8                | 94.1              | 94.6            |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 93.2                                      | 93.7                | 93.9              | 93.8            |  |
|               | $\xi_{2,lpha}$          | 93.1                                      | 94.6                | 93.5              | 94.1            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.1                                      | 93.9                | 93.9              | 94.1            |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 93.1                                      | 92.8                | 93.9              | 95.3            |  |
|               | $\xi_{2,lpha}$          | 93.0                                      | 92.4                | 93.8              | 94.2            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.4                                      | 93.5                | 93.9              | 94.2            |  |
|               |                         | ( <i>n</i> <sub>1</sub>                   | $(n_2, n_3, n_4) =$ | =(50, 60, 80, 80) |                 |  |
|               |                         | b =                                       | - 14                | b = 36            |                 |  |
| 5 - 8         | ξ1,α                    | 91.7                                      | 93.4                | 93.1              | 93.6            |  |
|               | $\xi_{2,lpha}$          | 92.7                                      | 94.2                | 93.5              | 94.0            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.9                                      | 94.2                | 93.7              | 93.4            |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 94.6                                      | 93.0                | 92.6              | 93.5            |  |
|               | $\xi_{2,lpha}$          | 94.3                                      | 92.9                | 92.6              | 94.0            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.6                                      | 94.1                | 94.2              | 94.1            |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 93.3                                      | 93.0                | 93.3              | 94.9            |  |
|               | $\xi_{2,lpha}$          | 93.1                                      | 93.6                | 93.1              | 94.5            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.1                                      | 93.9                | 94.2              | 94.8            |  |
|               |                         | $(n_1, n_2)$                              | $n_2, n_3, n_4) =$  | (63, 75, 100,     | 100)            |  |
|               |                         | <i>b</i> =                                | = 14                | <i>b</i> =        | = 36            |  |
| 5 - 8         | $\xi_{1,\alpha}$        | 92.0                                      | 93.7                | 94.4              | 94.7            |  |
|               | $\xi_{2,lpha}$          | 92.0                                      | 94.3                | 94.4              | 93.8            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.8                                      | 94.8                | 93.1              | 93.5            |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 93.5                                      | 94.0                | 93.4              | 93.2            |  |
|               | $\xi_{2,lpha}$          | 93.7                                      | 94.0                | 94.4              | 94.7            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.0                                      | 94.0                | 95.6              | 95.7            |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 93.4                                      | 93.0                | 93.9              | 93.4            |  |
|               | $\xi_{2,lpha}$          | 93.2                                      | 94.4                | 94.5              | 95.0            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.5                                      | 94.1                | 95.1              | 94.9            |  |

**Table 4.5:** Coverage probabilities (%) of two-sided 95% bootstrap confidenceintervals under Gamma models (type two)

## 4.5 Performance of DRM\_CEL Monitor Test when Model is Misspecified

In applications, the true distribution of observations is usually unknown. Any statistical model is just an approximation to the true distribution of data. Usually, instead of choosing a known basis function q(x), one may select q(x) as a rich class of functions and apply some classical model selection approaches to determine a subset of the class of functions as the basis function of DRM. The quantilequantile plot and density plot are two ways to judge the goodness-of-fit. Suppose that density ratio model is being applied to a data set. Without the full information of the true distribution of data, the basis function q(x) usually can not be chosen "correctly". Consequently, the influence of using an "incorrect" basis function for DRM is of big concern. For this reason, Fokianos and Kaimi [2006] study the impact of using an "incorrect" basis function on the estimation of  $\boldsymbol{\theta}$ , the parameters of DRM. Such an impact is serious. In Chen and Liu [2013], their simulation results based on independent data show that the quantile estimation are not as badly affected by the misspecified model as the parameter estimation. In this section, we study the same problem with clustered data. We would like to use simulation study to investigate the performance of DRM\_CEL quantile estimation when an "incorrect" basis function is used, that is, the model is misspecified. Our purpose is to evaluate the impact of misspecified model on the DRM\_CEL quantile estimations.

We have used simulation study to investigate the performance of cluster-based Wilcoxon test for quantile comparison. Even though in Rosner et al. [2003], the cluster-based Wilcoxon test has controlled type I error and good performance in the case of relative small sample size, there is a significant potential risk of using it in quantile comparison. Therefore, we do not consider this method as a comparison subject anymore.

We simulate random samples from the same Normal and Gamma models as before. As a trade-off between a well fitted model and a parsimonious model, we choose  $\mathbf{q}(x) = (1, x, \log(1 + |x|), |x|^{3/2})$ . Two Scenarios are considered for both models. In scenario 1, all clusters have a same size. In scenario 2, cluster sizes are not equal.

For each model set up, we generate 1000 samples. For each sample generated,

999 bootstrap repetitions are used to build two-sided 95% bootstrapping confidence intervals for multiple quantiles of all samples and their difference as well. We report the coverage probabilities of these confidence intervals based on 1000 repetitions. The simulation results are summarized in Tables 4.6 and 4.7.

| Cluster Sizes | Target                  | $\alpha = 0.05$                             | $\alpha = 0.10$                             | $\alpha = 0.05$                             | $\alpha = 0.10$                             |
|---------------|-------------------------|---|---|---|---|
|               |                         | $(n_1, n_2, n_3, n_4) = (38, 45, 60, 60)$   |   |   |   |
|               |                         | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ |
|               |                         | =(2.25, 2.2                                 | 5, 1.00, 1.00)                              | =(1.44, 1.4                                 | 4, 1.00, 1.00)                              |
| 5 - 8         | ξ1,α                    | 91.8  | 90.8  | 92.8  | 91.6  |
|               | $\xi_{2,lpha}$          | 93.5  | 93.1  | 93.6  | 93.1  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.6  | 92.2  | 94.6  | 93.3  |
| 4 - 9         | $\xi_{1,lpha}$          | 91.6  | 90.1  | 92.7  | 93.1  |
|               | $\xi_{2,lpha}$          | 92.9  | 91.6  | 93.5  | 94.0  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.4  | 93.8  | 94.7  | 94.8  |
| 3 - 10        | ξ1,α                    | 90.6  | 92.0  | 91.9  | 91.5  |
|               | ξ2,α                    | 91.8  | 93.8  | 92.0  | 92.5  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.6  | 93.4  | 94.3  | 93.7  |
|               |                         | ()  | $(n_1, n_2, n_3, n_4) =$                    | = (50, 60, 80,                              | 80)   |
|               |                         | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $(\sigma_{\gamma,3}^2,\sigma_{\gamma,4}^2)$ | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ |
|               |                         | =(2.25, 2.2                                 | 5, 1.00, 1.00)                              | =(1.44, 1.4                                 | 4, 1.00, 1.00)                              |
| 5 - 8         | ξ1,α                    | 92.1  | 91.8  | 93.9  | 91.6  |
|               | $\xi_{2,lpha}$          | 93.6  | 92.9  | 94.6  | 92.9  |
|               | $\Delta \xi_{1,2,lpha}$ | 95.3  | 94.8  | 95.1  | 95.1  |
| 4 - 9         | $\xi_{1,lpha}$          | 92.4  | 91.4  | 92.9  | 92.8  |
|               | $\xi_{2,lpha}$          | 94.4  | 92.5  | 94.1  | 93.1  |
|               | $\Delta \xi_{1,2,lpha}$ | 92.7  | 94.1  | 95.8  | 95.9  |
| 3 - 10        | $\xi_{1,lpha}$          | 92.8  | 91.8  | 92.2  | 92.5  |
|               | $\xi_{2,lpha}$          | 93.1  | 92.3  | 92.7  | 92.8  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.8  | 94.0  | 94.4  | 94.1  |
|               |                         | ( <i>n</i>                                  | $(n_1, n_2, n_3, n_4) =$                    | (63, 75, 100,                               | 100)  |
|               |                         | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ | $(\sigma_{\gamma,1}^2,\sigma_{\gamma,2}^2)$ | $\sigma_{\gamma,3}^2, \sigma_{\gamma,4}^2)$ |
|               |                         | =(2.25, 2.2                                 | 5, 1.00, 1.00)                              | =(1.44, 1.4                                 | 4, 1.00, 1.00)                              |
| 5 - 8         | $\xi_{1,\alpha}$        | 92.3  | 93.3  | 94.0  | 93.3  |
|               | $\xi_{2,lpha}$          | 92.7  | 93.7  | 94.3  | 93.3  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.8  | 93.0  | 94.6  | 94.5  |
| 4 - 9         | $\xi_{1,lpha}$          | 91.8  | 92.8  | 93.7  | 93.0  |
|               | $\xi_{2,lpha}$          | 92.8  | 93.0  | 93.8  | 92.3  |
|               | $\Delta \xi_{1,2,lpha}$ | 92.4  | 93.3  | 94.7  | 94.6  |
| 3 - 10        | $\xi_{1,lpha}$          | 92.5  | 92.4  | 93.0  | 92.9  |
|               | $\xi_{2,lpha}$          | 93.6  | 94.1  | 93.6  | 93.2  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.0  | 94.6  | 94.3  | 94.4  |

**Table 4.6:** Coverage probabilities (%) of two-sided 95% bootstrap confidence intervals under Normal models

| Cluster Sizes | Target                  | $\alpha = 0.05$                           | $\alpha = 0.10$     | $\alpha = 0.05$ | $\alpha = 0.10$ |  |
|---------------|-------------------------|---|---------------------|-----------------|-----------------|--|
|               |                         | $(n_1, n_2, n_3, n_4) = (38, 45, 60, 60)$ |                     |                 |                 |  |
|               |                         | b =                                       | = 36                | b =             | = 14            |  |
| 5 - 8         | $\xi_{1,\alpha}$        | 91.9                                      | 90.5                | 92.1            | 90.5            |  |
|               | $\xi_{2,lpha}$          | 92.9                                      | 92.0                | 93.2            | 91.4            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 92.9                                      | 93.2                | 94.1            | 94.9            |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 93.1                                      | 90.6                | 91.2            | 91.0            |  |
|               | $\xi_{2,lpha}$          | 93.2                                      | 92.2                | 91.6            | 91.4            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.3                                      | 93.6                | 94.3            | 94.4            |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 91.1                                      | 92.3                | 90.1            | 91.5            |  |
|               | $\xi_{2,lpha}$          | 92.7                                      | 92.9                | 91.2            | 91.9            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.2                                      | 94.2                | 91.7            | 91.8            |  |
|               |                         | ( <i>n</i> <sub>1</sub>                   | $(n_2, n_3, n_4) =$ | = (50,60,80,80) |                 |  |
|               |                         | <i>b</i> = 36                             |                     | b = 14          |                 |  |
| 5 - 8         | $\xi_{1,\alpha}$        | 93.3                                      | 91.1                | 90.2            | 92.3            |  |
|               | $\xi_{2,lpha}$          | 93.2                                      | 92.5                | 91.1            | 93.4            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.3                                      | 93.3                | 93.4            | 93.0            |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 93.7                                      | 91.8                | 91.4            | 91.7            |  |
|               | $\xi_{2,lpha}$          | 93.3                                      | 92.1                | 93.1            | 92.6            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.6                                      | 94.9                | 94.8            | 94.3            |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 92.6                                      | 92.0                | 91.9            | 91.6            |  |
|               | $\xi_{2,lpha}$          | 93.3                                      | 92.4                | 92.0            | 92.6            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.8                                      | 93.9                | 94.6            | 94.7            |  |
|               |                         | $(n_1, n_2)$                              | $n_2, n_3, n_4) =$  | (63, 75, 100,   | 53,75,100,100)  |  |
|               |                         | <i>b</i> =                                | = 36                | b = 14          |                 |  |
| 5 - 8         | $\xi_{1,\alpha}$        | 92.9                                      | 93.0                | 91.0            | 92.6            |  |
|               | $\xi_{2,lpha}$          | 93.8                                      | 93.2                | 93.0            | 93.8            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 92.7                                      | 93.2                | 93.7            | 93.4            |  |
| 4 - 9         | $\xi_{1,\alpha}$        | 94.2                                      | 94.0                | 91.0            | 90.6            |  |
|               | $\xi_{2,lpha}$          | 94.8                                      | 94.1                | 91.9            | 91.6            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 94.7                                      | 94.8                | 92.5            | 92.9            |  |
| 3 - 10        | $\xi_{1,\alpha}$        | 91.7                                      | 92.8                | 92.5            | 92.6            |  |
|               | $\xi_{2,lpha}$          | 92.8                                      | 92.9                | 92.6            | 93.4            |  |
|               | $\Delta \xi_{1,2,lpha}$ | 93.7                                      | 93.7                | 94.0            | 94.3            |  |

**Table 4.7:** Coverage probabilities (%) of two-sided 95% bootstrap confidence intervals under Gamma models

From Table 4.6 and 4.7, we notice that the impact of misspecified model on the

coverage probabilities of 5% and 10% quantiles of populations 1 and 2. Specifically, the coverage probabilities tend to be lower than the nominal level 95%. But the simulated coverage probabilities of quantile differences do not suffer too much from the misspecified model. This might be because the quantile estimates of both populations tend to be lower or larger than the true values simultaneously so that their differences are still close to the true values. Based on these simulation results, we can conclude that the choice of basis function indeed has some influence on the accuracy of quantile estimation. But as long as we choose an appropriate basis function such that the model fits data well, the interval estimation of population quantile is still in a reasonable range. Moreover, the interval estimation of quantile difference is still very reliable.

## **Chapter 5**

# Summary

The main interest of this thesis lies in the demonstration of our R package for DRM\_CEL quantile estimation and how to use the package to do further research. With the help of this package, we compare the performance of DRM\_CEL quantile estimation with cluster-based Wilcoxon rank sum test in different situations. The cluster-based Wilcoxon test as a nonparametric method, needs less assumptions on the underlying population distributions. But for the purpose of monitoring population quantiles change, this Wilcoxon test is not testing the hypothesis of interest and DRM\_CEL is more appropriate here. Taking advantage of the package, we use simulation study to investigate the performance of DRM\_CEL quantile in the case that cluster sizes are not equal, which is not considered in Chen et al. [2016]. We imitate the same cluster-based bootstrap procedure from that paper to conduct hypothesis tests and to build confidence intervals. Simulation results supports that the conclusions in Chen et al. [2016] still hold in the general situations.

As future research, adding the multi-parameter one-sided monitoring test by Zhu and Chen [2017] into our current R package is of interest. This multi-parameter one-sided monitoring test is also motivated by monitoring multiple quality indices in forestry products. In the literature, they find a novel way to derive a likelihood ratio test (LRT) type statistic for monitoring test of interest. Compared with the classical LRT, the new test retains good control of the type I error and is markedly more powerful. We would like to organize this new test into our package. Besides, the simulation results support our guess that the DRM\_CEL monitoring test should work in the general case. We would like to investigate the theory behind it.

# **Bibliography**

- J.A. Anderson. Multivariate logistic compounds. Biometrika, 66(1):17–26, 1979.  $\rightarrow$  pages 2, 3, 4
- J.D. Barrett and R.M. Kellogg. Bending strength and stiffness of second growth douglas-fir dimension lumber. Forest products journal (USA), 1991. → pages 2
- B.A. Bendtsen and J. Senft. Mechanical and anatomical properties in individual growth rings of plantation-grown eastern cottonwood and loblolly pine. Wood and Fiber Science, 18(1):23-38, 2007.  $\rightarrow$  pages 2
- D. Bertsimas, G.J. Lauprete, and A. Samarov. Shortfall as a risk measure: properties, optimization and applications. Journal of Economic Dynamics and control, 28(7):1353–1381, 2004. → pages 1
- H. Bier and M.J. Collins. Bending Properties of 100 x 50 mm Structural Timber from a 28-year-old Stand of New Zealand Radiata Pine. New Zealand Forest Service, 1985.  $\rightarrow$  pages 2
- R. S. Boone and M. Chudnoff. Compression wood formation and other characteristics of plantation-grown pinus caribaea. Forest Service Research Paper, Institute of Tropical Forestry, Puerto Rico, (ITF-13), 1972.  $\rightarrow$  pages 2
- H. Chen and J. Chen. Bahadur representations of the empirical likelihood quantile processes. Journal of Nonparametric Statistics, 12(5):645–660, 2000. → pages 5, 17
- J. Chen and Y. Liu. Quantile and quantile-function estimations under density ratio model. <u>The Annals of Statistics</u>, 41(3):1669–1692, 2013. → pages 6, 7, 10, 17, 20, 27, 53
- J. Chen, P. Li, Y. Liu, and J.V. Zidek. Monitoring test under nonparametric random effects model. <u>arXiv preprint arXiv:1610.05809</u>, 2016.  $\rightarrow$  pages iii, 1, 3, 6, 8, 9, 10, 11, 12, 15, 16, 17, 20, 37, 45, 58

- S.A. Czepiel. Maximum likelihood estimation of logistic regression models: theory and implementation. Available at czep. net/stat/mlelr. pdf, 2002.  $\rightarrow$  pages 20
- B. Efron. Bootstrap Methods: Another Look at the Jackknife. Springer, 1992.  $\rightarrow$  pages 28
- K. Fokianos. Density ratio model selection. Journal of Statistical Computation and Simulation, 77(9):805–819, 2007.  $\rightarrow$  pages 30
- K. Fokianos and I. Kaimi. On the effect of misspecifying the density ratio model. Annals of the Institute of Statistical Mathematics, 58(3):475–497, 2006.  $\rightarrow$  pages 30, 53
- P. Hall. On the bootstrap and confidence intervals. The Annals of Statistics, pages 1431-1452, 1986.  $\rightarrow$  pages 28
- P. Hall. Theoretical comparison of bootstrap confidence intervals. <u>The Annals of</u> Statistics, pages 927–953, 1988. → pages 28
- F.E. Harrell and C. Dupont. Hmisc: harrell miscellaneous. <u>R Package Version</u>, 3 (2), 2008. → pages 24, 28
- ASTM International. Standard practice for establishing allowable properties for visually-graded dimension lumber from in-grade tests of full-size specimens. astm d1990-07, 2007. → pages 1
- A. Keziou and S. Leoni-Aubin. On empirical likelihood for semiparametric two-sample density ratio models. Journal of Statistical Planning and Inference, 138 (4):915–928, 2008. → pages 20
- W.H. Kruskal and W.A. Wallis. Use of ranks in one-criterion variance analysis. Journal of the American Statistical Association, 47(260):583–621, 1952.  $\rightarrow$  pages 42
- E.L. Lehmann. Nonparametrics: Statistical Methods Based on Ranks. Springer, 2006.  $\rightarrow$  pages 11, 39
- B.G. Lindsay. Composite likelihood methods. Contemporary Mathematics, 80(1): 221–239, 1988.  $\rightarrow$  pages 7
- S. Nadarajah and A.K. Gupta. Some bivariate gamma distributions. <u>Applied</u> Mathematics Letters, 19(8):767–774, 2006. → pages 46

- H. Nakayama, H. Sayama, and Y. Sawaragi. A generalized lagrangian function and multiplier method. Journal of Optimization Theory and Applications, 17(3-4): 211–227, 1975. → pages 6
- A.B. Owen. Empirical Likelihood. Chapman and Hall/CRC, 2001.  $\rightarrow$  pages 5
- G.C. Pflug. Some emarks on the value-at-risk and the conditional value-at-risk. Probabilistic Constrained Optimization, pages 272–281, 2000.  $\rightarrow$  pages 1
- J. Qin and B. Zhang. A goodness-of-fit test for logistic regression models based on case-control data. Biometrika, 84(3):609–618, 1997. → pages 20
- B. Ripley, W. Venables, and M.B. Ripley. Package 'nnet'. <u>R package version</u>, pages 7–3, 2016.  $\rightarrow$  pages 20
- B. Rosner, R.J. Glynn, and M.L. Lee. Incorporation of clustering effects for the wilcoxon rank sum test: a large-sample approach. <u>Biometrics</u>, 59(4):1089–1098, 2003. → pages 11, 37, 38, 41, 42, 53
- R.J. Serfling. <u>Approximation Theorems of Mathematical Statistics</u>, volume 162. John Wiley Sons, 2009.  $\rightarrow$  pages 4, 12
- B.W. Silverman. Density Estimation for Statistics and Data Analysis. Routledge, 2018.  $\rightarrow$  pages 27
- I. Smith, S. Alkan, and Y.H. Chui. Variation of dynamic properties and static bending strength of a plantation-grown red pine. Journal of the Institute of Wood Science, 12(4):221–224, 1991. → pages 2
- C. Varin, N. Reid, and D. Firth. An overview of composite likelihood methods. Statistica Sinica, pages 5–42, 2011.  $\rightarrow$  pages 7
- S. Verrill, D.E. Kretschmann, and J.W. Evans. Simulations of strength property monitoring tests. <u>Unpublished manuscript</u>. Forest Products Laboratory, <u>Madison, Wisconsin. Available at http://www1. fpl. fs. fed. us/monit. pdf</u>, 2015. → pages 2, 3
- G. Zhu and J. Chen. Multi-parameter one-sided monitoring tests. Technometrics, (just-accepted), 2017.  $\rightarrow$  pages 58