2022-04-18

A Multidimensional Convolutional Bootstrapping Method for the Analysis of Degradation Data

Jared M. Clark
Brigham Young University

A Multidimensional Convolutional Bootstrapping Method for the Analysis of Degradation Data

Jared M. Clark

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Richard L. Warr, Chair
Nathan G. Sandholtz
Matthew J. Heiner

Department of Statistics

Brigham Young University

ABSTRACT

A Multidimensional Convolutional Bootstrapping
Method for the Analysis of Degradation Data

Jared M. Clark
Department of Statistics, BYU
Master of Science

While Monte Carlo methods for bootstrapping are typically easy to implement, they can be quite time intensive. This work aims to extend an established convolutional method of bootstrapping to work when convolutions in two or more dimensions are required. The convolutional method relies on efficient computational tools rather than Monte Carlo simulation which can greatly reduce the computation time. The proposed method is particularly well suited for the analysis of degradation data when the data are not collected on time intervals of equal length. The convolutional bootstrapping method is typically much faster than the Monte Carlo bootstrap and can be used to produce exact results in some simple cases. Even in more complicated applications, where it is not feasible to find exact results, mathematical bounds can be placed on the resulting distribution. With these benefits of the convolutional method, this bootstrapping approach has been shown to be a useful alternative to the traditional Monte Carlo bootstrap.

Keywords: discrete Fourier transform, Lévy process, Monte Carlo estimation, saddlepoint approximation

ACKNOWLEDGMENTS

There were numerous individuals who helped me succeed in preparing this document. I would first like to acknowledge Dr. Matthew Heiner, Dr. Robert Richardson and Dr. Nathan Sandholtz who have served as members of my committee. I appreciate the thoughtful suggestions provided by each of these individuals. Next, I would like to thank Dr. Richard Warr who has served my research advisor as I have worked on this document. I appreciate that he was always willing to help as I encountered unforeseen difficulties in this line of research. Finally, I would like to thank the faculty of the Department of Statistics who have constantly shown interest in my academic progress and research endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The analysis of degradation data is a field of active study. Degradation data analysis is a useful approach in reliability since it isn't always feasible to wait for a failure to occur. When the exact time of failure can't be observed, degradation measurements are often analyzed since degradation is usually closely associated with system failure. Furthermore, analyzing degradation data allows researchers to understand how degradation accumulates on a system over time. This allows researchers to estimate the remaining time until failure based on the current amount of degradation. For this reason, even when it is possible to record failure times, the study of degradation measurements is often worthwhile.

One common approach for the analysis of degradation data requires defining a degradation threshold. As soon as the cumulative degradation on a system has surpassed the threshold, the system is said to have achieved a soft failure. Once a soft failure has been defined in terms of a degradation threshold, the goal of analysis might be to better understand the distribution of failure times.

The analysis becomes somewhat complicated with the reality that it is not usually feasible to continually track the degradation. This means that the exact moment a soft failure is achieved probably won't be observed. When degradation is not continually recorded, analysts must employ some method of estimating the exact time of failure. In previous work, a discrete convolutional method has been developed to bootstrap failure times when the system is checked for degradation at equal time intervals (Clark and Warr 2021). The convolutional method, which will be explored in subsequent sections, provides an alternative to traditional Monte Carlo bootstrapping methods and does not introduce any Monte Carlo error. The convolutional method uses linear interpolation to estimate the exact failure time.

That is, there is an assumption made that on a relatively small time interval, the cumulative degradation increases linearly.

In the examples that will be considered in the simulation and application sections of this paper, there is also an assumption that degradation is independent of time; that is, the rate of degradation is not changing as the system ages. We are essentially assuming that the degradation accumulation on the system could be described as a Lévy process, which will be more formally defined in the subsequent sections. The convolutional method (Clark and Warr 2021) was shown to be useful due to a high level of accuracy (exact in many cases) and relatively short computation times. However, when degradation is measured on unequal time increments, it is more difficult to obtain the bootstrap failure time distribution from the data through convolutional approaches.

The condition that data are collected on equal time intervals certainly makes sense in some situations. In controlled studies, degradation is often recorded at the end of equal time intervals since the systems are closely monitored throughout the duration of the study. However, there are also common situations where the elapsed time between degradation measurements cannot be controlled. If the approach to data collection is observational, the data will have likely not been collected on equal time intervals. For example, if data are collected at a device repair shop, customers will visit the shop each having used their device a different amount of time. When the data are not collected on equal time intervals, the analysis approach must be able to incorporate the differing time measurements.

Although it might be simple enough to obtain degradation measurements on equal time intervals in many instances, there are certainly cases where it is not possible to collect the data in this manner. In previous research, one-dimensional methods have been established which work well with equal time interval degradation data. This occurs since regardless of the number of times convolved, all probability mass will be located at a single time for a particular convolution. Recognizing that degradation won't necessarily always

be recorded on equal time intervals supports the notion that existing methods should be expanded to better accommodate these situations.

When using Monte Carlo bootstrapping methods for analysis, the unequal time intervals don't add much to the difficulty of implementation. However, implementation of the aforementioned convolutional method sees a very noticeable increase in complexity. This arises since two random variables are simultaneously convolved: time and degradation. The goal of this work will be to extend the discrete convolutional method into multiple dimensions. The unequal time interval analysis, which is focused on in this work, is just one possible application of a multidimensional convolutional bootstrapping method.

In this paper, we will start by reviewing the relevant literature. Next, a multivariate convolutional method will be proposed. Relying on multivariate stochastic ordering, we will show that desired properties of the univariate convolutional method extend into higher dimensional problems. Simulation studies and applications will be used to show the benefits of employing the convolutional method in multidimensional bootstrapping scenarios.

CHAPTER 2

LITERATURE REVIEW

There are a number of researchers who have made noteworthy contributions to the study of degradation data. A selection of the most relevant works will be discussed in this section.

A pioneering work in the analysis of degradation data is Lu and Meeker (1993). This work described methods of using degradation information to estimate failure-time distributions. Since there is often an association between degradation and the functionality of a system, analysis of degradation path data is a useful alternative to typical failure-time analyses. That article also drew a connection between degradation data analysis and the field of pharmacokinetics. That connection suggested many of the methods already established in biostatistics could be readily applied, with minor modifications, to the field of reliability.

The analysis of degradation data was further extended by Shiau and Lin (1999) who developed non-parametric methods for analyzing accelerated degradation data. Accelerated lifetime testing imposes extreme conditions on a system in order to observe the entire lifetime in a relatively short period of time. The advancements of Shiau and Lin (1999) are interesting since they show the potential of non-parametric methods for the analysis of degradation data, which are typically analyzed using parametric models. The convolutional bootstrapping method proposed in this paper is another non-parametric approach for analyzing degradation data.

Yet another advancement to the analysis of degradation data came with a proposed method to model measurement error (Ye et al. 2013). Though modeling the measurement error is not the main focus of that article, there were a few points of particular interest that are relevant to this work. The method proposed in Ye et al. (2013) uses a model selection algorithm to determine an appropriate Wiener process (a special case of a Lévy process) for

modeling the data. Through simulation study, this approach was shown to be more efficient than competing methods. The Lévy process assumption seems to be reasonable for many data sets and when this assumption is made analyses are shown to be relatively efficient.

Many other methods for the analysis of degradation path data have assumed an underlying Wiener process. Zhang et al. (2018) compiles relevant literature exploring the Wiener process assumption for degradation data. While this assumption doesn't oversimplify analyses in most instances, there are certainly abnormalities that a Wiener process cannot account for. For example, if a system experiences a recovery state for any reason throughout the study, wherein the cumulative degradation actually decreases, a Wiener process cannot account for this negative trend in the degradation measurements.

Yet another approach for the analysis of degradation data is use of the bootstrap. Guo et al. (2018) propose a method of bootstrapping the remaining time until failure of a system. In bootstrapping remaining time until failure, we hope to incorporate uncertainty in making inference about how much longer a system will function. This approach is recommended when the degradation increments on a single system are not independent. Since this bootstrapping method is free of parametric assumptions, the method is particularly useful when no known distributions appear to fit the data well. Bootstrap approaches for the analysis of degradation data are the main focus of this work.

In addition to remaining system life, degradation data can also be used to bootstrap failure times which allows for the quantification of uncertainty without the need for parametric assumptions. Failure times have some unknown distribution which we hope to approximate from the data using bootstrapping methods. Although remaining system life and failure time are related, they differ in that the remaining system life is a function of the age of the system. Balakrishnan and Qin (2019) developed a saddlepoint method for the estimation of bootstrap distributions. This method is an alternative to the traditional Monte Carlo approach, which is often time intensive. The proposed method was applied to laser device degradation to show the possibility of analyzing degradation data with the sad-

dlepoint method. The saddlepoint approximation for the pdf of the bootstrap distribution of failure times, $\hat{f}(t)$, as defined by Balakrishnan and Qin (2019) is :

$$\hat{f}(t) = \frac{e^{K(\hat{s})-\hat{s}t}}{\sqrt{2\pi K''(\hat{s})}}$$

Here $K(s)$ is the cumulant generating function while $\hat{s}$ solves the equation $K'(\hat{s}) = t$. Since the moment generating function is unknown, an empirical saddlepoint approximation was proposed. The empirical saddlepoint approximation requires that the degradation data are measured on equal time intervals.

The more recent work of Palayangoda et al. (2020) aimed to improve the saddlepoint approximation of the bootstrap distribution of failure times. The method established in that work is also generalized to the unequal time interval problem. The use of a saddlepoint approximation in the analysis of degradation data is particularly viable when Monte Carlo methods are slow and inefficient. Palayangoda et al. (2020) also features the laser data application which will serve as a comparison in this work.

In addition to the saddlepoint approach, failure times can be bootstrapped by defining the distribution of the convolution of interest. As will be discussed in the methods section, some bootstrapped random variables can be defined as the convolution of data values. The discrete Fourier transform (DFT) can then be used to evaluate the convolution thereby defining the bootstrap distribution without the need for Monte Carlo samples. This approach is developed for one-dimensional problems in Clark and Warr (2021) where it is shown to produce exact or mathematically bounded results. These results can often be computed much faster than the stochastic estimates of a Monte Carlo approach. The convolutional approach is the main focus of this work, wherein it will be extended to handle a broader class of problems.

The convolutional method, which we propose here, relies heavily on the ability to bound the CDF of convolved distributions using the DFT. The theory for bounding one-dimensional convolutions was developed in Warr and Wight (2020). That work did not

generalize the results to multiple dimensions. The next section aims to extend this theory to multiple dimensions and outlines the basis for the convolutional method in more complicated analyses which require the simultaneous convolution of multiple random variables.

CHAPTER 3

METHODS

This work aims to extend the convolutional bootstrapping approach of Clark and Warr (2021) to multidimensional problems. First, the feasibility of placing mathematical bounds on the distribution of convolved random variables will be assessed by investigating the multidimensional properties of the DFT. Next, the convolutional theory will be applied directly to some bootstrap statistics of interest.

3.1 Discrete Fourier Transform

The discrete convolutional method for bootstrapping relies on the DFT. For our applications, we show that the DFT can be used to place mathematical bounds on convolutions. The discrete Fourier transform was chosen as the method of convolution due to the availability of its implementation, the fast Fourier transform (FFT), in computational environments. Furthermore, the FFT is simple to invert which makes the discrete Fourier transform an extremely practical choice.

Let $s_1 = \{s_{1.0}, s_{1.1}, ..., s_{1.N_1-1}\}$ and $s_2 = \{s_{2.0}, s_{2.1}, ..., s_{2.N_2-1}\}$ be sequences such that $s_1 \times s_2$ is the support for the two-dimensional random vector, $\mathbf{X}$. Notice that s_1 is a sequence of length N_1 while s_2 is a sequence of length N_2. Furthermore let $f_X(y, z)$ be the probability that $\mathbf{X}$ is equal to (y, z). Then the value of the discrete Fourier transform, denoted F_X, for (K_1, K_2), an index in $\mathbb{R}^2$, is:

$$F_X(K_1, K_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} f_X(s_{1.n_1}, s_{2.n_2}) e^{-i\frac{2\pi}{N_1} s_{1.n_1} K_1 - i\frac{2\pi}{N_2} s_{2.n_2} K_2}$$

This is a direct extension of the one-dimensional DFT. The one-dimensional DFT requires summing a univariate pmf over a single index. Similarly, the DFT can be extended to any higher finite dimension. For reference, the one-dimensional DFT is provided here:

$$F_X(K_1) = \sum_{n_1=0}^{N_1-1} f_X(s_{1.n_1}) e^{-i\frac{2\pi}{N_1} s_{1.n_1} K_1}$$

One reason the convolutional method is so practical for use in one dimension is the ease in inverting the DFT. The computational efficiency of the DFT inversion (using the fast Fourier transform) also holds in two dimensions. The inverse DFT is given as:

$$f_X(s_{1.n_1}, s_{2.n_2}) = \frac{1}{N_1 N_2} \sum_{K_1=0}^{N_1-1} \sum_{K_2=0}^{N_2-1} F_X(K_1, K_2) e^{i2\pi n_1 \frac{K_1}{N_1} + i2\pi n_2 \frac{K_2}{N_2}}$$

We aim to show that the CDF can be bounded in two dimensions. Here we consider mathematical bounds based on stochastic ordering. We are essentially creating bounds that will be guaranteed to contain a quantile of interest from the bootstrap distribution. There is one major source of error that will be considered here when creating these mathematical bounds. Bootstrapped statistics are discrete random variables. This means that the DFT can be used to find exact bootstrap distributions in many instances. However, it will not always be possible to create a grid for use with the DFT that has finite support matching the support of the bootstrap statistic. The error introduced by this discretization process will be referred to as discretization error.

It can be shown that the discretization error of the DFT can be bounded in multiple dimensions. Although the bootstrap statistic will be a discrete random variable, discretization error is introduced since it is often not feasible to create a grid that will match both the support of the bootstrap statistic and the support defined by the data. Warr and Wight (2020) relied on stochastic ordering to bound the discretization error. A similar approach will be taken here. Stochastic order is easily defined in one dimension, but there are various definitions for stochastic order in multiple dimensions. We adopt the multivariate stochastic

ordering definition from Shaked and Shanthikumar (2007). First, the concept of an upper set should be defined:

Definition 1. *Consider real valued vectors* $A = (a_1, a_2, \ldots, a_n)$ *and* $B = (b_1, b_2, \ldots, b_n)$, *such that* $a_1 \leq b_1$, $a_2 \leq b_2$, $\ldots$, $a_n \leq b_n$. *A subset of* $\mathbb{R}^n$, U, *is said to be an upper set if* $A \in U$ *implies* $B \in U$.

Now, the multivariate stochastic order has the following definition:

Definition 2. *The random vector* $\boldsymbol{X}$ *is stochastically less than the random vector* $\boldsymbol{Y}$ *(or* $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$*), if* $P(\boldsymbol{X} \in U) \leq P(\boldsymbol{Y} \in U)$ *for all upper sets,* U, *in* $\mathbb{R}^n$.

The following two theorems involving the multivariate stochastic order (Shaked and Shanthikumar 2007, p. 273) are used for bounding discretization error in bootstrapping scenarios:

Theorem 1. *Let* $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *be two n-dimensional random vectors. If* $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$ *and* $\boldsymbol{g}$: $\mathbb{R}^n \to \mathbb{R}^k$ *is any k-dimensional increasing function, for any positive integer k, then the k-dimensional vectors* $\boldsymbol{g}(\boldsymbol{X})$ *and* $\boldsymbol{g}(\boldsymbol{Y})$ *satisfy* $\boldsymbol{g}(\boldsymbol{X}) \leq_{st} \boldsymbol{g}(\boldsymbol{Y})$.

Theorem 2. *Let* $\boldsymbol{X}$, $\boldsymbol{Y}$, *and* $\boldsymbol{Z}$ *be random vectors such that* $(\boldsymbol{X}|\boldsymbol{Z} = \boldsymbol{z}) \leq_{st} (\boldsymbol{Y}|\boldsymbol{Z} = \boldsymbol{z})$ *for all realizations of* $\boldsymbol{Z}$, *denoted* $\boldsymbol{z}$. *Then* $\boldsymbol{X} \leq_{st} \boldsymbol{Y}$.

The following theorem is needed when bounding the bootstrap distribution of some statistics.

Theorem 3. *Consider nonnegative random vectors,* $\boldsymbol{X}$, $\boldsymbol{Y}$; *and random variable, Z, where:*

$$(\boldsymbol{X} \,|\, Z = z) = \begin{pmatrix} a_1 & a_2 & \ldots & a_n & b_1 & b_2 & \ldots & b_n \end{pmatrix}' \text{ and}$$
$$(\boldsymbol{Y} \,|\, Z = z) = \begin{pmatrix} c_1 & c_2 & \ldots & c_n & d_1 & d_2 & \ldots & d_n \end{pmatrix}'.$$

If $(\boldsymbol{X} \,|\, Z = z) \leq_{st} (\boldsymbol{Y} \,|\, Z = z)$, *then*

$$(\tilde{\boldsymbol{X}} \,|\, Z = z) = \begin{pmatrix} \sum_{i=1}^{z} a_i \\ \sum_{i=1}^{z} b_i \end{pmatrix} \leq_{st} (\tilde{\boldsymbol{Y}} \,|\, Z = z) = \begin{pmatrix} \sum_{i=1}^{z} c_i \\ \sum_{i=1}^{z} d_i \end{pmatrix}.$$

Proof. Theorem 3 follows as a direct result of Theorem 1 since summation on a nonnegative support is increasing. Likewise for the same random vectors defined in Theorem 3,

$$(A|Z=z)=\sum_{i=1}^{z}a_i \leq_{st} (C|Z=z)=\sum_{i=1}^{z}c_i$$

and

$$(B|Z=z)=\sum_{i=1}^{z}b_i \leq_{st} (D|Z=z)=\sum_{i=1}^{z}d_i$$

Finally, if $(A|Z=z) \leq_{st} (C|Z=z)$ for every z, then $A \leq_{st} C$ by Theorem 2. □

These properties will be applied directly to bootstrapped quantities of interest in the following section. There it will be shown that, in many cases, the bootstrap distribution of a statistic of interest can be mathematically bounded.

Another source of possible error comes from truncation. When bootstrapping there is typically a finite maximum for the bootstrapped statistic. This implies that as long as care is taken in choosing the support grid for use with the DFT, truncation should not occur.

3.2 Bootstrapping

In order to apply the one-dimensional convolutional theory developed in Warr and Wight (2020) to bootstrap applications, it was necessary to show that the bootstrap statistic of interest could be written as a convolution of known independent random variables defined by the sample.

Define a distribution, F, such that we have n mutually independent random vectors $\boldsymbol{X}_i \sim F$ for $i \in \{1,2,\ldots,n\}$. Once the random vectors are drawn from their distribution, they are no longer random and we denote the observed sample as $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ or just $\boldsymbol{x}$.

We now use the observed $\boldsymbol{x}$ to define a new distribution. Let $\boldsymbol{Y} \sim G$, where G is a discrete multivariate distribution with probability defined by the frequency of the vectors in $\boldsymbol{x}$.

For illustration, if we bootstrap the sample mean in a univariate scenario using Monte Carlo methods, we are taking draws from a distribution that can be easily defined as a scaled convolution of random variables $X_1, X_2, ..., X_n$ where the X_i $(i = 1, 2, ..., n)$ are independent and identically distributed with equal probability assigned to each observed value in the sample (assuming no ties). Since the bootstrap statistic can be recognized as a convolution of independent random variables, the convolutional theory applies.

Through use of the fast Fourier transform (FFT), the pmf of a convolution of random variables can be efficiently defined. In order to bootstrap the sample mean, let $Y_i = \frac{X_i}{n}$ where X_i are the independent and identically distributed random variables defined by the data. Now, since the Y_i's are independent and identically distributed, we remove the subscript and define its DFT, $F_Y(K)$, on an appropriate grid for which the bootstrap distribution of the sample mean can be easily defined. Start by convolving in the frequency domain using the FFT:

$$F_{\bar{X}}(K) = (F_Y(K))^n$$

Now, by implementing the inverse FFT, we have defined the bootstrap distribution of the sample mean. Bootstrapping other statistics with the convolutional method is possible when the bootstrap statistic can be written as the convolution of independent random variables defined by the data.

The bootstrap distribution of a mean vector can be found in a similar manner. Rather than treating X and Y as random variables, they can be thought of as random vectors. Using the multidimensional DFT in the same way will result in the bootstrap distribution of the mean vector. In the multivariate case, a support grid for each dimension of the random vector involved in the convolution is needed.

When it isn't possible to form a grid that matches the support of the bootstrap variable of interest, the DFT can be used to mathematically bound the distribution. In these instances, each vector in the original sample should have its values rounded down

to the closest grid point in each dimension. Rounding the data in this way produces a random vector that is stochastically greater than the distribution defined by the original data. Likewise, rounding each observed value up to the next grid point in each dimension will define a random vector that is stochastically less than the random vector defined by the data. Applying the convolutional method to both the upper and lower bound produces bounds on the convolution by Theorem 3. These are mathematical bounds on the CDF of the convolution. For any quantile we are able to generate an interval that is guaranteed to contain the truth, disregarding any numerical error.

An interesting application of the multivariate convolutional bootstrapping method is the analysis of degradation data. One assumption made in many degradation analyses is that the accumulation of degradation is a Lévy process. In the context of degradation measurements, a Lévy process would require that the amount degradation accumulated in one time interval is independent of the degradation accumulated in any disjoint time interval. Furthermore, the distribution of degradation accumulated is identical for any time intervals of equal length. These assumptions are needed when analyzing degradation data using the convolutional bootstrapping method.

Suppose we have both degradation and time measurements, and we are interested in the distribution of times that a soft failure occurs. To obtain the bootstrap distribution of failure times, one approach is to convolve the random vectors defined by the sample until the probability that the cumulative degradation has achieved the threshold is one. At each convolution the exact time of failure is approximated through linear interpolation.

In a simple scenario, assume that we observe two time-degradation pairs. For the first observation we observe 1 unit of degradation at the end of a 1 time unit interval. For the second observation we record 1.75 degradation units occurring by the end of a 2 time unit interval. Suppose a soft failure is declared at 2.5 units of cumulative degradation. To sample from the bootstrap distribution we take draws from our observations, with replacement, until the threshold of 2.5 is surpassed. Following the Monte Carlo approach we might

first sample (1, 1). Since the cumulative degradation has not surpassed the threshold we take another draw from our observations and select the (2, 1.75) measurement. Now the cumulative degradation is 2.75 which has surpassed the threshold of 2.5. We use our observed cumulative time-degradation measurements, (1, 1) and (3, 2.75), along with the threshold of 2.5 to estimate the failure time as 2.714. Notice that if we had drawn the (1, 1) measurement twice, our cumulative degradation would only be 2, so it is not enough to only consider convolving two random vectors. By convolving three random vectors, the minimum cumulative degradation is 3, which is greater than 2.5, and in this case we know that the soft failure threshold must be exceeded.

It will most likely not be possible to create a suitable grid for finding exact results through the convolutional method. However, by rounding all the data down, a random vector that is stochastically greater than the truth is formed. Similarly, when rounding all of the data up, a random vector that is stochastically less than the truth can be created. Consider:

$$L = \begin{pmatrix} T_{1l} & T_{2l} & \dots & T_{nl} & D_{1l} & D_{2l} & \dots & D_{nl} \end{pmatrix}'$$

$$U = \begin{pmatrix} T_{1u} & T_{2u} & \dots & T_{nu} & D_{1u} & D_{2u} & \dots & D_{nu} \end{pmatrix}'$$

In this case, L is a lower bound on the time and degradation and U is an upper bound on the time and degradation. The vectors (T_{il}, D_{il}) are independent and identically distributed as are the vectors (T_{iu}, D_{iu}) for $i = 1, 2, ..., n$. These pairs represent a possible increase in degradation and time. The maximum number of convolutions needed to surpass the degradation threshold is n.

Applying Theorem 3 to L and U, we are able to form an upper and lower bound on any convolution. This makes it possible to mathematically bound the distribution of failure times using the convolutional method. Since our cumulative time and cumulative degradation variables are bounded for each convolution, the distribution of failure times, independent of the number of convolutions, will also be bounded.

In addition to development of the convolutional method for particular types of analyses, the multidimensional convolutional method will be directly compared to Monte Carlo methods. The aim of comparison is to show that the convolutional method is often a relatively efficient approach that can provide exact, or mathematically bounded, results. The bounding properties of the DFT are particularly appealing when it is not feasible to define a grid that allows for probability mass at every possible realization of the bootstrap statistic. In these instances the convolutional method will produce exact mathematical bounds on the distribution of interest. The mathematical bounds of the convolutional method will be compared to the stochastic bounds of the Monte Carlo approach.

CHAPTER 4

SIMULATION STUDIES

The multidimensional convolutional approach to bootstrapping has a number of benefits. The comparisons of this section will focus on the accuracy and efficiency of this bootstrapping approach. Traditional Monte Carlo methods will serve as the main comparison. However, the saddlepoint approximation will act as a second comparison for the analysis of degradation data. Note that for all time comparisons a machine with 8 GB RAM and two 1.2 GHz Intel m3 cores was used. Although it would be possible for each method, no parallelization was utilized in these comparisons.

4.1 Exact Bootstrapping

The first simulation uses a data set of 25 observations, where each contains two measurements. The first variable, x, was generated by scaling (by a factor of 10) and rounding samples from a Gamma(1,1) distribution. The second variable, y, was taken as a linear function of x, where the measurements were divided by 5.5 and added to 6, with random noise added from a normal distribution. The measurements of y were also rounded. The resulting correlation between x and y was 0.74. This simulation will investigate the mean vector in a relatively simple scenario to demonstrate how the convolutional approach can generate exact results.

The results from the first comparison are found in Table 4.1. A total of 761,000 Monte Carlo resamples were used to generate the results in that table. The time needed to obtain Monte Carlo results is dependent on the number of resamples used. Here the number of Monte Carlo resamples was chosen so that the computation time would match that of the convolutional method. However, more computation time is generally needed to produce narrower confidence intervals.

Table 4.1: The results from both the Monte Carlo bootstrap and the convolutional bootstrap are similar. Results from both methods were obtained in 10.7s. Here the convolutional method is used to calculate the expected value of the bootstrap distribution of the sample mean vector.

Variable	**MC Mean**	**99% MC Int**	**Convolutional**
x	9.240	(9.236, 9.245)	9.240
y	8.599	(8.596, 8.602)	8.600

On the other hand, the convolutional method was able to generate the exact bootstrap distribution of the mean vector. In this rather simple example, there was no need to introduce any discretization error. Additional computation time will not improve the convolutional results in any way since this method has produced the exact distribution of interest, disregarding any numerical error introduced in computation. Since the data were recorded as integers and there were 25 observations, the grid for the convolutional method only needed precision to two decimal places in order to provide exact results.

Note that for ease of comparison, Table 4.1 compares the two variables in our vector of interest marginally. Both methods for bootstrapping preserve the dependence structure of the original data and could be used for multivariate inference.

4.2 Bounding Distributions

In many cases the data are not as easy to work with as they were in this first simulation. For the next simulation, the data were generated in the same manner, but the rounding to integer values did not occur. Without rounding the data, which were drawn from a continuous distribution, we would need an extremely fine grid in order to fully capture the support of the bootstrap sample mean vector. Since it is no longer plausible to find the exact bootstrap distribution of the mean vector, the convolutional method was used to bound the bootstrap distribution of the sample mean of the first variable, x.

Figure 4.1 displays the results of this second simulation. Here we adjusted both the number of Monte Carlo samples and the coarseness of the grid for the convolutional ap-

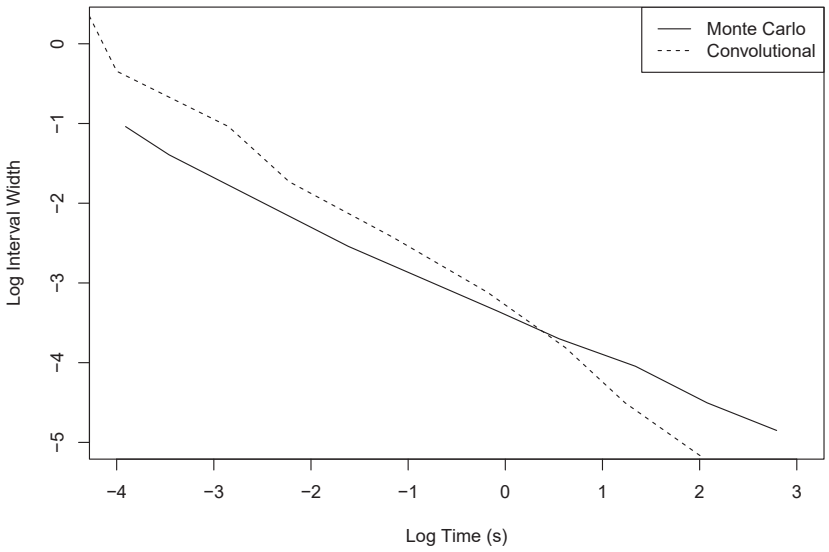

Figure 4.1: There are methods for assessing uncertainty in the bootstrap distribution for both approaches. The Monte Carlo approach quantifies Monte Carlo error through confidence intervals while the convolutional approach places mathematical bounds on the discretization error. As computation time increases it becomes more efficient to use the convolutional method.

proach. By adjusting these values we are able to control the width of the intervals produced, however narrower intervals typically require more computation time. Wider intervals could be obtained quicker using the Monte Carlo approach. However, the bounds from the convolutional method narrow at a faster rate. All Monte Carlo intervals had a confidence level of 99%. As the discrete grid on the variable, x, became finer, the grid on y was left unchanged. This parallels some of the later applications where it was relatively simple to place a grid on a second variable that matched the support. Changing the grid on the second variable in this problem would certainly have a large effect on the timing of the convolutional method.

These two simulations are meant to show the clear benefits of using the convolutional method. Monte Carlo simulation is subject to Monte Carlo error and it is not possible to generate results that are not stochastic in nature. On the other hand, the convolutional method is able to mathematically bound results. Naturally when high levels of accuracy

are needed in multiple dimensions, both the convolutional method and the Monte Carlo approach will require more computation time.

4.3 Unequal Time Intervals with Laser Data

The motivating application for the development of a multidimensional convolutional bootstrapping method can be found in Palayangoda et al. (2020). The data set used in that paper features degradation information on a set of 15 lasers. The degradation measurements were collected after intervals of 250 hours; however, the data were altered to serve as an example of an unequal time interval problem. Further information on the data set and how it was altered can be found in Palayangoda et al. (2020).

The main contribution of Palayangoda et al. (2020) was the development of a saddlepoint approximation for use in bootstrap analyses. We compare the results of our convolutional bootstrapping method to those of the saddlepoint approximation and traditional Monte Carlo methods in Table 4.2.

Table 4.2: Here we investigate the bootstrap distribution of failure times, with time recorded in hours. A separate bootstrap distribution is created with each degradation threshold. In general, the convolutional and Monte Carlo methods seem to be comparable. The saddlepoint differs substantially from the other approximations.

Threshold	**Quantile**	**99% Bootstrap CI**	**Convolutional**	**Saddlepoint**
2	0.05	(654.88, 682.78)	(664.77, 665.35)	722.50
2	0.10	(698.46, 731.67)	(708.59, 709.77)	782.50
2	0.90	(1264.91, 1297.37)	(1278.71, 1281.77)	1220.00
6	0.05	(2314.05, 2402.88)	(2331.63, 2388.42)	2495.00
6	0.10	(2445.94, 2520.90)	(2454.21, 2514.62)	2600.00
6	0.90	(3416.92, 3511.55)	(3413.82, 3519.00)	3357.50
10	0.05	(4120.70, 4150.73)	(4122.42, 4150.70)	4330.00
10	0.10	(4291.11, 4316.47)	(4287.98, 4318.93)	4465.00
10	0.90	(5558.81, 5583.98)	(5548.83, 5593.75)	5445.00

The Monte Carlo results were found by first bootstrapping 2,500 failure times for threshold 2, 1,525 failure times for threshold 6 and 8,000 failure times for threshold 10,

reporting the quantiles of interest and then repeating the process 1,200 times. The number of Monte Carlo samples was chosen such that the computation time would be the same across methods. While the Monte Carlo bootstrap intervals and the intervals from the convolutional method are generally comparable, the convolutional intervals place mathematical bounds on the bootstrap distribution of failure times. In contrast, the Monte Carlo intervals are stochastic in nature and do not necessarily contain the truth.

For the threshold of 2, there seems to be a clear benefit in using the convolutional method. For all three quantiles investigated here the convolutional method produced narrower intervals when compared to the Monte Carlo approach. With the larger number of convolutions needed to consider a threshold of 6, we don't see the same clear benefit to using the convolutional method, though the convolutional intervals are narrower for two of the quantiles.

The degradation threshold of 10 is computationally intensive for both methods. In this instance, the Monte Carlo intervals were narrower than the convolutional intervals for two of the quantiles. It should be noted that the convolutional method is producing mathematical bounds. The quantiles of interest must be within the convolutional intervals. Further research could aim to improve the efficiency of the convolutional method for large degradation threshold problems.

Notice that the saddlepoint approximation suggested by Palayangoda et al. (2020) differs largely from the other two methods. Point estimates are reported here since intervals were not provided in their paper.

In the majority of simulations here, the discrete convolutional method appears favorable when compared to the traditional Monte Carlo approach. In many instances, the traditional bootstrap will be preferred due to ease of implementation. However, it is often worthwhile to take advantage of the computational gains offered by the convolutional method.

CHAPTER 5

APPLICATION

One possible application of the multidimensional convolutional bootstrapping method comes in the form of an analysis of LED degradation data. The data used in this example were obtained from Meeker et al. (2022). Our goal is to understand the degrading light output from LEDs used in flashlights. The LEDs analyzed here were tested with a 130°C temperature and a 40mA current. In total, 30 LEDs were used and 9 light outputs were recorded on each LED. There are 270 LED degradation measurements in total. At the beginning of the LED lifetime, the cumulative degradation is fairly constant and a recorded change in the degradation is often just random noise. Furthermore, when bootstrapping degradation times we are making an inherent assumption that the rate of degradation is constant across the entire lifetime. The light outputs were measured as the relative change from the output at 138 hours which will be referred to as the initial output. Using the LED output at 138 hours ensures that the degradation paths are strictly decreasing and that the Lévy process assumption is met.

Once an LED output falls below 60% of the initial output, the LED device is considered to have failed. The analysis of this data set, presented in Meeker et al. (2022), aimed to estimate the 10th percentile of failure times. Here, we will find the entire distribution of bootstrap failure times, from which it will be simple to approximate any quantile of interest.

We start the analysis by first assessing whether the convolutional bootstrapping method is appropriate. Figure 5.1 does not provide evidence of recovery states, that is, the degradation paths of the LED units are strictly decreasing. Furthermore, there is not any strong evidence of a nonlinear trend. These characteristics support the assumptions needed for a bootstrap approach. In this analysis we treat the degradation observations

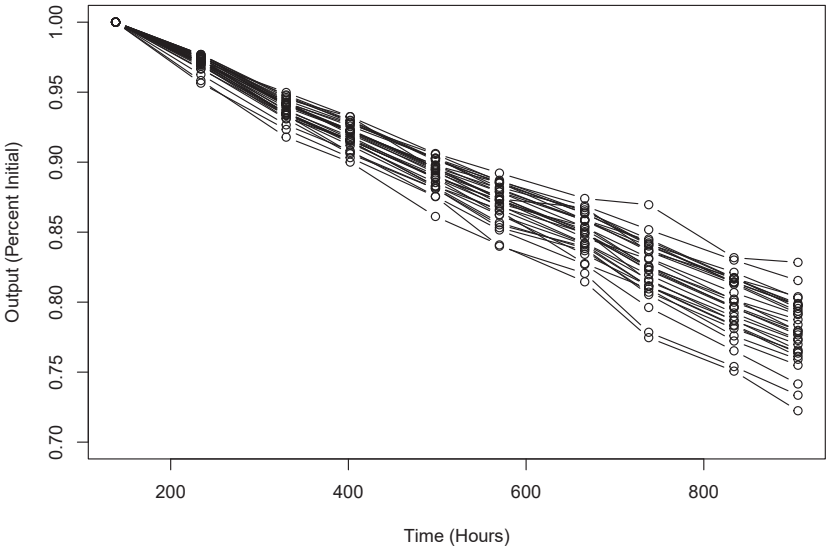

Figure 5.1: The degradation paths are strictly decreasing in a relatively linear fashion.

as independent. This means that the rate of degradation is not changing over time. We also do not take into account any possible device random effect, though this is a possible approach. Clark and Warr (2021) show how this could be implemented using the univariate discrete convolutional method. For the most part, the degradation paths of the lasers seem similar; however, addition of a random effect would be an interesting consideration in future analyses.

The resulting bootstrap distribution of failure times is shown in Figure 5.2. The Monte Carlo estimate from 20,000 samples is provided as well. Notice that the two methods are in close agreement. The distribution from the convolutional method is an estimate since generating an appropriate grid along the support of degradation was not feasible. However, the degradation was only ever recorded after intervals of 72 or 96 hours so an exact grid could be placed on the support for time.

Since both the convolutional and Monte Carlo approaches are only able to approximate the distribution, interval estimates are used to assess uncertainty. The 10th percentile is mathematically bounded by 1283.46 and 1303.33. This interval was found after 15.64

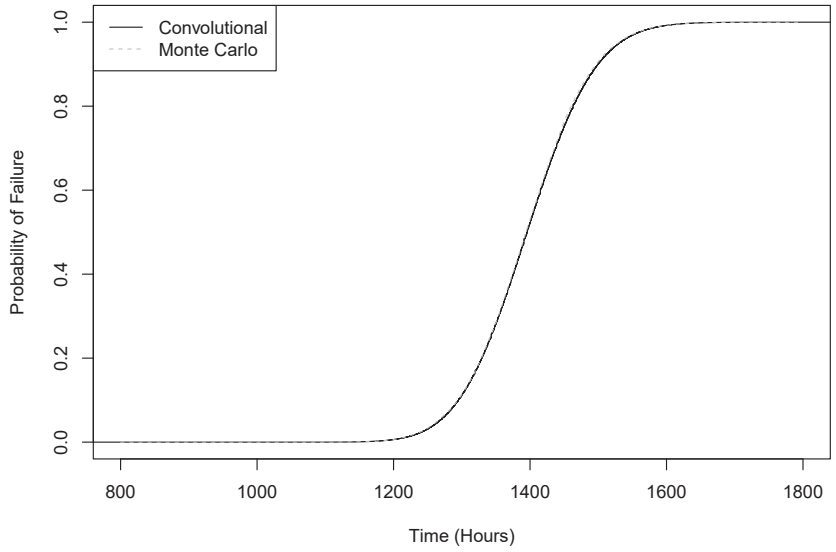

Figure 5.2: The CDF's of the soft failure times for the convolutional method and the traditional Monte Carlo method agree very closely.

minutes of computation time. A 99% bootstrap interval, also obtained in 15.64 minutes, is (1282.86, 1304.07). Although the convolutional interval is only slightly narrower in this instance, it has the advantage of mathematically bounding the quantile of interest.

In this application, the convolutional method has been shown to be a desirable alternative to the traditional Monte Carlo bootstrap. The two methods provide similar results, though the traditional approach introduces Monte Carlo error.

CHAPTER 6

CONCLUSION

Although use of Monte Carlo simulation is standard practice in bootstrap applications, these methods have some limitations. Any results generated from traditional bootstrapping approaches are subject to Monte Carlo error. This error is known to be rather large when estimating in the tails of distributions. Furthermore, bootstrapping can be computationally expensive for complicated analyses. In these situations obtaining narrow confidence bounds might not be feasible.

The discrete convolutional method provides one alternative to the Monte Carlo bootstrap. The multidimensional convolutional method relies on the notion that many bootstrap statistics are simply the convolution of independent discrete random vectors with distributions defined by the observed data. When it is necessary to bootstrap multiple measurements simultaneously, the multidimensional discrete Fourier transform can be employed as the method of computation. The fast Fourier transform provides an efficient algorithm for finding the distribution of convolutions. Desirable properties of the FFT include quick computation times and ease of invertibility. Use of the DFT is also justified since the distribution defined by the observed data is discrete.

The convolutional method for bootstrapping has the benefit of generating either exact results or mathematical bounds, thus avoiding the introduction of any Monte Carlo error. In the comparisons of this work, while holding the support grid for one variable constant, the mathematical bounds of the convolutional method are shown to narrow quicker than the stochastic bounds of the Monte Carlo approach. When high levels of accuracy are needed, the convolutional method is more efficient. In cases where bootstrapping procedures need to

be repeated many times, the computational gains of the convolutional method will compound resulting in a much more efficient mode of analysis.

The motivation for developing the multidimensional convolutional method came from the analysis of degradation data, though many other applications for the method exist. In degradation analyses with data which have unequal time interval lengths, it is often necessary to bootstrap both the time and degradation simultaneously. The multidimensional convolutional method is useful for producing a joint distribution on both time and degradation.

In future work, it would be of interest to investigate the inclusion of covariates in degradation analyses. While one temperature and current combination was analyzed here, the full LED data set contains degradation measurements from other temperatures and currents. Temperature and current could likely be incorporated in a more sophisticated bootstrap algorithm as covariates.

This work also focused on two-dimensional analyses. However, the theory has been laid for the possibility of higher dimensional analyses. In preliminary investigation, higher dimensional analyses were relatively inefficient due to the need to create a very large and extremely sparse array. These investigations were certainly not exhaustive and with improvements in computational resources, higher dimensional bootstrapping applications with the convolutional method might be of interest.

REFERENCES

Narayanaswamy Balakrishnan and Chengwei Qin. Nonparametric evaluation of the first passage time of degradation processes. *Applied Stochastic Models in Business and Industry*, 35(3):571–590, 2019.

Jared M Clark and Richard L Warr. Bootstrapping through discrete convolutional methods. *arXiv preprint arXiv:2107.08019*, 2021.

Jingbo Guo, Changxi Wang, Javier Cabrera, and Elsayed A Elsayed. Improved inverse gaussian process and bootstrap: Degradation and reliability metrics. *Reliability Engineering & System Safety*, 178:269–277, 2018.

C Joseph Lu and William O Meeker. Using degradation measures to estimate a time-to-failure distribution. *Technometrics*, 35(2):161–174, 1993.

William Q Meeker, Luis A Escobar, and Francis G Pascual. *Statistical methods for reliability data*. John Wiley & Sons, 2nd Edition, 2022.

Lochana K Palayangoda, Hon Keung Tony Ng, and Ronald W Butler. Improved techniques for parametric and nonparametric evaluations of the first-passage time for degradation processes. *Applied Stochastic Models in Business and Industry*, 36(4):730–753, 2020.

Moshe Shaked and J George Shanthikumar. *Stochastic orders*. Springer Science & Business Media, 2007.

Jyh-Jeu Horng Shiau and Hsin-Hua Lin. Analyzing accelerated degradation data by nonparametric regression. *IEEE Transactions on Reliability*, 48(2):149–158, 1999.

Richard L Warr and Cason J Wight. Error bounds for cumulative distribution functions of convolutions via the discrete fourier transform. *Methodology and Computing in Applied Probability*, 22(3):881–904, 2020.

Zhi-Sheng Ye, Yu Wang, Kwok-Leung Tsui, and Michael Pecht. Degradation data analysis using wiener processes with measurement errors. *IEEE Transactions on Reliability*, 62(4): 772–780, 2013.

Zhengxin Zhang, Xiaosheng Si, Changhua Hu, and Yaguo Lei. Degradation data analysis and remaining useful life estimation: A review on wiener-process-based methods. *European Journal of Operational Research*, 271(3):775–796, 2018.