UNIVERSITY OF CALIFORNIA, SAN DIEGO

Resampling Inhomogeneous Marked Point Processes

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics

by

William John Garner

Committee in charge:

Professor Dimitris Politis, Chair
Professor Ery Arias-Castro
Professor Anthony Gamst
Professor Elias Masry
Professor Ronghui Xu

2011

The dissertation of William John Garner is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2011

DEDICATION

To my parents for helping me start down this path and to my wife, Janette, for helping me reach the finish.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, my eternal gratitude goes out to my advisor Dimitris Politis for always believing in me. I am sure that I have tested his patience on more than one occasion. His thoughts and insights would help guide me when I would lose sight of my goals.

Len Haff has seen me transform from merely a math major into a true mathematician. I have had the honor of working with him half a dozen times and I cherish the friendship that we have formed. His support helped me overcome many challenges. Lance Small recognized my ability to complete a PhD and assisted me behind the scenes. I may never know the full extent of the help that he gave.

I also would like to thank Jason Schweinsberg, Patrick Fitzsimmons, Bruce Driver, and Ian Abramson for taking time out of their busy schedules to sit down with me and discuss my research. In particular, many thanks to Jason for helping me with a last minute issue that popped up. Each offered valuable insights in various aspects of this thesis.

During my time at UCSD, I have made valuable friendships with other graduate students and faculty members. I thank all of them for their presence in my life. I would also like to thank the department of mathematics for the opportunity to teach various classes. The experience was invaluable.

Thanks to my mom and dad for fostering my dreams to get a PhD. This has been quite a journey. And while not everyone is here today, I know they are both proud.

Finally, thanks to my wife, Janette, for her support and encouragement, especially in the past couple of months as everything managed to come together.

VITA

2006 B.A. in Applied Mathematics, University of California, San Diego

2008 M.S. in Statistics, University of California, San Diego

2011 Ph.D. in Mathematics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Resampling Inhomogeneous Marked Point Processes

by

William John Garner

Doctor of Philosophy in Mathematics

University of California, San Diego, 2011

Professor Dimitris Politis, Chair

We investigate methods for resampling inhomogeneous marked point processes, focusing on Poisson point processes. In Chapter 1 we introduce the problem and provide some background information. In Chapter 2 we adapt existing methods for resampling homogeneous marked point processes to the case of one-dimensional inhomogeneous marked point processes data. In Chapter 3 we extend theoretical results such as asymptotic normality from the homogeneous to the inhomogeneous setting. In Chapter 4 we establish the validity of our local block bootstrap procedure for one-dimensional inhomogeneous marked point processes data while in Chapter 5 we compare the performance of the one-dimensional methods. In Chapter 6 and Chapter 7, we extend the theory and validity of our local block bootstrap procedure to higher dimensions. We carry out simulations to test the performance of our methods under varying distributions of points as well as varying dependencies on the associated marks.

Chapter 1

Introduction

When criticized for the lack of motivation in his proofs, Carl Friedrich Gauss is quoted as saying, "The scaffold had to be removed before the edifice of a mathematical work should see the light of the public."

This is in stark contrast to the theme taken for this thesis, as great care has been taken to motivate the steps in the proofs. Most times, the final step was not known in advance and many of the figures and tables used to arrive at the end are included to aid the reader.

In many ways, this thesis reads like a narrative, which captures over three years of work. Pictures are included whenever they can help to visualize what is going on. In fact, this thesis contains a total of 76 figures!

All figures in this document were produced using `Mathematica`, Version 8.0 [81] while all simulations were carried out using `R`, Version 2.14 [61].

This paper discusses methods for resampling marked point process (MPP) data that arise from an inhomogeneous Poisson process. Some authors use the term nonhomogeneous Poisson process instead, and some care must be taken when researching the subject. A nice introduction to inhomogeneous Poisson processes can be found in [68]. A more formal treatment can be found in [13].

The difference between a point process and a marked point process is the fact that each observation of the point process carries with it an associated value (called a mark). Krickeberg [33] provides more details regarding marked point processes. On a very simple level, a marked point process can be thought of as a generalized time series. An example will help to illustrate this point.

Suppose we consider the price of a particular stock. In the case of a time series, we have an alarm clock on our desk that goes off at 4pm every day. At the moment, we record the stock price. Thus, we obtain equally spaced samples that can be indexed by an integer, k. (For the

purpose of this example, weekends and other days the stock market is closed are ignored.)

In the case of a marked point process, we are again concerned with the stock price. This time, however, our alarm clock goes off at random times throughout the day according to some probability distribution. (We can think of the time series situation as the case where the probability distribution assigns a probability of 0 for all times except for 4pm.)

Our objective is statistical inference of the common mean μ on the basis of measurements of a random, wide-sense stationary field $X(\cdot)$ at a finite number of non-lattice, irregularly spaced points. As part of our inference, we shall consider the following two estimators of the mean and assess their asymptotic properties.

$$\tilde{X}_K = \frac{1}{E\left[N(K)\right]} \int_K X(t) N(dt)$$

and

$$\bar{X}_K = \frac{1}{N(K)} \int_K X(t) N(dt)$$

We have adopted the convention that $0/0 = 1$. Here, N denotes the point process, K denotes the region of interest, and $N(K)$ denotes the number of observations in K. $N(dt)$ acts as a counting measure and the above integral is really just a sum over the observed values of N, namely the τ_i, which are themselves random. That is, we could express the above as:

$$\tilde{X}_K = \frac{1}{E\left[N(K)\right]} \sum_{i=1}^{N(K)} X(\tau_i)$$

and

$$\bar{X}_K = \frac{1}{N(K)} \sum_{i=1}^{N(K)} X(\tau_i)$$

The case where the data comes from a time series has been well studied. See, for example, [6] or [80]. In such a setting (where the data is regularly spaced), various methods have been developed to model (and resample) the data. Since we are concerned with obtaining confidence intervals for μ, we will need to employ bootstrap methods for resampling our data.

The simplest bootstrap technique is sampling with replacement where we associate each of n values with a chip in a bag, draw one chip at a time, record the value, replace the chip in the bag, and then repeat the process a total of n times.

This approach is often too naïve for the time-dependent data mentioned above. More sophisticated methods based on using blocks of data (as opposed to individual measurements) have been proposed.

Hall [23], Carlstein [7], Künsch [35], Lahiri [37], Liu and Singh [46], Politis and Romano [54, 55, 56, 57, 58], Raïs [62], and Sherman and Carlstein [76, 77] have developed different block-resampling techniques for the situations where data are of the form $\{X(\mathbf{t}), \mathbf{t} \in \mathbf{E}\}$, with $\mathbf{E}$ being a finite subset of the rectangular lattice $\mathbb{Z}^d$.

Politis, Romano, and Wolf [59] offer a nice overview of the subject, bringing together work from the above authors.

But in many instances such as queuing theory, spatial statistics, mining and geostatistics, and meterology, the data corresponding to the observations $X(\mathbf{t})$ occur at non-lattice, irregularly spaced points.

There are a myriad of possible distributions for such irregulary spaced data. A common assumption is that of an Poisson process, which is the focus of this work. Karr [31] provides some justification for such an assumption. Moreover, Karr puts together the theory for a homogeneous Poisson process with wide-sense stationary marks. Much of this work is based upon earlier work done by Masry [50] for nonparametric covariance estimation. Masry also explored spectral density estimation in [48, 49] and later with Lii in [44, 45].

One of the earliest papers on the subject of estimation of the mean of a continuous time stochastic process can be found in [5]. Brillinger considers continuous sampling, followed by discrete samples taken at regular intervals and irregularly spaced. He establishes asymptotic normality for an estimate of the mean and provides an approximation to the variance of this estimator.

For example, the rate at which customers arrive at a fast food restaurant would be well-modeled by a one-dimensional Poisson process (see Figure 1.1). If $d = 2$, one might consider the location of weather stations.

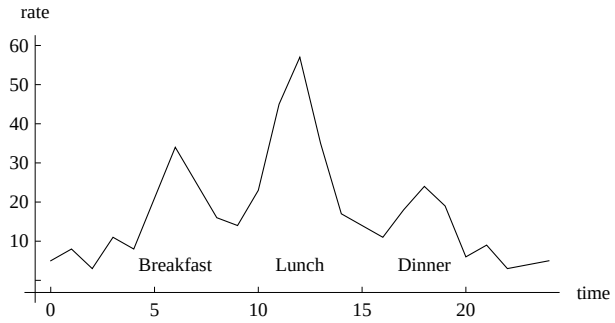

Figure 1.1: Customer arrivals at a fast food restaurant

There exists a vast amount of literature on inhomogeneous Poisson processes. Diggle, Lange, and Beneš [16], Schäbe [72], Cowling, Hall, and Phillips [9], Davison and Hinkley [14], Ventura, Davison, and Boniface [79], and Braun and Kulperger [4] provide excellent examples of data as well as provide different resampling methods and confidence intervals for the intensity function. Study of periodic or almost periodic intensities has been carried out by Lee, Wilson,

and Crawford [39], Kuhl, Wilson, and Johnson [34], Helmers and Mangku [24], Helmers, Mangku and Zitikis [25], and most recently, Shao and Lii [75].

A common approach to resample such data is a two-step method that first makes some parametric assumptions regarding the intensity function, $\lambda(t)$ or uses a non-parametric estimator and then resamples the data using a Monte Carlo method.

Snethlage [78] provides solid arguments against (blindly) applying a bootstrap procedure, such as a block bootstrap, to point process data as local structures will be destroyed. This is particular relevant for inhomogeneous point processes.

Our concern, however, is not on the point process itself, but rather the associated marks. Continuing with the above examples, if $d = 1$, then $X(t)$ might represent the amount of money a customer spends at a fast food restaurant. If $d = 2$, then $X(\mathbf{t})$ might represent the amount of rainfall at location $\mathbf{t}$ during a fixed time interval. Many authors have noted [10], [31], and [66] that irregularly spaced data seems to be the rule rather than the exception for $d > 1$.

A central assumption that we will make is that the location of our points is independent of the associated marks. As one might imagine, this is a fairly restrictive assumption as it precludes such data as the location and heights of trees in a forest. For example, as the density of trees increases, there is increased competition for resources and the heights will be reduced. [27] provides more discussion on this topic. Several tests have been proposed to test for the independence between the points and the marks. See [73], [8], [21], and [22] for more details.

The example of the fast food restaurant is a little complicated to begin with, so instead we shall consider a simpler example. Suppose we record the arrival time of customers at a liquor store. We might observations (plotted over time) that look like the following:

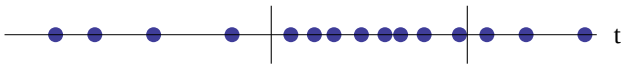

Figure 1.2: Distribution of points according to $\lambda(t)$

Notice that we have more data near the middle of the observation window than on the edges. Since we observe this clustering, we have reason to suspect that the data comes from an inhomogeneous point process. Indeed, the data comes from the following an Inhomogeneous Poisson Point Process with intensity function given as follows:

But as we said above, our primary focus is not the point process, but rather the associated marks. To that end, we are likely to encounter marked point process data like Figure 1.4.

Moreover, we will not limit ourselves to one-dimensional data. Indeed, in Chapter 7, we shall run simulations in two dimensions. Such a point process may look like Figure 1.5.

And the associated marks would look like Figure 1.6.

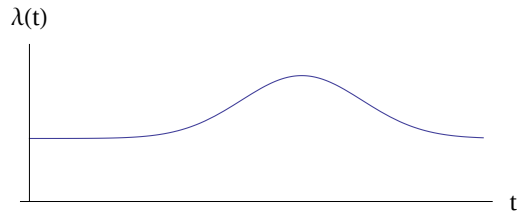

Figure 1.3: Intensity function, $\lambda(t)$

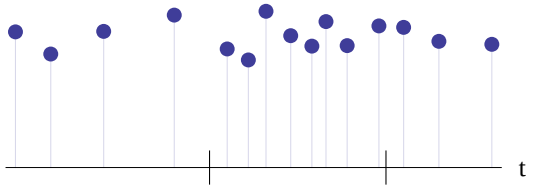

Figure 1.4: One-Dimensional Marked Point Process

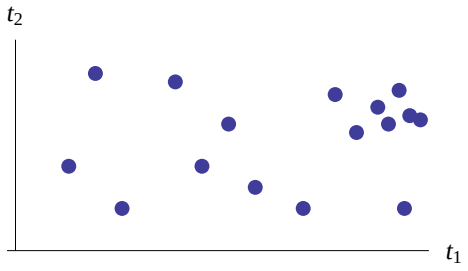

Figure 1.5: Two-Dimensional Point Process

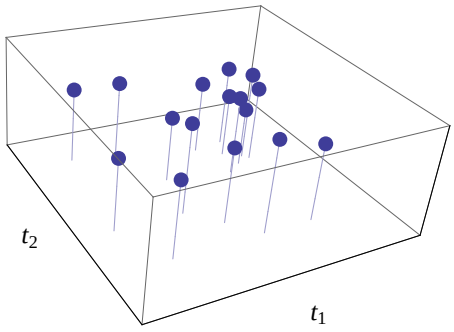

Figure 1.6: Two-Dimensional Marked Point Process

The central question of this thesis is how to accurately resample this data. To that end, we turn our attention to existing methods in the literature.

Chapter 2

Adaptation of Existing Methods

2.1 Introduction

As was mentioned in Chapter 1, our concern is with estimation of the mean, μ, of an inhomogeneous marked Poisson process. We have two estimators for the mean,

$$\tilde{X}_K = \frac{1}{\Lambda(K)} \int_K X(t) N(dt)$$

and

$$\bar{X}_K = \frac{1}{N(K)} \int_K X(t) N(dt)$$

In particular, our focus is on obtaining confidence intervals for the mean. To that end, we need to resample our data. Currently, there are no methods for resampling an inhomogeneous marked (Poisson) point process. There are many references to resampling the inhomogeneous Poisson process itself, however. See, for example, [40].

The closest related works are by Politis, Paparoditis, and Romano for resampling homogeneous marked point processes [53] and by Politis and Sherman for estimating moments from marked point processes [60]. The available literature all pertain to homogeneous point processes, though. As such, it would seem natural to try to use this approach to resample our data. As we shall use this method shortly, we record it below.

2.2 Block Bootstrap for a Homogeneous Marked Poisson Point Process

In their paper, Politis, Paparoditis, and Romano propose a block resampling scheme that would fill a block with *any* block of equal width from the entire window. We state their circular bootstrap algorithm below. (Note: This algorithm performs wrapping. A similar algorithm is also presented in the paper that does not wrap, but instead changes the probabilities near the boundary.)

1. Begin by imagining that K is "wrapped around" on a strip; in other words, we interpret the index t as being modulo K. If $t \notin K$ we will redefine $t = t$ (modulo K). With this definition, we have data $X(t)$ even if $t \notin K$.

2. Let $c = c(K)$ be a number in $(0, 1)$ depending on K such that $c \to 0$ but $cK \to \infty$. Define a scaled-down replica of K by $B = \{ct : t \in K\}$. B has the same shape as K but smaller dimensions. Also, define the displaced sets $B + y$ and let $L = \lfloor 1/c \rfloor$.

3. Generate random points $Y_1, Y_2, \ldots, Y_L$ independent and identically distributed from a uniform distribution on K and define

$$\tilde{X}^* \equiv \frac{1}{L} \sum_{i=1}^{L} \frac{1}{\lambda |B|} \int_{B+Y_i} X(t) N(dt)$$

and

$$\bar{X}^* \equiv \frac{1}{L} \sum_{i=1}^{L} \frac{1}{N(B + Y_i)} \int_{B+Y_i} X(t) N(dt)$$

Unfortunately, the algorithm above (as stated) is not appropriate in our setting, as the following sequence of figures illustrates.

Suppose we have the following distribution of points (see Figure 2.1) and our goal is to resample the second block. (Blocks are separated by the vertical lines.)

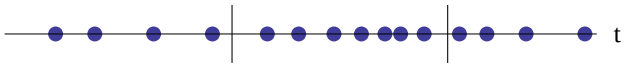

Figure 2.1: Original distribution of points

Notice that the second block is fairly dense (this would correspond to a larger value of $\lambda(t)$). Our block resampling scheme might choose the block illustrated by a red dash (see Figure 2.2).

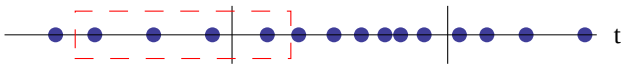

Figure 2.2: Choosing our resampling block

Notice that this block contains only 4 points whereas the original block contained 7. After we replace the second block with this block, our data has lost the clustering it had before (see Figure 2.3). In fact, this set of points looks like it could have come from a homogeneous Poisson process.

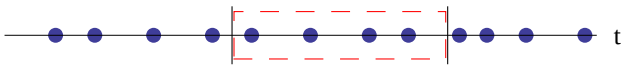

Figure 2.3: Resampled points

One of Politis' formers students, Arif Dowla, considered problems in time series [18] and since resampling techniques exist for time series data, it seemed like a good place to start. (It was thoughts of time series that motivated the opening example in Chapter 1). That is, suppose that we have marked point process data such as that in Figure 2.4.

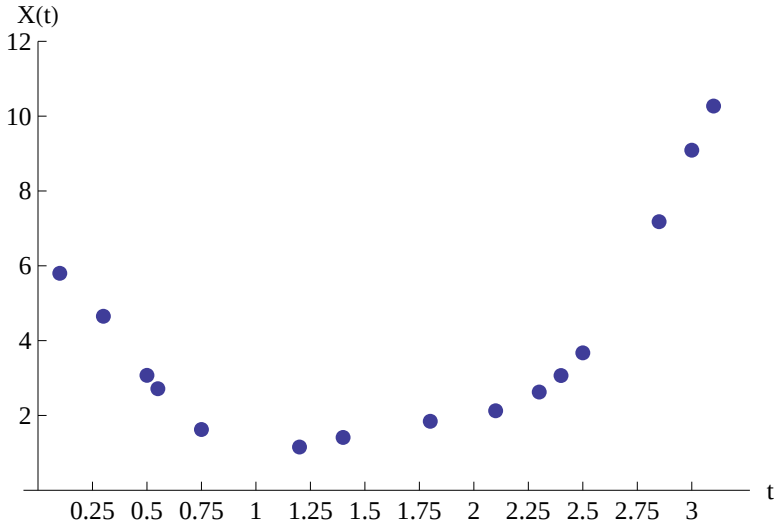

Figure 2.4: Example Marked Point Process Data

The data does not appear *too* different from time series data. If we impose a restriction that the points are not "too dense" and "somewhat" equally spaced, then they could be mapped to a grid, such as in Figure 2.5.

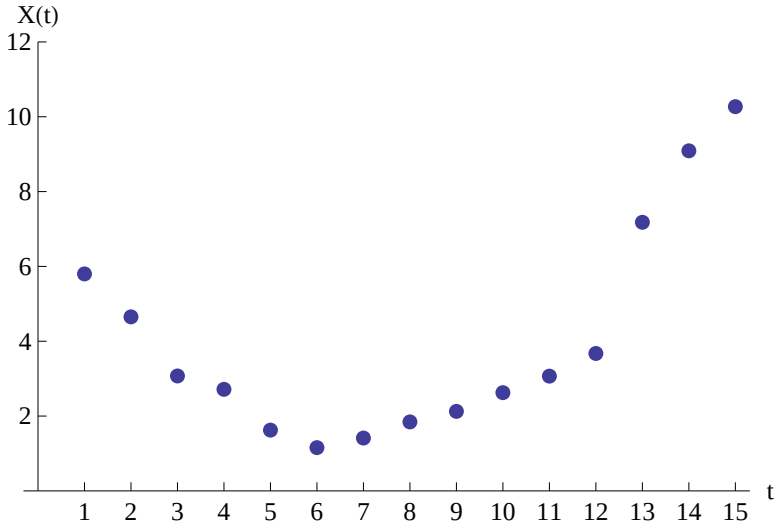

Figure 2.5: Mapping Marked Point Process Data to a Grid

While this would shift the values, with the (vague) assumptions, the impact would be minimal and it should be possible to get a handle on the problem. The issue is that these assumptions are not well-specified, nor realistic. It even precludes homogeneous Poisson process data, which is more closely related to the actual data that we would expect.

However, thinking about trying to reindex the data did lead to the discovery of one way to resample inhomgeneous Marked Poisson process data (for the one-dimensional case).

2.3 The Method

An inhomogeneous Poisson process $X(t)$ with rate $\lambda(t)$ can be transformed into a homogeneous Poisson process $Y(u(t))$ with rate 1 via the time transformation $u(t) = \Lambda(t)$. See [13] for an overview. In essence, the time axis is being stretched to spread the points out.

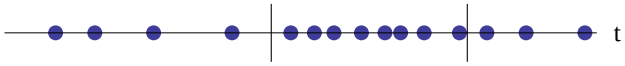

Figure 2.6: Inhomogeneous Poisson Process Before Remapping

The vertical lines in Figure 2.6 denote an area where the intensity, $\lambda(t)$ increases. Notice that in Figure 2.7, that area has been expanded.

But for this method to work, we need to know the cumulative intensity function, $\Lambda(t)$. The literature mentioned in Chapter 1 provide many ways to estimate the cumulative intensity function. Law and Kelton [38] suggest a non-parametric procedure for estimating $\lambda(t)$ with a

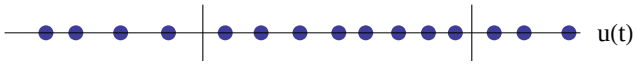

Figure 2.7: Inhomogeneous Poisson Process After Remapping

piecewise-constant function.

This requires one to divide the interval $[0, K]$ into non-overlapping pieces on which the intensity is assumed to be (fairly) constant and estimate a single rate for each interval.

Issues arise when determining the size of the pieces, which we shall denote w. (The method described allows for unequal widths, but for the sake of simplicity, we assume equal widths of our pieces, which is a divisor of K.) Suppose $w = 1$ and we have points as seen below in Figure 2.8.

Figure 2.8: Estimating $\Lambda(t)$ with $w = 1$

To estimate the intensity, we have $\hat{\lambda}(t) = \dfrac{\#\text{ of points}}{w}$. This gives us estimates for $\lambda(t)$ in Table 2.1.

Table 2.1: Estimates of $\hat{\lambda}(t)$ for $w = 1$

Interval	$(0,1]$	$(1,2]$	$(2,3]$	$(3,4]$	$(4,5]$	$(5,6]$	$(6,7]$	$(7,8]$	$(8,9]$	$(9,10]$
$\hat{\lambda}(t)$	2	2	1	0	3	3	1	1	2	1

If we use $w = 2$, instead, we arrive at a similar result.

Figure 2.9: Estimating $\Lambda(t)$ with $w = 2$

Table 2.2: Estimates of $\hat{\lambda}(t)$ for $w = 2$

Interval	$(0,2]$	$(2,4]$	$(4,6]$	$(6,8]$	$(8,10]$
$\hat{\lambda}(t)$	2	0.5	3	1	1.5

If we were to use too large of an w, say $w = 5$, then our estimate will suffer.

Figure 2.10: Estimating $\Lambda(t)$ with $w = 5$

Table 2.3: Estimates of $\hat{\lambda}(t)$ for $w = 5$

Interval	$(0, 5]$	$(5, 10]$
$\hat{\lambda}(t)$	1.6	1.6

In the last example, it would appear as though the intensity is constant on the entire interval.

So, the question remains as to what constitutes the optimal choice of w. Unfortunately, that answer changes depending on the data.

Lewis and Shedler [42] suggest a non-parametric kernel estimate of the form

$$\hat{\lambda}(t; K) = \frac{1}{h(K)} \sum_{j=1}^{N(K)} W\left(\frac{t - \tau_j}{h(K)}\right)$$

where $N(K)$ is the number of observations on $[0, K]$, $\tau_1, \ldots \tau_{N(K)}$ are the observed values, $W(\cdot)$ is a bounded non-negative integrable weight function with $\int_{-\infty}^{\infty} W(u)du = 1$ and $h(K)$ is a positive bandwidth function which tends to zero as $K \to \infty$, but in such a way that $o(h(K)) = 1/K$. Thus, it might be that $h(K) \sim K^{-1/2}$, for example.

The first method described can be thought of a special case of this kernel estimate, in which the weight function is rectangular (it applies equal weight to all observations in a particular window and no weight to any other observations).

But similar to above, this problem requires the user to specify both a bandwidth function and a weight function.

We shall also consider a second approach where we break the interval into a fine grid and estimate the intensity at each point by taking a local average (on $[-w, w]$, say). Both methods give us a reasonable approximation to the true intensity function, $\lambda(t)$.

Once we have an estimate for $\lambda(t)$, we have access to a way to resample an inhomogeneous marked (Poisson) point process. The idea is as follows:

1. Estimate the cumulative intensity function, $\Lambda(t)$.

2. Use the fact that an inhomogeneous Poisson process $N(t)$ with rate $\lambda(t)$ can be transformed to a homogeneous Poisson process $Y(u)$ with rate 1 via the time transformation $u = \Lambda(t)$. See [13] for an overview.

3. Apply the block bootstrap algorithm for resampling a homogeneous marked point process to $Y(u)$.

4. Invert $\Lambda(t)$ to map back to the inhomogeneous setting, if so desired.

The last step is is not necessary, as we are just concerned with the value of the marks, not necessarily of the original location of the points. Also, it should be noted that the methods we used above are not exhaustive. Many other techniques for estimating $\lambda(t)$ and $\Lambda(t)$ exist. For example, Leemis [40] provides another nonparametric technique for estimating $\Lambda(t)$ that does not require the user to specify any parameters or weighting functions.

2.4 Simulations

In the simulations that follow, we use accept-reject algorithms specified by Lewis and Shedler [43] to generate one-dimensional inhomogeneous Poisson process data. Pasupathy [52] is another excellent reference for generating (in)homogeneous Poisson processes data.

We run our simulations using the statistical software `R` [61]. The `R` code used for these simulations can be found in the appendix.

We consider five different intensity functions to generate the points.

First, we consider $\lambda_1(t) = 2 + \sin \frac{2\pi}{250} t$ for $0 \leq t \leq 1000$. This is shown in Figure 2.11.

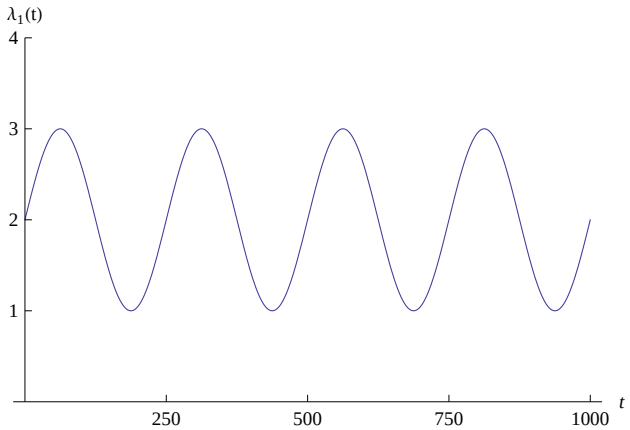

Figure 2.11: One-Dimensional Inhomogeneous Simulation Intensity Function 1

Second, we consider $\lambda_2(t) = 1 + \frac{1}{500} t$ for $0 \leq t \leq 1000$. This is shown in Figure 2.12.

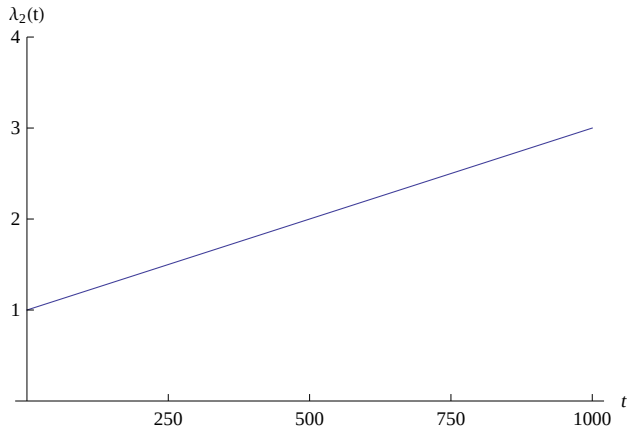

Figure 2.12: One-Dimensional Inhomogeneous Simulation Intensity Function 2

Third, we consider $\lambda_3(t) = 3 - \frac{1}{250}|t - 500|$ for $0 \leq t \leq 1000$. This is shown in Figure 2.13.

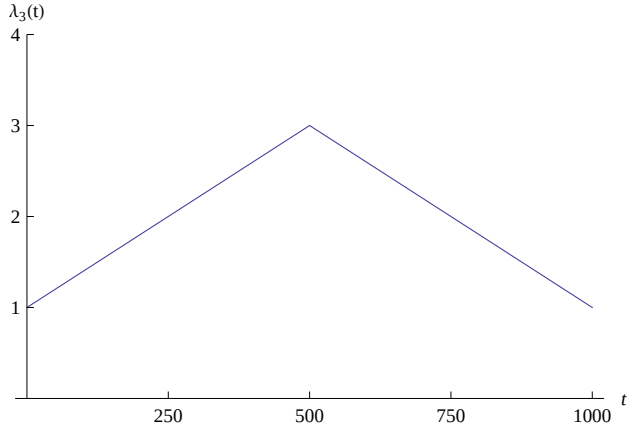

Figure 2.13: One-Dimensional Inhomogeneous Simulation Intensity Function 3

Fourth, we consider $\lambda_4(t) = 1 + \frac{3}{1000000}t^2$ for $0 \leq t \leq 1000$. This is shown in Figure 2.14.

Fifth, we consider $\lambda_5(t) = 4 - 2\left(\frac{t - 500}{450}\right)^6$ for $0 \leq t \leq 1000$. This is shown in Figure 2.15.

Once we have the points, we use two different covariance functions to generate associated marks. First, we consider $R(t) = \exp(-|t|)$ and second, we consider $R(t) = \exp(-|t|/3)$.

Since we need to transform the inhomogeneous Poisson process into a homogeneous Poisson process, we use three different transformation methods.

In Method 1, we estimate $\lambda(t)$ as a piecewise constant function on intervals of length w. By finding the area under these rectangles, we are able to estimate $\Lambda(t)$.

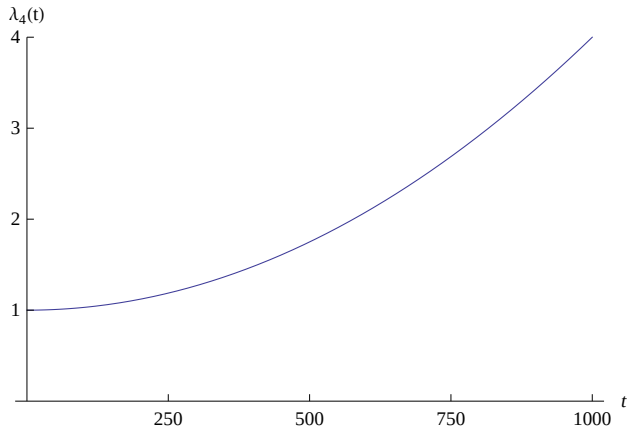

Figure 2.14: One-Dimensional Inhomogeneous Simulation Intensity Function 4

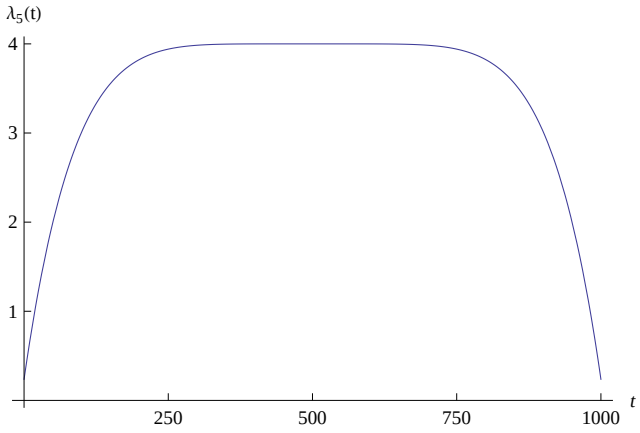

Figure 2.15: One-Dimensional Inhomogeneous Simulation Intensity Function 5

In Method 2, we estimate $\lambda(t)$ by taking a local average over a window $[-w, w]$. We construct a fine grid (with points every 0.1) and use this as our approximation to $\lambda(t)$. Again, by considering the area under the curve, we estimate $\Lambda(t)$.

In Method 3, we use the knowledge of $\lambda(t)$ to integrate the function and obtain an exact expression for $\Lambda(t)$. That is, $\Lambda(t) = \int_0^t \lambda(s)ds$. Notice that in the third method, we do not need an additional parameter, w, but is not a feasible method in practice.

Once we have obtained our homogeneous Poisson process data, we apply the block bootstrap techniques in [53] to resample our (now homogeneous) marked point process. This requires us to specify a block size, b, for which we consider three different choices ($b = 2, 5, 10$).

Putting all of the above together, we have a total of 210 models under consideration. Each model was simulated 1,000 times and each time, a 95% confidence interval was constructed.

At this point, a note on the choice of b is in order. The choice of block size was determined by the assumptions that appear in Section 4.3. Simply put, we need to choose a b that is of a much smaller order than K. Also, for the sake of simplicity, we consider b that divide K.

In terms of applying these methods to actual data, one needs estimates for w, the bin width and b, the block size. Estimating w for our piecewise constant and local averaging approaches has been well-studied. For example, see [71] and [36]. An optimal choice for b can be selected by modifying the method used for selecting blocks using subsampling methods in [28] for time-series data, which is itself an extension of work done by [20] and [3] for the case of i.i.d. data. This will be the focus of future work.

It was brought to the author's attention that Brillinger [5] considered the problem of estimation of the mean and showed how one may estimate the variance of $\bar{X}_K$. In particular, using our blocks of size b, we have that the variance of $\bar{X}_K$ is approximated by

$$\frac{1}{L(L-1)} \sum_{j=1}^{L} \left(\gamma_X^{(b)}(j) - \bar{X}_K \right)^2$$

where $L = K/b$ (which is assumed to be an integer) and

$$\gamma_X^{(b)}(j) = \frac{1}{N(jb) - N((j-1)b)} \int_{(j-1)b}^{jb} X(t)N(dt)$$

Note that we have corrected the typo in formula (25), where $\gamma_X^{(S)}$ was incorrectly listed as $\gamma_X^{(T)}$. Also, Brillinger makes use of assumptions based upon cumulants while we make assumptions on α-mixing. Some work is needed to show that our assumptions imply Brillinger's.

In the tables that follow, we present the mean and standard deviation of the constructed confidence intervals as well as the percentage of confidence intervals containing the true mean, 0

(the coverage probability).

When considering all of the tables, we see that the mean confidence interval length is around 0.1 while the standard deviation is around 0.025. Also, the average number of resampled observations varies, but stays within 5% of the true counts.

Examining Tables 2.4 - 2.8, we see that $b = 2$ seems to give the best coverage in all instances. For various choices of w, the results are roughly the same, though in Tables 2.4 - 2.6, $w = 10$ seems best while in Table 2.7, $w = 5$ seems best and in Table 2.8, $w = 20$ is best.

In Table 2.8, we observe overall poor coverage. This could be attributed to the fast changing intensities near the sides of the interval. However, this would not explain why such a large choice of w was optimal.

When examining Tables 2.9 - 2.13, we see some differences. First, the choice of $b = 2$ and $w = 10$ is optimal across all tables. Moreover, changes in b have an appreciable impact on the overall coverage in Tables 2.9 and 2.10. That is, we see a 5% decrease when increasing from $b = 2$ to $b = 5$ and an additional 5% decrease when increasing from $b = 5$ to $b = 10$. This is not present in the other tables, however.

Instead, the overall coverage drops significantly in Tables 2.11 - 2.13. It is not clear why only three out of five tables exhibit this behavior, though.

Table 2.4: Transformation Method with Intensity function $\lambda_1(t)$ and $R(t) = \exp(-|t|)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	2092.908	0.09316	0.02215	85.8%
		10	1947.576	0.09468	0.02344	87.9%
		20	2007.023	0.09425	0.02151	83.6%
	2	5	2058.042	0.09351	0.02175	83.7%
		10	2005.433	0.09376	0.02381	84.8%
		20	2038.794	0.09393	0.02188	82.0%
	3	NA	1946.399	0.09637	0.02618	95.0%
5	1	5	2024.743	0.09420	0.02164	78.8%
		10	1937.223	0.95000	0.02365	80.0%
		20	2068.244	0.09335	0.02203	78.2%
	2	5	2053.537	0.09373	0.02197	78.1%
		10	1951.529	0.09437	0.02315	81.4%
		20	2066.126	0.09337	0.02187	76.4%
	3	NA	1969.763	0.09597	0.02636	90.7%
10	1	5	2045.831	0.09374	0.02168	75.2%
		10	1961.638	0.09431	0.02345	78.0%
		20	2068.244	0.09303	0.02151	73.4%
	2	5	2075.651	0.09299	0.02162	77.1%
		10	1931.079	0.09510	0.02358	76.3%
		20	2036.970	0.09378	0.02159	74.4%
	3	NA	1933.351	0.09624	0.02574	87.0%

Table 2.5: Transformation Method with Intensity function $\lambda_2(t)$ and $R(t) = \exp(-|t|)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	1967.525	0.10034	0.02396	70.7%
		10	2063.676	0.09794	0.02338	85.4%
		20	2006.179	0.09981	0.02437	68.3%
	2	5	2006.785	0.09910	0.02358	73.7%
		10	2063.376	0.09694	0.02371	84.6%
		20	2030.599	0.09838	0.02339	73.1%
	3	NA	2038.185	0.09899	0.02414	94.9%
5	1	5	2008.460	0.09900	0.02368	72.2%
		10	2022.507	0.09733	0.02380	82.6%
		20	2047.619	0.09814	0.02366	71.2%
	2	5	2043.625	0.09851	0.02389	70.7%
		10	2074.556	0.09750	0.02403	82.7%
		20	2009.071	0.09929	0.02376	67.7%
	3	NA	2045.142	0.09773	0.02514	93.1%
10	1	5	2081.618	0.09751	0.02372	65.8%
		10	2076.823	0.09649	0.02322	82.7%
		20	2024.016	0.09869	0.02376	69.0%
	2	5	2033.735	0.09841	0.02386	69.1%
		10	2008.519	0.09885	0.02389	82.3%
		20	2057.804	0.09844	0.02376	67.7%
	3	NA	2044.047	0.09849	0.02487	92.0%

Table 2.6: Transformation Method with Intensity function $\lambda_3(t)$ and $R(t) = \exp(-|t|)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	2037.237	0.09326	0.02255	93.6%
		10	2062.927	0.09320	0.02299	95.1%
		20	2098.390	0.09232	0.02296	93.8%
	2	5	2041.594	0.09294	0.02239	93.7%
		10	2086.367	0.09314	0.02343	94.2%
		20	2094.782	0.09281	0.02339	93.3%
	3	NA	2075.337	0.09337	0.02371	93.5%
5	1	5	2106.712	0.09199	0.02277	87.2%
		10	2030.061	0.09387	0.02286	89.2%
		20	2042.671	0.09355	0.02313	86.5%
	2	5	1996.850	0.09447	0.02296	88.1%
		10	2032.069	0.09346	0.02262	88.6%
		20	2078.169	0.09285	0.02294	85.7%
	3	NA	2005.316	0.09400	0.02261	86.1%
10	1	5	2073.468	0.09279	0.02306	83.8%
		10	2097.910	0.09186	0.02242	87.3%
		20	2057.125	0.09330	0.02312	83.4%
	2	5	2010.870	0.09391	0.02263	84.6%
		10	2083.004	0.09278	0.02387	86.5%
		20	1983.265	0.09496	0.02306	80.8%
	3	NA	2034.399	0.09360	0.02270	80.8%

Table 2.7: Transformation Method with Intensity function $\lambda_4(t)$ and $R(t) = \exp(-|t|)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	2001.551	0.09284	0.02234	93.4%
		10	2006.680	0.09229	0.02195	93.1%
		20	2034.798	0.09197	0.02217	92.5%
	2	5	2041.922	0.09229	0.02268	94.4%
		10	2068.085	0.09137	0.02251	87.8%
		20	2033.748	0.09155	0.02185	93.1%
	3	NA	1994.686	0.09230	0.02150	93.5%
5	1	5	2043.547	0.09201	0.02254	89.4%
		10	1962.131	0.09347	0.02220	85.3%
		20	2019.791	0.92277	0.02221	88.1%
	2	5	1995.500	0.09198	0.02137	87.9%
		10	2077.124	0.09108	0.02216	87.8%
		20	1989.552	0.09265	0.02193	87.5%
	3	NA	1975.456	0.09308	0.02158	88.9%
10	1	5	2091.278	0.09027	0.02159	87.3%
		10	2017.657	0.09182	0.02172	84.7%
		20	1952.381	0.09322	0.02164	85.5%
	2	5	1980.919	0.09278	0.02183	83.7%
		10	2054.709	0.09137	0.02185	84.1%
		20	2024.077	0.09206	0.02220	85.0%
	3	NA	1974.671	0.09245	0.02114	85.5%

Table 2.8: Transformation Method with Intensity function $\lambda_5(t)$ and $R(t) = \exp(-|t|)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	3430.586	0.08431	0.03161	56.7%
		10	3528.629	0.08380	0.03210	57.8%
		20	3667.806	0.08116	0.03123	61.7%
	2	5	3509.555	0.08216	0.03079	61.5%
		10	3442.563	0.08387	0.03157	59.9%
		20	3487.787	0.08227	0.03082	62.3%
	3	NA	3380.220	0.08675	0.03446	59.6%
5	1	5	3484.546	0.08264	0.03079	56.4%
		10	3495.254	0.08314	0.03141	58.3%
		20	3510.250	0.08294	0.03148	58.1%
	2	5	3596.883	0.08175	0.03104	59.6%
		10	3615.060	0.08121	0.03084	61.4%
		20	3547.707	0.08245	0.03131	59.8%
	3	NA	3612.726	0.08390	0.03434	60.9%
10	1	5	3453.226	0.08332	0.03113	55.9%
		10	3400.524	0.08340	0.03073	56.0%
		20	3469.936	0.08272	0.03089	60.2%
	2	5	3558.986	0.08195	0.03079	59.0%
		10	3495.875	0.08223	0.03056	58.7%
		20	3585.228	0.08090	0.03033	59.1%
	3	NA	3392.772	0.08581	0.03384	60.3%

Table 2.9: Transformation Method with Intensity function $\lambda_1(t)$ and $R(t) = \exp(-|t|/3)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	2023.552	0.10275	0.02638	86.8%
		10	1910.069	0.10572	0.02635	87.8%
		20	2042.987	0.10255	0.02676	87.1%
	2	5	1965.352	0.10385	0.02623	84.2%
		10	2038.069	0.10257	0.02680	86.5%
		20	2017.845	0.10373	0.02737	83.4%
	3	NA	2045.079	0.10147	0.02552	88.3%
5	1	5	1952.009	0.10491	0.02671	75.1%
		10	2042.744	0.10194	0.02629	73.5%
		20	1957.214	0.10478	0.02687	74.9%
	2	5	2022.115	0.10311	0.02699	74.9%
		10	2040.431	0.10219	0.02674	73.4%
		20	1982.684	0.10354	0.02646	71.7%
	3	NA	2049.844	0.10158	0.02570	78.3%
10	1	5	2020.047	0.10274	0.02670	66.9%
		10	1950.018	0.10431	0.02635	66.7%
		20	1988.093	0.10332	0.02637	65.7%
	2	5	2009.881	0.10240	0.02590	65.9%
		10	2000.272	0.10322	0.02656	64.7%
		20	1953.765	0.10436	0.02644	62.6%
	3	NA	2032.483	0.10184	0.02559	66.9%

Table 2.10: Transformation Method with Intensity function $\lambda_2(t)$ and $R(t) = \exp(-|t|/3)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	1878.512	0.10018	0.02409	90.8%
		10	1907.633	0.09929	0.02395	92.6%
		20	1941.822	0.09836	0.02389	90.2%
	2	5	1933.702	0.09901	0.02437	90.2%
		10	1939.827	0.09843	0.02387	89.5%
		20	1920.165	0.09940	0.02413	89.5%
	3	NA	1928.076	0.09960	0.02534	87.4%
5	1	5	1900.411	0.09957	0.02410	80.5%
		10	1942.919	0.09853	0.02415	82.0%
		20	1959.518	0.09783	0.02371	80.7%
	2	5	1940.572	0.09816	0.02357	81.8%
		10	1953.456	0.09792	0.02374	81.7%
		20	1895.902	0.09936	0.02361	78.0%
	3	NA	1915.697	0.10000	0.02550	77.5%
10	1	5	1942.785	0.09831	0.02374	71.5%
		10	1916.959	0.09846	0.02322	71.5%
		20	1912.716	0.09887	0.02369	71.4%
	2	5	1948.358	0.09849	0.02414	70.9%
		10	1929.970	0.09807	0.02317	71.8%
		20	1932.286	0.09826	0.02332	70.6%
	3	NA	1956.429	0.09869	0.02541	72.0%

Table 2.11: Transformation Method with Intensity function $\lambda_3(t)$ and $R(t) = \exp(-|t|/3)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	2051.810	0.09073	0.02198	46.2%
		10	2000.916	0.09121	0.02119	50.3%
		20	1961.677	0.09216	0.02134	47.1%
	2	5	1999.232	0.09195	0.02217	45.8%
		10	2021.857	0.09133	0.02182	46.9%
		20	2003.470	0.09188	0.02203	43.4%
	3	NA	1980.930	0.09175	0.02077	47.2%
5	1	5	1961.745	0.09270	0.02213	46.0%
		10	1992.937	0.09182	0.02186	48.7%
		20	2039.625	0.09016	0.02119	46.5%
	2	5	1955.352	0.09281	0.02206	43.6%
		10	2015.492	0.09122	0.02192	46.1%
		20	1920.793	0.09351	0.02185	44.9%
	3	NA	2027.315	0.09017	0.02044	47.1%
10	1	5	1989.511	0.09203	0.02247	45.1%
		10	1985.328	0.09230	0.02232	45.1%
		20	1982.043	0.09239	0.02240	42.4%
	2	5	2009.688	0.09153	0.02205	45.4%
		10	2021.135	0.09136	0.02222	44.5%
		20	1993.055	0.09175	0.02186	45.6%
	3	NA	1972.788	0.09158	0.02077	43.9%

Table 2.12: Transformation Method with Intensity function $\lambda_4(t)$ and $R(t) = \exp(-|t|/3)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	2003.359	0.09976	0.02831	35.6%
		10	2007.652	0.09951	0.02826	36.4%
		20	2031.888	0.09892	0.02829	36.2%
	2	5	1993.151	0.09890	0.02734	38.4%
		10	2010.916	0.09923	0.02793	38.4%
		20	2005.910	0.09954	0.02825	37.4%
	3	NA	2039.875	0.09858	0.02797	36.6%
5	1	5	2023.721	0.09816	0.02721	35.4%
		10	2049.873	0.09853	0.02848	32.6%
		20	2003.407	0.09918	0.02808	35.2%
	2	5	2032.391	0.09898	0.02833	31.7%
		10	2001.556	0.09931	0.02792	33.3%
		20	2015.623	0.09919	0.02817	35.0%
	3	NA	1987.632	0.09928	0.02733	35.5%
10	1	5	2005.652	0.09885	0.02755	32.0%
		10	1970.179	0.10014	0.02787	32.0%
		20	2017.393	0.09872	0.02756	34.5%
	2	5	2012.970	0.09872	0.02766	32.3%
		10	2019.836	0.09833	0.02720	33.0%
		20	1948.918	0.09988	0.02690	34.9%
	3	NA	2018.217	0.09864	0.02766	31.3%

Table 2.13: Transformation Method with Intensity function $\lambda_5(t)$ and $R(t) = \exp(-|t|/3)$

b	Method	w	N Obs.	CI Len. Mean	CI Len. SD	Cover %
2	1	5	3532.622	0.07621	0.02929	48.8%
		10	3360.993	0.07680	0.02849	53.4%
		20	3431.828	0.07646	0.02870	53.9%
	2	5	3352.252	0.07747	0.02880	51.4%
		10	3425.007	0.07622	0.02855	54.1%
		20	3426.325	0.07664	0.02882	52.0%
	3	NA	3381.245	0.07949	0.03193	51.6%
5	1	5	3495.971	0.07539	0.02838	53.7%
		10	3395.725	0.07679	0.02859	52.7%
		20	3311.624	0.07741	0.02879	53.6%
	2	5	3508.099	0.07602	0.02902	52.1%
		10	3384.467	0.07712	0.02889	52.2%
		20	3446.214	0.07645	0.02900	52.9%
	3	NA	3475.020	0.07825	0.03157	53.5%
10	1	5	6435.853	0.07561	0.02854	51.0%
		10	3393.209	0.07695	0.02882	50.2%
		20	3411.386	0.07677	0.02903	49.7%
	2	5	3422.886	0.07606	0.02856	50.1%
		10	3415.915	0.07582	0.02815	53.7%
		20	3501.770	0.07595	0.02896	49.7%
	3	NA	3398.694	0.07932	0.03209	50.4%

Chapter 3

Theory for a One-Dimensional Inhomogeneous Poisson Process

3.1 Introduction

Our goal is statistical inference of the mean of the random process, $\mu = E[X(t)]$ on the basis of observing $\{X(t)\}$ for the points generated by an inhomogeneous Poisson process N over the closed interval $[0, K]$. There are two natural ways to estimate μ:

$$\tilde{X}_K = \frac{1}{\Lambda(K)} \int_K X(t) N(dt)$$

and

$$\bar{X}_K = \frac{1}{N(K)} \int_K X(t) N(dt)$$

Note: The only difference between the two formulas is the division by either the expected or actual sample size, respectively.

Karr [31] explored the case where N is a homogeneous Poisson process and established the fact that the above estimators are consistent and asymptotically normal at rate $\sqrt{|K|}$ with the same asymptotic variance. We state Karr's Theorem below as it is the springboard for our extension to the inhomogeneous setting.

Theorem 3.1.1 (Karr's Theorem)**.** *Let* $R(t) = Cov(X(0), X(t))$ *and assume that* $\int R(t)dt < \infty$

Also, let $\frac{1}{\sqrt{|K|}} \int_K (X(t) - \mu)dt \xrightarrow{d} N\left(0, \int R(t)dt\right)$

as $K \to \infty$*. Then, as* $K \to \infty$*, we have*

$$\sqrt{|K|}\left(\tilde{X}_K - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right) \text{ and } \sqrt{|K|}\left(\bar{X}_K - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right)$$

and furthermore,

$$\lim_{K\to\infty} \text{var}\left(\sqrt{|K|}\tilde{X}_K\right) = \lim_{K\to\infty} \text{var}\left(\sqrt{|K|}\bar{X}_K\right) = \sigma^2 \equiv \int R(t)dt + \frac{R(0)+\mu^2}{\lambda}$$

A key assumption Karr makes is the independence of the point process, N, and the associated marks, X. This is also the case in [47]. However, one can imagine in a forest, where the placement of trees follows a point process and their heights are the associated marks. Due to limited resources, we would expect the marks to be dependent on the point process. This complicates matters significantly. [27] is a great resource on this subject.

The goal of this chapter is to extend Karr's Theorem [31] to the inhomogeneous setting. Along the way we shall mimic many of the proof techniques that are employed by Politis, Paparoditis and Romano [53]. To that end, we begin with some background on mixing coefficients.

3.2 One-Dimensional Mixing Coefficients

We shall assume that our random process $\{X(t), t \in \mathbb{R}\}$ satisfies a certain weak dependence condition that will be quantified in terms of mixing coefficients.

Let $\rho(\cdot,\cdot)$ denote the distance in the l_∞-norm on $\mathbb{R}$. The strong mixing coefficients of Rosenblatt [67] are then defined as:

$$\alpha_X(k) \equiv \sup_{E_1,E_2\subset\mathbb{R}} \{|P(A_1\cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), i = 1,2, \rho(E_1,E_2) \geq k\}$$

where $\mathcal{F}(E_i)$ is the σ-algebra generated by $\{X(t), t \in E_i\}$.

In a similar manner, Doukhan [17] defined mixing conditions that also depend on the size of the sets considered.

$$\alpha_X(k;l_1,l_2) \equiv$$
$$\sup_{E_1,E_2\subset\mathbb{R}} \{|P(A_1\cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), |E_i| \leq l_i, i = 1,2, \rho(E_1,E_2) \geq k\}$$

Notice that $\alpha_X(k;l_1,l_2) \leq \alpha_X(k)$, and that in essence $\alpha_X(k) = \alpha_X(k;\infty,\infty)$.

Definition 3.2.1. A random field is said to be **$\boldsymbol{\alpha}$-strong mixing** if $\lim_{k\to\infty} \alpha_X(k) = 0$.

Definition 3.2.2. We shall define a weaker set of mixing called **$\bar{\boldsymbol{\alpha}}$-strong mixing**. Define

$$\begin{aligned}\bar{\alpha_X}(k;l) \quad \equiv \quad & \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), i = 1, 2, \\ & E_2 = E_1 + \mathbf{t}, |E_1| \leq l, \rho(E_1, E_2) \geq k\}\end{aligned}$$

where the supremum is taken over all closed intervals $E_1 \subset \mathbb{R}$, and over all $t \in \mathbb{R}$ such that $\rho(E_1, E_1 + t) \geq k$. Also, we define $\bar{\alpha}_X(k) = \bar{\alpha}_X(k; \infty)$.

Notice that $\bar{\alpha_X}(k) \leq \alpha_X(k)$, so that if the random field is α-strong mixing, then it will necessarily be $\bar{\alpha}$-strong mixing as well, and thus $\lim_{k\to\infty} \bar{\alpha}_X(k) = 0$.

More discussion and references on strong mixing coefficients can be found in Doukhan [17], Roussas and Ioannides [69], and Leonenko and Ivanov [41].

3.3 Assumptions

For our theoretical results, we make an assumption regarding the process $X(t)$ and an assumption regarding the mixing coefficients.

A.3.1. $0 < \lambda_{min} \leq \lambda(t) \leq \lambda_{max}$ (That is, $\lambda(t)$ is bounded above and below.)

A.3.2. $X(t)$ is wide-sense stationary with $E[|X(t)|^{2+\delta}] < \infty$ for some $\delta > 0$

A.3.3. $\bar{\alpha}(k; 0) < const.k^{-1-\varepsilon}$, where $\varepsilon > \dfrac{2}{\delta}$ for the δ specified in Assumption A.3.2

3.4 Main Results

We shall begin by showing that our estimators from Section 3.1 are unbiased for μ.

$$\begin{aligned}E\left[\tilde{X}_K\right] &= \frac{1}{\Lambda(K)} E\left[\int_K X(t)N(dt)\right] \\ &= \frac{1}{\Lambda(K)} E\left[E\left[\int_K X(t)N(dt)\,\middle|\, N\right]\right] \\ &= \frac{1}{\Lambda(K)} E\left[\int_K E\left[X(t)\right] N(dt)\right] \\ &= \frac{1}{\Lambda(K)} E\left[\int_K \mu N(dt)\right] \\ &= \frac{1}{\Lambda(K)} E\left[\mu N(K)\right] \\ &= \mu\end{aligned}$$

$$\begin{aligned}
E\left[\bar{X}_K\right] &= E\left[\frac{1}{N(K)}\int_K X(t)N(dt)\right] \\
&= E\left[\frac{1}{N(K)}E\left[\int_K X(t)N(dt)\middle| N\right]\right] \\
&= E\left[\frac{1}{N(K)}\int_K E\left[X(t)\right]N(dt)\right] \\
&= E\left[\frac{1}{N(K)}\int_K \mu N(dt)\right] \\
&= E\left[\frac{1}{N(K)}\mu N(K)\right] \\
&= \mu
\end{aligned}$$

Note: In both calculations above, we used the stationarity of $X(t)$. Also, we have adopted the convention that $0/0 = 1$. Next, we wish to discuss the variance of our estimators. The following lemma will prove useful as we go between $\tilde{X}_K$ and $\bar{X}_K$.

Lemma 3.4.1. *Assuming A.3.1, as* $K \to \infty$, $\dfrac{N(K)}{\Lambda(K)} \xrightarrow{p} 1$.

Note: There is actually a stronger result, namely that the convergence happens almost surely. For our needs, though, convergence in probability is enough. Armed with the above lemma, Slutsky's Theorem (see [74]) implies that convergence results that hold for $\tilde{X}_K$ also apply to $\bar{X}_K$. Some care needs to be taken when using this fact, though, as we shall see shortly.

To compute the variance of $\tilde{X}_K$, we begin by considering the following:

$$\begin{aligned}
E\left[\int_K\int_K X(t)N(dt)X(s)N(ds)\right] &= \int_K\int_K E\left[X(t)N(dt)X(s)N(ds)\right] \\
&= \int_K\int_K E\left[X(t)X(s)\right]E\left[N(dt)N(ds)\right]
\end{aligned}$$

where the second equality follows from the assumed independence of $X(\cdot)$ and $N(\cdot)$.

Observe, though, that for a wide-sense stationary process, we have:

$$E\left[X(t)X(s)\right] = \mathrm{cov}(X(t), X(s)) + E[X(t)]E[X(s)] = R(s-t) + \mu^2$$

and

$$E\left[N(dt)N(ds)\right] = \begin{cases} E\left[N(dt)\right]E[N(ds)] & \text{if } t \neq s \\ \left(E\left[N(dt)\right]\right)^2 & \text{if } t = s \end{cases}$$
$$= \begin{cases} \lambda(t)\lambda(s)dtds & \text{if } t \neq s \\ (\lambda(t)dt)^2 + \lambda(t)dt & \text{if } t = s \end{cases}$$

This follows since $N(dt)$ is a Poisson$(\lambda(t)dt)$ random variable. Thus, its first moment is equal to $\lambda(t)dt$ and its second moment is equal to square of the first moment plus the variance, i.e. $(\lambda(t)dt)^2 + \lambda(t)dt$.

Thus, we have that

$$E\left[\int_K \int_K X(t)N(dt)X(s)N(ds)\right] = \int_K \int_K \left(R(s-t)+\mu^2\right)\lambda(t)\lambda(s)dtds + \int_K \left(R(0)+\mu^2\right)\lambda(t)dt$$

And so, it follows that

$$\begin{aligned} \text{var}(\tilde{X}_K) &= \frac{1}{\left[\Lambda(K)\right]^2}\text{var}\left(\int_K X(t)N(dt)\right) \\ &= \frac{1}{\left[\Lambda(K)\right]^2}\left(\begin{array}{l} E\left[\int_K \int_K X(t)N(dt)X(s)N(ds)\right] \\ -E\left[\int_K X(t)N(dt)\right]E\left[\int_K X(s)N(ds)\right] \end{array}\right) \\ &= \frac{1}{\left[\Lambda(K)\right]^2}\left(\begin{array}{l} \int_K \int_K \left(R(s-t)+\mu^2\right)\lambda(t)\lambda(s)dtds \\ +\int_K \left(R(0)+\mu^2\right)\lambda(t)dt - \mu^2[\Lambda(K)]^2 \end{array}\right) \\ &= \frac{1}{\left[\Lambda(K)\right]^2}\left(\int_K \int_K R(s-t)\lambda(t)\lambda(s)dtds + \int_K \left(R(0)+\mu^2\right)\lambda(t)dt\right) \end{aligned}$$

Just as a check to ensure that our formula is correct, suppose that instead we had a homogeneous Poisson process.

In that instance, we have that $\lambda(t) = \lambda$ for all t, at which point the above formula reduces to:

$$\begin{aligned}
\mathrm{var}(\tilde{X}_K) &= \frac{1}{(\lambda K)^2}\left(\lambda^2 \int_K \int_K R(s-t)dtds + \left(R(0)+\mu^2\right)\lambda K\right) \\
&= \frac{1}{K^2}\int_K \int_K R(s-t)dtds + \frac{R(0)+\mu^2}{\lambda K}
\end{aligned}$$

Letting $u = s - t$, then $du = ds$, and we have

$$\begin{aligned}
\frac{1}{K^2}\int_K \int_K R(s-t)dtds &= \frac{1}{K^2}\int_0^K \int_{-t}^{K-t} R(u)dudt \\
&= \frac{1}{K^2}\left(\int_{-K}^{0}\int_{-u}^{K} R(u)dtdu + \int_0^K \int_0^{K-u} R(u)dtdu\right) \\
&= \frac{1}{K^2}\left(\int_{-K}^{0}(K+u)R(u)du + \int_0^K (K-u)R(u)du\right) \\
&= \frac{1}{K^2}\left(K\int_{-K}^{0}\left(1+\frac{u}{K}\right)R(u)du + K\int_0^K\left(1-\frac{u}{K}\right)R(u)du\right) \\
&= \frac{1}{K}\int_{-K}^{K}\left(1-\frac{|u|}{K}\right)R(u)du
\end{aligned}$$

which gives us precisely Prop 10.8 in [31] (pg 391).

At this point, it is worth addressing some confusion that exists in the literature. Karr [31] gives the formula listed above. However, in their paper, Politis, Paparoditis, and Romano [53] "correct" the formula by eliminating the μ^2. This might come from the following:

$$\begin{aligned}
\mathrm{var}(\tilde{X}_K) &= \frac{1}{[\Lambda(K)]^2}\mathrm{var}\left(\int_K X(t)N(dt)\right) \\
&= \frac{1}{[\Lambda(K)]^2}E\left[\int_K\int_K (X(t)-\mu)(X(s)-\mu)N(dt)N(ds)\right] \\
&= \frac{1}{[\Lambda(K)]^2}\int_K\int_K E\left[(X(t)-\mu)(X(s)-\mu)N(dt)N(ds)\right] \\
&= \frac{1}{[\Lambda(K)]^2}\int_K\int_K E\left[(X(t)-\mu)(X(s)-\mu)\right]E\left[N(dt)N(ds)\right] \\
&= \frac{1}{[\Lambda(K)]^2}\left(\int_K\int_K R(s-t)\lambda(t)\lambda(s)dtds + \int_K R(0)\lambda(t)dt\right)
\end{aligned}$$

So, what happened? The steps above seem compelling. The fallacy is with the second line. Suppose, for the sake of contradiction, that the second line were true and that $X(t) = \mu$ for all t. Since $E\left[(X(t) - \mu)\right] = 0$, the reasoning above would give us that the variance would be equal to 0.

But Campbell's Theorem [32] (stated below without proof) gives us $\text{var}(\tilde{X}_K) = \dfrac{\mu^2}{\Lambda(K)}$.

Theorem 3.4.2 (Campbell's Theorem)**.** *Let N be a Poisson process on K with mean measure $\lambda(t)$, and let $X(t) : K \to \mathbb{R}$ be measurable. Then the sum $\Sigma = \sum_{t \in N} X(t)$ is absolutely convergent in probability if and only if $\int_K \min(|X(t)|, 1)\lambda(t)dt < \infty$.*

If this condition holds, then $E[e^{\theta\Sigma}] = e^{\int_K (e^{\theta X(t)} - 1)\lambda(t)dt}$ for any complex θ for which the integral on the right converges. Moreover,

$$E[\Sigma] = \int_K X(t)\lambda(t)dt$$

in the sense that the expectation exists if and only if the integral converges, and they are equal. If the expected value converges, then

$$\text{var}[\Sigma] = \int_K [X(t)]^2 \lambda(t)dt,$$

finite or infinite.

Under Assumption A.3.1, we have that

$$\int_K \min(|\mu|, 1)\lambda(t)dt \leq \max(\mu K, \lambda_{max} K) < \infty$$

Also, we have that

$$E\left[\int_K \mu N(dt)\right] = \int_K \mu\lambda(t)dt = \mu\Lambda(K),$$

so we have that the expected value exists and the integral converges. Thus, we have that

$$\text{var}\left[\int_K \mu N(dt)\right] = \int_K \mu^2\lambda(t)dt = \mu^2\Lambda(K).$$

The question remains: what is wrong with subtracting μ from $X(t)$? The reason is there are two sources of variability: $X(t)$ and $N(t)$. So, even if $X(t)$ is fixed, there will still be some variability in our estimator, $\tilde{X}_K$.

We can see this another way, by conditioning on N.

$$\begin{aligned}
\mathrm{var}(\tilde{X}_K) &= \frac{1}{[\Lambda(K)]^2}\mathrm{var}\left(\int_K \mu N(dt)\right) \\
&= \frac{1}{[\Lambda(K)]^2}\left(\mathrm{var}\left[E\left[\int_K \mu N(dt)\middle| N\right]\right] + E\left[\mathrm{var}\left[\int_K \mu N(dt)\middle| N\right]\right]\right) \\
&= \frac{1}{[\Lambda(K)]^2}\left(\mathrm{var}\left[E\left[\mu\right| N\right]\int_K N(dt)\right] + E\left[\int_K \mathrm{var}\left[\mu\right| N\right] N(dt)\right]\right) \\
&= \frac{1}{[\Lambda(K)]^2}\left(\mu^2\Lambda(K) + 0\right) \\
&= \frac{\mu^2}{\Lambda(K)}
\end{aligned}$$

This, in turn, led to an examination of the variance of $\bar{X}_K$. If $X(t) = \mu$, then $Var\left[\bar{X}_K\right] = 0$, and μ is strangely absent. It is possible that Politis, Paparoditis, and Romano [53] were considering $\bar{X}_K$ when they argued that μ^2 should not be present.

This belief is further supported by considering the proof of their Theorem 3. They claim that

$$\sqrt{|K|}\left(\tilde{X}_K - \mu + \mu - \bar{X}_K\right) = \frac{N(K) - \lambda|K|}{N(K)\lambda\sqrt{|K|}}\int_K (X(t) - \mu)N(dt)$$

which is used as justification as to why $\sqrt{|K|}\left(\tilde{X}_K - \mu\right)$ and $\sqrt{|K|}\left(\bar{X}_K - \mu\right)$ have the same asymptotic distribution and same asymptotic variance.

However, the statement is not true. Indeed,

$$\begin{aligned}
&\frac{N(K) - \lambda|K|}{N(K)\lambda\sqrt{|K|}}\int_K (X(t) - \mu)N(dt) \\
&= \sqrt{|K|}\left(\frac{1}{\lambda|K|}\int_K (X(t) - \mu)N(dt) - \frac{1}{N(K)}\int_K (X(t) - \mu)N(dt)\right) \\
&= \sqrt{|K|}\left(\frac{1}{\lambda|K|}\int_K X(t)N(dt) - \frac{N(K)}{\lambda|K|}\mu + \mu - \frac{1}{N(K)}\int_K X(t)N(dt)\right) \\
&= \sqrt{|K|}\left(\tilde{X}_K - \frac{N(K)}{\lambda|K|}\mu + \mu - \bar{X}_K\right)
\end{aligned}$$

As they point out in the lines preceeding this equality, $\frac{N(K)}{\lambda|K|} = 1 + O_p\left(1/\sqrt{|K|}\right)$. Thus, $\sqrt{|K|}\left(\tilde{X}_K - \bar{X}_K\right) \neq o_p(1)$ as claimed and hence $\sqrt{|K|}\left(\tilde{X}_K - \mu\right)$ and $\sqrt{|K|}\left(\bar{X}_K - \mu\right)$ do not necessarily share the same asymptotic properties.

These differing results as well as the confusion in the literature led to a deeper exploration of the two estimators and a discovery that there is some truth and fallacy in both [31] and [53]. In particular, Theorem 3.1.1 is not correct as stated.

More precisely, the result stated are true for $\tilde{X}_K$, but are not correct for $\bar{X}_K$ as the following simple example illustrates. Suppose that the marks are i.i.d. $N(\mu, \sigma^2)$. Let us compute the variance of both of our appropriately centered and scaled estimators. We begin with $\tilde{X}_K$.

$$\begin{aligned} Var\left[\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right)\right] &= Var\left[E\left[\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right)\middle| N(K)\right]\right] \\ &\quad + E\left[Var\left[\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right)\middle| N(K)\right]\right] \\ &= Var\left[\frac{N(K) - \Lambda(K)}{\sqrt{\Lambda(K)}}\mu\right] + E\left[\frac{N(K)}{[\Lambda(K)]^2}\sigma^2\right] \\ &= \frac{\mu^2 + \sigma^2}{\Lambda(K)} \end{aligned}$$

where the last line holds since $E[N(K)] = Var[N(K)] = \Lambda(K)$ gives us that $Var\left[\frac{N(K) - \Lambda(K)}{\sqrt{\Lambda(K)}}\right] = 1$.

Turning our attention to $\bar{X}_K$, we have that

$$\begin{aligned} Var\left[\sqrt{\Lambda(K)}\left(\bar{X}_K - \mu\right)\right] &= Var\left[E\left[\sqrt{\Lambda(K)}\left(\bar{X}_K - \mu\right)\middle| N(K)\right]\right] \\ &\quad + E\left[Var\left[\sqrt{\Lambda(K)}\left(\bar{X}_K - \mu\right)\middle| N(K)\right]\right] \\ &= Var\left[0\right] + E\left[\frac{N(K)}{[N(K)]^2}\sigma^2\right] \\ &\approx \frac{\sigma^2}{\Lambda(K)} \end{aligned}$$

Strictly speaking, the last expectation is infinite since $P(N(K) = 0) > 0$. However, since $E\left[N(K)\right] / \sqrt{Var\left[N(K)\right]}$ tends towards infinity as $K \to \infty$, the region where $1/N(K)$ is large has an extremely small probability and thus can be ignored. The approximation comes from an asymptotic series expansion where the higher order terms have been omitted. Slightly more formal, the approximation holds since $N(K) \approx \Lambda(K)$ for large K by the Law of Large Numbers and thus $E\left[N(K)\right] \approx 1/\Lambda(K)$.

This also calls into question Karr's claim regarding the asymptotic normality of $\bar{X}_K$. However, we can appeal to work done by Brillinger [5]. In particular, we can transform our inhomogeneous Poisson process to a homogeneous Poisson process using the transformation,

$\Lambda(t)$ discussed in Chapter 2. Brillinger established the asymptotic normality of $\bar{X}_K$ and showed that the variance does not contain a μ^2, which is in agreement with our work above. However, he bases his proofs on assumptions of cumulants while we make assumptions on α-mixing. Additional work is needed to show that our assumptions imply Brillinger's.

Notice, however, that in the case of $\mu = 0$, we may apply Slutsky's Theorem [74] both $\tilde{X}_K$ and $\bar{X}_K$ are asymptotically normal with the same variance, as Karr claimed.

Our last goal is to establish a central limit theorem. Suppose that $X(t)$ is wide-sense stationary with finite second moment and the covariance function is absolutely integrable, i.e.

$$\int |R(t)|\, dt < \infty. \tag{3.4.1}$$

Let

$$\frac{1}{\sqrt{\Lambda(K)}} \int_K (X(t) - \mu)\lambda(t)dt \xrightarrow{d} N\left(0, \theta^2\right) \tag{3.4.2}$$

as $K \to \infty$, for a θ^2 that will be defined below.

Lemma 3.4.3. *If there exists a $\delta > 0$ such that Assumptions A.3.2 and A.3.3 hold, then equations (3.4.1) and (3.4.2) hold.*

Note: Different sufficient conditions for equations (3.4.1) and (3.4.2) can be found in Rozanov [70].

Above, we made reference to θ^2. Necessarily, it will be the asymptotic variance of

$$\frac{1}{\sqrt{\Lambda(K)}} \int_K (X(t) - \mu)\lambda(t)dt$$

To that end, we compute it. Let $\hat{X}(t) = X(t) - \mu$. Notice that

$$
\begin{aligned}
&\operatorname{var}\left(\frac{1}{\sqrt{\Lambda(K)}}\int_K \hat{X}(t)\lambda(t)dt\right)\\
&=\frac{1}{\Lambda(K)}\left(\begin{array}{l} E\left[\left(\int_K \hat{X}(s)\lambda(s)ds\right)\left(\int_K \hat{X}(t)\lambda(t)dt\right)\right]\\ -E\left[\left(\int_K \hat{X}(s)\lambda(s)ds\right)\right]E\left[\left(\int_K \hat{X}(t)\lambda(t)dt\right)\right]\end{array}\right)\\
&=\frac{1}{\Lambda(K)}\int_K\int_K E[\hat{X}(s)\hat{X}(t)]\lambda(s)\lambda(t)dsdt-0\\
&=\frac{1}{\Lambda(K)}\int_K\int_K R(s-t)\lambda(s)\lambda(t)dsdt
\end{aligned}
$$

So, we have that $\theta^2=\lim_{K\to\infty}\frac{1}{\Lambda(K)}\int_K\int_K R(s-t)\lambda(s)\lambda(t)dsdt$.

By Assumption A.3.1, we have that

$$
\begin{aligned}
\frac{1}{\Lambda(K)}\int_K\int_K R(s-t)\lambda(s)\lambda(t)dsdt &\leq \frac{(\lambda_{max})^2}{\Lambda(K)}\int_K\int_K |R(s-t)|\,dsdt\\
&= \frac{(\lambda_{max})^2K}{\Lambda(K)}\int_{-K}^{K}\left(1-\frac{|u|}{K}\right)|R(u)|\,du\\
&\leq \frac{(\lambda_{max})^2K}{\Lambda(K)}\int_{-\infty}^{\infty}|R(u)|\,du
\end{aligned}
$$

By equation (3.4.1), we have that this integral is finite, so we are assured that $\theta^2<\infty$.

Theorem 3.4.4. *Assuming equations (3.4.1) and (3.4.2), then as $K\to\infty$, we have that*

$$\sqrt{\Lambda(K)}\left(\tilde{X}_K-\mu\right)\xrightarrow{d}N\left(0,\sigma^2\right)$$

and

$$\sqrt{\Lambda(K)}\left(\bar{X}_K-\mu\right)\xrightarrow{d}N\left(0,\phi^2\right),$$

where $\phi^2=\theta^2+R(0)$ and $\sigma^2=\phi^2+\mu^2$.

For $\tilde{X}_K$, our result is analogous to Karr [31]. And for $\bar{X}_K$, our result is analogous to Politis, Paparoditis, and Romano [53]. The reason we are able to extend results from the

homogeneous setting to the inhomogeneous setting is the generality of Theorem 3.4.2 which applies to both homogeneous and inhomogeneous point processes.

3.5 Proofs

Proof of Lemma 3.4.1: We begin by establishing the result for the homogeneous case, then generalizing to the inhomogeneous setting. Note: $N(t)$ is short-hand for $N(0,t]$, the number of points on the interval $(0,t]$.

Suppose $N(K)$ is a homogeneous Poisson process with intensity λ (so we have that $N(K) \sim \text{Poisson}(\lambda K)$). Then

$$E[N(K)] = \lambda K$$

and

$$Var[N(K)] = \lambda K$$

By Markov's inequality, for any $\varepsilon > 0$,

$$P\left(\left|\frac{N(K)}{K} - \lambda\right| \geq \varepsilon\right) \leq \frac{\lambda}{\varepsilon^2 K}$$

Now, let $t_j = j^2$. Then the above inequality implies that

$$\sum_{j=1}^{\infty} P\left(\left|\frac{N(j^2)}{j^2} - \lambda\right| \geq \varepsilon\right) < \infty$$

So, by the Borel-Cantelli lemma, with probability 1, we have that

$$\left|\frac{N(j^2)}{j^2} - \lambda\right| \geq \varepsilon$$

for at most a finite number of integer values of j. Since $\varepsilon > 0$ is arbitrary, we have

$$\lim_{K\to\infty} \frac{N(j^2)}{j^2} = \lambda$$

with probability 1. To obtain our desired result, take j to be the integer part of $K^{1/2}$, so that for $K > 1$,

$$N(j^2) \leq N(K) \leq N((j+1)^2)$$

and

$$j^2 \leq K \leq (j+1)^2$$

Since $(j+1)^2/j^2 \to 1$, the proof is complete for the homogeneous case.

Now, suppose that N is an inhomogeneous Poisson process. Define $N_1 = \Lambda(N)$. Then N_1 is a homogeneous Poisson process with rate 1. (Note: This is the same transformation used in Chapter 2.) And so, with probability 1, we have that $\lim_{K\to\infty} \frac{N(K)}{\Lambda(K)} = 1$.

□

Proof of Lemma 3.4.3: We appeal to Theorem 1.7.1 of Leonenko and Ivanov [41] where they employ a weaker notion than our $\bar{\alpha}$-mixing.

We only need to verify the finiteness of $\int |R(t)|\, dt < \infty$, as that is enough to ensure that the limiting variance of $\frac{1}{\sqrt{\Lambda(K)}} \int_K (X(t) - \mu)\lambda(t)dt$ is finite.

A result in Roussas and Ioannides [69] (which requires Assumption A.3.2) states

$$|\mathrm{cov}(X(0), X(t)| \leq \mathrm{const.} \bar{\alpha}_X(|t|\,; 0)^{1-2/(2+\delta)}$$

Thus, we have that

$$\begin{aligned}
\int |R(t)|\, dt &= O\left(\int \bar{\alpha}_X(|t|\,;0)^{1-2/(2+\delta)} dt \right) \\
&= O\left(\int_1^\infty \left(\frac{1}{y^{1+\varepsilon}} \right)^{1-2/(2+\delta)} dy \right) \\
&= O\left(\int_1^\infty \frac{1}{y^{1+(\varepsilon\delta-2)/(2+\delta)}} dy \right) \\
&< \infty
\end{aligned}$$

The first line follows from the inequality from Roussas and Ioannides.

The second line follows by letting $y = |t|$, using the bound on the mixing coefficients (Assumption A.3.3), and recognizing that the contribution of the integral from 0 to 1 does not affect the order. (We can bound the integrand by $|R(0)|$.)

The third line is simple algebra and the finiteness comes from the assumption that $\varepsilon > \frac{2}{\delta}$ so we have that $\frac{\varepsilon\delta - 2}{2+\delta} > 0$, and thus the integral converges.

□

Proof of Theorem 3.4.4: Our strategy is to show that the characteristic function of

$$\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right)$$

converges to that of a $N(0, \sigma^2)$ random variable.

$$\begin{aligned}
&E\left[\exp\left(i\alpha\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right)\right)\right] \\
&= E\left[E\left[\exp\left(i\alpha\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right)\right)\middle| X\right]\right] \\
&= E\left[E\left[\exp\left(\tfrac{i\alpha}{\sqrt{\Lambda(K)}}\int_K X(t)N(dt) - \tfrac{i\alpha}{\sqrt{\Lambda(K)}}\int_K \mu\lambda(t)dt\right)\middle| X\right]\right] \\
&= E\left[\exp\left(\int_K\left(\exp\left(\tfrac{i\alpha}{\sqrt{\Lambda(K)}}X(t)\right) - 1\right)\lambda(t)dt - \tfrac{i\alpha}{\sqrt{\Lambda(K)}}\int_K \mu\lambda(t)dt\right)\right]
\end{aligned}$$

The second line follows from the properties of iterated expectation.

The third line follows by substituting $\tilde{X}_K$, using the fact that $\mu = \dfrac{1}{\sqrt{\Lambda(K)}}\displaystyle\int_K \mu\lambda(t)dt$, and distributing $i\alpha\sqrt{\Lambda(K)}$ through both terms inside of the parentheses.

The fourth line follows by applying Campbell's Theorem (Theorem 3.4.2) to the first term (which applies since X is deterministic after we condition on it) and recognizing that the second term is a constant with respect to the inner expectation.

From here, we consider a Taylor Series expansion of $\displaystyle\int_K\left(e^{\frac{i\alpha}{\sqrt{\Lambda(K)}}X(t)} - 1\right)\lambda(t)dt$

Recall, $e^x = 1 + x + x^2/2 + O(x^3)$. Thus, $e^x - 1 \approx x + x^2/2$. It should be noted that the error term converges in probability to 0. Hence, we have:

$$\begin{aligned}
&\int_K\left(e^{\frac{i\alpha}{\sqrt{\Lambda(K)}}X(t)} - 1\right)\lambda(t)dt \\
&\approx \int_K \frac{i\alpha}{\sqrt{\Lambda(K)}}X(t)\lambda(t)dt + \int_K \frac{1}{2}\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}X(t)\right)^2\lambda(t)dt \\
&= \int_K \frac{i\alpha}{\sqrt{\Lambda(K)}}X(t)\lambda(t)dt + \int_K \frac{i^2\alpha^2}{2\Lambda(K)}X^2(t)\lambda(t)dt \\
&= \frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K X(t)\lambda(t)dt - \frac{\alpha^2}{2\Lambda(K)}\int_K X^2(t)\lambda(t)dt
\end{aligned}$$

And so, we have that

$$E\left[\exp\left(\int_K \left(e^{\frac{i\alpha}{\sqrt{\Lambda(K)}}X(t)}-1\right)\lambda(t)dt - \frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K \mu\lambda(t)dt\right)\right]$$
$$\approx E\left[\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K X(t)\lambda(t)dt - \frac{\alpha^2}{2\Lambda(K)}\int_K X^2(t)\lambda(t)dt - \frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K \mu\lambda(t)dt\right)\right]$$
$$= E\left[\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K [X(t)-\mu]\,\lambda(t)dt - \frac{\alpha^2}{2\Lambda(K)}\int_K X^2(t)\lambda(t)dt\right)\right]$$

Recall, we assumed that $\frac{1}{\sqrt{\Lambda(K)}}\int_K (X(t)-\mu)\lambda(t)dt \xrightarrow{d} N\left(0,\theta^2\right)$. In terms of characteristic functions, we have

$$\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K (X(t)-\mu)\lambda(t)dt\right)$$
$$= \exp\left(i\alpha\left(\frac{1}{\sqrt{\Lambda(K)}}\int_K (X(t)-\mu)\lambda(t)dt\right)\right)$$
$$\xrightarrow{d} e^{-\frac{\alpha^2}{2}\theta^2}$$

By consistency, we have that

$$\frac{1}{\Lambda(K)}\int_K X^2(t)\lambda(t)dt \to E[X(t)^2] = Var[X(t)] + (E[X(t)])^2 = R(0)+\mu^2.$$

Thus,

$$E\left[\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K [X(t)-\mu]\,\lambda(t)dt - \frac{\alpha^2}{2\Lambda(K)}\int_K X^2(t)\lambda(t)dt\right)\right]$$
$$\to \exp\left(-\frac{\alpha^2}{2}\theta^2 - \frac{\alpha^2}{2}(R(0)+\mu^2)\right)$$

So, we have that $\sqrt{\Lambda(K)}\left(\tilde{X}_K-\mu\right) \xrightarrow{d} N\left(0,\sigma^2\right)$, where $\sigma^2 = \theta^2 + R(0) + \mu^2$.

Brillinger [5] establishes the asymptotic normality and variance of $\bar{X}_K$. As we noted above, the main difference is that unlike $\tilde{X}_K$, there is no μ^2.

□

Chapter 4

Local Block Bootstrap for a One-Dimensional Inhomogeneous Poisson Process

4.1 Introduction

In Chapter 3, we established the asymptotic normality of the sample mean. In order to construct confidence intervals for the mean μ, though, the asymptotic variance would need to be explicitly estimated. Difficulties arise when trying to consistently estimate $\int R(t)dt$ as we have irregularly spaced data.

The resampling method discussed in the next section is able to yield confidence intervals for the mean without the need for explicit estimation of the asymptotic variance. Alternatively, the resampling method may provide an estimate of the asymptotic variance to be used in connection with the asymptotic normality result established in Chapter 3.

4.2 Local Block Bootstrap Algorithm in One Dimension

Our observation interval K can be any closed interval in $\mathbb{R}$. For the sake of simplicity, though, we suppose that the interval starts at the origin. Thus, we are considering $[0, K]$. For our asymptotic results, K will be assumed to expand towards ∞.

Suppose we obtain observations $\tau_1, \tau_2, \ldots, \tau_{N(K)}$ from an inhomogeneous Poisson process with intensity function $\lambda(t)$. And suppose that for each τ_i, we observe a corresponding mark, $X(\tau_i)$. Together, this gives us an inhomogeneous marked point process.

We shall employ a local block bootstrap method to resample such data, as was introduced in [51]. Essentially, we resample the data in blocks (just like a block bootstrap), but when filling a particular block, we only consider blocks that are in a "local" neighborhood of the original block. The block size itself is a parameter while the proximity to the original block is determined by a second parameter.

There is not much literature on the local block bootstrap. It is first introduced in [51], and is used by Dowla in his thesis [18] and in a subsequent paper [19].

The local block bootstrap algorithm for generating $X^*(\tau_1), X^*(\tau_2), \ldots, X^*(\tau_{N^*(K)})$ will necessarily depend on the bootstrap point process, N^*. In particular, we have that $\Lambda^*(K) = E^*[N^*(K)]$. Our algorithm can be described as follows:

1. Select an integer block size b such that $b^{5/2}/K \to 0$ as $K \to \infty$
2. Choose an $\alpha > 2$.
3. Define the bandwidth parameter to be $h = O(b^\alpha)$
4. Define the total number of blocks to be $L = K/b$
5. For each of the $j = 1, 2, \ldots, L$ blocks, denote the coordinate of the point closest to the origin as c_j. That is, let $c_j = (j-1)b$.
6. Add a perturbation d_j to c_j where d_j are i.i.d. Uniform$[-h, h]$ random variables.
7. Form a new block of width b starting at $c_j + d_j$ and record which of the $X(\tau_1), \ldots, X(\tau_{N(K)})$ occur in this block.

   Note: It is possible that parts of the block may lay outside of $[0, K]$. To correct this, imagine that K is "wrapped around" on itself. That is, our calculations are done modulo K.
8. Let

   $\tilde{X}^* = \dfrac{1}{\Lambda^*(K)} \displaystyle\sum_{i=1}^{N(K)} W_i X(\tau_i)$, where W_i is the number of times that $X(\tau_i)$ occurs in the resampled data

   and

   $\bar{X}^* = \dfrac{1}{N^*(K)} \displaystyle\sum_{i=1}^{N(K)} W_i X(\tau_i)$

   Also, let

   $\approx{X}^* = \dfrac{1}{\Lambda(K)} \displaystyle\sum_{i=1}^{N(K)} W_i X(\tau_i)$

   and

$$\overline{\overline{X}}^* = \frac{1}{N(K)} \sum_{i=1}^{N(K)} W_i X(\tau_i)$$

This generation of points $X^*(\tau_1), ..., X^*(\tau_{N^*(K)})$ and subsequently of $\tilde{X}^*$, $\bar{X}^*$, $\tilde{\tilde{X}}^*$, and $\overline{\overline{X}}^*$ is governed by a probability mechanism which we will denote by P^*, with moments denoted by E^*, Var^*, etc. This generation is done conditionally on the marked point process data observed; thus P^* is really a conditional probability. Notice that the first two estimates involve the bootstrap point process N^*, while the last two do not.

9. Let $P^*(\sqrt{\Lambda(K)}(\tilde{\tilde{X}}^* - E^*[\tilde{\tilde{X}}^*]) \leq x)$ and $P^*(\sqrt{\Lambda(K)}(\overline{\overline{X}}^* - E^*[\overline{\overline{X}}^*]) \leq x)$ denote the conditional (given the marked point process data) distribution functions of the bootstrap sample means.

Note: We will not concern ourselves with the trivial matter of divisibility, and issues like K/b being an integer. The reason for this is that for a practical application with a finite sample, we can truncate the window size K and obtain perfect divisibility. As for the asymptotic case, we can always ignore truncations which are clearly of negligible order.

4.3 Assumptions

For our bootstrap results, we need to impose some various restrictions on the process $X(t)$, the mixing coefficients, and the parameters.

A.4.1. $0 < \lambda_{min} \leq \lambda(t) \leq \lambda_{max}$ (That is, $\lambda(t)$ is bounded above and below.)

A.4.2. $X(t)$ is wide-sense stationary with $E[|X(t)|^{6+\delta}] < \infty$ for some $\delta > 0$

A.4.3. $\bar{\alpha}(k;0) < const.k^{-1-\varepsilon}$, where $\varepsilon > \max\left\{\frac{2}{\delta}, \frac{8+\delta}{4+\delta}\right\}$ for the δ specified in Assumption A.4.2

A.4.4. $\int_{-\infty}^{\infty} |Q(u,v,v-w)|\, dv = C_Q < \infty$, for all u, w

A.4.5. $h/K \to 0$ as $K \to \infty$

A.4.6. $b^{5/2}/K \to 0$ as $K \to \infty$

Note: If we assume $\mu = 0$, we can use (A1) in the proof Theorem 3.3 in Kunsch [35] (which appeals to Theorem 17.2.3 of Ibragimov and Linnik [26]). That is, if we assume

A.4.7. $\int t^2 \alpha(t)^{\delta/(6+\delta)} dt < \infty$

then in conjunction with Assumption A.4.2 (with the same δ), we have that Assumption A.4.4 holds.

4.4 Main Results

Since the resampling process can be difficult to keep track of, it will help to consider examples (with illustrations). A few words are in order, though.

First, the α used in these examples is less than 2, but that does not affect the theory that we develop. Indeed, the results would hold for any $\alpha > 1$, but as we will see later, in order to ensure that our bootstrap variance tends to the true variance, we need to impose an additional restriction on α.

Second, the examples used have $h = \lceil b^{\alpha-1} \rceil b$. (Note that this is of order b^α, as required.) The reason for this is for the ease of the display of the probability calculations. While true in a more general setting (that is, $h = const.b^\alpha$), by requiring h to be an integer multiple of b, we have some nice geometric interpretations for many of the formulas.

Our goal is fairly straightforward: We want to show that our bootstrap mean and variance are consistent for their real-world counterparts and that we have a corresponding Central Limit Theorem in the bootstrap world.

Suppose we begin with a window $[0, K]$ and we define $h = \lceil b^{\alpha-1} \rceil b$ for $\alpha > 1$. Let $b = 3$ and $\alpha = 3/2$, so that $h = 6$.

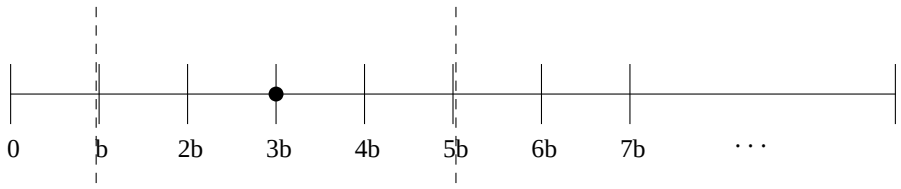

Figure 4.1: Resampling a Block Using the Local Block Bootstrap

As was mentioned in the algorithm, suppose we index the blocks that we wish to fill by their starting point. So, above, we wish to fill the block $3b$ (the fourth block). The vertical dashed lines show how far the starting point for the $3b$-block can move. (It can move left or right at most h, and we use toroidal wrapping if we reach an edge.)

Recall that $\tilde{X}^* = \dfrac{1}{\Lambda^*(K)} \displaystyle\sum_{i=1}^{N(K)} W_i X(\tau_i)$, where W_i is the number of times that $X(\tau_i)$ occurs in the resampled data.

We can think of W_i as a sum of Bernoulli random variables, where the sum runs over all

of the blocks. That is, suppose we have L blocks. Then

$$W_i = \sum_{j=1}^{L} Y_{ij}, \text{ where } Y_{ij} \sim \text{Bernoulli}(p_{ij})$$

Here, p_{ij} represents the probability that τ_i is contained in block j. Recall, τ_i denotes the position of the mark of the ith datapoint, while $X(\tau_i)$ is the mark at τ_i.

Lemma 4.4.1. $E^*[W_i] = 1$.

To see why the claim is true, let us work through an example. Suppose that we have eight blocks (with $b = 3$ and $\alpha = 3/2$, so $K = 24$) and only a single data point, τ_1, at $t = 7$.

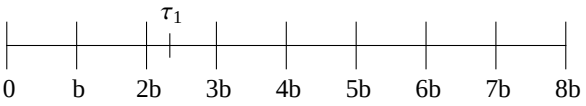

Figure 4.2: A single data point τ_1

Following our resampling scheme, the probability that τ_1 is included in block j (for $j = 0, 1, \ldots, 7$) is given in Table 4.1:

Table 4.1: Probabilities that τ_1 is included in block j

j	0	1	2	3	4	5	6	7
p_{1j}	2/12	3/12	3/12	3/12	1/12	0	0	0

Notice that all of the probabilities sum to 1. Since $E^*[Y_{ij}] = p_{ij}$ if $Y_{ij} \sim \text{Bernoulli}(p_{ij})$, we have that $E^*[W_i] = 1$.

The reason this works is because of our choice of h. Our resampling scheme allows us to choose a block uniformly from an integer multiple of b blocks (in fact, $2\left\lceil b^{\alpha-1}\right\rceil$ blocks).

Notice that the point τ_1 can only be include in the resampling if the block we choose comes from the shaded region that is within b units to the left of τ_1.

Figure 4.3: Visualizing how we include τ_1

The grey area is spread across two original blocks. (In fact, it can only be spread across at most two blocks.) How many blocks can contain the part in between b and $2b$? There are $2\left\lceil b^{\alpha-1}\right\rceil$ such blocks. The same is true for how many blocks can contain the part in between $2b$ and $3b$.

Thus, the entire grey area will appear $2\left\lceil b^{\alpha-1}\right\rceil$ times in the sum over all of the blocks.

The probabilities in Table 4.1 actually are just given by

$$p_{ij} = \frac{\text{part of grey area in block } j}{2h}$$

Since the entire grey area appears $2\left\lceil b^{\alpha-1}\right\rceil$ in the sum over all of the blocks and its area is equal to b, the total area is given by $2\left\lceil b^{\alpha-1}\right\rceil b = 2h$. Thus, we have that

$$\sum p_{ij} = \sum \frac{\text{part of grey area in block } j}{2h} = \frac{2h}{2h} = 1$$

as we claimed.

In Table 4.1, we did not specify a formula for computing the probabilities p_{1j}. Notice that the probabilities depend on both the block j that we wish to fill and the relative location of the point τ_1 to the nearest blocks.

By the division algorithm, we can express τ_1 as $\tau_1 = J_{1b} + R_1$, where $0 \leq R_1 < b$. Then we claim the following is true:

Claim 4.4.2. *If τ_1 is at least h units from the boundary, then*

$$p_{1j} = \frac{(b - R_1)1_{\{|\tau_1 - (j+1)b| < h\}} + R_1 1_{\{|\tau_1 - jb| < h\}}}{2h}.$$

Notice that $\tau_1 - (j+1)b < h$ corresponds to the event that block j will contain the portion of the grey area that appears in the block to the left of the block that τ_1 appeared in.

In our example above, τ_1 appears in the $[2b, 3b]$ block, so this represents the portion of the grey area in the previous figure that appears in the $[b, 2b]$ block.

That block has an area of $b - R_1$ by its construction. Thus, we have the first part of our formula, namely $\dfrac{(b - R_1)1_{\{|\tau_1 - (j+1)b| < h\}}}{2h}$.

We also have that $\tau_1 - jb < h$ corresponds to the event that block j will contain the portion of the grey area that appears in the same block as τ_1.

In our example above, τ_1 appears in the $[2b, 3b]$ block, so this represents the portion of the grey area from in the previous figure that appears in the $[2b, 3b]$ block.

That block has an area of R_1 by its construction. Thus, we have the second part of our formula, namely $\dfrac{R_1 1_{\{|\tau_1 - jb| < h\}}}{2h}$.

To see this numerically, consider Table 4.2, which is an expanded version of Table 4.1.

Those entries in the fourth row that have a value less than h ($h = 6$) represent the blocks for which the grey area in the $[b, 2b]$ block will be included. (The area in that block is precisely $b - R_1 = 3 - 1 = 2$.)

Table 4.2: Probabilities that τ_1 is included in block j, expanded

j	0	1	2	3	4	5	6	7
p_{1j}	2/12	3/12	3/12	3/12	1/12	0	0	0
jb	0	3	6	9	12	15	18	21
$\lvert\tau_1 - (j+1)b\rvert$	$\lvert 4\rvert$	$\lvert 1\rvert$	$\lvert -2\rvert$	$\lvert -5\rvert$	$\lvert -8\rvert$	$\lvert -11\rvert$	$\lvert -14\rvert$	$\lvert -17\rvert$
$\lvert\tau_1 - jb\rvert$	$\lvert 7\rvert$	$\lvert 4\rvert$	$\lvert 1\rvert$	$\lvert -2\rvert$	$\lvert -5\rvert$	$\lvert -8\rvert$	$\lvert -11\rvert$	$\lvert -14\rvert$

Similarly, those entries in the fifth row that have a value less than h represent the blocks for which the grey area in the $[2b, 3b]$ block will be included. (The area in that block is precisely $R_1 = 1$.)

Our formula fails to hold when we have a point, τ_1, that is within h units of the boundary. In that case, we have to consider the fact that the block on the opposite side can wrap around to include the point. Consider the following examples.

Suppose $\tau_1 = 2$, $b = 3$ and $\alpha = 3/2$, so $h = 6$. We visualize this in Figure 4.4, which is analogous to Figure 4.3.

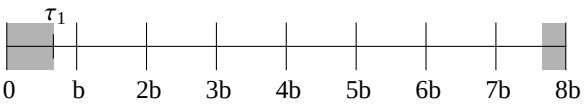

Figure 4.4: Visualizing how we include τ_1 near the edge (1st case)

Again, the grey area is spread across two blocks. But our criteria from above for determining probabilities fails. Consider Table 4.3.

Table 4.3: Probabilities that τ_1 is included in block j near the edge (1st case)

j	0	1	2	3	4	5	6	7
p_{1j}	3/12	3/12	2/12	0	0	0	1/12	3/12
jb	0	3	6	9	12	15	18	21
$\lvert\tau_1 - (j+1)b\rvert$	$\lvert -1\rvert$	$\lvert -4\rvert$	$\lvert -7\rvert$	$\lvert -10\rvert$	$\lvert -13\rvert$	$\lvert -16\rvert$	$\lvert -19\rvert$	$\lvert -22\rvert$
$\lvert\tau_1 - jb\rvert$	$\lvert 2\rvert$	$\lvert -1\rvert$	$\lvert -4\rvert$	$\lvert -7\rvert$	$\lvert -10\rvert$	$\lvert -13\rvert$	$\lvert -16\rvert$	$\lvert -19\rvert$

Notice that if $|\tau_1 - (j+1)b| < h$, we add $(b - R_1)/2h$ to our probability. At the same time, though, if $K - |\tau_1 - (j+1)b| < h$, then we also add $(b - R_1)/2h$ to our probability.

Similarly, if $|\tau_1 - jb| < h$, we add $R_1/2h$ to our probability. At the same time, though, if $K - |\tau_1 - jb| < h$, then we also add $R_1/2h$ to our probability.

Something similar happens when the point τ_1 is near the right-boundary of K. Suppose $\tau_1 = 2$, $b = 3$ and $\alpha = 3/2$, so $h = 6$. We visualize this in Figure 4.5.

Again, the grey area is spread across two blocks. But our criteria from above for determining probabilities fails. Consider Table 4.4.

Again, if $|\tau_1 - (j+1)b| < h$, we add $(b - R_1)/2h$ to our probability. At the same time,

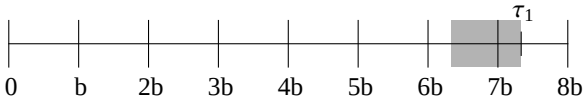

Figure 4.5: Visualizing how we include τ_1 near the edge (2nd case)

Table 4.4: Probabilities that τ_1 is included in block j near the edge (2nd case)

j	0	1	2	3	4	5	6	7
p_{1j}	3/12	1/12	0	0	0	2/12	3/12	3/12
jb	0	3	6	9	12	15	18	21
$\lvert\tau_1 - (j+1)b\rvert$	$\lvert 19\rvert$	$\lvert 16\rvert$	$\lvert 13\rvert$	$\lvert 10\rvert$	$\lvert 7\rvert$	$\lvert 4\rvert$	$\lvert 1\rvert$	$\lvert -2\rvert$
$\lvert\tau_1 - jb\rvert$	$\lvert 22\rvert$	$\lvert 19\rvert$	$\lvert 16\rvert$	$\lvert 13\rvert$	$\lvert 10\rvert$	$\lvert 7\rvert$	$\lvert 4\rvert$	$\lvert 1\rvert$

though, if $K - |\tau_1 - (j+1)b| < h$, then we also add $(b - R_1)/2h$ to our probability.

Similarly, if $|\tau_1 - jb| < h$, we add $R_1/2h$ to our probability. At the same time, though, if $K - |\tau_1 - jb| < h$, then we also add $R_1/2h$ to our probability.

As we shall see shortly, we will not need to worry about these last two cases, as they will constitute a minority of cases when considering all possible points.

Next, we need to determine $Var^*\left[\sqrt{\Lambda^*(K)}\tilde{X}^*\right] = Var^*\left[\frac{1}{\sqrt{\Lambda^*(K)}}\sum_{i=1}^{N(K)} W_i X(\tau_i)\right]$. To that end, we need to first compute $Cov^*(W_i, W_k)$.

Recall $W_i = \sum_{j=1}^{L} Y_{ij}$, where L is the total number of blocks and $Y_{ij} \sim$ Bernoulli(p_{ij}), where p_{ij} is the probability that block j will contain τ_i.

Thus, we have $Cov^*(W_i, W_k) = Cov^*\left(\sum_{j=1}^{L} Y_{ij}, \sum_{m=1}^{L} Y_{km}\right) = \sum_{j=1}^{L}\sum_{m=1}^{L} Cov^*(Y_{ij}, Y_{km})$.

So, we have reduced the problem above to that of computing $Cov^*(Y_{ij}, Y_{km})$. But since the covariance of two indicator random variables 1_A and 1_B is given by $P^*(A \cap B) - P^*(A)P^*(B)$, we are left computing the joint probabilities of Y_{ij} and Y_{km}.

Definition 4.4.3. We say that two points τ_i and τ_k are **sufficiently close** to each other provided that $|\tau_k - \tau_i| < b$. Otherwise, we say that τ_i and τ_k are **far apart**.

Definition 4.4.4. Given $[a_1, a_2]$, we say that a point τ_i is **removed from the boundary** by h units provided that $|\tau_i - a_1| > h$ and $|\tau_i - a_2| > h$.

Claim 4.4.5. *If τ_i and τ_k are far apart, then the probability that a particular block contains both τ_i and τ_k will be zero. On the otherhand, if τ_i and τ_k are sufficiently close, then the probability that a particular block contains both τ_i and τ_k will be positive.*

The reason is because if τ_i and τ_k are far apart, then both cannot appear in the same block. Moreover, if, for example, the probability for τ_i is non-zero, then necessarily the probability for τ_j will be zero.

We also know that the probability that block j contains τ_i and block m contains τ_k are independent, because of the independence of the resampling of the blocks.

Thus, $P^*(Y_{ij}Y_{km}) = P^*(Y_{ij})P^*(Y_{km})$ for all $j \neq m$, so we have that

$$Cov^*(Y_{ij}, Y_{km}) = P^*(Y_{ij}Y_{km}) - P^*(Y_{ij})P^*(Y_{km}) = 0.$$

Hence,

$$\begin{aligned} Cov^*(W_i, W_k) &= Cov^*\left(\sum_{j=1}^{L} Y_{ij}, \sum_{m=1}^{L} Y_{km}\right) \\ &= \sum_{j=1}^{L}\sum_{m=1}^{L} Cov^*(Y_{ij}, Y_{km}) \\ &= \sum_{j=1}^{L} Cov^*(Y_{ij}, Y_{kj}) \end{aligned}$$

And by Claim 4.4.5, we only need to consider the case where τ_i and τ_k are sufficiently close. That is, when τ_i and τ_k are within b units of each other.

Moreover, we only need to consider those pairs τ_i and τ_k that are removed from the boundary by h units. The reason can be seen Claim 4.4.6.

Claim 4.4.6. *Assuming A.4.1 and A.4.5, the number of τ_i that are within h units of the boundary is $O(h)$.*

Recall, if $N \sim$ Poission(λ), then on a window of size w, we would expect to see $w\lambda$ observations in that window. Under Assumption A.4.1, we have that the expected number of points within h units of the boundary would be between $2\lambda_{min}h$ and $2\lambda_{max}h$. Thus, the number of observed points within h units of the boundary is $O(h)$.

Since $K/h \to \infty$ (by Assumption A.4.5), the number of points that are within h units of the boundary is asymptotically negligible.

We can say a little more by showing that their contribution to the overall variance is negligible. To bound those covariances near the boundary, we make the following observations:

$P^*(Y_{ij}) = p_{ij} \leq \dfrac{b}{2h}$, $P^*(Y_{kj}) = p_{kj} \leq \dfrac{b}{2h}$ for each j, and

$$P^*(Y_{ij}Y_{kj}) \leq \frac{b - |\tau_i - \tau_k|}{2h} \leq \frac{b}{2h}.$$

Thus, we have that $Cov^*(Y_{ij}, Y_{kj}) \leq \dfrac{b}{2h} - \left(\dfrac{b}{2h}\right)^2 = O\left(\dfrac{b}{h}\right)$.

Hence, the contribution to $Var^*\left[\sqrt{\Lambda^*(K)}\tilde{X}^*\right]$ is $O\left(\dfrac{b}{K}\right)$, which tends to 0 by Assumption A.4.5 (since b is of smaller order than h).

Let us turn our attention to another example. This time, suppose that $\tau_1 = 8$ and $\tau_2 = 9.5$. (Again, suppose that $b = 3$ and $\alpha = 3/2$, so that $h = 6$.)

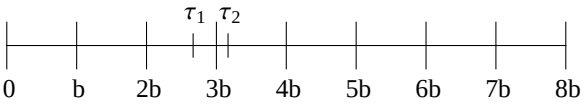

Figure 4.6: Placements of τ_1 and τ_2

Again, as was the case with Figure 4.3, we can shade in area b units to the left of τ_1 and τ_2 to visualize the probability that a particular block will include both τ_1 and τ_2.

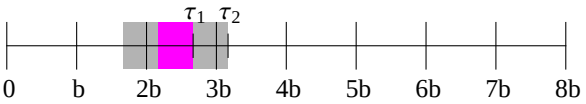

Figure 4.7: The probability of choosing τ_1 and τ_2

Notice that only if we choose a block which begins inside the purple area will we include both τ_1 and τ_2 in our resampled block. The width of that purple area is given by 0.5 (which, in general is given by $b - |\tau_2 - \tau_1|$.

Thus, we have the following table for the joint probabilities (Table 4.5).

Table 4.5: Joint probabilities for τ_1 and τ_2

j	0	1	2	3	4	5	6	7
$p_{1j} \cap p_{2j}$	0	1.5/12	1.5/12	1.5/12	1.5/12	0	0	0

In general, there will be a total of $2\left\lceil b^{\alpha-1}\right\rceil$ blocks that will contain the purple area, which has a probability of $\dfrac{b - |\tau_2 - \tau_1|}{2h}$.

By itself, though, this is not so useful. When we are computing the covariances, we need both the joint probabilities and the individual probabilities. To that end, we construct a

modified version of Table 4.5.

Table 4.6: All probabilities for τ_1 and τ_2 in different blocks

j	0	1	2	3	4	5	6	7
p_{1j}	1/12	3/12	3/12	3/12	2/12	0	0	0
p_{2j}	0	2.5/12	3/12	3/12	3/12	0.5/12	0	0
$p_{1j} \cap p_{2j}$	0	1.5/12	1.5/12	1.5/12	1.5/12	0	0	0

Notice that whenever the joint probability is equal to 0, either $p_{1j} = 0$ or $p_{2j} = 0$. This simplifies the calculation of the covariance quite a bit.

Letting $\tau_i = J_{ib} + R_i$, we have a clean expression for the covariance.

There are $2\left\lceil b^{\alpha-1}\right\rceil - 2$ blocks where $p_{1j} \cap p_{2j} = \dfrac{b - |\tau_2 - \tau_1|}{2h}$ and $p_{1j} = p_{2j} = \dfrac{b}{2h}$.

There is one block where $p_{1j} \cap p_{2j} = \dfrac{b - |\tau_2 - \tau_1|}{2h}$, but $p_{1j} = \dfrac{b}{2h}$ and $p_{2j} = \dfrac{R_2}{2h}$.

And there is one block where $p_{1j} \cap p_{2j} = \dfrac{b - |\tau_2 - \tau_1|}{2h}$, but $p_{1j} = \dfrac{R_1}{2h}$ and $p_{2j} = \dfrac{b - R_2}{2h}$.

For every other block, the covariance will be zero. When we compute $Cov^*(W_1, W_2)$, we sum up $Cov^*(Y_{1j}, Y_{2j})$ over all the j blocks. This leads to the following (general) formula:

Lemma 4.4.7. *For τ_i and τ_k sufficiently close, in different blocks (with $\tau_i < \tau_k$), τ_i and τ_k removed from the boundary by h units, we have that*

$$\begin{aligned}
Cov^*(W_i, W_k) &= \sum_{j=1}^{L} Cov^*(Y_{ij}, Y_{kj}) \\
&= \left(2\left\lceil b^{\alpha-1}\right\rceil - 2\right)\left(\frac{b - |\tau_i - \tau_k|}{2h} - \left(\frac{b}{2h}\right)^2\right) \\
&\quad + \left(\frac{b - |\tau_i - \tau_k|}{2h}\right) - \left(\frac{b}{2h}\right)\left(\frac{b - R_k}{2h}\right) \\
&\quad + \left(\frac{b - |\tau_i - \tau_k|}{2h}\right) - \left(\frac{R_i}{2h}\right)\left(\frac{b}{2h}\right) \\
&= \left(2\left\lceil b^{\alpha-1}\right\rceil\right)\left(\frac{b - |\tau_i - \tau_k|}{2h} - \left(\frac{b}{2h}\right)^2\right) + \frac{b(R_k + b - R_i)}{(2h)^2} \\
&= \frac{2h}{b}\left(\frac{b - |\tau_i - \tau_k|}{2h} - \left(\frac{b}{2h}\right)^2\right) + O\left(\frac{b^2}{h^2}\right) \\
&= 1 - \frac{|\tau_i - \tau_k|}{b} - \frac{b}{2h} + O\left(\frac{b^2}{h^2}\right)
\end{aligned}$$

where the $O(\cdot)$ term is uniform as it does not depend on i and k (since $R_k \approx b$).

Suppose that $\tau_1 = 7$ and $\tau_2 = 8.5$. (Again, suppose that $b = 3$ and $\alpha = 3/2$, so that $h = 6$.) We have the following analogous figures to Figure 4.6 and Figure 4.7.

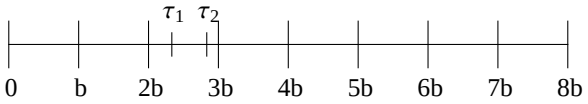

Figure 4.8: Different τ_1 and τ_2

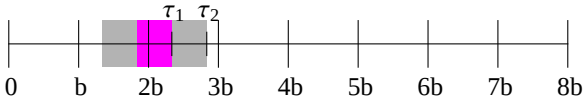

Figure 4.9: The probability of choosing τ_1 and τ_2 again

This time, notice that the purple area is split across two different blocks. Thus, there is a chance that a particular block will include both τ_1 and τ_2, but not with a probability that corresponds to the entire purple area.

Let us construct a table similar to that of Table 4.6.

j	0	1	2	3	4	5	6	7
p_{1j}	2/12	3/12	3/12	3/12	1/12	0	0	0
p_{2j}	0.5/12	3/12	3/12	3/12	2.5/12	0	0	0
$p_{1j} \cap p_{2j}$	0.5/12	1.5/12	1.5/12	1.5/12	1/12	0	0	0

Table 4.7: All probabilities for τ_1 and τ_2 in the same block

Notice that when τ_1 and τ_2 appear in the same block to begin with, their resampling probabilities are contained in the same blocks as well.

In fact, if we write $\tau_1 = J_{1b} + R_1$ and $\tau_2 = J_{2b} + R_2$, then we will have that $J_1 = J_2$. Moreover, we have that for $2\left\lceil b^{\alpha-1}\right\rceil - 1$ blocks,

$$P^*(Y_{1j}Y_{2j}) = \frac{b - |\tau_1 - \tau_2|}{2h} = \frac{b - |R_1 - R_2|}{2h}$$

and

$$P^*(Y_{1j}) = P^*(Y_{2j}) = \frac{b}{2h}.$$

Thus, $Cov^*(Y_{1j}, Y_{2j}) = \dfrac{b - |R_1 - R_2|}{2h} - \left(\dfrac{b}{2h}\right)^2$.

Notice that on the edges, though, the probabilities are specified differently. Without loss of generality, suppose that $\tau_1 < \tau_2$. Then on the left edge, we will have that

$$P^*(Y_{1j}Y_{2j}) = P^*(Y_{2j}) = \frac{b - R_2}{2h}$$

and

$$P^*(Y_{1j}) = \frac{b - R_1}{2h}.$$

So, $Cov^*(Y_{1j}, Y_{2j}) = \left(\frac{b - R_2}{2h}\right)\left(1 - \left(\frac{b - R_1}{2h}\right)\right)$.

On the right edge, we will have that

$$P^*(Y_{1j}Y_{2j}) = P^*(Y_{1j}) = \frac{R_1}{2h} \text{ and } P^*(Y_{2j}) = \frac{R_2}{2h}.$$

So, $Cov^*(Y_{1j}, Y_{2j}) = \left(\frac{R_1}{2h}\right)\left(1 - \frac{R_2}{2h}\right)$.

For every other block, the covariance will be zero. When we compute $Cov^*(W_1, W_2)$, we sum up $Cov^*(Y_{1j}, Y_{2j})$ over all the j blocks. This leads to the following (general) formula:

Lemma 4.4.8. *For τ_i and τ_k sufficiently close, in the same block (with $\tau_i < \tau_k$), τ_i and τ_k removed from the boundary by h units, we have that*

$$\begin{aligned}
Cov^*(W_i, W_k) &= \sum_{j=1}^{L} Cov^*(Y_{ij}, Y_{kj}) \\
&= \left(2\left\lceil b^{\alpha-1}\right\rceil - 1\right)\left(\frac{b - |R_i - R_k|}{2h} - \left(\frac{b}{2h}\right)^2\right) \\
&\quad + \left(\frac{b - R_k}{2h}\right)\left(1 - \left(\frac{b - R_i}{2h}\right)\right) \\
&\quad + \left(\frac{R_i}{2h}\right)\left(1 - \frac{R_k}{2h}\right) \\
&= \left(2\left\lceil b^{\alpha-1}\right\rceil\right)\left(\frac{b - |\tau_i - \tau_k|}{2h} - \left(\frac{b}{2h}\right)^2\right) + \frac{bR_i + bR_k - 2R_iR_k}{(2h)^2} \\
&= \frac{2h}{b}\left(\frac{b - |\tau_i - \tau_k|}{2h} - \left(\frac{b}{2h}\right)^2\right) + O\left(\frac{b^2}{h^2}\right) \\
&= 1 - \frac{|\tau_i - \tau_k|}{b} - \frac{b}{2h} + O\left(\frac{b^2}{h^2}\right)
\end{aligned}$$

where the $O(\cdot)$ term is uniform as it does not depend on i and k (since $R_k \approx b$).

Here, we used the fact that $\tau_i < \tau_k$ (so, $R_i < R_k$) to express $\dfrac{b - R_k + R_i}{2h}$ as $\dfrac{b - |R_i - R_k|}{2h}$ and $|R_i - R_k| = |\tau_i - \tau_k|$ (since $J_i = J_k$) to obtain the formula on the third line.

Let us take a moment to collect everything we have done up until this point.

Our goal is to compute $Var^*\left[\sqrt{\Lambda^*(K)}\tilde{X}^*\right] = Var^*\left[\dfrac{1}{\sqrt{\Lambda^*(K)}}\displaystyle\sum_{i=1}^{N(K)} W_i X(\tau_i)\right]$.

In order to compute this formula, we needed a formula for $Cov^*(W_i, W_k)$. We began by recognizing that each W_i can be thought of as a sum of Bernoulli random variables, where the sum runs over all of the blocks.

We found a formula for the probability p_{ij} that the point τ_i is resampled in block j, but discovered that our formula fails to hold when τ_i is within h units of the boundary.

That was not a problem, though; as we were able to show that the contribution of those points is asymptotically negligible.

From there, we developed a formula for the covariance that depended on the placement of the points τ_i and τ_k, but only differed in an error term of insignificant order.

And so, we have that

$$
\begin{aligned}
&Var^*[\sqrt{\Lambda^*(K)}\tilde{X}^*] \\
&= \frac{1}{\Lambda^*(K)} Var^*\left[\sum_{i=1}^{N(K)} W_i X(\tau_i)\right] \\
&= \frac{1}{\Lambda^*(K)} \sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)} Cov^*(W_i, W_k) X(\tau_i) X(\tau_k) \\
&= \frac{1}{\Lambda^*(K)} \sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)} \left(1 - \frac{|\tau_i - \tau_k|}{b} - \frac{b}{2h} + O\left(\frac{b^2}{h^2}\right)\right) 1_{\{|\tau_i - \tau_k| < b\}} X(\tau_i) X(\tau_k) \\
&= \frac{1}{\Lambda^*(K)} \sum_{i \neq k} \left(1 - \frac{|\tau_i - \tau_k|}{b}\right) 1_{\{|\tau_i - \tau_k| < b\}} X(\tau_i) X(\tau_k) \\
&\quad + \frac{1}{\Lambda^*(K)} \sum_{i=1}^{N(K)} X^2(\tau_i) \\
&\quad + \frac{1}{\Lambda^*(K)} \sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)} O\left(\frac{b}{h}\right) 1_{\{|\tau_i - \tau_k| < b\}} X(\tau_i) X(\tau_k)
\end{aligned}
$$

With all of the relevant theory established, we are ready to state our results. But first, we make an important observation.

In the algorithm described in Section 4.2, references were made to four estimators. The

first two contain random quantities, $\Lambda^*(K)$, in the denominator (in the bootstrap world) while the last two do not (they use $\Lambda(K)$ instead). Since the division by a random quantity complicates things, we shall use the last two estimators to prove our results. The following lemma, Lemma 4.4.9, when coupled with Slutsky's Theorem (see [74]), allows us to do this.

Lemma 4.4.9. *Assuming A.4.1, as* $K \to \infty$, $\dfrac{\Lambda^*(K)}{\Lambda(K)} \xrightarrow{p} 1$.

And so, we shall state results pertaining to $\hat{\hat{X}}^*$ instead. (That is, we will use $\Lambda(K)$ in the denominator instead of the quantity $\Lambda^*(K)$.)

Lemma 4.4.10. *Assuming A.4.1, A.4.5, and A.4.6, we have that*

$$\frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} O\left(\frac{b}{h}\right) 1_{\{|\tau_i - \tau_k| < b\}} X(\tau_i) X(\tau_k) \xrightarrow{p} 0$$

Thus, we see that the third line of the variance formula above does not contribute anything asymptotically.

Lemma 4.4.11. *Let* $\hat{R}(0) = \dfrac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} X^2(\tau_i)$. *Then, under Assumptions A.4.1-A.4.3,*

i. $E[\hat{R}(0)] = R(0) + \mu^2$

ii. $Var[\hat{R}(0)] \to 0$ *as* $K \to \infty$

To establish Lemma 4.4.11(ii), the following lemma from [53] will prove useful:

Lemma 4.4.12. *Let* N_g *be a general Poisson process (not necessarily homogeneous) on* $\mathbb{R}^d$, *possessing mean measure* Λ_g, *and assumed to be independent of the random field* $\{X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^{\mathbf{d}}\}$. *Let* E_1, E_2 *be two subsets of* $\mathbb{R}^d$ *such that* $\rho(E_1, E_2) = k > 0$, *and define*

$$\tilde{Y}_i = \frac{1}{N_g(E_i)} \int_{E_i} X(\mathbf{t}) N(d\mathbf{t})$$

and

$$\bar{Y}_i = \frac{1}{\Lambda_g(E_i)} \int_{E_i} X(\mathbf{t}) N(d\mathbf{t})$$

for $i = 1, 2$; *also assume that* $E[|X(t)|^p] = C_p < \infty$ *for some* $p > 2$. *Then*

$$\left| Cov(\tilde{Y}_1, \tilde{Y}_2) \right| \leq 10 C_p^{2/p} (\alpha_X(k; |E_1|, |E_2|))^{1-2/p}$$

and

$$\left|Cov(\bar{Y}_1, \bar{Y}_2)\right| \leq 10C_p^{2/p}(\alpha_X(k; |E_1|, |E_2|))^{1-2/p}$$

If E_1, E_2 are compact, convex, and are translates of one another, that is, if $E_1 = E_2 + \mathbf{t}$, then we also have

$$\left|Cov(\tilde{Y}_1, \tilde{Y}_2)\right| \leq 10C_p^{2/p}(\bar{\alpha}_X(k; |E_1|))^{1-2/p}$$

and

$$\left|Cov(\bar{Y}_1, \bar{Y}_2)\right| \leq 10C_p^{2/p}(\bar{\alpha}_X(k; |E_1|))^{1-2/p}$$

where $|\cdot|$ denotes the Lebesgue measure (volume) of the set.

In anticipation of the component of the variance with unequal indices, the following seven lemmas will prove useful. The motivation for this approach is based upon a similar decomposition carried out in [50].

Lemma 4.4.13a. Under Assumption A.4.4, we have that

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\int_0^K \left[\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|v-u|}{b}\right)^+ \\ \times |Q(u-v, t-v, s-v)|\end{array}\right] dsdtdudv = O\left(\frac{b^2}{K}\right)$$

Lemma 4.4.13b. Under Assumption A.4.2 and Assumption A.4.3, we have that

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\int_0^K \left[\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|v-u|}{b}\right)^+ \\ \times |R(t-v)R(s-u)|\end{array}\right] dsdtdudv = O\left(\frac{b^2}{K}\right)$$

Lemma 4.4.13c. Under Assumption A.4.2 and Assumption A.4.3, we have that

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\int_0^K \left[\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|v-u|}{b}\right)^+ \\ \times |R(s-v)R(u-t)|\end{array}\right] dsdtdudv = O\left(\frac{b^2}{K}\right)$$

Lemma 4.4.13d. Under Assumption A.4.4, we have that

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K \left[\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|t-u|}{b}\right)^+ \\ \times |Q(t-t, u-t, s-t)|\end{array}\right] dsdtdu = O\left(\frac{b}{K}\right)$$

Lemma 4.4.13e. We have that

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K \left[\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|t-u|}{b}\right)^+ \\ \times |R(0)R(s-u)|\end{array}\right] dsdtdu = O\left(\frac{b^2}{K}\right)$$

Lemma 4.4.13f. We have that

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K \left[\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|t-u|}{b}\right)^+ \\ \times |R(u-t)R(s-t)|\end{array}\right] dsdtdu = O\left(\frac{b^2}{K}\right)$$

Lemma 4.4.13g. We have that

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K \left[\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|t-u|}{b}\right)^+ \\ \times |R(s-t)R(t-u)|\end{array}\right] dsdtdu = O\left(\frac{b^2}{K}\right)$$

We have the following result pertaining to the unequal terms in the variance formula.

Lemma 4.4.14. *Suppose Assumptions A.4.1-A.4.4, and A.4.6 hold.*

Let $\hat{\theta}^2 = \frac{1}{\Lambda(K)}\sum_{i\neq k}\left(1-\frac{|\tau_i-\tau_k|}{b}\right)1_{\{|\tau_i-\tau_k|<b\}}X(\tau_i)X(\tau_k)$. *Then*

i. $E[\hat{\theta}^2] \xrightarrow{p} \theta^2$

ii. $Var[\hat{\theta}^2] \to 0$ *as* $K \to \infty$

where $\theta^2 = \lim_{K\to\infty}\frac{1}{\Lambda(K)}\int_K\int_K R(s-t)\lambda(s)\lambda(t)dsdt$

The preceding lemmas give us the following result.

Lemma 4.4.15. *If Assumptions A.4.1-A.4.6 hold, then* $Var^*[\sqrt{\Lambda(K)}\tilde{\tilde{X}}^*] \xrightarrow{p} \theta^2 + R(0) + \mu^2$ *as* $K \to \infty$.

Now that we have established that the bootstrap variance tends (asymptotically) to the true variance, we are ready to state our main theorem.

Theorem 4.4.16. *Suppose Assumptions A.4.1-A.4.6 hold. Then we have the following:*

i. $E^*\left[\tilde{\tilde{X}}^*\right] = \tilde{X}_K$

ii. $\frac{Var^*[\tilde{\tilde{X}}^*]}{Var[\tilde{X}_K]} \xrightarrow{p} 1$

iii. $\sup_x \left| P^* \left(\sqrt{\Lambda(K)}(\tilde{\tilde{X}}^* - \tilde{X}_K) \leq x \right) - P\left(\sqrt{\Lambda(K)}(\tilde{X}_K - \mu) \leq x \right) \right| \xrightarrow{p} 0$

Instead of using torodial wrapping, we could take a different approach for edge effects. That is, we only take a one-sided average. This approach may be preferrable in cases where the intensity at the edges differs significantly.

The drawback for this approach, though, is that not all points will be resampled once (on average). Indeed, those points within h units of the boundary will appear (on average) less than once. The solution for this is to modify our theorems not to subtract $\tilde{X}_K$, but rather subtract $E^*[\tilde{X}^*]$. This recenters the data around the bootstrap mean. (Note: In the case of torodial wrapping, $E^*[\tilde{X}^*] = \tilde{X}_K$.)

As a simple example, suppose we have $b = 3$, $h = 5$ and five blocks (so $K = 15$). Suppose we have data $\tau_1 = 1$, $\tau_2 = 4.5$, and $\tau_3 = 10$. This is illustrated in Figure 4.10.

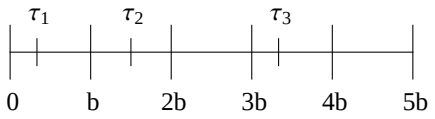

Figure 4.10: Three points to be resampled without torodial wrapping

Let us examine τ_1. As before, we shade a region b units to the left of the point, stopping when we hit the left boundary. In Figure 4.11, τ_1 will only be included in a resample block if the shifted block starts in the shaded region.

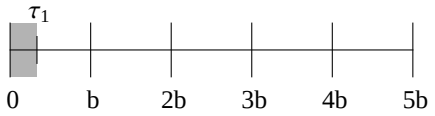

Figure 4.11: Probability of including τ_1 without wrapping

Notice that unlike before, since τ_1 is near the left boundary, the probability of its inclusion is reduced from 3 to 1. We repeat the same for τ_2 and τ_3. These appear in Figure 4.12 and Figure 4.13, respectively.

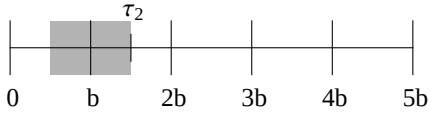

Figure 4.12: Probability of including τ_2 without wrapping

In Table 4.8, we compute a table of probabilities for the inclusion of the points in a block.

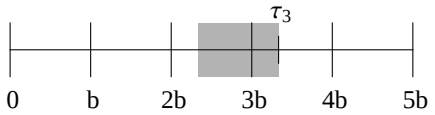

Figure 4.13: Probability of including τ_3 without wrapping

Table 4.8: Probabilities when not using torodial wrapping

j	0	1	2	3	4
p_{1j}	1/5	1/8	0/10	0/8	0/5
p_{2j}	3/5	3/8	3/10	0.5/8	0/5
p_{3j}	0/5	1/8	3/10	3/8	3/5

Notice that $\sum_{j=0}^{4} p_{1j} = \frac{26}{80}$, $\sum_{j=0}^{4} p_{2j} = \frac{107}{80}$, and $\sum_{j=0}^{4} p_{3j} = \frac{102}{80}$.

In Section 4.5, we resample our data using torodial wrapping and no wrapping.

As was discussed in Chapter 3, the results for $\tilde{X}^*$ do not extend immediately to $\bar{X}^*$. However, as we saw, the results were nearly identical though the asymptotic variance was different. We conjecture that analogous results hold for our bootstrap estimators and this will be investigated further.

4.5 Simulations

The first step is to generate data from a one-dimensional marked point process. Since we have assumed independence between the point process and the associated marks, we begin by generating an inhomogeneous Poisson process.

We do this using the Accept-Reject Method as specified by Lewis and Shedler ([43]). The idea is as follows:

- Choose λ such that $\lambda(t) \leq \lambda$ for all $t \leq K$.
- Generate events according to a Poisson process with rate λ
- Accept the event, say at time t, with probability $\lambda(t)/\lambda$, independently of what has happened before.
- The process of the counted events then forms an inhomogeneous Poisson process with intensity function $\lambda(t)$, $t \leq K$

Formally, we have the following algorithm:

1. Set $i = 0$ and $t = 0$

2. Generate a random number $U_1 \sim U(0,1)$
3. Set $t = t - \ln(U_1)/\lambda$. If $t > K$, stop.
4. Generate another random number $U_2 \sim U(0,1)$
5. If $U_2 \leq \lambda(t)/\lambda$, set $i = i + 1$ and $S(i) = t$
6. Go to step 2.

The output will consist of the counter i, which is the number of events that occur up to time K and $S(1), \ldots, S(i)$ will constitute the event times.

Next, we need to generate the corresponding marks. We assume that $X(\cdot)$ is a stationary, Gaussian process with mean 0 and covariance function $R(t)$. Our algorithm for generating the marks is as follows:

1. Initialize a matrix of covariances, Σ with dimensions given by $N(K)$, the number of points found in the algorithm above
2. Populate Σ with values $R(S[i] - S[j])$ for all pairs i, j
3. Consider the eigenvalue decomposition of $\Sigma = VDV^T$ and express $\Sigma^{1/2} = VD^{1/2}V^T$
4. Generate $N(K)$ i.i.d. $N(0,1)$ random variables and place them into the vector Z
5. Set the marks equal to $X = \Sigma^{1/2}Z$

Finally, we record the pair $\{S(j), X(S(j))\}$. This is our randomly generated marked point process.

We run simulations using two resampling methods (torodial wrapping and no wrapping), two different covariance functions, five different intensity functions, three choices of b, and two corresponding choices of h for a total of 120 models. The covariance functions, intensity functions, and block sizes are the same as those considered in Chapter 2. Each model was simulated 1,000 times and each time, a 95% confidence interval was constructed.

As was mentioned in Chapter 2, in practice, one needs to determine optimal choices for b and h. Since there is a strong relationship in purpose of the local bandwidth parameter h and the window parameter w from Chapter 2, we suggest using the same value of h as would be determined for w. Again, methods from [28] can be adapted to choose an optimal b and will be the focus of future work.

In the tables that follow, we present the mean and standard deviation of the constructed confidence intervals as well as the percentage of confidence intervals containing the true mean, 0 (the coverage probability).

When considering all of the tables, we see that the mean confidence interval length is around 0.1 while the standard deviation is around 0.0025. This is ten times smaller than the standard deviations observed in Chapter 2, suggesting that the local block bootstrap procedure is more stable than the transformation methods. Also, the average number of resampled observations varies, but stays within 5% of the true counts.

Again, looking at all of the tables, we see that the choice of using torodial wrapping or not using wrapping leads to differences around 1-2%. This is likely due to edge effects. And in all of the tables, it would appear that the best choice for b and h was given by $b = 2$ and $h = 8$. In general, as b increases, the coverage decreases.

In Tables 4.9 - 4.13 increasing from $b = 2$ to $b = 5$ led to a 5% reduction in coverage. A similar reduction was seen when increasing from $b = 5$ to $b = 10$.

In Tables 4.14 - 4.18, the decrease in coverage is more pronounced. Increasing from $b = 2$ to $b = 5$ led to approximately a 15% reduction and increasing from $b = 5$ to $b = 10$ led to an additional 10% reduction.

In all of the tables, we see that increasing the size of h led to a reduction in coverage. This makes sense, as we are venturing further and further away from the original source block if h is large. Notice that the coverage is poorest for $b = 10$ and $h = 200$. With such a large h, it is almost as though we are performing a block bootstrap on the data.

Investigating further, with $b = 2$ and $h = 64$, coverage probabilities dropped from around 90% (with $b = 2$ and $h = 8$) to under 60%. Again, that is to be expected. With such a small block size and a large area to choose a block from, we are essentially performing an i.i.d. bootstrap which ignores the dependencies between the marks.

Overall, we see that the method performs as well as the transformation methods discussed in Chapter 2, if not better in Tables 4.9 - 4.12. In Table 4.13, there is almost a 20% improvement. It is not clear why the difference is so profound. This may have to do with the small b and h better capturing the changes in intensity compared to the averaging techniques.

4.6 Proofs

Proof of Lemma 4.4.9: Recall, we defined $\Lambda(K) = E[N(K)]$. The bootstrap analog is given by $\Lambda^*(K) = E^*[N^*(K)]$. That is, $\Lambda^*(K)$ represents the expected number of points in our resampled data.

Suppose we are given data $X(\tau_1), \ldots, X(\tau_{N(K)})$ from an inhomogeneous marked point processes. We can express the number of points in the resampled data as $N^*(K) = \sum_{i=1}^{N(K)} W_i$, where $W_i = \sum_{j=1}^{L} Y_{ij}$, with L being the total number of blocks and each $Y_{ij} \sim$ Bernoulli(p_{ij}), where p_{ij} is the probability that block j will contain τ_i.

Table 4.9: LBB with Intensity function $\lambda_1(t)$ and $R(t) = \exp(-|t|)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	1954.318	0.08609	0.00206	93.7%
		16	1952.737	0.08621	0.00220	89.6%
	5	32	1950.369	0.08616	0.00254	89.2%
		64	1951.194	0.08589	0.00262	86.3%
	10	125	1953.077	0.08586	0.00298	82.8%
		200	1956.268	0.08578	0.00314	80.3%
Yes	2	8	1950.289	0.08616	0.00207	92.0%
		16	1950.606	0.08621	0.00219	92.0%
	5	32	1951.117	0.08619	0.00245	88.1%
		64	1950.563	0.08607	0.00258	84.7%
	10	125	1950.539	0.08624	0.00298	80.8%
		200	1953.133	0.08605	0.00302	81.8%

Table 4.10: LBB with Intensity function $\lambda_2(t)$ and $R(t) = \exp(-|t|)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	2043.184	0.08630	0.00213	93.8%
		16	2040.180	0.08658	0.00227	90.4%
	5	32	2043.399	0.08638	0.00225	87.4%
		64	2042.942	0.08649	0.00238	83.7%
	10	125	2041.969	0.08619	0.00245	83.4%
		200	2049.441	0.08584	0.00248	80.2%
Yes	2	8	2043.876	0.08623	0.00219	95.2%
		16	2042.437	0.08613	0.00230	90.9%
	5	32	2044.057	0.08629	0.00239	86.8%
		64	2042.413	0.08635	0.00249	85.2%
	10	125	2039.997	0.08640	0.00275	82.8%
		200	2043.570	0.08630	0.00285	80.6%

Table 4.11: LBB with Intensity function $\lambda_3(t)$ and $R(t) = \exp(-|t|)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	2061.352	0.08496	0.00220	93.7%
		16	2060.136	0.08509	0.00226	93.1%
	5	32	2063.514	0.08480	0.00218	88.7%
		64	2065.587	0.08467	0.00218	84.1%
	10	125	2078.238	0.08438	0.00211	85.3%
		200	2100.423	0.08374	0.00206	83.3%
Yes	2	8	2058.005	0.08493	0.00220	92.8%
		16	2059.290	0.08494	0.00226	91.7%
	5	32	2061.326	0.08484	0.00216	87.5%
		64	2059.865	0.08491	0.00216	85.2%
	10	125	2059.506	0.08484	0.00209	82.7%
		200	2058.204	0.08492	0.00214	84.7%

Table 4.12: LBB with Intensity function $\lambda_4(t)$ and $R(t) = \exp(-|t|)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	2029.325	0.08421	0.00241	91.2%
		16	2028.551	0.08410	0.00253	86.2%
	5	32	2030.243	0.08441	0.00258	82.3%
		64	2028.417	0.08433	0.00280	80.9%
	10	125	2020.363	0.08420	0.00258	78.9%
		200	2011.100	0.08417	0.00258	79.4%
Yes	2	8	2026.474	0.08419	0.00240	93.8%
		16	2025.551	0.08434	0.00252	88.4%
	5	32	2025.267	0.08411	0.00260	81.7%
		64	2023.321	0.08434	0.00254	80.2%
	10	125	2025.233	0.08415	0.00274	77.0%
		200	2026.084	0.08419	0.00287	76.0%

Table 4.13: LBB with Intensity function $\lambda_5(t)$ and $R(t) = \exp(-|t|)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	3536.550	0.06726	0.00155	83.8%
		16	3541.459	0.06721	0.00161	81.6%
	5	32	3545.025	0.06717	0.00156	78.5%
		64	3563.186	0.06694	0.001652	74.4%
	10	125	3610.997	0.06653	0.00175	71.9%
		200	3656.953	0.06594	0.00184	75.8%
Yes	2	8	3540.276	0.06730	0.00159	85.5%
		16	3539.508	0.06723	0.00163	83.7%
	5	32	3540.601	0.06726	0.00157	76.4%
		64	3537.530	0.06729	0.00167	76.6%
	10	125	3540.439	0.06712	0.00183	73.5%
		200	3541.537	0.06708	0.00193	73.2%

Table 4.14: LBB with Intensity function $\lambda_1(t)$ and $R(t) = \exp(-|t|/3)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	2001.073	0.09348	0.00267	90.8%
		16	2002.933	0.09348	0.00283	86.6%
	5	32	2001.081	0.09353	0.00333	74.4%
		64	2003.042	0.09357	0.00342	75.1%
	10	125	2003.474	0.09388	0.00412	67.3%
		200	1999.383	0.09407	0.00439	66.6%
Yes	2	8	2003.642	0.09329	0.00263	90.3%
		16	2000.942	0.09355	0.00273	85.5%
	5	32	2001.530	0.09324	0.00328	75.0%
		64	2002.528	0.09329	0.00348	72.9%
	10	125	2003.680	0.09314	0.00428	70.7%
		200	1999.656	0.09332	0.00431	69.7%

Table 4.15: LBB with Intensity function $\lambda_2(t)$ and $R(t) = \exp(-|t|/3)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	1939.197	0.09009	0.00247	94.2%
		16	1939.619	0.09019	0.00270	90.2%
	5	32	1944.503	0.09009	0.00313	73.0%
		64	1945.081	0.08994	0.00322	71.7%
	10	125	1945.487	0.08981	0.00312	63.9%
		200	1942.283	0.08939	0.00333	66.4%
Yes	2	8	1937.296	0.09010	0.00268	93.9%
		16	1936.591	0.09006	0.00265	85.7%
	5	32	1936.508	0.08993	0.00302	75.5%
		64	1938.310	0.08994	0.00323	73.3%
	10	125	1937.765	0.09002	0.00330	64.0%
		200	1937.570	0.08980	0.00328	67.4%

Table 4.16: LBB with Intensity function $\lambda_3(t)$ and $R(t) = \exp(-|t|/3)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	1990.864	0.08456	0.00234	92.9%
		16	1993.150	0.08449	0.00243	88.9%
	5	32	1992.792	0.08449	0.00288	75.4%
		64	1998.173	0.08455	0.00304	75.6%
	10	125	2002.634	0.08417	0.00324	69.1%
		200	2023.374	0.08304	0.00337	68.8%
Yes	2	8	1994.675	0.08443	0.00242	94.8%
		16	1993.568	0.08462	0.00238	85.2%
	5	32	1996.786	0.08426	0.00298	79.7%
		64	1993.825	0.08431	0.00303	76.3%
	10	125	1995.872	0.08410	0.00339	69.5%
		200	1995.825	0.08412	0.00336	68.0%

Table 4.17: LBB with Intensity function $\lambda_4(t)$ and $R(t) = \exp(-|t|/3)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	2003.180	0.08907	0.00279	92.4%
		16	2005.325	0.08894	0.00273	89.4%
	5	32	2003.860	0.08875	0.00335	73.3%
		64	1999.573	0.08881	0.00352	73.3%
	10	125	1996.452	0.08903	0.00375	66.9%
		200	1987.529	0.08928	0.00402	66.8%
Yes	2	8	2005.069	0.08900	0.00270	92.4%
		16	2001.746	0.08870	0.00286	86.1%
	5	32	2003.815	0.08862	0.00337	74.0%
		64	2006.300	0.08859	0.00354	70.8%
	10	125	2001.181	0.08855	0.00409	66.4%
		200	2000.023	0.08857	0.00420	64.9%

Table 4.18: LBB with Intensity function $\lambda_5(t)$ and $R(t) = \exp(-|t|/3)$

Wrap?	b	h	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	2	8	3419.962	0.06293	0.00175	81.7%
		16	3419.356	0.06272	0.00187	66.2%
	5	32	3426.088	0.06257	0.00218	68.8%
		64	3442.950	0.06253	0.00219	64.2%
	10	125	3484.721	0.06207	0.00221	60.3%
		200	3512.568	0.06230	0.00215	60.8%
Yes	2	8	3416.265	0.06287	0.00171	79.2%
		16	3416.322	0.06288	0.00187	77.6%
	5	32	3410.394	0.06295	0.00223	67.5%
		64	3412.456	0.06270	0.00221	65.4%
	10	125	3412.946	0.06272	0.00224	55.8%
		200	3416.548	0.06293	0.00238	60.4%

In Lemma 4.4.1, we saw that $E^*[W_i] = 1$, so we have that $\Lambda^*(K) = E^*[N^*(K)] = N(K)$. Thus, we need only show that $\dfrac{N(K)}{E[N(K)]} = \dfrac{N(K)}{\Lambda(K)} \xrightarrow{p} 1$.

However, we have already established this result in Lemma 3.4.1.

□

Proof of Lemma 4.4.10: To get a handle on this expression, we shall first compute its expected value.

Recall, a sequence of random variables $\{X_n\}$ is said to be bounded in probability if, given $\varepsilon > 0$, one can find a constant k such that $P(|X_n| > k) \leq \varepsilon$ for all $n \geq n_0 = n_0(\varepsilon)$. If X_n is bounded in probability, then we write $X_n = O_p(1)$.

Thus, we have that $X(\tau_i) = O_p(1)$. Moreover, by Assumption A.4.1, the number of observed points grows linearly with K. That is, $\Lambda(K) = O(K)$.

The only piece left is to determine an expression for $\displaystyle\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} 1_{\{|\tau_i - \tau_k| < b\}}$.

Consider its expected value. We shall condition on the value of $N(K)$, so that the summations are no longer random. Thus,

$$E\left[\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} 1_{\{|\tau_i - \tau_k| < b\}}\right] = E\left[E\left[\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} 1_{\{|\tau_i - \tau_k| < b\}} \middle| N(K) = n\right]\right]$$

We shall also make use of the following fact:

Since the marks are independent of the point process, we have that

$\{(\tau_1, X(\tau_1)), (\tau_2, X(\tau_2)), \ldots, (\tau_{N(K)}, X(\tau_{N(K)}))|N(t) = n\}$

has the same distribution as

$\{U_1, X(\tau_1)), (U_2, X(\tau_2)), \ldots, (U_n, X(\tau_n))\}$,

where $\{U_i\}$ are i.i.d. with $P\left(U_i \in (t, t+dt]\right) = \dfrac{\lambda(t)}{\Lambda(K)} 1_{[0,K]}(t)dt$.

Thus, $f_{U_i}(t) = \dfrac{\lambda(t)}{\Lambda(K)} 1_{[0,K]}(t)$.

And so, we have that

$$\begin{aligned}
E\left[\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}\right] &= E\left[E\left[\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}\middle| N(K)=n\right]\right] \\
&= E\left[E\left[\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)} 1_{\{|U_i-U_k|<b\}}\middle| N(K)=n\right]\right] \\
&= E\left[E\left[N(K) + \sum_{i\neq k} 1_{\{|U_i-U_k|<b\}}\middle| N(K)=n\right]\right] \\
&= E[N(K)] + \Big(E[N(K)^2] - E[N(K)]\Big)\rho_2^b(K) \\
&= \Lambda(K) + [\Lambda(K)]^2\rho_2^b(K)
\end{aligned}$$

where

$$\begin{aligned}
\rho_2^b(K) &= E\left[1_{\{|U_i-U_k|<b\}}\right] \\
&= \int_0^K\int_0^K 1_{\{|t-s|<b\}} f_{U_i}(s) f_{U_k}(t) dsdt \\
&= \frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K 1_{\{|t-s|<b\}}\lambda(t)\lambda(s)dsdt
\end{aligned}$$

Note: The first term above, $E[N(K)]$ comes from the case where $i = k$. (Necessarily, the difference will be less than b, and there are n such pairs.) Since there are n^2 possible pairings, the other term, $E[N(K)^2] - E[N(K)]$, takes those into consideration.

Let us take a closer look at $\rho_2^b(K)$. By Assumption A.4.1, we have

$$\left|\rho_2^b(K)\right| \leq \frac{(\lambda_{max})^2}{[\Lambda(K)]^2}\int_0^K\int_0^K 1_{\{|t-s|<b\}} dsdt$$

But notice that $\int_0^K \int_0^K 1_{\{|t-s|<b\}} ds dt = K^2 - (K-b)^2 = 2Kb - b^2$.

And so, it follows that $\rho_2^b(K) = O\left(\frac{b}{K}\right)$. Thus, we have that

$$\begin{aligned} E\left[\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} 1_{\{|\tau_i - \tau_k|<b\}}\right] &= \Lambda(K) + [\Lambda(K)]^2 \rho_2^b(K) \\ &= O(K) + O(bK) \\ &= O(bK) \end{aligned}$$

Finally, we see that

$$E\left[\frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} O\left(\frac{b}{h}\right) 1_{\{|\tau_i - \tau_k|<b\}} X(\tau_i) X(\tau_k)\right] = O\left(\frac{b^2}{h}\right)$$

But by Assumptions A.4.5 and A.4.6, we have that this tends to 0.

Next, though, we need to verify that the variance tends to 0 as well. To that end, it suffices to show that the second moment tends to 0.

We claim that $E\left[\left(\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} 1_{\{|\tau_i - \tau_k|<b\}}\right)^2\right] = O\left(b^2 K^2\right)$.

We have that

$$\begin{aligned} &E\left[\left(\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} 1_{\{|\tau_i - \tau_k|<b\}}\right)^2\right] \\ &= E\left[\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} \sum_{j=1}^{N(K)} \sum_{l=1}^{N(K)} 1_{\{|\tau_i - \tau_k|<b\}} 1_{\{|\tau_j - \tau_l|<b\}}\right] \\ &= E\left[E\left[\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} \sum_{j=1}^{N(K)} \sum_{l=1}^{N(K)} 1_{\{|\tau_i - \tau_k|<b\}} 1_{\{|\tau_j - \tau_l|<b\}} \,\middle|\, N(K) = n\right]\right] \end{aligned}$$

Before we continue, we need to consider all four-tuples. There are fifteen such arrangements.

Suppose $i = j = k = l$. Then $1_{\{|\tau_i - \tau_k|<b\}} 1_{\{|\tau_j - \tau_l|<b\}} = 1$, and the sum reduces to

$$\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\sum_{j=1}^{N(K)}\sum_{l=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}1_{\{|\tau_j-\tau_l|<b\}} = \sum_{i=1}^{N(K)} 1 = N(K).$$

If three of them are the same (there are four ways this happens), then one of indicators will always be true, and so the sum reduces to

$$\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\sum_{j=1}^{N(K)}\sum_{l=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}1_{\{|\tau_j-\tau_l|<b\}} = \sum_{i\neq k} 1_{\{|\tau_i-\tau_k|<b\}}$$

If $i = k$ and $j = l$, then again the indicator is 1 and the sum reduces to

$$\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\sum_{j=1}^{N(K)}\sum_{l=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}1_{\{|\tau_j-\tau_l|<b\}} = \sum_{i\neq k} 1 = N(K)(N(K)-1)$$

If $i = j$ and $k = l$ or $i = l$ and $j = k$, then the indicators are the same and so the sum reduces to

$$\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\sum_{j=1}^{N(K)}\sum_{l=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}1_{\{|\tau_j-\tau_l|<b\}} = \sum_{i\neq k} 1_{\{|\tau_i-\tau_k|<b\}}$$

If $i = k$ (with j and l different) or $j = l$ (with i and k different), then one of the indicators is 1 and the sum reduces to

$$\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\sum_{j=1}^{N(K)}\sum_{l=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}1_{\{|\tau_j-\tau_l|<b\}} = \sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}$$

If $i = j$ (with k and l different), or $i = l$ (with j and k different), or $j = k$ (with i and l different) or $k = l$ (with i and j different), then the sum reduces to

$$\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\sum_{j=1}^{N(K)}\sum_{l=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}1_{\{|\tau_j-\tau_l|<b\}} = \sum_{i\neq j\neq k} 1_{\{|\tau_i-\tau_k|<b\}}1_{\{|\tau_i-\tau_j|<b\}}$$

And finally, if all four are different, there is no simplification in the summation. Thus, we have that

$$
\begin{aligned}
&E\left[\left(\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}\right)^2\right] \\
&= E\left[\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\sum_{j=1}^{N(K)}\sum_{l=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}1_{\{|\tau_j-\tau_l|<b\}}\right] \\
&= E\left[E\left[\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\sum_{j=1}^{N(K)}\sum_{l=1}^{N(K)} 1_{\{|\tau_i-\tau_k|<b\}}1_{\{|\tau_j-\tau_l|<b\}}\,\middle|\, N(K)=n\right]\right] \\
&= E\left[\begin{array}{l} N(K)+8\left(N(K)(N(K)-1)\right)\rho_2^b(K) \\ +N(K)(N(K)-1)+4\left(N(K)(N(K)-1)(N(K)-2)\right)\rho_3^b(K) \\ +\left(N(K)(N(K)-1)(N(K)-2)(N(K)-3)\right)\rho_4^b(K)\end{array}\right] \\
&= \Lambda(K)+8[\Lambda(K)]^2\rho_2^b(K)+[\Lambda(K)]^2+4[\Lambda(K)]^3\rho_3^b(K)+[\Lambda(K)]^4\rho_4^b(K)
\end{aligned}
$$

where

$$\rho_2^b(K)=\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K 1_{\{|t-s|<b\}}\lambda(t)\lambda(s)dsdt$$

$$\rho_3^b(K)=\frac{1}{[\Lambda(K)]^3}\int_0^K\int_0^K\int_0^K 1_{\{|t-s|<b\}}1_{\{|t-u|<b\}}\lambda(u)\lambda(t)\lambda(s)dsdtdu$$

$$\rho_4^b(K)=\frac{1}{[\Lambda(K)]^4}\int_0^K\int_0^K\int_0^K\int_0^K 1_{\{|t-s|<b\}}1_{\{|u-v|<b\}}\lambda(v)\lambda(u)\lambda(t)\lambda(s)dsdtdudv$$

Above, we saw that $\rho_2^b(K)=O\left(\frac{b}{K}\right)$.

A similar calculation as above yields

$$\int_0^K\int_0^K\int_0^K 1_{\{|t-s|<b\}}1_{\{|t-u|<b\}}dsdtdu=K^3-(K-b)^2K=2K^2b-b^2K$$

So, we see that $\rho_3^b(K)=O\left(\frac{b}{K}\right)+O\left(\frac{b^2}{K^2}\right)=O\left(\frac{b^2}{K^2}\right)$. Notice that

$$
\begin{aligned}
&\rho_4^b(K) \\
&= \frac{1}{[\Lambda(K)]^4} \int_0^K \int_0^K \int_0^K \int_0^K 1_{\{|t-s|<b\}} 1_{\{|u-v|<b\}} \lambda(v)\lambda(u)\lambda(t)\lambda(s) ds dt du dv \\
&= \frac{1}{[\Lambda(K)]^4} \int_0^K \int_0^K 1_{\{|u-v|<b\}} \lambda(v)\lambda(u) \left(\int_0^K \int_0^K 1_{\{|t-s|<b\}} \lambda(t)\lambda(s) ds dt \right) du dv \\
&= \frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K 1_{\{|t-s|<b\}} \lambda(t)\lambda(s) ds dt \frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K 1_{\{|u-v|<b\}} \lambda(v)\lambda(u) du dv \\
&= \left(\rho_2^b(K)\right)^2
\end{aligned}
$$

So, we have that $\rho_4^b(K) = O\left(\frac{b^2}{K^2}\right)$.

Putting this together, we have that

$$
\begin{aligned}
&E\left[\left(\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} 1_{\{|\tau_i - \tau_k|<b\}} \right)^2 \right] \\
&= \Lambda(K) + 8[\Lambda(K)]^2 \rho_2^b(K) + [\Lambda(K)]^2 + 4[\Lambda(K)]^3 \rho_3^b(K) + [\Lambda(K)]^4 \rho_4^b(K) \\
&= O(K) + O(K^2) O\left(\frac{b}{K}\right) + O(K^2) + O(K^3) O\left(\frac{b^2}{K^2}\right) + O(K^4) O\left(\frac{b^2}{K^2}\right) \\
&= O(K) + O(bK) + O(K^2) + O(b^2 K) + O(b^2 K^2) \\
&= O(b^2 K^2)
\end{aligned}
$$

Thus, it follows that $\sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} 1_{\{|\tau_i - \tau_k|<b\}} = O_p(bK)$, since dividing through by bK would give us constant values for the mean and the variance.

Hence, we have that

$$
\begin{aligned}
&\frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} O\left(\frac{b}{h}\right) 1_{\{|\tau_i - \tau_k|<b\}} X(\tau_i) X(\tau_k) \\
&= O\left(\frac{1}{K}\right) O\left(\frac{b}{h}\right) O_p(bK) O_p(1) O_p(1) \\
&= O_p\left(\frac{b^2}{h}\right)
\end{aligned}
$$

But by Assumptions A.4.5 and A.4.6, we have that this tends to 0, completing the proof.

□

Proof of Lemma 4.4.11: To establish (i), notice that

$$\begin{aligned}
E\left[\hat{R}(0)\right] &= E\left[\frac{1}{\Lambda(K)}\sum_{i=1}^{N(K)} X^2(\tau_i)\right] \\
&= \frac{1}{\Lambda(K)} E\left[E\left[\sum_{i=1}^{N(K)} X^2(\tau_i)\middle| N\right]\right] \\
&= \frac{1}{\Lambda(K)} E\left[N(K)E[X^2(t)|N]\right] \\
&= \frac{1}{\Lambda(K)} E\left[N(K)\left(Var[X(t)|N] + (E[X(t)|N])^2\right)\right] \\
&= \frac{R(0)+\mu^2}{\Lambda(K)} \underbrace{E\left[N(K)\right]}_{=\Lambda(K)} \\
&= R(0)+\mu^2
\end{aligned}$$

In integral form, we have that

$$E\left[\hat{R}(0)\right] = E\left[E\left[\hat{R}(0)\middle| N\right]\right] = E\left[\frac{1}{\Lambda(K)}\int_K \left(R(0)+\mu^2\right) N(dt)\right] = R(0)+\mu^2.$$

To compute the variance, we use the formula $Var\left[\hat{R}(0)\right] = E\left[\hat{R}(0)^2\right] - \left(E\left[\hat{R}(0)\right]\right)^2$. To that end, we compute

$$
\begin{aligned}
& E\left[\hat{R}(0)^2\right] \\
&= \frac{1}{[\Lambda(K)]^2} E\left[\sum_{i=1}^{N(K)} \sum_{j=1}^{N(K)} X^2(\tau_i) X^2(\tau_j)\right] \\
&= \frac{1}{[\Lambda(K)]^2} E\left[\int_K \int_K X^2(t) X^2(s) N(dt) N(ds)\right] \\
&= \frac{1}{[\Lambda(K)]^2} E\left[E\left[\int_K \int_K X^2(t) X^2(s) N(dt) N(ds) | N\right]\right] \\
&= \frac{1}{[\Lambda(K)]^2} E\left[\int_K \int_K E\left[X^2(t) X^2(s) | N\right] N(dt) N(ds)\right] \\
&= \frac{1}{[\Lambda(K)]^2} E\left[\int_K \int_K \left(Cov(X^2(t), X^2(s)) + \left(R(0) + \mu^2\right)^2\right) N(dt) N(ds)\right] \\
&= \frac{1}{[\Lambda(K)]^2} \int_K \int_K \left(Cov(X^2(t), X^2(s)) + \left(R(0) + \mu^2\right)^2\right) \lambda(t) \lambda(s) dt ds \\
&= \frac{1}{[\Lambda(K)]^2} \left(\int_K \int_K Cov(X^2(t), X^2(s)) \lambda(t) \lambda(s) dt ds + \left(R(0) + \mu^2\right)^2 [\Lambda(K)]^2\right) \\
&= \left(R(0) + \mu^2\right)^2 + \frac{1}{[\Lambda(K)]^2} \int_K \int_K Cov(X^2(t), X^2(s)) \lambda(t) \lambda(s) dt ds
\end{aligned}
$$

So, we have that

$$
\begin{aligned}
Var\left[\hat{R}(0)\right] &= E\left[\hat{R}(0)^2\right] - \left(E\left[\hat{R}(0)\right]\right)^2 \\
&= \frac{1}{[\Lambda(K)]^2} \int_K \int_K Cov(X^2(t), X^2(s)) \lambda(t) \lambda(s) dt ds
\end{aligned}
$$

By Assumption A.4.1, we have that

$$
Var\left[\hat{R}(0)\right] \leq \frac{(\lambda_{max})^2}{[\Lambda(K)]^2} \int_K \int_K \left|Cov(X^2(t), X^2(s))\right| dt ds
$$

Let $R_2(u) = Cov(X^2(0), X^2(u))$. (Here, we use the stationarity of $X(t)$). Now, making the change of variables $u = t - s$, we can reduce the above to

$$\begin{aligned}\int_K \int_K \left|Cov(X^2(t), X^2(s))\right| dtds &= \int_0^K \int_0^K |R_2(t-s)|\, dtds \\ &= K \int_{-K}^{K} \left(1 - \frac{|u|}{K}\right) |R_2(u)|\, du \\ &= 2K \int_0^K \left(1 - \frac{|u|}{K}\right) |R_2(u)|\, du\end{aligned}$$

Appealing to Lemma 4.4.12, with $E_1 = \{t\}$ and $E_2 = \{s\}$, we see that E_1 and E_2 are translations of each other, $k = \rho(E_1, E_2) = |t-s|$, and $|E1| = |E2| = 0$.

Consider the random field $\{X^2(t)\}$ instead of $\{X(t)\}$. It is still independent of the Poisson process. Then, $\tilde{Y}_1 = X^2(t)$ and $\tilde{Y}_2 = X^2(s)$.

Under the assumption that $E[|X^2(t)|^p] = C_p < \infty$ for some $p > 2$ (which is satisfied by the stronger Assumption A.4.2), we have that

$$\left|Cov(X^2(t), X^2(s))\right| \leq 10 C_p^{2/p} (\bar{\alpha}_X(|t-s|\,; 0))^{1-2/p}$$

Thus, we have that

$$\begin{aligned}Var\left[\hat{R}(0)\right] &\leq \frac{(\lambda_{max})^2}{[\Lambda(K)]^2} \left(2K \int_0^K \left(1 - \frac{|u|}{K}\right) |R_2(|u|)|\, du\right) \\ &\leq \frac{2K(\lambda_{max})^2}{[\Lambda(K)]^2} \left(\int_0^K |R_2(|u|)|\, du\right) \\ &\leq \frac{2K(\lambda_{max})^2}{[\Lambda(K)]^2} \left(|R_2(|0|)| + \int_1^K |R_2(|u|)|\, du\right)\end{aligned}$$

But, $\left|Cov(X^2(t), X^2(s))\right| = |R_2(|t-s|)| = |R_2(|u|)| \leq 10 C_p^{2/p} (\bar{\alpha}_X(|u|\,; 0))^{1-2/p}$, so we have that

$$\int_d^K |R_2(|u|)|\, du \leq \text{const.} \int_d^K (\bar{\alpha}_X(|u|\,; 0))^{1-2/(2+\delta)} du, \text{ for some } \delta > 0.$$

By Assumption A.4.3, we have that

$$\begin{aligned}\int_1^K (\bar{\alpha}_X(|u|\,;0))^{1-2/(2+\delta)}du &\leq \int_1^K \left(\frac{1}{u^{1+\varepsilon}}\right)^{1-2/(2+\delta)} du \\ &= \int_1^K \left(\frac{1}{u^{(1+2/\delta+\xi)(\delta/(2+\delta))}}\right) du \\ &= \int_1^K \left(\frac{1}{u^{1+\xi}}\right) du\end{aligned}$$

As $K \to \infty$, we have that this integral will converge.

Putting this all together, we have that $Var\left[\hat{R}(0)\right] \leq O\left(\frac{1}{K}\right)$, and thus we have verified that the variance tends to 0 as $K \to \infty$, which completes the proof of (ii).

□

Proof of Lemma 4.4.13a: Begin by making the substitutions $u_1 = u - v$, $u_2 = t - v$, $u_3 = s - v$, and $u_4 = v$.

Notice that u_1, u_2 and u_3 now range from $-K$ to K while u_4 still ranges from 0 to K.

Also, $|t - s| = |u_2 - u_3|$ and $|v - u| = |u_1|$. Moreover, the absolute value of the Jacobian of the transformation is 1.

And so, after making this change of variables, we have

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\int_0^K \left[\left(1 - \frac{|t-s|}{b}\right)^+\left(1 - \frac{|v-u|}{b}\right)^+ |Q(u-v, t-v, s-v)|\right] dsdtdudv$$

$$\leq \frac{1}{[\Lambda(K)]^2}\int_0^K\int_{-K}^K\int_{-K}^K\int_{-K}^K \left(1 - \frac{|u_2-u_3|}{b}\right)^+\left(1 - \frac{|u_1|}{b}\right)^+ |Q(u_1, u_2, u_3)|\, du_1du_2du_3du_4$$

$$= \frac{K}{[\Lambda(K)]^2}\int_{-K}^K\int_{-K}^K\int_{-K}^K \left(1 - \frac{|u_2-u_3|}{b}\right)^+\left(1 - \frac{|u_1|}{b}\right)^+ |Q(u_1, u_2, u_3)|\, du_1du_2du_3$$

Now, let $v_1 = u_2 - u_3$ and $v_2 = u_2$. Then we have that

$$\frac{K}{[\Lambda(K)]^2}\int_{-K}^{K}\int_{-K}^{K}\int_{-K}^{K}\left(1-\frac{|u_2-u_3|}{b}\right)^+\left(1-\frac{|u_1|}{b}\right)^+|Q(u_1,u_2,u_3)|\,du_1du_2du_3$$

$$\leq \frac{K}{[\Lambda(K)]^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(1-\frac{|v_1|}{b}\right)^+\left(1-\frac{|u_1|}{b}\right)^+\int_{-\infty}^{\infty}|Q(u_1,v_2,v_2-v_1)|\,dv_2du_1dv_1$$

$$\leq \frac{C_QK}{[\Lambda(K)]^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(1-\frac{|v_1|}{b}\right)^+\left(1-\frac{|u_1|}{b}\right)^+du_1dv_1$$

$$=\frac{C_QKb^2}{[\Lambda(K)]^2}$$

$$=O\left(\frac{b^2}{K}\right)$$

□

Proof of Lemma 4.4.13b: We begin with the same substitutions as in the proof of Lemma 4.4.13a. Let $u_1 = u - v$, $u_2 = t - v$, $u_3 = s - v$, and $u_4 = v$. Notice that u_1, u_2 and u_3 now range from $-K$ to K while u_4 still ranges from 0 to K. Also, $|t-s| = |u_2 - u_3|$, $|v-u| = |u_1|$, $t - v = u_2$ and $s - u = u_3 - u_1$. Again, the absolute value of the Jacobian of the transformation is 1.

And so, after making this change of variables, we have

$$\frac{1}{[\Lambda(K)]^2}\int_{0}^{K}\int_{0}^{K}\int_{0}^{K}\int_{0}^{K}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|v-u|}{b}\right)^+|R(t-v)R(s-u)|\,dsdtdudv$$

$$\leq \frac{K}{[\Lambda(K)]^2}\int_{-K}^{K}\int_{-K}^{K}\int_{-K}^{K}\left(1-\frac{|u_2-u_3|}{b}\right)^+\left(1-\frac{|u_1|}{b}\right)^+|R(u_2)R(u_3-u_1)|\,du_1du_2du_3$$

Now, let $v_1 = u_3 - u_1$ and $v_2 = u_2 - u_3$ and $v_3 = u_1$. Again, the Jacobian is 1. This gives us

$$\frac{K}{[\Lambda(K)]^2}\int_{-K}^{K}\int_{-K}^{K}\int_{-K}^{K}\left(1-\frac{|u_2-u_3|}{b}\right)^+\left(1-\frac{|u_1|}{b}\right)^+|R(u_2)R(u_3-u_1)|\,du_1du_2du_3$$

$$=\frac{K}{[\Lambda(K)]^2}\int_{-K}^{K}\int_{-2K}^{2K}\int_{-2K}^{2K}\left(1-\frac{|v_2|}{b}\right)^+\left(1-\frac{|v_3|}{b}\right)^+|R(v_1+v_2+v_3)R(v_1)|\,dv_1dv_2dv_3$$

$$\leq\frac{|R(0)|\,K}{[\Lambda(K)]^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(1-\frac{|v_2|}{b}\right)^+\left(1-\frac{|v_3|}{b}\right)^+\int_{-\infty}^{\infty}|R(v_1)|\,dv_1dv_2dv_3$$

$$=\frac{|R(0)|\,C_RK}{[\Lambda(K)]^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(1-\frac{|v_2|}{b}\right)^+\left(1-\frac{|v_3|}{b}\right)^+dv_2dv_3$$

$$=\frac{|R(0)|\,C_RKb^2}{[\Lambda(K)]^2}$$

$$=O\left(\frac{b^2}{K}\right)$$

Note: we used the fact that $|R(v_1+v_2+v_3)|\leq|R(0)|$ on the third-to-last line above. Also, Assumption A.4.2 and Assumption A.4.3 give us $\int_{-\infty}^{\infty}|R(u)|\,du=C_R<\infty$ on the second-to-last line above.

□

Proof of Lemma 4.4.13c: The proof is almost identical to the proof of Lemma 4.4.13b.

We begin by making the same substitutions as in the proof of Lemma 4.4.13a. Let $u_1=u-v$, $u_2=t-v$, $u_3=s-v$, and $u_4=v$. Notice that u_1, u_2 and u_3 now range from $-K$ to K while u_4 still ranges from 0 to K. Also, $|t-s|=|u_2-u_3|$, $|v-u|=|u_1|$, $s-v=u_3$ and $u-t=u_1-u_2$. Again, the absolute value of the Jacobian of the transformation is 1.

$$\frac{1}{[\Lambda(K)]^2}\int_{0}^{K}\int_{0}^{K}\int_{0}^{K}\int_{0}^{K}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|v-u|}{b}\right)^+|R(s-v)R(u-t)|\,dsdtdudv$$

$$\leq\frac{K}{[\Lambda(K)]^2}\int_{-K}^{K}\int_{-K}^{K}\int_{-K}^{K}\left(1-\frac{|u_2-u_3|}{b}\right)^+\left(1-\frac{|u_1|}{b}\right)^+|R(u_3)R(u_1-u_2)|\,du_1du_2du_3$$

Now, let $v_1=u_1-u_2$ and $v_2=u_2-u_3$ and $v_3=u_1$. Then we have that

$$\frac{K}{[\Lambda(K)]^2}\int_{-K}^{K}\int_{-K}^{K}\int_{-K}^{K}\left(1-\frac{|u_2-u_3|}{b}\right)^+\left(1-\frac{|u_1|}{b}\right)^+|R(u_3)R(u_1-u_2)|\,du_1du_2du_3$$

$$=\frac{K}{[\Lambda(K)]^2}\int_{-K}^{K}\int_{-2K}^{2K}\int_{-2K}^{2K}\left(1-\frac{|v_2|}{b}\right)^+\left(1-\frac{|v_3|}{b}\right)^+|R(v_3-v_1-v_2)R(v_1)|\,dv_1dv_2dv_3$$

$$\leq\frac{|R(0)|\,K}{[\Lambda(K)]^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(1-\frac{|v_2|}{b}\right)^+\left(1-\frac{|v_3|}{b}\right)^+\int_{-\infty}^{\infty}|R(v_1)|\,dv_1dv_2dv_3$$

$$=\frac{|R(0)|\,C_RK}{[\Lambda(K)]^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(1-\frac{|v_2|}{b}\right)^+\left(1-\frac{|v_3|}{b}\right)^+dv_2dv_3$$

$$=\frac{|R(0)|\,C_RKb^2}{[\Lambda(K)]^2}$$

$$=O\left(\frac{b^2}{K}\right)$$

Note: we used the fact that $|R(v_3-v_1-v_2)|\leq|R(0)|$ on the third-to-last line above. Also, Assumption A.4.2 and Assumption A.4.3 give us $\int_{-\infty}^{\infty}|R(u)|\,du=C_R<\infty$ on the second-to-last line above.

□

Proof of Lemma 4.4.13d: We begin with the substitutions $u_1=u-t$, $u_2=s-t$, $u_3=t$. Notice that u_1 and u_2 now range from $-K$ to K while u_3 still ranges from 0 to K. Also, $|t-s|=|u_2|$ and $|t-u|=|u_1|$. The Jacobian of the transformation is 1. The above becomes

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|t-u|}{b}\right)^+|Q(0,u-t,s-t)|\,dsdtdu$$

$$\leq\frac{1}{[\Lambda(K)]^2}\int_0^K\int_{-K}^K\int_{-K}^K\left(1-\frac{|u_2|}{b}\right)^+\left(1-\frac{|u_1|}{b}\right)^+|Q(0,u_1,u_2)|\,du_1du_2du_3$$

$$=\frac{K}{[\Lambda(K)]^2}\int_{-K}^K\int_{-K}^K\left(1-\frac{|u_2|}{b}\right)^+\left(1-\frac{|u_1|}{b}\right)^+|Q(0,u_1,u_2)|\,du_1du_2$$

$$\leq\frac{K}{[\Lambda(K)]^2}\int_{-\infty}^\infty\left(1-\frac{|u_2|}{b}\right)^+\int_{-\infty}^\infty|Q(0,u_1,u_2)|\,du_1du_2$$

$$\leq\frac{C_QK}{[\Lambda(K)]^2}\int_{-\infty}^\infty\left(1-\frac{|u_2|}{b}\right)^+du_2$$

$$=\frac{C_QKb}{[\Lambda(K)]^2}$$

$$=O\left(\frac{b}{K}\right)$$

□

Proof of Lemma 4.4.13e: Let $u_1=s-t$, $u_2=u-t$, and $u_3=t$. Notice that u_1 and u_2 now range from $-K$ to K while u_3 still ranges from 0 to K. Also, $|t-s|=|u_1|$, $|t-u|=|u_2|$ and $s-u=u_1-u_2$. The Jacobian of the transformation is 1. This gives us

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|t-u|}{b}\right)^+|R(0)R(s-u)|\,dsdtdu$$

$$\leq\frac{|R(0)|}{[\Lambda(K)]^2}\int_0^K\int_{-K}^K\int_{-K}^K\left(1-\frac{|u_1|}{b}\right)^+\left(1-\frac{|u_2|}{b}\right)^+|R(u_1-u_2)|\,du_1du_2du_3$$

$$=\frac{|R(0)|\,K}{[\Lambda(K)]^2}\int_{-K}^K\int_{-K}^K\left(1-\frac{|u_1|}{b}\right)^+\left(1-\frac{|u_2|}{b}\right)^+|R(u_1-u_2)|\,du_1du_2$$

$$\leq\frac{|R(0)|^2K}{[\Lambda(K)]^2}\int_{-\infty}^\infty\int_{-\infty}^\infty\left(1-\frac{|u_1|}{b}\right)^+\left(1-\frac{|u_2|}{b}\right)^+du_1du_2$$

$$=\frac{|R(0)|^2Kb^2}{[\Lambda(K)]^2}$$

$$=O\left(\frac{b^2}{K}\right)$$

Note: We used the fact that $|R(u_1-u_2)|\leq|R(0)|$ in the third-to-last line above.

□

Proof of Lemma 4.4.13f: Let $u_1 = s - t$, $u_2 = u - t$, and $u_3 = t$. Notice that u_1 and u_2 now range from $-K$ to K while u_3 still ranges from 0 to K. Also, $|t - u| = |u_2|$. The Jacobian of the transformation is 1. This gives us

$$\begin{aligned}
&\frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|t-u|}{b}\right)^+ |R(u-t)R(s-t)|\, dsdtdu \\
&\leq \frac{1}{[\Lambda(K)]^2} \int_0^K \int_{-K}^K \int_{-K}^K \left(1 - \frac{|u_1|}{b}\right)^+ \left(1 - \frac{|u_2|}{b}\right)^+ |R(u_2)R(u_1)|\, du_1 du_2 du_3 \\
&= \frac{K}{[\Lambda(K)]^2} \int_{-K}^K \int_{-K}^K \left(1 - \frac{|u_1|}{b}\right)^+ \left(1 - \frac{|u_2|}{b}\right)^+ |R(u_2)R(u_1)|\, du_1 du_2 \\
&\leq \frac{|R(0)|^2 K}{[\Lambda(K)]^2} \int_{-\infty}^\infty \int_{-\infty}^\infty \left(1 - \frac{|u_1|}{b}\right)^+ \left(1 - \frac{|u_2|}{b}\right)^+ du_1 du_2 \\
&= \frac{|R(0)|^2 K b^2}{[\Lambda(K)]^2} \\
&= O\left(\frac{b^2}{K}\right)
\end{aligned}$$

Note: We used the fact that $|R(u_2)R(u_1)| \leq |R(u_2)||R(u_1)| \leq |R(0)|^2$ in the third-to-last line above.

□

Proof of Lemma 4.4.13g: The proof of Lemma 4.4.13g is almost identical to Lemma 4.4.13f. The only change is in the substitution used.

Let $u_1 = s - t$, $u_2 = t - u$, and $u_3 = t$. Notice that u_1 and u_2 now range from $-K$ to K while u_3 still ranges from 0 to K. Again, the Jacobian of the transformation is 1. This gives us

$$\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|t-u|}{b}\right)^+|R(s-t)R(t-u)|\,dsdtdu$$

$$=\frac{1}{[\Lambda(K)]^2}\int_0^K\int_{-K}^K\int_{-K}^K\left(1-\frac{|u_1|}{b}\right)^+\left(1-\frac{|u_2|}{b}\right)^+|R(u_1)R(u_2)|\,du_1du_2du_3$$

$$=\frac{K}{[\Lambda(K)]^2}\int_{-K}^K\int_{-K}^K\left(1-\frac{|u_1|}{b}\right)^+\left(1-\frac{|u_2|}{b}\right)^+|R(u_1)R(u_2)|\,du_1du_2$$

$$\leq\frac{|R(0)|^2K}{[\Lambda(K)]^2}\int_{-\infty}^\infty\int_{-\infty}^\infty\left(1-\frac{|u_1|}{b}\right)^+\left(1-\frac{|u_2|}{b}\right)^+du_1du_2$$

$$=\frac{|R(0)|^2Kb^2}{[\Lambda(K)]^2}$$

$$=O\left(\frac{b^2}{K}\right)$$

Note: We used the fact that $|R(u_1)R(u_2)|\leq|R(u_2)||R(u_1)|\leq|R(0)|^2$ in the third-to-last line above.

□

Proof of Lemma 4.4.14: Our estimator is given by

$$\hat{\theta}^2=\frac{1}{\Lambda(K)}\sum_{i\neq k}\left(1-\frac{|\tau_i-\tau_k|}{b}\right)1_{\{|\tau_i-\tau_k|<b\}}X(\tau_i)X(\tau_k)$$

We can express this in integral form as

$$\hat{\theta}^2=\frac{1}{\Lambda(K)}\int_0^K\int_0^K\left(1-\frac{|t-s|}{b}\right)^+X(t)X(s)N^{(2)}(ds,dt),$$

where $N^{(2)}(dt,ds)=N(dt)N(ds)1_{\{t\neq s\}}$.

We begin by considering the expected value. Our strategy is to condition on N.

$$
\begin{aligned}
E\left[\hat{\theta}^2\right] &= \frac{1}{\Lambda(K)} E\left[\int_K \int_K \left(1-\frac{|t-s|}{b}\right)^+ X(t)X(s)N^{(2)}(dt,ds)\right] \\
&= \frac{1}{\Lambda(K)} E\left[E\left[\int_K \int_K \left(1-\frac{|t-s|}{b}\right)^+ X(t)X(s)N^{(2)}(dt,ds)\middle| N\right]\right] \\
&= \frac{1}{\Lambda(K)} E\left[\int_K \int_K \left(1-\frac{|t-s|}{b}\right)^+ E\left[X(t)X(s)|N\right] N^{(2)}(dt,ds)\right] \\
&= \frac{1}{\Lambda(K)} \int_K \int_K \left(1-\frac{|t-s|}{b}\right)^+ R(t-s)\lambda(t)\lambda(s)dtds
\end{aligned}
$$

Thus, we have that

$$
\begin{aligned}
\theta^2 - E\left[\hat{\theta}^2\right] &= \frac{1}{\Lambda(K)} \int_K \int_K R(t-s)\lambda(t)\lambda(s)dtds \\
&\quad - \frac{1}{\Lambda(K)} \int_K \int_K R(t-s)\lambda(t)\lambda(s)dtds \\
&= \frac{1}{\Lambda(K)} \int_K \int_K \left(1-\left(1-\frac{|t-s|}{b}\right)^+\right) R(t-s)\lambda(t)\lambda(s)dtds
\end{aligned}
$$

And so,

$$
\begin{aligned}
\left|\theta^2 - E\left[\hat{\theta}^2\right]\right| &= \left|\frac{1}{\Lambda(K)} \int_K \int_K \left(1-\left(1-\frac{|t-s|}{b}\right)^+\right) R(t-s)\lambda(t)\lambda(s)dtds\right| \\
&\leq \frac{1}{\Lambda(K)} \int_K \int_K \left(1-\left(1-\frac{|t-s|}{b}\right)^+\right) |R(t-s)|\,|\lambda(t)|\,|\lambda(s)|\,dtds
\end{aligned}
$$

Under Assumption A.4.1, we have

$$
\left|\theta^2 - E\left[\hat{\theta}^2\right]\right| \leq \frac{(\lambda_{max})^2}{\Lambda(K)} \int_K \int_K \left(1-\left(1-\frac{|t-s|}{b}\right)^+\right) |R(t-s)|\,dtds
$$

Now, let $u = t - s$, so that $du = dt$. After making the change of variables, we can reduce the double integral into a single integral as

$$
\begin{aligned}
&\left|\theta^2 - E\left[\hat{\theta}^2\right]\right| \\
&= \frac{(\lambda_{max})^2 K}{\Lambda(K)} \int_{-K}^{K} \left(1 - \frac{|u|}{K}\right)\left(1 - \left(1 - \frac{|u|}{b}\right)^{+}\right) |R(u)|\, du \\
&= \frac{(\lambda_{max})^2 K}{\Lambda(K)} \int_{-b}^{b} \left(1 - \frac{|u|}{K}\right) \frac{|u|}{b} |R(u)|\, du \\
&+ \frac{(\lambda_{max})^2 K}{\Lambda(K)} \int_{-K}^{-b} \left(1 - \frac{|u|}{K}\right) |R(u)|\, du + \frac{(\lambda_{max})^2 K}{\Lambda(K)} \int_{b}^{K} \left(1 - \frac{|u|}{K}\right) |R(u)|\, du \\
&= 2\frac{(\lambda_{max})^2 K}{\Lambda(K)} \int_{0}^{b} \left(1 - \frac{u}{K}\right) \frac{u}{b} |R(u)|\, du + 2\frac{(\lambda_{max})^2 K}{\Lambda(K)} \int_{b}^{K} \left(1 - \frac{u}{K}\right) |R(u)|\, du
\end{aligned}
$$

where the last line is due to symmetry.

In order to ensure that these integrals tend to 0, we require

$$|R(u)| \leq \min\left\{|R(0)|, \frac{const.}{|u|^{\eta}}, \right\}$$

for some $\eta > 2$. We can connect this with mixing coefficients, though. Using Assumption A.4.2, we can strengthen the lemma from Roussas and Ioannides to obtain the bound

$$|R(t)| = |Cov(X(0), X(t))| \leq const.\bar{\alpha}_X(t;0)^{1-2/(6+\delta)}$$

But by Assumption A.4.3 (appealing to the second condition), we have that

$$
\begin{aligned}
|R(u)| &\leq const.(\bar{\alpha}_X(u;0))^{1-2/(6+\delta)} \\
&\leq const.\left(u^{-1-\varepsilon}\right)^{1-2/(6+\delta)} \\
&< const.\left(u^{-1-(8+\delta)/(4+\delta)}\right)^{1-2/(6+\delta)} \\
&= const.\left(u^{-2}\right)
\end{aligned}
$$

And so, our assumption on $|R(u)|$ is valid. Hence,

$$
\begin{aligned}
&\left|\theta^2 - E\left[\hat{\theta}^2\right]\right| \\
&\leq 2\frac{(\lambda_{max})^2 K}{\Lambda(K)} \int_0^b \left(1 - \frac{u}{K}\right) \frac{u}{b} \left|R(u)\right| du + 2\frac{(\lambda_{max})^2 K}{\Lambda(K)} \int_b^K \left(1 - \frac{u}{K}\right) \left|R(u)\right| du \\
&\leq \frac{2(\lambda_{max})^2 K}{\Lambda(K)} \int_0^d \left(1 - \frac{u}{K}\right) \frac{u}{b} \left|R(0)\right| du + \frac{2(\lambda_{max})^2 K}{\Lambda(K)} \int_d^b \left(1 - \frac{u}{K}\right) \frac{u}{b} \frac{C}{u^\eta} du \\
&+ \frac{2(\lambda_{max})^2 K}{\Lambda(K)} \int_b^K \left(1 - \frac{u}{K}\right) \frac{C}{u^\eta} du \\
&= O\left(\frac{1}{b}\right) + O\left(\frac{1}{Kb^{\eta-2}} + \frac{1}{b^{\eta-1}}\right) + O\left(\frac{1}{Kb^{\eta-2}} + \frac{1}{K^{\eta-1}} + \frac{1}{b^{\eta-1}}\right) \\
&= O\left(\frac{1}{b}\right) + O\left(\frac{1}{b^{\eta-1}}\right) \\
&= O\left(\frac{1}{b}\right)
\end{aligned}
$$

Here, we used the fact that $K > b$ and $\eta > 2$ to simplify the above expressions.
Note: d is specified to be the value for which $|R(0)| = const./|u|^\eta$. Thus, we see that

$$E[\hat{\theta}^2] = \theta^2 + O\left(\frac{1}{b}\right),$$

and (i) is proven.

Next, we wish to consider the second moment.

$$
\begin{aligned}
&E\left[\left(\hat{\theta}^2\right)^2\right] \\
&= \frac{1}{\left[\Lambda(K)\right]^2} E\left[\left(\sum_{i \neq k} \left(1 - \frac{|\tau_i - \tau_k|}{b}\right) 1_{\{|\tau_i - \tau_k| < b\}} X(\tau_i) X(\tau_k)\right)^2\right] \\
&= \frac{1}{\left[\Lambda(K)\right]^2} E\left[\begin{array}{l} \left(\sum_{i \neq k} \left(1 - \frac{|\tau_i - \tau_k|}{b}\right) 1_{\{|\tau_i - \tau_k| < b\}} X(\tau_i) X(\tau_k)\right) \\ \times \left(\sum_{j \neq l} \left(1 - \frac{|\tau_j - \tau_l|}{b}\right) 1_{\{|\tau_j - \tau_l| < b\}} X(\tau_j) X(\tau_l)\right) \end{array}\right] \\
&= \frac{1}{\left[\Lambda(K)\right]^2} E\left[\int_K \int_K \int_K \int_K \left[\begin{array}{l} \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|v-u|}{b}\right)^+ \\ \times X(v) X(u) X(t) X(s) \end{array}\right] N^{(2)}(dv, du) N^{(2)}(dt, ds)\right]
\end{aligned}
$$

There are seven cases to consider, depending on what variables are equal to each other. So, we have that

$$E\left[\left(\hat{\theta}^2\right)^2\right] = V_1 + V_2 + V_3 + V_4 + V_5 + V_6 + V_7,$$

where

$$V_1 = \frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \int_0^K \left[\begin{array}{l} \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|v-u|}{b}\right)^+ \\ \times E\left[X(v)X(u)X(t)X(s)\right] \lambda(v)\lambda(u)\lambda(t)\lambda(s) \end{array} \right] dsdtdudv$$

$$V_2 = \frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left[\begin{array}{l} \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|t-u|}{b}\right)^+ \\ \times E\left[X(t)^2 X(u)X(s)\right] \lambda(u)\lambda(t)\lambda(s) \end{array} \right] dsdtdu$$

$$V_3 = \frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left[\begin{array}{l} \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|s-u|}{b}\right)^+ \\ \times E\left[X(s)^2 X(u)X(t)\right] \lambda(u)\lambda(t)\lambda(s) \end{array} \right] dsdtdu$$

$$V_4 = \frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left[\begin{array}{l} \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|v-t|}{b}\right)^+ \\ \times E\left[X(t)^2 X(v)X(s)\right] \lambda(v)\lambda(t)\lambda(s) \end{array} \right] dsdtdv$$

$$V_5 = \frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left[\begin{array}{l} \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|v-s|}{b}\right)^+ \\ \times E\left[X(s)^2 X(v)X(t)\right] \lambda(v)\lambda(t)\lambda(s) \end{array} \right] dsdtdv$$

$$V_6 = \frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K \left[\begin{array}{l} \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|t-s|}{b}\right)^+ \\ \times E\left[X(s)^2 X(t)^2\right] \lambda(t)\lambda(s) \end{array} \right] dsdt$$

$$V_7 = \frac{1}{[\Lambda(K)]^2} \int_0^K \int_0^K \left[\begin{array}{l} \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|s-t|}{b}\right)^+ \\ \times E\left[X(s)^2 X(t)^2\right] \lambda(t)\lambda(s) \end{array} \right] dsdt$$

Of the seven components listed, there are quite a few repetitions due to symmetry. In particular, V_2, V_3, V_4, and V_5 are all the same (just with a change of variables) and V_6 and V_7 are the same, since $|t-s| = |s-t|$.

Thus, we can simplify the above expression into a more compact form as

$$E\left[\left(\hat{\theta}^2\right)^2\right] = V_1 + 4V_2 + 2V_6$$

Recall that

$$\begin{aligned}\left(E[\hat{\theta}^2]\right)^2 &= \frac{1}{[\Lambda(K)]^2}E\left[\int_0^K\int_0^K\left(1-\frac{|t-s|}{b}\right)^+X(t)X(s)N^{(2)}(ds,dt)\right]^2\\ &= \frac{1}{[\Lambda(K)]^2}\left(\int_K\int_K\left(1-\frac{|t-s|}{b}\right)^+R(t-s)\lambda(t)\lambda(s)dtds\right)^2\\ &= \frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\int_0^K\left[\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|v-u|}{b}\right)^+\\ \times R(t-s)R(v-u)\lambda(v)\lambda(u)\lambda(t)\lambda(s)\end{array}\right]dsdtdudv\end{aligned}$$

Notice that this has the same form as V_1, so we may merge this to obtain a formula for the variance of $\hat{\theta}^2$, since $Var[\hat{\theta}^2] = E[(\hat{\theta}^2)^2] - (E[\hat{\theta}^2])^2$.

Namely, we have that

$$Var[\hat{\theta}^2] = \tilde{V}_1 + 4V_2 + 2V_6,$$

where

$$\tilde{V}_1 = \frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\int_0^K\left[\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|v-u|}{b}\right)^+\times\\ \left(\begin{array}{l}E\left[X(v)X(u)X(t)X(s)\right]\\ -R(t-s)R(v-u)\end{array}\right)\lambda(v)\lambda(u)\lambda(t)\lambda(s)\end{array}\right]dsdtdudv$$

Now, define

$$Q(s,t,u) = E[X(0)X(s)X(t)X(u)] - R(s)R(u-t) - R(t)R(u-s) - R(u)R(t-s),$$

the fourth-order cumulant of X. From that definition, we have that

$$\begin{aligned}&E[X(v)X(u)X(t)X(s)]\\ &= Q(u-v,t-v,s-v) + R(u-v)R(s-t)\\ &+ R(t-v)R(s-u) + R(s-v)R(u-t)\end{aligned}$$

Note: This definition does not require s, t, u, and v to be distinct. Thus, it applies to all terms listed above.

Consider $\tilde{V}_1$. By Assumption A.4.1, we have that

$$\begin{aligned}\tilde{V}_1 &\leq \frac{(\lambda_{max})^4}{[\Lambda(K)]^2}\int_0^K\int_0^K\int_0^K\int_0^K\left|\begin{array}{l}\left(1-\frac{|t-s|}{b}\right)^+\left(1-\frac{|v-u|}{b}\right)^+\times\\ \left(\begin{array}{l}E\left[X(v)X(u)X(t)X(s)\right]\\ -R(t-s)R(v-u)\end{array}\right)\end{array}\right|dsdtdudv\\ &\leq \tilde{V}_{11} + \tilde{V}_{12} + \tilde{V}_{13}\end{aligned}$$

where

$$\tilde{V}_{11}$$
$$= \frac{(\lambda_{max})^4}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \int_0^K \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|v-u|}{b}\right)^+ |Q(u-v, t-v, s-v)| \, dsdtdudv$$

$$\tilde{V}_{12} = \frac{(\lambda_{max})^4}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \int_0^K \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|v-u|}{b}\right)^+ |R(t-v)R(s-u)| \, dsdtdudv$$

$$\tilde{V}_{13} = \frac{(\lambda_{max})^4}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \int_0^K \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|v-u|}{b}\right)^+ |R(s-v)R(u-t)| \, dsdtdudv$$

Note: The second and fifth term inside the absolute value canceled each other out.
By Lemma 4.4.13a, we have that under Assumption A.4.4,

$$\tilde{V}_{11} = O\left(\frac{b^2}{K}\right)$$

By Lemma 4.4.13b and Lemma 4.4.13c, we have that under Assumption A.4.2 and Assumption A.4.3,

$$\tilde{V}_{12} = \tilde{V}_{13} = O\left(\frac{b^2}{K}\right)$$

Thus, we see that $\tilde{V}_1 = O\left(\frac{b^2}{K}\right)$

Now, we turn our attention to V_2. We shall use the fact that the decomposition of $E[X(s)X(t)X(u)X(v)]$ does not require s, t, u, and v to be distinct. Thus, we can write V_2 as four pieces, similar to what we did with $\tilde{V}_1$. The main difference is that we do not have a cancellation like above. Thus, we have that

$$\begin{aligned} V_2 &\leq \frac{(\lambda_{max})^3}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left| \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|t-u|}{b}\right)^+ E\left[X(t)^2 X(u) X(s)\right] \right| dsdtdu \\ &\leq V_{21} + V_{22} + V_{23} + V_{24} \end{aligned}$$

where

$$V_{21} = \frac{(\lambda_{max})^3}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|t-u|}{b}\right)^+ |Q(t-t, u-t, s-t)| \, dsdtdu$$

$$V_{22} = \frac{(\lambda_{max})^3}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|t-u|}{b}\right)^+ |R(t-t)R(s-u)|\, dsdtdu$$

$$V_{23} = \frac{(\lambda_{max})^3}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|t-u|}{b}\right)^+ |R(u-t)R(s-t)|\, dsdtdu$$

$$V_{24} = \frac{(\lambda_{max})^3}{[\Lambda(K)]^2} \int_0^K \int_0^K \int_0^K \left(1 - \frac{|t-s|}{b}\right)^+ \left(1 - \frac{|t-u|}{b}\right)^+ |R(s-t)R(t-u)|\, dsdtdu$$

By Lemma 4.4.13d, we have that under Assumption A.4.4,

$$V_{21} = O\left(\frac{b}{K}\right)$$

By Lemmas 4.4.13e, 4.4.13f, and 4.4.13g, we have that,

$$V_{22} = V_{23} = V_{24} = O\left(\frac{b^2}{K}\right)$$

Thus, we see that $V_2 = O\left(\frac{b^2}{K}\right)$

To recap, we have established that

$$Var\left[\hat{\theta}^2\right] = \frac{2}{[\Lambda(K)]^2} \int_0^K \int_0^K \left(\left(1 - \frac{|t-s|}{b}\right)^2\right)^+ E\left[X(s)^2 X(t)^2\right] \lambda(t)\lambda(s) dsdt + O\left(\frac{b^2}{K}\right)$$

At this point, we appeal to Theorem 17.2.2 in Ibragimov and Linnik [26].

Theorem. *Let the random variables ξ, η be measureable with respect to $M_{-\infty}^t$ and $M_{t+\tau}^{\infty}$ respectively, and suppose that for some $\delta > 0$, $E[|\xi|^{2+\delta}] < c_1 < \infty$, $E[|\eta|^{2+\delta}] < c_2 < \infty$. Then $|E[\xi\eta] - E[\xi]E[\eta]| \leq [4 + 3(c_1^{\beta} c_2^{1-\beta} + c_1^{1-\beta} c_2^{\beta})]\alpha(\tau)^{1-2\beta}$, where $\beta = (2+\delta)^{-1}$, where $\alpha(\tau)$ is the strong-mixing coefficient.*

Here, let $\xi = X(s)^2$ and $\eta = X(t)^2$, and notice that we can use the same bound for c_1 and c_2, which is valid because of Assumption A.4.2. Thus, we have that

$$\begin{aligned}
\left|E\left[X(s)^2 X(t)^2\right]\right| &= \left|E\left[X(s)^2 X(t)^2\right] - E[X(s)^2]E[X(t)^2] + E[X(s)^2]E[X(t)^2]\right| \\
&\leq \left|E\left[X(s)^2 X(t)^2\right] - E[X(s)^2]E[X(t)]^2\right| + \left|E[X(s)^2]E[X(t)^2]\right| \\
&\leq const.\alpha(|t-s|)^{\delta/(2+\delta)} + \left(\left|E[X(t)^2]\right|\right)^2 \\
&< \infty
\end{aligned}$$

Also, we have that

$$\begin{aligned}\frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\left(\left(1-\frac{|t-s|}{b}\right)^+\right)^2 dsdt &\leq \frac{1}{[\Lambda(K)]^2}\int_0^K\int_0^K\left(1-\frac{|t-s|}{b}\right)^+ dsdt \\ &= \frac{Kb}{[\Lambda(K)]^2} \\ &= O\left(\frac{b}{K}\right)\end{aligned}$$

So, we see that

$$Var\left[\hat{\theta}^2\right] = O\left(\frac{b}{K}\right) + O\left(\frac{b^2}{K}\right) = O\left(\frac{b^2}{K}\right)$$

By Assumption A.4.6, we see that this tends to 0 as $K \to \infty$, completing the proof of (ii).

□

Proof of Lemma 4.4.15: We saw that

$$\begin{aligned}Var^*[\sqrt{\Lambda^*(K)}\tilde{X}^*] &= \frac{1}{\Lambda^*(K)}\sum_{i\neq k}\left(1-\frac{|\tau_i-\tau_k|}{b}\right)1_{\{|\tau_i-\tau_k|<b\}}X(\tau_i)X(\tau_k) \\ &\quad + \frac{1}{\Lambda^*(K)}\sum_{i=1}^{N(K)}X^2(\tau_i) \\ &\quad + \frac{1}{\Lambda^*(K)}\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}O\left(\frac{b}{h}\right)1_{\{|\tau_i-\tau_k|<b\}}X(\tau_i)X(\tau_k)\end{aligned}$$

So, we have that

$$\begin{aligned}Var^*[\sqrt{\Lambda(K)}\,\tilde{\tilde{X}}^*] &= \frac{1}{\Lambda(K)}\sum_{i\neq k}\left(1-\frac{|\tau_i-\tau_k|}{b}\right)1_{\{|\tau_i-\tau_k|<b\}}X(\tau_i)X(\tau_k) \\ &\quad + \frac{1}{\Lambda(K)}\sum_{i=1}^{N(K)}X^2(\tau_i) \\ &\quad + \frac{1}{\Lambda(K)}\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}O\left(\frac{b}{h}\right)1_{\{|\tau_i-\tau_k|<b\}}X(\tau_i)X(\tau_k)\end{aligned}$$

By Lemma 4.4.14, we have that first term tends in probability to θ^2. By Lemma 4.4.11, we have that the second term tends to $R(0)+\mu^2$. By Lemma 4.4.10, we have that the third term tends to 0.

□

Proof of Theorem 4.4.16:

Proof of (i):

We have that $\overset{\approx}{X}^* = \frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} W_i X(\tau_i)$ and, by Lemma 4.4.1, $E^*[W_i] = 1$. In the bootstrap world, the value of $N(K)$ as well as the value of the τ_i's is known; the only randomness comes from the W_i. Thus, we have that:

$$E^*\left[\overset{\approx}{X}^*\right] = \frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} E^*[W_i] X(\tau_i) = \frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} X(\tau_i) = \frac{1}{\Lambda(K)} \int_K X(t) N(dt) = \tilde{X}_K$$

Proof of (ii):

$$\frac{Var^*[\overset{\approx}{X}^*]}{Var[\tilde{X}_K]} = \frac{Var^*[\sqrt{\Lambda(K)}\,\overset{\approx}{X}^*]}{Var[\sqrt{\Lambda(K)}\tilde{X}_K]}$$

We saw in Theorem 3.4.4 and Lemma 4.4.15 that both the numerator and denominator tend to $\theta^2 + R(0) + \mu^2$ as $K \to \infty$. Thus, we have that their ratio tends to 1 in probability.

Proof of (iii):

We have already seen in Theorem 3.4.4 that

$$\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right)$$

Note: Assumptions A.4.2 and A.4.3 are sufficient to establish this result. Thus, all that remains to be shown is that

$$\frac{\sqrt{\Lambda(K)}\left(\overset{\approx}{X}^* - \tilde{X}_K\right)}{\sqrt{Var^*\left[\sqrt{\Lambda(K)}\overset{\approx}{X}^*\right]}} \xrightarrow{d} N(0,1)$$

But recall that

$$\begin{aligned} \overset{\approx}{X}^* &= \frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} W_i X(\tau_i) \\ &= \frac{1}{\Lambda(K)} \sum_{j=1}^{L} \sum_{i=1}^{N(K)} Y_{ij} X(\tau_i) \end{aligned}$$

Also, we have that

$$\begin{aligned}\tilde{X}_K &= \frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} X(\tau_i) \\ &= \frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} \left(\sum_{j=1}^{L} p_{ij} \right) X(\tau_i) \\ &= \frac{1}{\Lambda(K)} \sum_{j=1}^{L} \sum_{i=1}^{N(K)} p_{ij} X(\tau_i)\end{aligned}$$

since $\sum_{j=1}^{L} p_{ij} = 1$

So, we may rewrite

$$\sqrt{\Lambda(K)} \left(\tilde{\tilde{X}}^* - \tilde{X}_K \right)$$

as

$$\sqrt{\Lambda(K)} \left(\tilde{\tilde{X}}^* - \tilde{X}_K \right) = \sum_{j=1}^{L} \left(\frac{1}{\sqrt{\Lambda(K)}} \sum_{i=1}^{N(K)} (Y_{ij} - p_{ij}) X(\tau_i) \right)$$

where $p_{ij} = E^*[Y_{ij}]$ is the probability that $X(\tau_i)$ appears in block j in the resampled data. Also, notice that by our resampling scheme, each block ($j = 1, 2, \ldots, L$) is independent.

Recall, the Lindeberg-Feller Central Limit Theorem with the Liapunov Condition (see Theorem 9.8.1 in [64]).

Theorem. *Let $\{T_j\}$ be a sequence of independent random variables satisfying*

$$E^*[T_j] = 0,\ S_L = \sum_{j=1}^{L} T_j,\ Var^*[T_j] = \sigma_j^2 < \infty,\ \text{and}\ s_L^2 = \sum_{j=1}^{L} \sigma_j^2$$

If for some $\gamma > 0$,

$$\frac{\sum_{j=1}^{L} E^* \left[|T_j|^{2+\gamma} \right]}{s_L^{2+\gamma}} \to 0,$$

then

$$\frac{S_L}{s_L} \xrightarrow{d} N(0,1)$$

Here, we define $T_j = \frac{1}{\sqrt{\Lambda(K)}} \sum_{i=1}^{N(K)} (Y_{ij} - p_{ij})X(\tau_i)$, for $j = 1, 2, \ldots, L$ (so each is independent).

Notice that by its construction, we have that

$$E^*[T_j] = 0,\ S_L = \sqrt{\Lambda(K)}\left(\tilde{\tilde{X}}^* - \tilde{X}_K\right), \text{ and } s_L^2 = Var^*\left[\sqrt{\Lambda(K)}\tilde{\tilde{X}}^*\right]$$

Using $\gamma = 4$, all that we need to show is that

$$\frac{\sum_{j=1}^{L} E^*\left[|T_j|^6\right]}{s_L^6} \to 0$$

as $L \to \infty$ (where $L = K/b$)

By Assumption A.4.1, we have that in a block of size b, we expect to find $O(b)$ points. Thus, we have that there are $O_p(b)$ points in each of the L resampled blocks. Moreover, we have that $X(\tau_i) = O_p(1)$. Thus, we see that

$$\sum_{i=1}^{N(K)} (Y_{ij} - p_{ij})X(\tau_i) = O_p(b)$$

And so, it follows that

$$T_j = \frac{1}{\sqrt{\Lambda(K)}} \sum_{i=1}^{N(K)} (Y_{ij} - p_{ij})X(\tau_i) = O_p\left(\frac{b}{K^{1/2}}\right)$$

In Lemma 4.4.15, we saw that

$$s_L^2 = Var^*\left[\sqrt{\Lambda(K)}\tilde{\tilde{X}}^*\right] \xrightarrow{p} \theta^2 + R(0) + \mu^2$$

Hence,

$$s_L^6 = (s_L^2)^3\left(\theta^2 + R(0) + \mu^2\right)^3,$$

which is a constant. So, all that remains to be shown is that $\sum_{j=1}^{L} E^*\left[|T_j|^6\right] \to 0$.

But we just saw above that $T_j = O_p\left(\frac{b}{K^{1/2}}\right)$, so it follows immediately that

$$E^*\left[|T_j|^6\right] = O_p\left(\frac{b^6}{K^3}\right)$$

And so, we see that

$$\sum_{j=1}^{L} E^* \left[|T_j|^6 \right] = O_p \left(\frac{Lb^6}{K^3} \right) = O_p \left(\frac{b^5}{K^2} \right)$$

since $L = K/b$. But by Assumption A.4.6, we see that this tends to 0. Thus, by the Lindeberg-Feller Central Limit Theorem, we have that

$$\frac{S_L}{s_L} = \frac{\sqrt{\Lambda(K)} \left(\tilde{\tilde{X}}^* - \tilde{X}_K \right)}{\sqrt{Var^* \left[\sqrt{\Lambda(K)} \tilde{\tilde{X}}^* \right]}} \xrightarrow{d} N(0,1)$$

□

Chapter 5

Simulation Comparison for One-Dimensional Methods

Here, we compare the performance of the various one-dimensional methods that we have introduced thus far. Many of the models were rerun with different parameters so that an accurate comparison can be made. For example, in Chapter 2, the window size w referred to two different quantities. For Method 1, it referred to the width we used on which we considered $\lambda(t)$ to be a constant.

But in Method 2, the w was the parameter that specified the distance from the point that we could consider a local average over. That is, about each point, we considered the interval $[-w, w]$. (In the case of a point near the boundary, the window was reduced accordingly.) We considered three choices of w, namely $w = 5, 10, 20$ for each of the three block sizes $b = 2, 5, 10$.

In Chapter 4, we again considered the same models and the same block sizes, but we used another parameter h to denote the proximity of the resampled block. (We cast a net on the interval $[-h, h]$ to shift the block, and in the case of no wrapping, we reduced the size accordingly.) The choices of h was made so that the requirement of $h = O(b^\alpha)$ where $\alpha > 5/2$ would be satisfied.

In order to compare our methods, we re-ran the local block bootstrap simulations using $h = 5, 10, 20$, matching those up with $b = 2, 5, 10$, respectively.

Also, in order to ensure that we can compare our methods as best as possible, the w used for Method 1 was changed to $w = 10, 20, 40$.

Since w and h share a similar purpose, it seems only natural to use the same values in both approaches. The choice of h or w has been combined into the column of "Band" that appears in the tables below. That is, Band refers to the width of the interval used to determine how "close" we are to the original point. This required the generation of an additional 70 models not included in Chapters 2 or 4.

We consider the five approaches discussed thus far: Local Block Bootstrap with Wrapping, Local Block Bootstrap without Wrapping, Block Bootstrap with Transformation using Piecewise Constant estimates for $\lambda(t)$ (Method 1), Block Bootstrap with Transformation using Local Averaging for $\lambda(t)$ (Method 2), Block Bootstrap with Transformation using True $\lambda(t)$ (Method 3).

Unlike the tables in Chapters 2 and 4, here we present only the coverage probabilities. (We omit the average number of observations as well as the mean and standard deviation of the length of the confidence intervals.)

We consider the same two choices of covariance function, the same five intensity functions, as well as three choices of b (again, $b = 2, 5, 10$) and "Band" (Band $= 10, 20, 40$) for a total of 150 models. As before, each model was simulated 1,000 times and each time, a 95% confidence interval was constructed. The coverage probability refers to the proportion of the confidence intervals that contained the true mean.

As was noted in Chapters 2 and 4, $b = 2$ seems to be the best choice for all methods, in general. When considering Table 5.1, we see that an increase in the block size results in a mild reduction (around 5%) in coverage probabilities.

In Table 5.2, however, increases in b result in more noticeable reductions. This was seen in Chapter 4 and is likely attributable to the increased dependence among the marks.

Looking closer at Table 5.1, we see that with $\lambda_1(t)$ and $\lambda_2(t)$, the local block bootstrap methods perform approximately 10% better than the two averaging transformation methods (Methods 1 and 2), and comparable to the exact transformation method (Method 3).

For $\lambda_3(t)$ and $\lambda_4(t)$, both sets of methods yield similar results, with the local block bootstrap performing slightly better. For $\lambda_5(t)$, though, there is a significant improvement (20-30%) using the local block bootstrap. The reason for this is not clear.

Similar results hold in Table 5.2 when examining $\lambda_1(t)$ and $\lambda_2(t)$. However, *significant* differences appear for $\lambda_3(t)$ and $\lambda_4(t)$, with the local block bootstrap maintaining its coverage probabilities, but the transformation method failing to work. (The differences are 30-60%!) The reason for this difference is not clear and the models were re-run and the results similar. This will be a source of future study.

The coverage with $\lambda_5(t)$ drops for all methods and the effects are more noticeable when considering a larger b. This overall reduction is more inline with expectations.

Table 5.1: Comparison of One-Dimensional Resampling Methods with $R(t) = \exp(-|t|)$

			LBB Method		Transformation Method		
$\lambda_i(t)$	b	Band	Wrap	No Wrap	Method 1	Method 2	Method 3
1	2	10	95.6%	94.8%	87.9%	83.7%	95.0%
	5	20	91.3%	91.6%	78.2%	81.4%	90.7%
	10	40	93.7%	92.8%	74.5%	74.4%	87.0%
2	2	10	95.6%	95.2%	85.4%	73.7%	94.9%
	5	20	93.1%	92.8%	71.2%	82.7%	93.1%
	10	40	87.2%	88.4%	67.3%	67.9%	92.0%
3	2	10	94.4%	95.7%	95.1%	93.7%	93.5%
	5	20	88.9%	91.2%	86.5%	88.6%	86.1%
	10	40	92.4%	93.5%	86.4%	85.7%	84.8%
4	2	10	94.4%	95.2%	93.1%	94.4%	93.5%
	5	20	89.8%	90.8%	88.1%	87.8%	88.9%
	10	40	85.0%	83.8%	83.8%	87.5%	85.5%
5	2	10	88.8%	89.6%	57.8%	61.5%	65.6%
	5	20	80.7%	85.0%	58.1%	61.4%	66.9%
	10	40	76.7%	78.7%	58.7%	59.8%	66.3%

Table 5.2: Comparison of One-Dimensional Resampling Methods with $R(t) = \exp(-|t|/3)$

			LBB Method		Transformation Method		
$\lambda_i(t)$	b	Band	Wrap	No Wrap	Method 1	Method 2	Method 3
1	2	10	93.2%	94.2%	87.8%	84.2%	88.3%
	5	20	80.5%	83.7%	74.9%	73.4%	78.3%
	10	40	74.2%	72.1%	65.6%	62.6%	66.9%
2	2	10	93.2%	94.9%	92.6%	90.2%	87.4%
	5	20	85.5%	87.0%	80.7%	81.7%	77.5%
	10	40	74.2%	73.9%	67.7%	70.6%	72.0%
3	2	10	92.2%	94.1%	50.3%	45.8%	47.2%
	5	20	82.3%	81.7%	46.5%	46.1%	47.1%
	10	40	73.6%	80.7%	45.7%	45.6%	43.9%
4	2	10	94.1%	94.1%	36.4%	38.4%	36.6%
	5	20	82.3%	83.8%	35.2%	33.3%	35.5%
	10	40	71.9%	72.4%	32.7%	34.9%	31.3%
5	2	10	84.4%	86.2%	53.4%	51.4%	51.6%
	5	20	72.7%	70.8%	53.6%	52.2%	53.5%
	10	40	66.1%	64.4%	48.7%	49.7%	50.4%

Chapter 6

Theory for Higher Dimensional Inhomogeneous Poisson Processes

6.1 Introduction

In Chapter 3, we established the theory necessary for a Central Limit Theorem for an Inhomogeneous Marked Poisson Process. Now, we turn our attention to generalizing the results to higher dimensions.

This is fairly straightforward, though, as the main theorems cited in Chapter 3 also hold for higher dimensions. Recall, our goal is to estimate the mean of the random process, $\mu = E[X(\mathbf{t})]$ on the basis of observing $\{X(\mathbf{t})\}$ for the points generated by an inhomogeneous Poisson process N over a compact subset $K \in \mathbb{R}^d$. There are two natural ways to estimate μ:

$$\tilde{X}_K = \frac{1}{\Lambda(K)} \int_K X(\mathbf{t}) N(d\mathbf{t})$$

and

$$\bar{X}_K = \frac{1}{N(K)} \int_K X(\mathbf{t}) N(d\mathbf{t})$$

Note: The only difference between the two formulas is the division by either the expected or actual sample size, respectively.

Karr [30] explored the case where N is a homogeneous Poisson process and established the fact that the above estimators are consistent and asymptotically normal at rate $\sqrt{|K|}$ with the same asymptotic variance. Again, we note that the statement is not correct, as was pointed out in Chapter 3. Here, $|\cdot|$ denotes the Lebesgue measure (volume).

We state the correted, multivariate form of Karr's Theorem below (this is analogous to Theorem 3.1.1, but has been updated to reflect the differing variances discussed in Chapter 3) as it is the springboard for our extension to the inhomogeneous setting. We should note that the statements pertaining to $\bar{X}_K$ are conjectured based upon the one-dimensional results. The work done by Brillinger [5] only applied to one-dimensional data. As we shall see below, there is a general central limit theorem by Jenish and Prucha [29] which would give us asymptotic normality. However, this would require some additional work to use.

Theorem 6.1.1 (Karr's Theorem). *Let* $R(\boldsymbol{t}) = Cov(X(0), X(\boldsymbol{t}))$ *and assume that* $\int R(\boldsymbol{t})dt < \infty$

Also, let $$\frac{1}{\sqrt{|K|}} \int_K (X(\boldsymbol{t}) - \mu)dt \xrightarrow{d} N\left(0, \int R(\boldsymbol{t})dt\right)$$

as $diam(K) \to \infty$. *Then, as* $diam(K) \to \infty$, *we have*

$$\sqrt{|K|}\left(\tilde{X}_K - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right) \text{ and } \sqrt{|K|}\left(\bar{X}_K - \mu\right) \xrightarrow{d} N\left(0, \phi^2\right)$$

and furthermore,

$$\lim_{diam(K)\to\infty} \text{var}\left(\sqrt{|K|}\tilde{X}_K\right) = \sigma^2 \equiv \int R(\boldsymbol{t})dt + \frac{R(0) + \mu^2}{\lambda}$$

and

$$\lim_{diam(K)\to\infty} \text{var}\left(\sqrt{|K|}\bar{X}_K\right) = \phi^2 \equiv \int R(\boldsymbol{t})dt + \frac{R(0)}{\lambda}$$

Here, $diam(K)$ denotes the supremum of the diameters of all l_∞ balls contained in K.

The goal of this chapter is to extend Karr's Theorem [30] to the inhomogeneous setting. We shall mimic many of the proof techniques that are employed by Politis, Paparoditis and Romano [53]. To that end, we begin with some background on mixing coefficients.

6.2 Higher Dimensional Mixing Coefficients

We shall assume that our random field $\{X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d\}$ satisfies a certain weak dependence condition that will be quantified in terms of mixing coefficients.

Let $\rho(\cdot, \cdot)$ denote the distance in the l_∞-norm on $\mathbb{R}^d$. The strong mixing coefficients of Rosenblatt [67] are then defined as:

$$\alpha_X(k) \equiv \sup_{E_1,E_2\subset\mathbb{R}^d} \{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), i = 1, 2, \rho(E_1, E_2) \geq k\}$$

where $\mathcal{F}(E_i)$ is the σ-algebra generated by $\{X(\mathbf{t}), \mathbf{t} \in E_i\}$.

In a similar manner, Doukhan [17] defined mixing conditions that also depend on the size of the sets considered.

$$\alpha_X(k; l_1, l_2) \equiv \sup_{E_1,E_2\subset\mathbb{R}^d} \left\{ \begin{array}{l} |P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), \\ \\ |E_i| \leq l_i, i = 1, 2, \rho(E_1, E_2) \geq k \end{array} \right\}$$

Notice that $\alpha_X(k; l_1, l_2) \leq \alpha_X(k)$, and that in essence $\alpha_X(k) = \alpha_X(k; \infty, \infty)$.

Definition 6.2.1. A random field is said to be ***α*-strong mixing** if $\lim_{k\to\infty} \alpha_X(k) = 0$.

Definition 6.2.2. We shall define a weaker set of mixing called **$\bar{\alpha}$-strong mixing**. Define

$$\begin{aligned} \bar{\alpha_X}(k; l) \quad \equiv \quad & \sup\{|P(A_1 \cap A_2) - P(A_1)P(A_2)| : A_i \in \mathcal{F}(E_i), i = 1, 2, \\ & E_2 = E_1 + \mathbf{t}, |E_1| \leq l, \rho(E_1, E_2) \geq k\} \end{aligned}$$

where the supremum is taken over all compact and convex sets $E_1 \subset \mathbb{R}^d$, and over all $\mathbf{t} \in \mathbb{R}^d$ such that $\rho(E_1, E_1 + \mathbf{t}) \geq k$. Also, we define $\bar{\alpha}_X(k) = \bar{\alpha}_X(k; \infty)$.

Notice that $\bar{\alpha_X}(k) \leq \alpha_X(k)$, so that if the random field is α-strong mixing, then it will necessarily be $\bar{\alpha}$-strong mixing as well, and thus $\lim_{k\to\infty} \bar{\alpha}_X(k) = 0$.

More discussion and references on strong mixing coefficients can be found in Doukhan [17], Roussas and Ioannides [69], and Leonenko and Ivanov [41].

6.3 Assumptions

For our theoretical results, we make an assumption regarding the process $X(\mathbf{t})$ and an assumption regarding the mixing coefficients.

A.6.1. $0 < \lambda_{min} \leq \lambda(\mathbf{t}) \leq \lambda_{max}$ (That is, $\lambda(\mathbf{t})$ is bounded above and below.)

A.6.2. $X(\mathbf{t})$ is wide-sense stationary with $E[|X(\mathbf{t})|^{2+\delta}] < \infty$ for some $\delta > 0$

A.6.3. $\bar{\alpha}(k; 0) < const.k^{-d-\varepsilon}$, where $\varepsilon > \dfrac{2d}{\delta}$ for the δ specified in Assumption A.6.2

6.4 Main Results

We shall begin by showing that our estimators from Section 6.1 are unbiased for μ.

$$\begin{aligned}
E\left[\tilde{X}_K\right] &= \frac{1}{\Lambda(K)} E\left[\int_K X(\mathbf{t}) N(d\mathbf{t})\right] \\
&= \frac{1}{\Lambda(K)} E\left[E\left[\int_K X(\mathbf{t}) N(d\mathbf{t}) \middle| N\right]\right] \\
&= \frac{1}{\Lambda(K)} E\left[\int_K E\left[X(\mathbf{t})\right] N(d\mathbf{t})\right] \\
&= \frac{1}{\Lambda(K)} E\left[\int_K \mu N(d\mathbf{t})\right] \\
&= \frac{1}{\Lambda(K)} E\left[\mu N(K)\right] \\
&= \mu \\
E\left[\bar{X}_K\right] &= E\left[\frac{1}{N(K)} \int_K X(\mathbf{t}) N(d\mathbf{t})\right] \\
&= E\left[\frac{1}{N(K)} E\left[\int_K X(\mathbf{t}) N(d\mathbf{t}) \middle| N\right]\right] \\
&= E\left[\frac{1}{N(K)} \int_K E\left[X(\mathbf{t})\right] N(d\mathbf{t})\right] \\
&= E\left[\frac{1}{N(K)} \int_K \mu N(d\mathbf{t})\right] \\
&= E\left[\frac{1}{N(K)} \mu N(K)\right] \\
&= \mu
\end{aligned}$$

Note: In both calculations above, we used the stationarity of $X(\mathbf{t})$. Also, the above is nearly identical to the derivation in Section 3.4. For this reason, the formulas for the variance of our estimators should not come as a surprise.

To compute the variance of $\tilde{X}_K$, we begin by considering the following:

$$E\left[\int_K\int_K X(\mathbf{t})N(d\mathbf{t})X(\mathbf{s})N(d\mathbf{s})\right] = \int_K\int_K E\left[X(\mathbf{t})N(d\mathbf{t})X(\mathbf{s})N(d\mathbf{s})\right]$$
$$= \int_K\int_K E\left[X(\mathbf{t})X(\mathbf{s})\right]E\left[N(d\mathbf{t})N(d\mathbf{s})\right]$$

where the second equality follows from the assumed independence of $X(\cdot)$ and $N(\cdot)$.

Observe, though, that for a wide-sense stationary process, we have:

$$E\left[X(\mathbf{t})X(\mathbf{s})\right] = \mathrm{cov}(X(\mathbf{t}), X(\mathbf{s})) + E[X(\mathbf{t})]E[X(\mathbf{s})] = R(\mathbf{s}-\mathbf{t}) + \mu^2$$

and

$$E\left[N(dt)N(ds)\right] = \begin{cases} E\left[N(d\mathbf{t})\right]E[N(d\mathbf{s})] & \text{if } \mathbf{t} \neq \mathbf{s} \\ \left(E\left[N(d\mathbf{t})\right]\right)^2 & \text{if } \mathbf{t} = \mathbf{s} \end{cases}$$
$$= \begin{cases} \lambda(\mathbf{t})\lambda(\mathbf{s})d\mathbf{t}d\mathbf{s} & \text{if } \mathbf{t} \neq \mathbf{s} \\ \left(\lambda(\mathbf{t})d\mathbf{t}\right)^2 + \lambda(\mathbf{t})d\mathbf{t} & \text{if } \mathbf{t} = \mathbf{s} \end{cases}$$

This follows since $N(d\mathbf{t})$ is a Poisson$(\lambda(\mathbf{t})d\mathbf{t})$ random variable. Thus, its first moment is equal to $\lambda(\mathbf{t})d\mathbf{t}$ and its second moment is equal to square of the first moment plus the variance, i.e. $(\lambda(\mathbf{t})d\mathbf{t})^2 + \lambda(\mathbf{t})d\mathbf{t}$.

Thus, we have that

$$E\left[\int_K\int_K X(\mathbf{t})N(d\mathbf{t})X(\mathbf{s})N(d\mathbf{s})\right] = \int_K\int_K \left(R(\mathbf{s}-\mathbf{t}) + \mu^2\right)\lambda(\mathbf{t})\lambda(\mathbf{s})d\mathbf{t}d\mathbf{s}$$
$$+ \int_K \left(R(\mathbf{0}) + \mu^2\right)\lambda(\mathbf{t})d\mathbf{t}$$

And so, it follows that

$$
\begin{aligned}
\mathrm{var}(\tilde{X}_K) &= \frac{1}{[\Lambda(K)]^2}\mathrm{var}\left(\int_K X(\mathbf{t})N(d\mathbf{t})\right) \\
&= \frac{1}{[\Lambda(K)]^2}\left(\begin{array}{l} E\left[\int_K\int_K X(\mathbf{t})N(d\mathbf{t})X(\mathbf{s})N(d\mathbf{s})\right] \\ -E\left[\int_K X(\mathbf{t})N(d\mathbf{t})\right]E\left[\int_K X(\mathbf{s})N(d\mathbf{s})\right]\end{array}\right) \\
&= \frac{1}{[\Lambda(K)]^2}\left(\begin{array}{l}\int_K\int_K \left(R(\mathbf{s}-\mathbf{t})+\mu^2\right)\lambda(\mathbf{t})\lambda(\mathbf{s})d\mathbf{t}d\mathbf{s} \\ +\int_K \left(R(\mathbf{0})+\mu^2\right)\lambda(\mathbf{t})d\mathbf{t}-\mu^2[\Lambda(K)]^2\end{array}\right) \\
&= \frac{1}{[\Lambda(K)]^2}\left(\int_K\int_K R(\mathbf{s}-\mathbf{t})\lambda(\mathbf{t})\lambda(\mathbf{s})d\mathbf{t}d\mathbf{s}+\int_K \left(R(\mathbf{0})+\mu^2\right)\lambda(\mathbf{t})d\mathbf{t}\right)
\end{aligned}
$$

The same issues that we saw in Chapter 3 also appear here. That is, $\bar{X}_K$ has a different asymptotic variance than $\tilde{X}_K$. But unlike in Chapter 3, we cannot use the transformation method to establish asymptotic normality. However, Theorem 1 of [29] gives us a central limit theorem for random fields with irregularly spaced observations. (Note: Assumptions A.6.2 and A.6.3 are stronger than Assumptions 2 and 3 needed for the theorem.)

The only concern with the application of the theorem is with their Assumption 1, which requires that points be at least ρ_0 apart. This assumption would be violated with Poisson data. Despite this fact, we conjecture that the set of points for which this minimal distance is violated does not contribute an appreciable amount to the sum (over all of the points). This may require additional assumptions on $\lambda(t)$ and will be the source of further investigation. (It is worth noting that in one dimension, [83] gives an expression for the distribution of the inter-arrival times for an inhomogeneous Poisson process, so we might quantify how many points are likely to violate the condition.)

Our last goal is to establish a central limit theorem. Suppose that $X(\cdot)$ is a wide-sense stationary process with $E[X(\mathbf{t})^2] < \infty$ for each $\mathbf{t}$. Also suppose that the covariance function $R(\mathbf{t}) = Cov(X(\mathbf{0}), X(\mathbf{t}))$ is absolutely integrable, i.e.

$$\int |R(\mathbf{t})|\, d\mathbf{t} < \infty. \tag{6.4.1}$$

where $\int$ is short-hand for $\int_{\mathbb{R}^d}$. Also, let

$$\frac{1}{\sqrt{\Lambda(K)}} \int_K (X(\mathbf{t}) - \mu)\lambda(\mathbf{t})d\mathbf{t} \xrightarrow{d} N\left(0, \theta^2\right) \tag{6.4.2}$$

as $diam(K) \to \infty$.

Similar to the one-dimensional case, θ^2 is the asymptotic variance of

$$\frac{1}{\sqrt{\Lambda(K)}} \int_K (X(t) - \mu)\lambda(t)dt$$

So, we have that $\theta^2 = \lim_{diam(K)\to\infty} \frac{1}{\Lambda(K)} \int_K \int_K R(\mathbf{s} - \mathbf{t})\lambda(\mathbf{s})\lambda(\mathbf{t})d\mathbf{s}d\mathbf{t}$.

By Assumption A.6.1, we have that

$$\begin{aligned} \frac{1}{\Lambda(K)} \int_K \int_K R(s-t)\lambda(s)\lambda(t)dsdt &\leq \frac{(\lambda_{max})^2}{\Lambda(K)} \int_K \int_K |R(s-t)|\, dsdt \\ &\leq \frac{(\lambda_{max})^2|K|}{\Lambda(K)} \int |R(u)|\, du \end{aligned}$$

By equation (6.4.1), we have that this integral is finite, so we are assured that $\theta^2 < \infty$.

Lemma 6.4.1. *If there exists a $\delta > 0$ such that Assumptions A.6.2 and A.6.3 hold, then equations (6.4.1) and (6.4.2) hold.*

Note: Different sufficient conditions for equations (6.4.1) and (6.4.2) can be found in Yadrenko [82].

Theorem 6.4.2. *Assuming equations (6.4.1) and (6.4.2), then as $diam(K) \to \infty$, we have that*

$$\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right)$$

and

$$\sqrt{\Lambda(K)}\left(\bar{X}_K - \mu\right) \xrightarrow{d} N\left(0, \phi^2\right),$$

where $\phi^2 = \theta^2 + R(\mathbf{0})$ and $\sigma^2 = \phi^2 + \mu^2$.

The similarity of this result to Karr's Theorem lies with Campbell's Theorem (Theorem 3.4.2) which is applicable in a multivariate setting with an inhomogeneous point process. As was the case with Theorem 6.1.1, the results pertaining to $\bar{X}_K$ are only a conjecture at this point. The same subtle issues from Chapter 3, but unlike before, we cannot use a transformation to side-step the issue. This will be a source of further study.

6.5 Proofs

Proof of Lemma 6.4.1: We appeal to Theorem 1.7.1 of Leonenko and Ivanov [41] where they employ a weaker notion than our $\bar{\alpha}$-mixing.

We only need to verify the finiteness of $\int |R(\mathbf{t})|\, d\mathbf{t} < \infty$, as that is enough to ensure that the limiting variance of $\frac{1}{\sqrt{\Lambda(K)}} \int_K (X(\mathbf{t}) - \mu)\lambda(\mathbf{t}) d\mathbf{t}$ is finite.

A result in Roussas and Ioannides [69] (which requires Assumption A.6.2) states

$$|\mathrm{cov}(X(\mathbf{0}), X(\mathbf{t})| \leq \mathrm{const.} \bar{\alpha}_X(\max_i |t_i|\,; 0)^{1-2/(2+\delta)}$$

Thus, we have that

$$\begin{aligned}
\int |R(\mathbf{t})|\, d\mathbf{t} &= O\left(\int \bar{\alpha}_X(\max_i |t_i|\,; 0)^{1-2/(2+\delta)} d\mathbf{t} \right) \\
&= O\left(\int_0^\infty y^{d-1} \left(\frac{1}{y^{d+\varepsilon}} \right)^{1-2/(2+\delta)} dy \right) \\
&= O\left(\int_0^\infty \frac{1}{y^{1-2d/(2+\delta)+\varepsilon(1-2/(2+\delta))}} dy \right) < \infty
\end{aligned}$$

The first line follows from the inequality from Roussas and Ioannides.

The second line follows by letting $y = \max_i |t_i|$, using the bound on the mixing coefficients (Assumption A.6.3), and reducing the multiply integrals to a single integral (since this is the pdf of the maximum of d i.i.d. random variables).

The third line is simple algebra and the finiteness comes from the fact that in the denominator

$$\begin{aligned}
y^{1-2d/(2+\delta)+\varepsilon(1-2/(2+\delta))} &> y^{1-\varepsilon\delta/(2+\delta)+\varepsilon-2\varepsilon/(2+\delta)} \\
&= y^{1+\varepsilon(1-(\delta+2)/(2+\delta))} \\
&= y^{1+\varepsilon} \\
&> y
\end{aligned}$$

□

Proof of Theorem 6.4.2: As was mentioned above, we shall focus on establishing results for $\tilde{X}_K$ and leave $\bar{X}_K$ as a conjecture. Our strategy is to show the characteristic function of $\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right)$ converges to that of a $N(0, \sigma^2)$ random variable. The proof is nearly identical to that of Theorem 3.4.4.

$$
\begin{aligned}
&E\left[\exp\left(i\alpha\sqrt{\Lambda(K)}\left(\tilde{X}_K-\mu\right)\right)\right]\\
&=E\left[E\left[\exp\left(i\alpha\sqrt{\Lambda(K)}\left(\tilde{X}_K-\mu\right)\right)\Big|X\right]\right]\\
&=E\left[E\left[\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K X(\mathbf{t})N(d\mathbf{t})-\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K \mu\lambda(\mathbf{t})d\mathbf{t}\right)\Big|X\right]\right]\\
&=E\left[\exp\left(\int_K\left(\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}X(t)\right)-1\right)\lambda(t)dt-\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K\mu\lambda(t)dt\right)\right]
\end{aligned}
$$

The second line follows from the properties of iterated expectation.

The third line follows by substituting $\tilde{X}_K$, using the fact that $\mu=\dfrac{1}{\sqrt{\Lambda(K)}}\displaystyle\int_K\mu\lambda(\mathbf{t})d\mathbf{t}$, and distributing $i\alpha\sqrt{\Lambda(K)}$ through both terms inside of the parentheses.

The fourth line follows by applying Campbell's Theorem (Theorem 3.4.2) to the first term and recognizing that the second term is a constant with respect to the inner expectation.

From here, consider a Taylor Series expansion of $\displaystyle\int_K\left(\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}X(\mathbf{t})\right)-1\right)\lambda(\mathbf{t})d\mathbf{t}$

Recall, $e^x=1+x+x^2/2+O(x^3)$. Thus, $e^x-1\approx x+x^2/2$. It should be noted that the error term converges in probability to 0. Thus, we have:

$$
\begin{aligned}
&\int_K\left(e^{\frac{i\alpha}{\sqrt{\Lambda(K)}}X(\mathbf{t})}-1\right)\lambda(\mathbf{t})d\mathbf{t}\\
&\approx\int_K\frac{i\alpha}{\sqrt{\Lambda(K)}}X(\mathbf{t})\lambda(\mathbf{t})d\mathbf{t}+\int_K\frac{1}{2}\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}X(\mathbf{t})\right)^2\lambda(\mathbf{t})d\mathbf{t}\\
&=\int_K\frac{i\alpha}{\sqrt{\Lambda(K)}}X(\mathbf{t})\lambda(\mathbf{t})d\mathbf{t}+\int_K\frac{i^2\alpha^2}{2\Lambda(K)}X^2(\mathbf{t})\lambda(\mathbf{t})d\mathbf{t}\\
&=\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K X(\mathbf{t})\lambda(\mathbf{t})d\mathbf{t}-\frac{\alpha^2}{2\Lambda(K)}\int_K X^2(\mathbf{t})\lambda(\mathbf{t})d\mathbf{t}
\end{aligned}
$$

And so, we have that

$$
\begin{aligned}
&E\left[\exp\left(\int_K\left(e^{\frac{i\alpha}{\sqrt{\Lambda(K)}}X(\mathbf{t})}-1\right)\lambda(\mathbf{t})d\mathbf{t}-\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K\mu\lambda(\mathbf{t})d\mathbf{t}\right)\right]\\
&\approx E\left[\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K\left[X(\mathbf{t})-\mu\right]\lambda(\mathbf{t})d\mathbf{t}-\frac{\alpha^2}{2\Lambda(K)}\int_K X^2(\mathbf{t})\lambda(\mathbf{t})d\mathbf{t}\right)\right]
\end{aligned}
$$

Recall, we assumed that $\dfrac{1}{\sqrt{\Lambda(K)}}\displaystyle\int_K(X(\mathbf{t})-\mu)\lambda(\mathbf{t})d\mathbf{t}\xrightarrow{d}N\left(0,\theta^2\right)$. In terms of characteristic functions, we have

$$\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K (X(\mathbf{t})-\mu)\lambda(\mathbf{t})d\mathbf{t}\right)$$
$$= \exp\left(i\alpha\left(\frac{1}{\sqrt{\Lambda(K)}}\int_K (X(\mathbf{t})-\mu)\lambda(\mathbf{t})d\mathbf{t}\right)\right)$$
$$\xrightarrow{d} e^{-\frac{\alpha^2}{2}\theta^2}$$

By consistency,

$$\frac{1}{\Lambda(K)}\int_K X^2(\mathbf{t})\lambda(\mathbf{t})d\mathbf{t} \to E[X(\mathbf{t})^2] = Var[X(\mathbf{t})] + (E[X(\mathbf{t})])^2 = R(\mathbf{0}) + \mu^2$$

Thus,

$$E\left[\exp\left(\frac{i\alpha}{\sqrt{\Lambda(K)}}\int_K [X(t)-\mu]\,\lambda(t)dt - \frac{\alpha^2}{2\Lambda(K)}\int_K X^2(t)\lambda(t)dt\right)\right]$$
$$\to \exp\left(-\frac{\alpha^2}{2}\theta^2 - \frac{\alpha^2}{2}(R(0)+\mu^2)\right)$$

So, we have that $\sqrt{\Lambda(K)}\left(\tilde{X}_K - \mu\right) \xrightarrow{d} N\left(0, \sigma^2\right)$, where $\sigma^2 = \theta^2 + R(0) + \mu^2$.

□

Chapter 7

Local Block Bootstrap for Higher Dimensional Inhomogeneous Poisson Processes

7.1 Introduction

In Chapter 6, we established the asymptotic normality of the sample mean. In order to construct confidence intervals for the mean μ, though, the asymptotic variance would need to be explicitly estimated. Difficulties arise when trying to consistently estimate $\int R(\mathbf{t})d\mathbf{t}$ as we have irregularly spaced data.

The resampling method discussed in the next section is able to yield confidence intervals for the mean without the need for explicit estimation of the asymptotic variance. Alternatively, the resampling method may provide an estimate of the asymptotic variance to be used in connection with the asymptotic normality result in Chapter 6.

Many of the results are natural extensions of the one-dimensional local block bootstrap developed in Chapter 4. For illustrative purposes, work in two dimensions will be carried out. The theory extends readily to higher dimensions and in those instances, formulas are stated without rigorous proof.

From a practical point of view, one needs to verify the assumptions before resampling the data. In particular, the stationarity asusmption may be violated. The greater the nonstationarity, the less reliable the resampling method will be. If our data was from a homogeneous Poisson process, then we use Ripley's K-function (see [65], [15], and [2]) to assess the Poisson assumption.

However, the inhomogeneous assumption permits many situations where the data will

fail for such tests. Baddeley, Møller, and Waagepetersen [1] extend Ripley's K-function Tests to the inhomogeneous setting and their paper serves as the reference on the subject.

7.2 Local Block Bootstrap Algorithm in Higher Dimensions

Our observation region K can be any compact set in $\mathbb{R}^d$. For the sake of simplicity, though, we suppose that K is a d-dimensional "box" given by $[0, K_1] \times [0, K_2] \times \ldots \times [0, K_d]$. For our asymptotic results, K will be assumed to expand uniformly in all dimensions towards ∞.

Suppose we obtain observations $\tau_1, \tau_2, \ldots, \tau_{N(K)}$ from an inhomogeneous Poisson process with intensity function $\lambda(\mathbf{t})$. And suppose that for each τ_i, we observe a corresponding mark, $X(\tau_i)$. Together, this gives us an inhomogeneous marked point process.

We shall employ a local block bootstrap method to resample such data, as was introduced in [51]. Essentially, we resample the data in blocks (just like a block bootstrap), but when filling a particular block, we only consider blocks that are in a "local" neighborhood of the original block. The block size itself is a parameter while the proximity to the original block is determined by a second parameter.

The main difference in higher dimensions is the fact that a neighboring block need not be on the left or right. In essence, for each block, we have a "box" that constitutes all nearby regions in $\mathbb{R}^d$.

Also, the transformation method mentioned in Chapter 2 cannot be used here, as the critical component to that method relied upon the fact that the cumulative intensity function $\Lambda(t)$ was invertible. Thus, we could identify points in the original (inhomogeneous space) or the transformed (homogeneous space).

Thus, this approach appears to be the only method available to resample inhomogeneous Poisson process data in higher dimensions.

The local block bootstrap algorithm for generating $X^*(\tau_1), X^*(\tau_2), \ldots, X^*(\tau_{N^*(K)})$ will necessarily depend on the bootstrap point process, N^*. In particular, we have that $\Lambda^*(K) = E^*[N^*(K)]$. Our algorithm can be described as follows:

1. For each dimension ($i = 1, 2, \ldots, d$), select a block size b_i such that $b_i^{5/2}/K_i \to 0$ as $K_i \to \infty$

   Note: This is the same as requiring $diam(K) \to \infty$.

2. Choose an $\alpha > 2$.

3. For each dimension ($i = 1, 2, \ldots, d$), define the bandwidth parameter to be $h_i = O(b_i^\alpha)$

4. Define the total number of blocks to be $L = \prod_{i=1}^{d} \frac{K_i}{b_i}$

5. For each of the $j = 1, 2, \ldots, L$ blocks, denote the "bottom-left" (thinking in $\mathbb{R}^2$) coordinates of the block as c_j.

6. Add a perturbation d_j to c_j where the components of d_j are i.i.d. Uniform$[-h_i, h_i]$ random variables.

7. Form a new block of size $b_1 \times b_2 \times \cdots \times b_d$ starting at $c_j + d_j$ and record which of the $X(\tau_1)$, $\ldots$, $X(\tau_{N(K)})$ occur in this block.

   Note: It is possible that parts of the block may lay outside of K. To correct this, imagine that K is "wrapped around" on itself. That is, our calculations are done modulo K.

8. Let
   $\tilde{X}^* = \frac{1}{\Lambda^*(K)} \sum_{i=1}^{N(K)} W_i X(\tau_i)$, where W_i is the number of times that $X(\tau_i)$ occurs in the resampled data

   and
   $\bar{X}^* = \frac{1}{N^*(K)} \sum_{i=1}^{N(K)} W_i X(\tau_i)$
   Also, let
   $\tilde{\tilde{X}}^* = \frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} W_i X(\tau_i)$
   and
   $\bar{\bar{X}}^* = \frac{1}{N(K)} \sum_{i=1}^{N(K)} W_i X(\tau_i)$

   This generation of points $X^*(\tau_1), ..., X^*(\tau_{N^*(K)})$ and subsequently of $\tilde{X}^*$, $\bar{X}^*$, $\tilde{\tilde{X}}^*$, and $\bar{\bar{X}}^*$ is governed by a probability mechanism which we will denote by P^*, with moments denoted by E^*, Var^*, etc. This generation is done conditionally on the marked point process data observed; thus P^* is really a conditional probability. Notice that the first two estimates involve the bootstrap point process N^*, while the last two do not.

9. Let $P^*(\sqrt{\Lambda(K)}(\tilde{\tilde{X}}^* - E^*[\tilde{\tilde{X}}^*]) \leq x)$ and $P^*(\sqrt{\Lambda(K)}(\bar{\bar{X}}^* - E^*[\bar{\bar{X}}^*]) \leq x)$ denote the conditional (given the marked point process data) distribution functions of the bootstrap sample means.

Note: We will not concern ourselves with the trivial matter of divisibility, and issues like K_i/b_i being an integer. The reason for this is that for a practical application with a finite sample, we can truncate K_i and obtain perfect divisibility. As for the asymptotic case, we can always ignore truncations which are clearly of negligible order.

7.3 Assumptions

For our bootstrap results, we need to impose some various restrictions on the process $X(t)$, the mixing coefficients, and the parameters.

A.7.1. $0 < \lambda_{min} \leq \lambda(\mathbf{t}) \leq \lambda_{max}$

A.7.2. $X(t)$ is wide-sense stationary with $E[|X(\mathbf{t})|^{6+\delta}] < \infty$ for some $\delta > 0$

A.7.3. $\bar{\alpha}(k;\mathbf{0}) < const.k^{-1-\varepsilon}$, where $\varepsilon > \max\left\{\frac{2}{\delta}, \frac{8+\delta}{4+\delta}\right\}$ for the δ specified in Assumption A.7.2

A.7.4. $\int |Q(u,v,v-w)|\,dv = C_Q < \infty$, for all u, w

A.7.5. $h_i/K_i \to 0$ as $K_i \to \infty$ for $i = 1, 2, \ldots d$.

A.7.6. $b_i^{5/2}/K_i \to 0$ as $K_i \to \infty$ for $i = 1, 2, \ldots d$.

Notice that these assumptions are nearly identical to the one-dimensional case. We require the same relationships between b_i, h_i and K_i in each coordinate that existed in one dimension.

7.4 Main Results

Since the resampling process can be difficult to keep track of, it will help to consider examples (with illustrations). A few words are in order, though.

First, the α used in these examples is less than 2, but that does not affect the theory that we develop. Indeed, the results would hold for any $\alpha > 1$, but as we will see later, in order to ensure that our bootstrap variance tends to the true variance, we need to impose an additional restriction on α.

Second, the examples used have $h = \lceil b^{\alpha-1} \rceil\, b$. (Note that this is of order b^α, as required.) The reason for this is for the ease of the display of the probability calculations. While true in a more general setting (that is, $h = const.b^\alpha$), by requiring h to be an integer multiple of b, we have some nice geometric interpretations for many of the formulas.

Our goal is fairly straightforward: We want to show that our bootstrap mean and variance are consistent for their real-world counterparts and that we have a corresponding Central Limit Theorem in the bootstrap world.

Suppose we begin with a region $K = [0, K_1] \times [0, K_2]$ and we define $h_i = \lceil b_i^{\alpha-1} \rceil\, b_i$ for $\alpha > 1$. Let $b_1 = 3$, $b_2 = 2$ and $\alpha = 3/2$, so that $h_1 = 6$ and $h_2 = 4$.

Our region may look like the following:

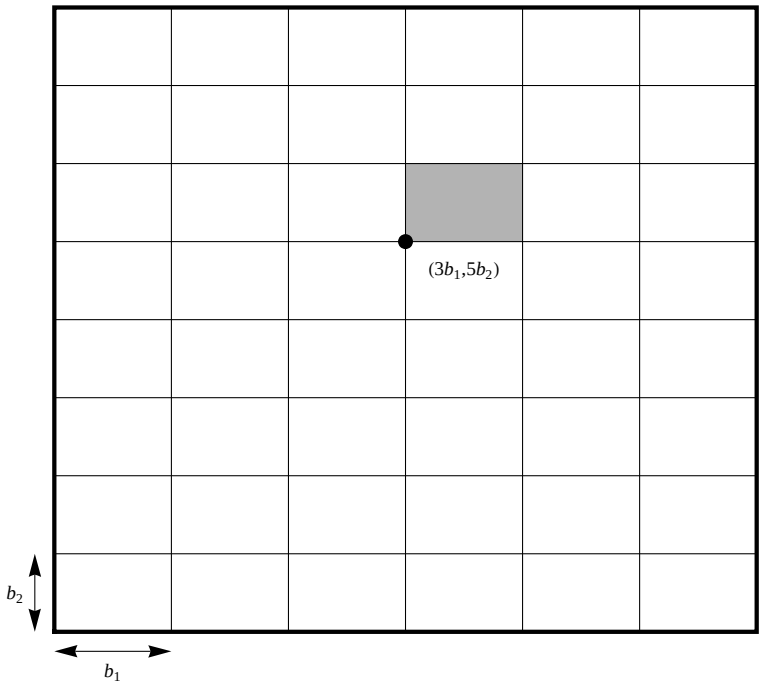

Figure 7.1: Two-Dimensional Region With Shaded Block

Each block size has dimensions b_1 and b_2. There are a total of $L = 48$ blocks in Figure 7.1. For the sake of identifying the blocks, we shall suppose that the bottom-left corner of K is the origin and that each block is indexed by its lower-left coordinate.

For example, the shaded block above is located at $(3b_1, 5b_2)$. To fill this block, we shall chose a block of the same size uniformly from the original data with a lower-left coordinate "close" to $(3b_1, 5b_2)$. That is, in each coordinate, we choose a number uniformly on $[-h_i, h_i]$ by which to perturb the point $(3b_1, 5b_2)$. Some possible examples are drawn in Figure 7.2.

There is a potential problem with edge effects, when we cannot choose a point uniformly to the left or right. We can deal with that edge effect by using either a reflection or torodial wrapping. For the reflection technique, we reflect the values on the side we do have and then draw uniformly from our slightly extended interval. For torodial wrapping, we think about gluing the bottom left corner to the top-right, so if we spill over, we end up on the opposite side of the region. For our purposes, we shall use torodial wrapping.

Recall that $\tilde{X}^* = \dfrac{1}{\Lambda^*(K)} \displaystyle\sum_{i=1}^{N(K)} W_i X(\tau_i)$, where W_i is the number of times that $X(\tau_i)$ occurs in the resampled data.

We can think of W_i as a sum of Bernoulli random variables, where the sum runs over all L blocks. That is, $W_i = \displaystyle\sum_{j=1}^{L} Y_{ij}$, where $Y_{ij} \sim \text{Bernoulli}(p_{ij})$. Here, p_{ij} represents the probability

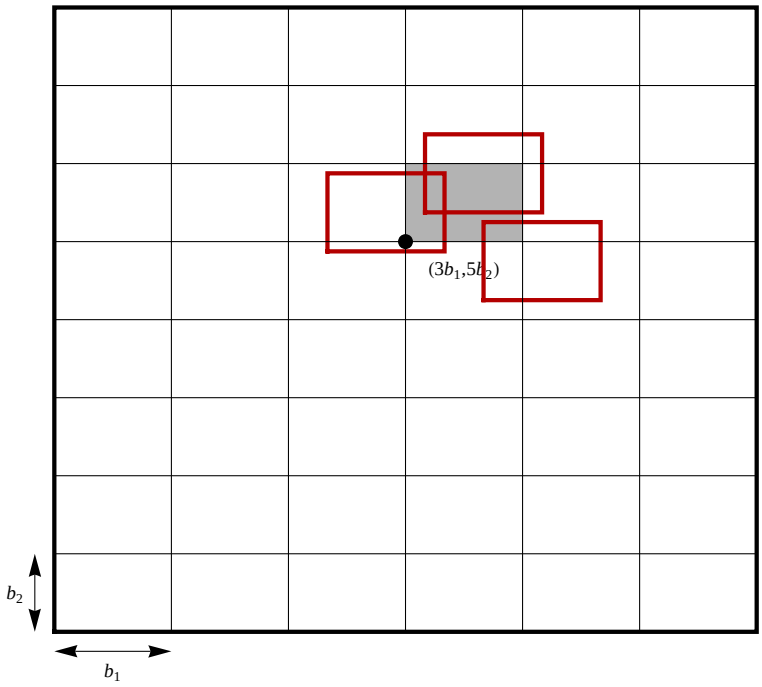

Figure 7.2: Two-Dimensional Region With Possible Resampled Blocks

that τ_i is contained in block j. Recall, τ_i denotes the position of the mark of the ith datapoint, while $X(\tau_i)$ is the mark at τ_i.

As was the case in Chapter 4, we make the following claim.

Lemma 7.4.1. $E^*[W_i] = 1$.

Again, the reason this is the case is because of our use of torodial wrapping. It will help to see this with some examples, though.

As a first step, we need to determine probabilities that a particular point, τ_i, will be included in a particular block, j. That is, we are looking for p_{ij}. Consider Figure 7.3.

Recall that each block is determined by its lower-left coordinate. Consider the point P. It defines the pink shaded block. Notice that the point P, like the point τ_1, can only move left or right by at most h_1 and up or down by at most h_2. (The possible locations of point P are shown with a thick black line while the possible locations for the point τ_1 are shown with a dashed line.) Where the point P ends up determines what block we will use to fill in the currently shaded block.

Notice that only if the point P moves to the grey region will we include the point τ_1 in our chosen block. Hence, the probability that our block defined by P will include the point τ_1 is equal to the area of the grey region divided by $4h_1h_2$. In this instance, the area of the grey region is b_1b_2.

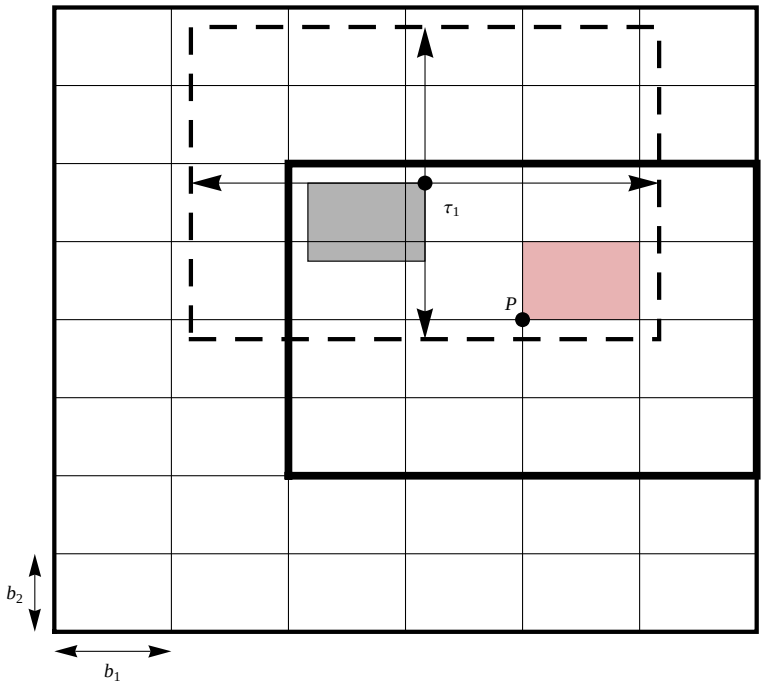

Figure 7.3: The Resampling Procedure

Consider another example. Suppose we will to fill the block indexed by the point Q.

This is the same set up as above. That is, our block will only contain the point τ_1 if our point Q ended up in grey shaded region. However, this time, not all of the grey region is contained inside the points available for Q to move. (That is, we do not have the full area of $b_1 b_2$ like we did when considering P.)

The only part of the shaded block that Q can migrate to is given by the red region in Figure 7.5. This time, the area of the shaded region will depend on the location of the point τ_1 relative to the block which contains it.

That is, we can focus on the block originally containing τ_1 since we have designed our h_1 and h_2 to be multiples of b_1 and b_2, respectively. This leads to the area in purple being equal to the area in red. While it may be challenging to directly determine the red area, it seems to be more straight-forward to determine the purple area.

Consider the point $\tau_1 = (\tau_{11}, \tau_{12})$. If we consider blocks by their lower-left coordinate, then we have that a block which is i copies to the right and j copies up from the lower-left corner will have coordinates (ib_1, jb_2).

If $|\tau_{11} - (i+1)b_1| \geq h_1$ or $|\tau_{12} - (j+1)b_2| \geq h_2$, then the area of the shaded region will be 0. That is, only if $|\tau_{11} - (i+1)b_1| < h_1$ and $|\tau_{12} - (j+1)b_2| < h_2$, will there be some non-zero probability.

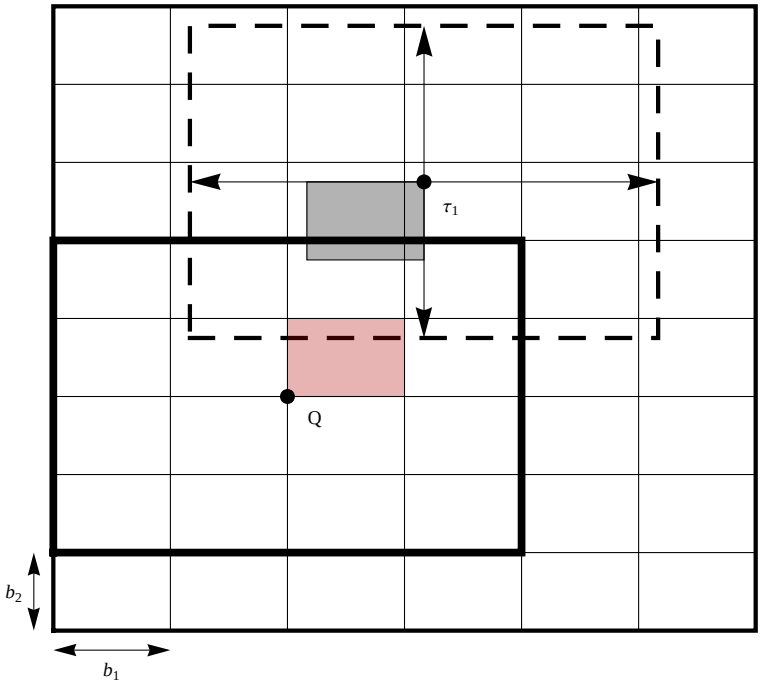

Figure 7.4: The Resampling Procedure Again

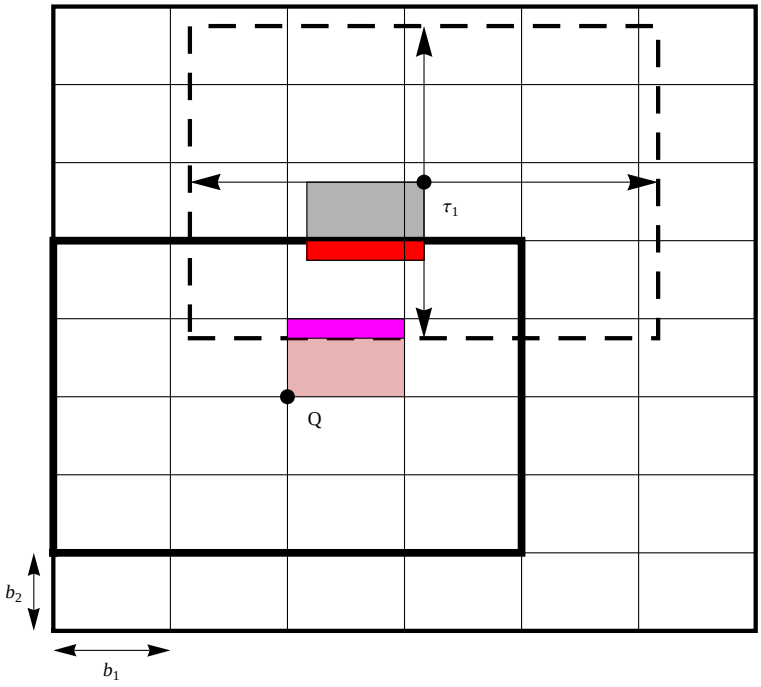

Figure 7.5: The Resampling Procedure Again (more detail)

To better see this, consider the first example (from Figure 7.3). As we said, suppose $b_1 = 3$ and $b_2 = 2$ so that the dimensions of the figure are 18×16. Suppose that the point τ_1 is given by $(9.5, 11.5)$. Again, we set $\alpha = 3/2$ so that $h_1 = 6$ and $h_2 = 4$.

Consider the block specified by P (see Figure 7.3). It has coordinates $P = (4 \cdot 3, 3 \cdot 2) = (12, 6)$.

Notice that $|9.5 - (4 + 1) \cdot 3| = 5.5$, which is less than 6 and $|11.5 - (3 + 1) \cdot 2| = 3.5$, which is less than 4. Thus, we have a non-zero probability.

Consider the block specified by Q. It has coordinates $(2 \cdot 3, 3 \cdot 2) = (6, 6)$.

Notice that $|9.5 - (2 + 1) \cdot 3| = 0.5$, which is less than 6 and $|11.5 - (3 + 1) \cdot 2| = 3.5$, which is less than 4. Thus, we have a non-zero probability.

Consider the block directly below Q. It has coordinates $(2 \cdot 3, 2 \cdot 2) = (6, 4)$.

Notice that $|9.5 - (2 + 1) \cdot 3| = 0.5$, which is less than 4. However, $|11.5 - (2 + 1) \cdot 2| = 5.5 > 4$, and so that block cannot contain τ_1.

Now, let us shift our attention and determine how many points end up in the resampled data. That is, a particular point τ_1 can end up in multiple resampled blocks. (In the above example, the point τ_1 could potentially occur in as many as 25 blocks.)

In general, a point can end up in as many as $J = \left(2\left\lceil {b_1}^{\alpha-1}\right\rceil + 1\right)\left(2\left\lceil {b_2}^{\alpha-1}\right\rceil + 1\right)$ blocks. These blocks are not created equally, though.

Those in the interior have a probability of $\dfrac{b_1 b_2}{4h_1h_2} = \dfrac{1}{4\left\lceil {b_1}^{\alpha-1}\right\rceil\left\lceil {b_2}^{\alpha-1}\right\rceil}$ of containing the point. There will be $\left(2\left\lceil {b_1}^{\alpha-1}\right\rceil - 1\right)\left(2\left\lceil {b_2}^{\alpha-1}\right\rceil - 1\right)$ such blocks.

Suppose the point τ_1 has coordinates (τ_{11}, τ_{12}). Using the division algorithm, we can express $\tau_{11} = J_{11}b_1 + R_{11}$ and $\tau_{12} = J_{12}b_2 + R_{12}$. (In general, we can define $\tau_{ij} = J_{ij}b_j + R_{ij}$, for $i = 1, \ldots, n$ and $j = 1, 2$.)

With this notation, we have that there will be $2\left\lceil {b_1}^{\alpha-1}\right\rceil - 1$ blocks that have a probability of $\dfrac{b_1 R_{12}}{4h_1h_2}$ of containing τ_1 and $2\left\lceil {b_1}^{\alpha-1}\right\rceil - 1$ blocks that have a probability $\dfrac{b_1(b_2 - R_{12})}{4h_1h_2}$ of containing τ_1.

Similarly, there will be $2\left\lceil {b_2}^{\alpha-1}\right\rceil - 1$ blocks that have a probability of $\dfrac{R_{11}b_2}{4h_1h_2}$ of containing τ_1 and $2\left\lceil {b_2}^{\alpha-1}\right\rceil - 1$ blocks that have a probability $\dfrac{(b_1 - R_{11})b_2}{4h_1h_2}$ of containing τ_1.

Finally, there are four corners to consider. The four corners will have probabilities

$$\frac{R_{11}R_{12}}{4h_1h_2},\ \frac{R_{11}(b_2 - R_{12})}{4h_1h_2},\ \frac{(b_1 - R_{11})R_{12}}{4h_1h_2},\ \text{and}\ \frac{(b_1 - R_{11})(b_2 - R_{12})}{4h_1h_2}.$$

Notice that we specified a total of $\left(2\left\lceil {b_1}^{\alpha-1}\right\rceil + 1\right)\left(2\left\lceil {b_2}^{\alpha-1}\right\rceil + 1\right)$ blocks, as we claimed.

More importantly, we claim that the probabilities sum to 1. To see this, let us determine the contribution from each of the blocks.

We have $\left(2\left\lceil {b_1}^{\alpha-1}\right\rceil - 1\right)\left(2\left\lceil {b_2}^{\alpha-1}\right\rceil - 1\right)$ blocks, each with an area of $b_1 b_2$.
This gives a total area of $\left(2\left\lceil {b_1}^{\alpha-1}\right\rceil - 1\right)\left(2\left\lceil {b_2}^{\alpha-1}\right\rceil - 1\right) b_1 b_2$.
We have $2\left(2\left\lceil {b_1}^{\alpha-1}\right\rceil - 1\right)$ blocks with a total area of

$$\left(2\left\lceil {b_1}^{\alpha-1}\right\rceil - 1\right) b_1 R_{12} + \left(2\left\lceil {b_1}^{\alpha-1}\right\rceil - 1\right) b_1 (b_2 - R_{12}) = \left(2\left\lceil {b_1}^{\alpha-1}\right\rceil - 1\right) b_1 b_2$$

We have $2\left(2\left\lceil {b_2}^{\alpha-1}\right\rceil - 1\right)$ blocks with a total area of

$$\left(2\left\lceil {b_2}^{\alpha-1}\right\rceil - 1\right) R_{11} b_2 + \left(2\left\lceil {b_2}^{\alpha-1}\right\rceil - 1\right) (b_1 - R_{11}) b_2 = \left(2\left\lceil {b_2}^{\alpha-1}\right\rceil - 1\right) b_1 b_2$$

Finally, we have 4 blocks (the corners) with a total area of

$$R_{11}R_{12} + R_{11}(b_2 - R_{12}) + (b_1 - R_{11})R_{12} + (b_1 - R_{11})(b_2 - R_{12}) = b_1 b_2$$

Adding these areas together, we have precisely $4h_1h_2$.

But recall that the total possible area (of all blocks) was $(2h_1) \cdot (2h_2) = 4h_1h_2$. Thus, the probabilities sum to 1 as claimed.

Notice, though, that our development would hold in d dimensions as well, by the assumed independence in each coordinate. So, even though we have illustrated the formula for two-dimensions, the result holds for any dimension.

As in the one-dimensional case, we shall develop formulas for $Cov^*(W_i, W_k)$ based upon various cases. The transition is fairly straight-forward, though there are significantly more cases to consider. In the end, though, we shall see that they all yield the same result. Each one will be explained in detail.

Again, we need some definitions before we continue.

Definition 7.4.2. Suppose $\tau_i = (\tau_{i1}, \tau_{i2})$ and $\tau_k = (\tau_{k1}, \tau_{k2})$. If $|\tau_{i1} - \tau_{k1}| < b_1$ and $|\tau_{i2} - \tau_{k2}| < b_2$, we say that τ_i and τ_k are **sufficiently close**.

This is natural extension of our definition from the one-dimensional setting. That is, we require the two points to be sufficiently close (using the one-dimensional definition) in each coordinate.

Definition 7.4.3. Suppose $(a_{11}, a_{21}), (a_{12}, a_{21}), (a_{12}, a_{22}), (a_{11}, a_{22})$ specify the vertices of a rectangle. We say that a point $\tau_i = (\tau_{i1}, \tau_{i2})$ is removed from the boundary by h_1 units in its first coordinate provided that $|\tau_{i1} - a_{11}| > h_1$ and $|\tau_{i1} - a_{12}| > h_1$. We say that a point $\tau_i = (\tau_{i1}, \tau_{i2})$ is removed from the boundary by h_2 units in its second coordinate provided that $|\tau_{i2} - a_{21}| > h_2$

and $|\tau_{i2} - a_{22}| > h_2$. If both conditions above are satisfied, then the point $\tau_i = (\tau_{i1}, \tau_{i2})$ is **removed from the boundary**.

Again, this is a natural extension from the one-dimensional setting. We just require the coordinates to be removed from the boundary by the window size in each dimension.

Similar to the one-dimensional case, those points that are not removed from the boundary will not contribute an appreciable amount to the summation and thus may be ignored.

Finally, we use the division algorithm to express each of the points, τ_i, as multiples of the block widths in each dimension, plus a remainder term. That is, suppose $\tau_i = (\tau_{i1}, \tau_{i2})$. Using the division algorithm, we can express $\tau_{i1} = J_{i1}b_1 + R_{i1}$ and $\tau_{i2} = J_{i2}b_2 + R_{i2}$.

This definition will be useful as we express our probabilities visually in the following figures.

Now, we are ready to consider the various cases that $Cov^*(W_i, W_k)$ may take. As before, we shall not consider this problem directly, but rather recognize that

$$\begin{aligned} Cov^*(W_i, W_k) &= Cov^*\left(\sum_{j=1}^{L} Y_{ij}, \sum_{m=1}^{L} Y_{km}\right) \\ &= \sum_{j=1}^{L}\sum_{m=1}^{L} Cov^*(Y_{ij}, Y_{km}) \\ &= \sum_{j=1}^{L} Cov^*(Y_{ij}, Y_{kj}) \end{aligned}$$

where $Y_{ij} \sim \text{Bernoulli}(p_{ij})$, with p_{ij} equal to the probability that τ_i occurs in block j.

There are a total of five cases which we shall now illustrate.

Case (i): τ_i and τ_k occur in the same block, with relative positions given below.

Notice that there are a total of $\left(2\left\lceil b_1^{\alpha-1}\right\rceil + 1\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil + 1\right)$ blocks contained within the boundaries of the dashed rectangles in Figure 7.6.

Recall, only those blocks (which are indexed by their lower-left coordinate) which intersect the dashed rectangle will have a non-zero probability of containing the point τ_i.

We can break these blocks (in Figure 7.6 there are 25 such blocks) into six different groups. We shall illustrate each section we are discussing with a corresponding diagram.

First off, we have the middle blocks.

Notice that there are a total of $\left(2\left\lceil b_1^{\alpha-1}\right\rceil - 1\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)$ such blocks highlighted in Figure 7.7. Each of these blocks completely contains the grey rectangles that correspond to τ_i and τ_k, as well as their intersection (the purple area).

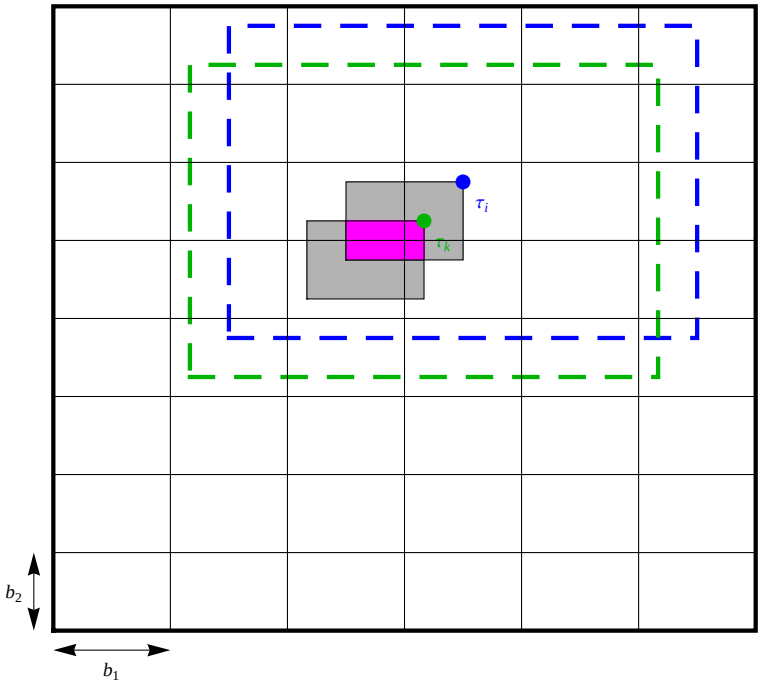

Figure 7.6: Covariance formula for two dimensions, Case (i)

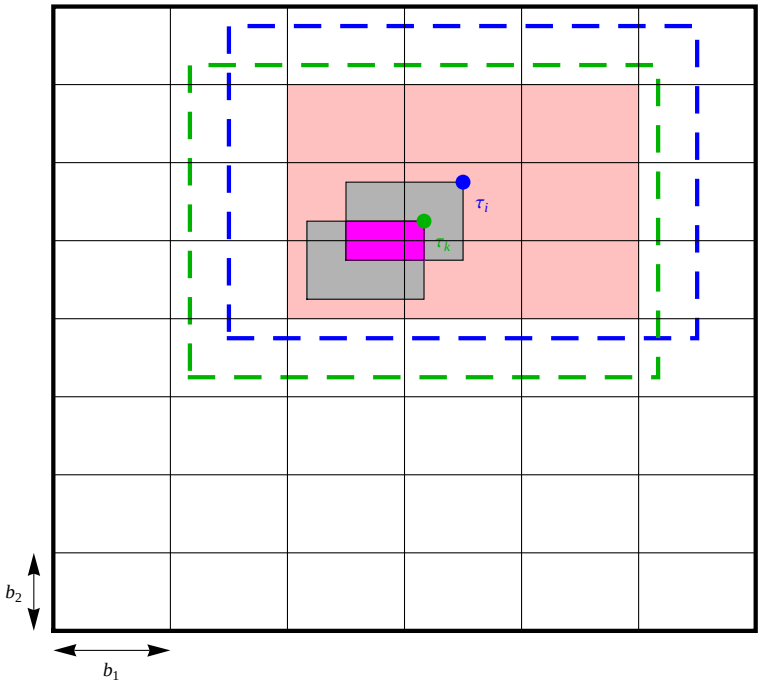

Figure 7.7: Covariance formula for two dimensions, Case (i), middle blocks

We have $p_{ij} = \dfrac{b_1 b_2}{4h_1 h_2} = p_{kj}$ and $p_{ij} \cap p_{kj} = \dfrac{(b_1 - (R_{i1} - R_{k1}))(b_2 - (R_{i2} - R_{k2}))}{4h_1 h_2}$.

Here, we abuse notation slightly and use $p_{ij} \cap p_{kj}$ to represent the joint probability that both τ_i and τ_k are contained in block j.

Also, notice that we could have expressed the joint probability as

$$p_{ij} \cap p_{kj} = \frac{(b_1 - |\tau_{i1} - \tau_{k1}|)(b_2 - |\tau_{i2} - \tau_{k2}|)}{4h_1 h_2}.$$

Ultimately, we will return to this form, but with the orientation of the points such that they are, it is easier to work exclusively with b_i's and R_{im}'s.

Next, we have the bottom blocks.

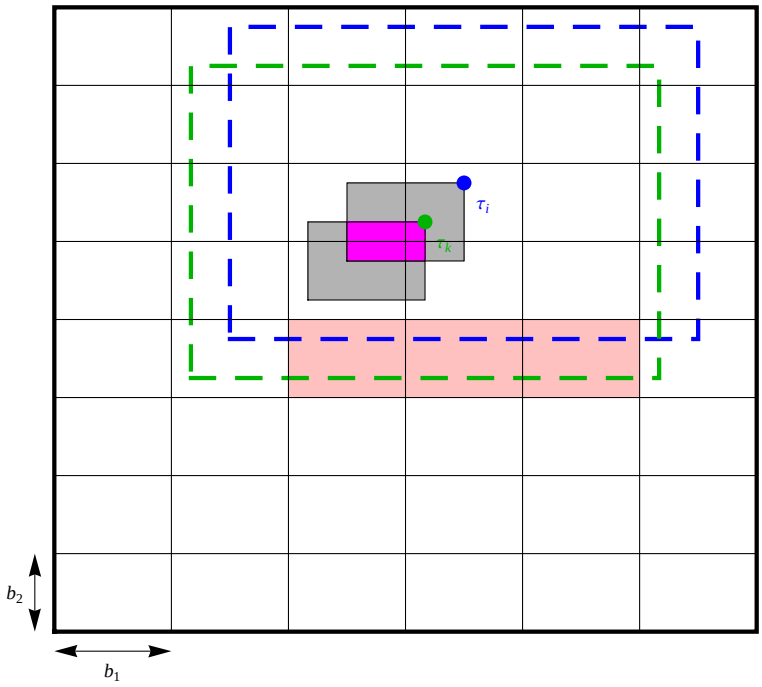

Figure 7.8: Covariance formula for two dimensions, Case (i), bottom blocks

Notice that there are a total of $2\left\lceil b_1^{\alpha-1} \right\rceil - 1$ such blocks highlighted in Figure 7.8. Each of these blocks contains the complete width, b_1, of the grey blocks, but the heights are not the full b_2.

Indeed, we have that $p_{ij} = \dfrac{b_1(b_2 - R_{i2})}{4h_1 h_2}$ and $p_{kj} = \dfrac{b_1(b_2 - R_{k2})}{4h_1 h_2}$.

For the joint probability, we have $p_{ij} \cap p_{kj} = \dfrac{(b_1 - (R_{i1} - R_{k1}))(b_2 - R_{i2})}{4h_1 h_2}$ as that constitutes the subset of the purple block that is captured by these blocks. (The maximum width is

$b_1 - (R_{i1} - R_{k1})$ while the maximum height is $b_2 - (R_{i2} - R_{k2})$. Here, the height is only $b_2 - R_{i2}$.)

Next, we have the top blocks.

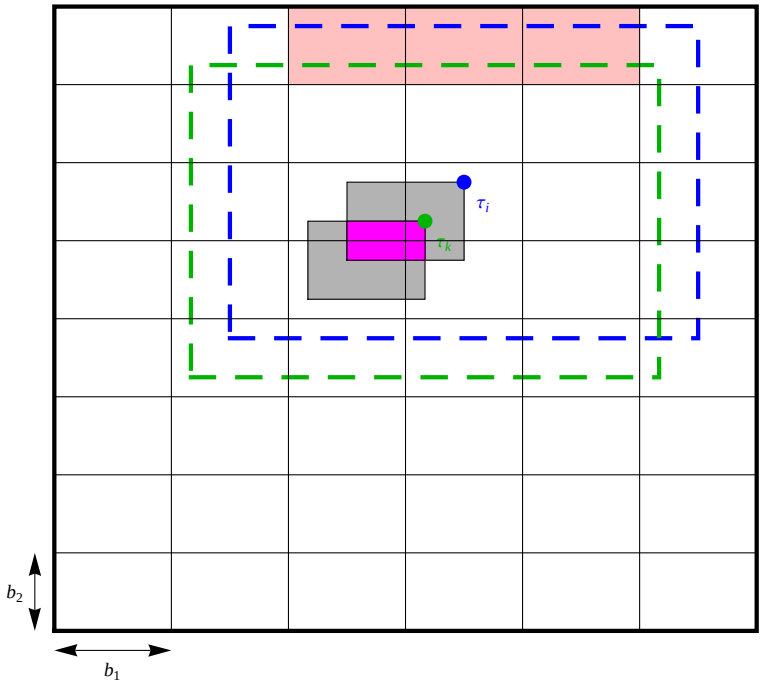

Figure 7.9: Covariance formula for two dimensions, Case (i), top blocks

Notice that there are a total of $2\left\lceil b_1^{\alpha-1}\right\rceil - 1$ such blocks highlighted in Figure 7.9. Each of these blocks contains the complete width, b_1, of the grey blocks, but the heights are not the full b_2.

Indeed, we have that $p_{ij} = \dfrac{b_1 R_{i2}}{4h_1h_2}$ and $p_{kj} = \dfrac{b_1 R_{k2}}{4h_1h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{(b_1 - (R_{i1} - R_{k1}))R_{k2}}{4h_1h_2}$ as that constitutes the subset of the purple block that is captured by these blocks.

Next, we have the left blocks.

Notice that there are a total of $2\left\lceil b_2^{\alpha-1}\right\rceil - 1$ such blocks highlighted in Figure 7.10. Each of these blocks contains the complete height, b_2, of the grey blocks, but the widths are not the full b_1.

Indeed, we have that $p_{ij} = \dfrac{(b_1 - R_{i1})b_2}{4h_1h_2}$ and $p_{kj} = \dfrac{(b_1 - R_{k1})b_2}{4h_1h_2}$.

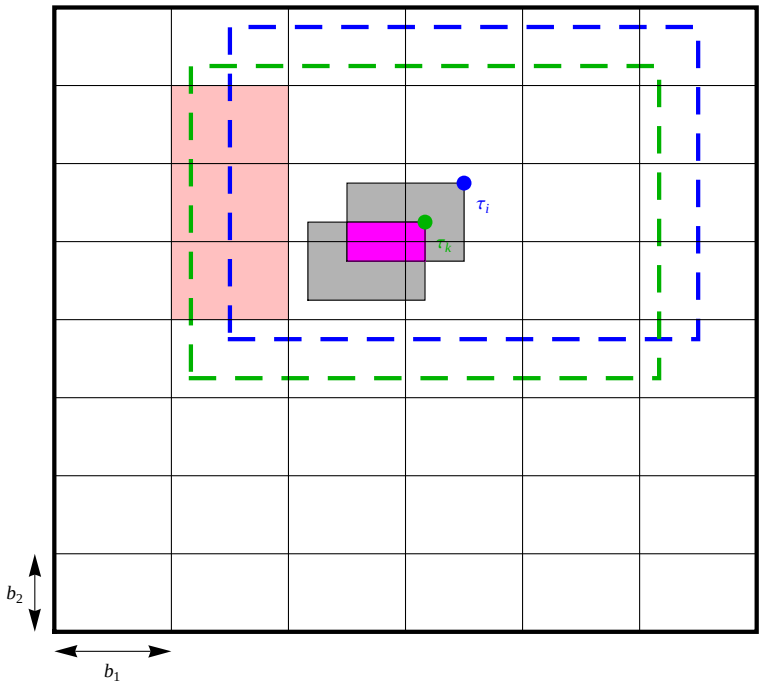

Figure 7.10: Covariance formula for two dimensions, Case (i), left blocks

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{(b_1 - R_{i1})(b_2 - (R_{i2} - R_{k2}))}{4h_1h_2}$ as that constitutes the subset of the purple block that is captured by these blocks. (The maximum width is $b_1 - (R_{i1} - R_{k1})$ while the maximum height is $b_2 - (R_{i2} - R_{k2})$. Here, the width is only $b_1 - R_{i1}$.)

Next, we have the right blocks.

Notice that there are a total of $2\left\lceil b_2^{\alpha-1}\right\rceil - 1$ such blocks highlighted in Figure 7.11. Each of these blocks contains the complete height, b_2, of the grey blocks, but the widths are not the full b_1.

Indeed, we have that $p_{ij} = \dfrac{R_{i1}b_2}{4h_1h_2}$ and $p_{kj} = \dfrac{R_{k1}b_2}{4h_1h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{R_{k1}(b_2 - (R_{i2} - R_{k2}))}{4h_1h_2}$ as that constitutes the subset of the purple block that is captured by these blocks.

Finally, there are four corners to deal with as highlighted in Figure 7.12. We shall compute those probabilities working clockwise from the bottom-left.

In the bottom-left corner, we have that

$$p_{ij} = \frac{(b_1 - R_{i1})(b_2 - R_{i2})}{4h_1h_2},\ p_{kj} = \frac{(b_1 - R_{k1})(b_2 - R_{k2})}{4h_1h_2},$$

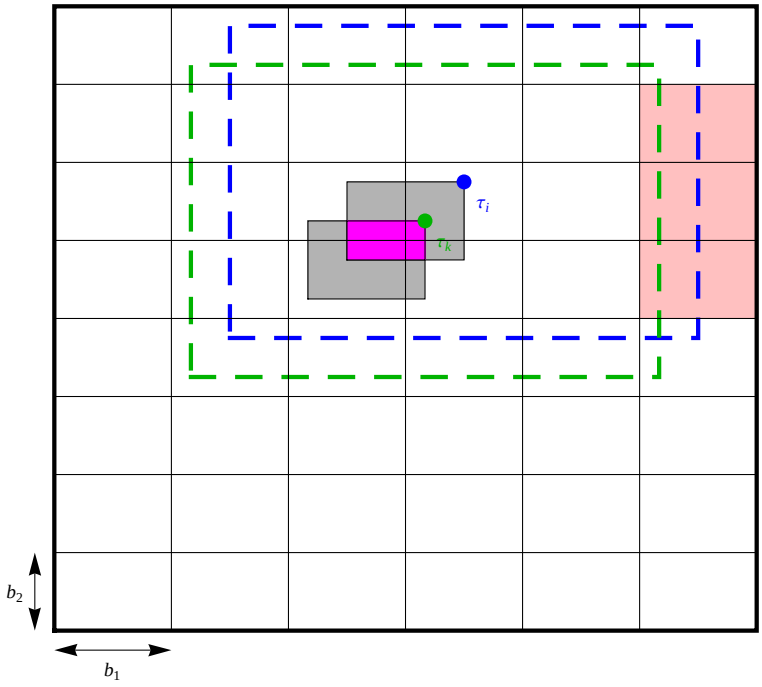

Figure 7.11: Covariance formula for two dimensions, Case (i), right blocks

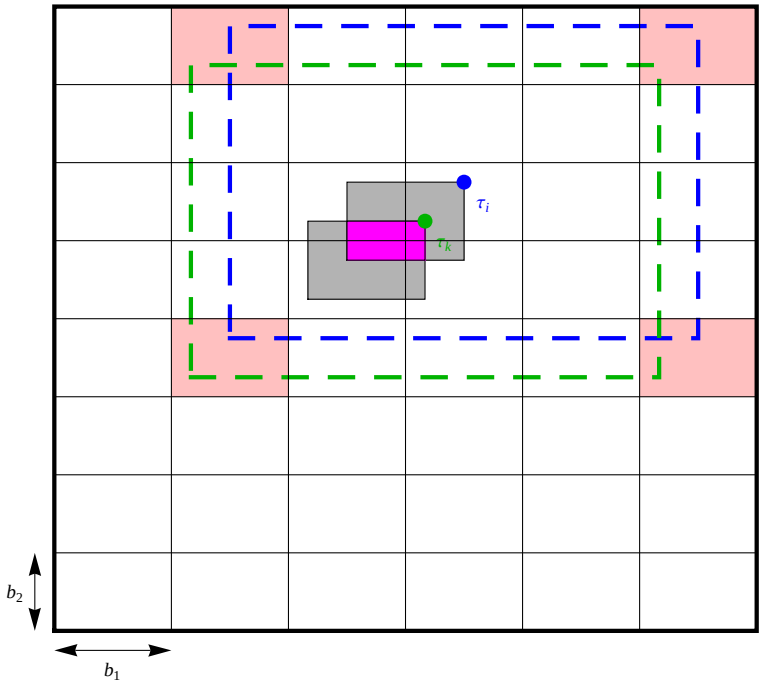

Figure 7.12: Covariance formula for two dimensions, Case (i), corner blocks

and $p_{ij} \cap p_{kj} = \dfrac{(b_1 - R_{i1})(b_2 - R_{i2})}{4h_1h_2}$.

In the top-left corner, we have that

$p_{ij} = \dfrac{(b_1 - R_{i1})R_{i2}}{4h_1h_2}$, $p_{kj} = \dfrac{(b_1 - R_{k1})R_{k2}}{4h_1h_2}$,

and $p_{ij} \cap p_{kj} = \dfrac{(b_1 - R_{i1})R_{k2}}{4h_1h_2}$.

In the top-right corner, we have that

$p_{ij} = \dfrac{R_{i1}R_{i2}}{4h_1h_2}$, $p_{kj} = \dfrac{R_{k1}R_{k2}}{4h_1h_2}$,

and $p_{ij} \cap p_{kj} = \dfrac{R_{k1}R_{k2}}{4h_1h_2}$.

In the bottom-right corner, we have that

$p_{ij} = \dfrac{R_{i1}(b_2 - R_{i2})}{4h_1h_2}$, $p_{kj} = \dfrac{R_{k1}(b_2 - R_{k2})}{4h_1h_2}$,

and $p_{ij} \cap p_{kj} = \dfrac{R_{k1}(b_2 - R_{i2})}{4h_1h_2}$.

Putting this all together, we have the following covariance formula:

$$
\begin{aligned}
&Cov^*(W_i, W_k) \\
&= \sum_{j=1}^{l} Cov^*(Y_{ij}, Y_{kj}) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 1\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)\left[\frac{(b_1 - (R_{i1} - R_{k1}))(b_2 - (R_{i2} - R_{k2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 1\right)\left[\frac{(b_1 - (R_{i1} - R_{k1}))(b_2 - R_{i2})}{4h_1h_2} - \left(\frac{b_1(b_2 - R_{i2})}{4h_1h_2}\right)\left(\frac{b_1(b_2 - R_{k2})}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 1\right)\left[\frac{(b_1 - (R_{i1} - R_{k1}))R_{k2}}{4h_1h_2} - \left(\frac{b_1R_{i2}}{4h_1h_2}\right)\left(\frac{b_1R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)\left[\frac{(b_1 - R_{i1})(b_2 - (R_{i2} - R_{k2}))}{4h_1h_2} - \left(\frac{(b_1 - R_{i1})b_2}{4h_1h_2}\right)\left(\frac{(b_1 - R_{k1})b_2}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)\left[\frac{R_{k1}(b_2 - (R_{i2} - R_{k2}))}{4h_1h_2} - \left(\frac{R_{i1}b_2}{4h_1h_2}\right)\left(\frac{R_{k1}b_2}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(b_1 - R_{i1})(b_2 - R_{i2})}{4h_1h_2} - \left(\frac{(b_1 - R_{i1})(b_2 - R_{i2})}{4h_1h_2}\right)\left(\frac{(b_1 - R_{k1})(b_2 - R_{k2})}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(b_1 - R_{i1})R_{k2}}{4h_1h_2} - \left(\frac{(b_1 - R_{i1})R_{i2}}{4h_1h_2}\right)\left(\frac{(b_1 - R_{k1})R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{R_{k1}R_{k2}}{4h_1h_2} - \left(\frac{R_{i1}R_{i2}}{4h_1h_2}\right)\left(\frac{R_{k1}R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{R_{k1}(b_2 - R_{i2})}{4h_1h_2} - \left(\frac{R_{i1}(b_2 - R_{i2})}{4h_1h_2}\right)\left(\frac{R_{k1}(b_2 - R_{k2})}{4h_1h_2}\right)\right] \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(b_1 - (R_{i1} - R_{k1}))(b_2 - (R_{i2} - R_{k2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + O\left(\frac{b_1b_2}{h_1h_2}\right)
\end{aligned}
$$

Notice, we have that $R_{i1} - R_{k1} = |\tau_{i1} - \tau_{k1}|$ and $R_{i2} - R_{k2} = |\tau_{i2} - \tau_{k2}|$, so we can reduce the above formula to the following lemma:

Lemma 7.4.4. *For τ_i and τ_k sufficiently close and in the same block (with $\tau_{k1} < \tau_{i1}$ and $\tau_{k2} < \tau_{i2}$), τ_i and τ_k removed from the boundary, we have that*

$$
\begin{aligned}
&Cov^*(W_i, W_k) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(b_1 - |\tau_{i1} - \tau_{k1}|)(b_2 - |\tau_{i2} - \tau_{k2}|)}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + O\left(\frac{b_1b_2}{h_1h_2}\right)
\end{aligned}
$$

The second case is a slight variant of the first in that both τ_i and τ_k occur in the same block, but their relative positions have changed.

Case (ii): τ_i and τ_k occur in the same block, with relative positions given below.

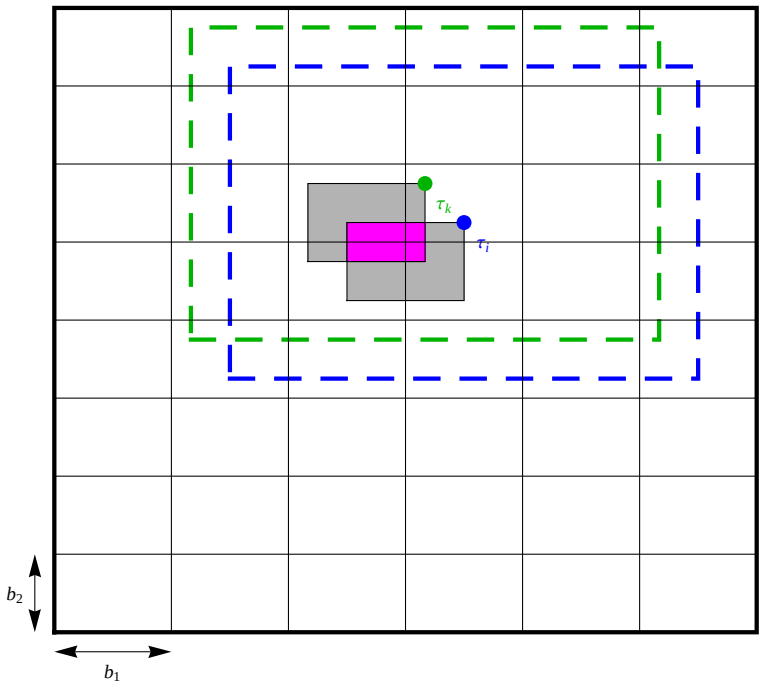

Figure 7.13: Covariance formula for two dimensions, Case (ii)

Again, notice that there are a total of $\left(2\left\lceil b_1^{\alpha-1}\right\rceil+1\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil+1\right)$ blocks contained within the boundaries of the dashed rectangles in Figure 7.13.

Recall, only those blocks (which are indexed by their lower-left coordinate) which intersect the dashed rectangle will have a non-zero probability of containing the point τ_i.

We can break these blocks (in Figure 7.13, there are 25 such blocks) into six different groups. We shall illustrate each section we are discussing with a corresponding diagram.

First, we have the middle blocks. There are a total of $\left(2\left\lceil b_1^{\alpha-1}\right\rceil-1\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil-1\right)$ such blocks highlighted in Figure 7.14. Each of these blocks completely contains the grey rectangles that correspond to τ_i and τ_k, as well as their intersection (the purple area).

Thus, we have that $p_{ij}=\dfrac{b_1 b_2}{4h_1h_2}=p_{kj}$

and $p_{ij}\cap p_{kj}=\dfrac{(b_1-(R_{i1}-R_{k1}))(b_2-(R_{k2}-R_{i2}))}{4h_1h_2}$.

Next, we have the bottom blocks. Notice that there are a total of $2\left\lceil b_1^{\alpha-1}\right\rceil-1$ such

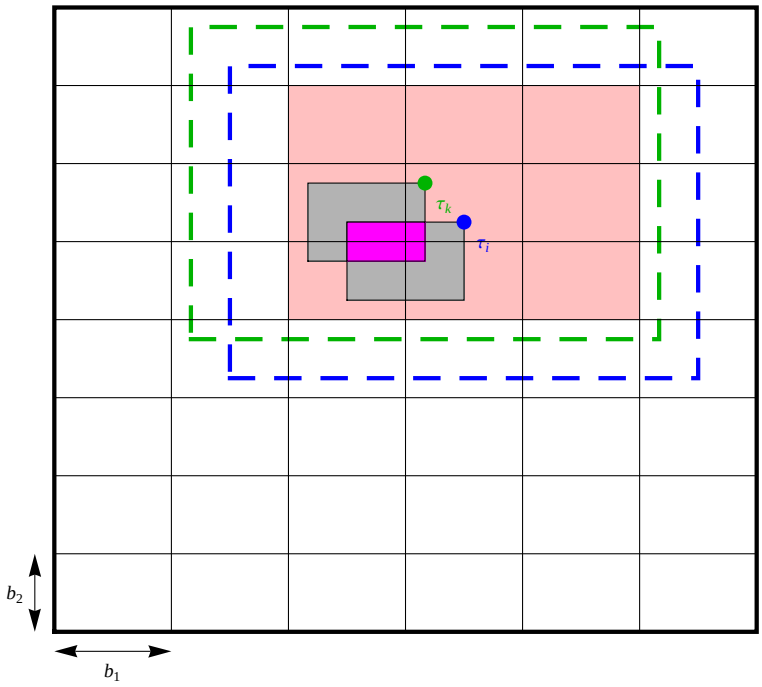

Figure 7.14: Covariance formula for two dimensions, Case (ii), middle blocks

blocks highlighted in Figure 7.15. Each of these blocks contains the complete width, b_1, of the grey blocks, but the heights are not the full b_2.

Indeed, we have that $p_{ij} = \dfrac{b_1(b_2 - R_{i2})}{4h_1h_2}$ and $p_{kj} = \dfrac{b_1(b_2 - R_{k2})}{4h_1h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{(b_1 - (R_{i1} - R_{k1}))(b_2 - R_{k2})}{4h_1h_2}$ as that constitutes the subset of the purple block that is captured by these blocks. (The maximum width is $b_1 - (R_{i1} - R_{k1})$ while the maximum height is $b_2 - (R_{k2} - R_{i2})$. Here, the height is only $b_2 - R_{k2}$.)

Next, we have the top blocks.

Notice that there are a total of $2\left\lceil b_1^{\alpha-1}\right\rceil - 1$ such blocks highlighted in Figure 7.16. Each of these blocks contains the complete width, b_1, of the grey blocks, but the heights are not the full b_2.

Indeed, we have that $p_{ij} = \dfrac{b_1R_{i2}}{4h_1h_2}$ and $p_{kj} = \dfrac{b_1R_{k2}}{4h_1h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{(b_1 - (R_{i1} - R_{k1}))R_{i2}}{4h_1h_2}$ as that constitutes the subset of the purple block that is captured by these blocks.

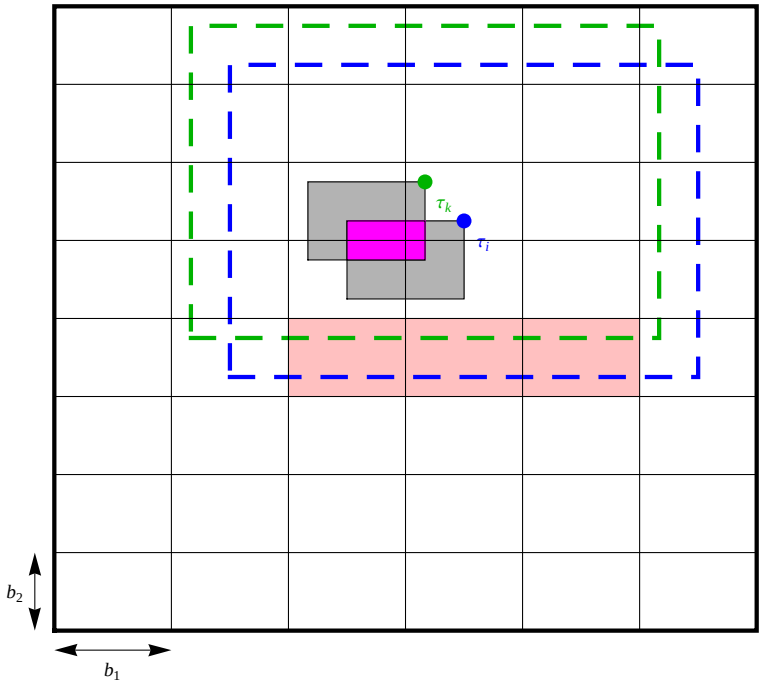

Figure 7.15: Covariance formula for two dimensions, Case (ii), bottom blocks

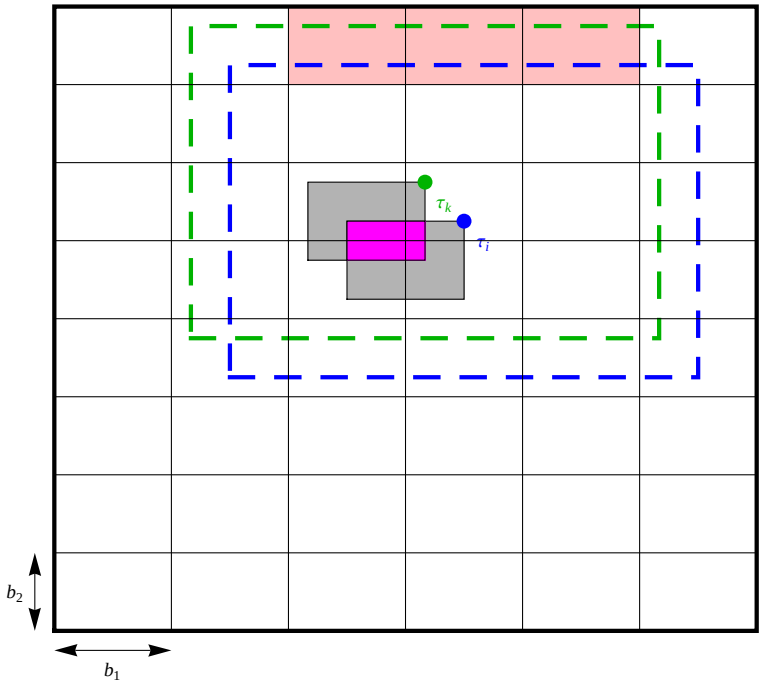

Figure 7.16: Covariance formula for two dimensions, Case (ii), top blocks

Next, we have the left blocks. Notice that there are a total of $2\left\lceil b_2^{\alpha-1}\right\rceil - 1$ such blocks highlighted in Figure 7.17. Each of these blocks contains the complete height, b_2, of the grey blocks, but the widths are not the full b_1.

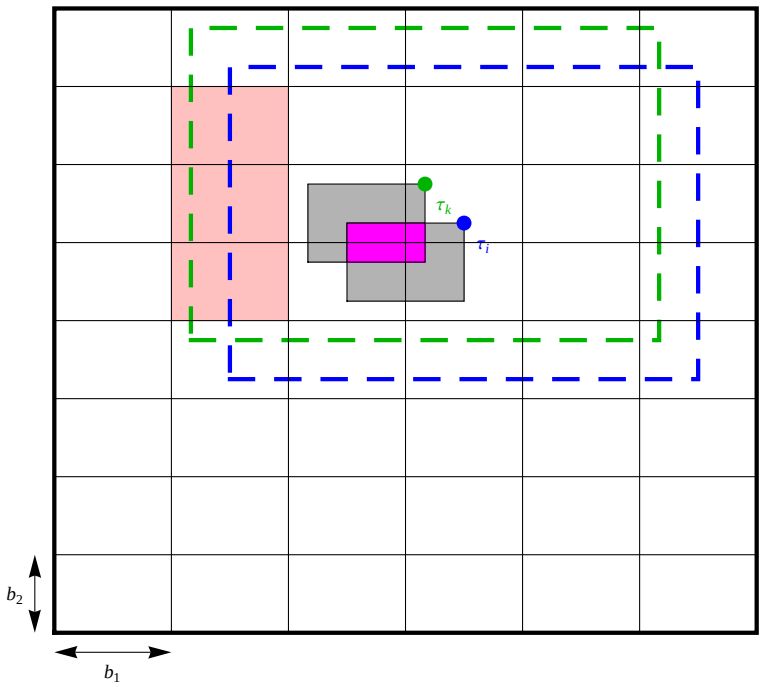

Figure 7.17: Covariance formula for two dimensions, Case (ii), left blocks

Indeed, we have that $p_{ij} = \dfrac{(b_1 - R_{i1})b_2}{4h_1h_2}$ and $p_{kj} = \dfrac{(b_1 - R_{k1})b_2}{4h_1h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{(b_1 - R_{i1})(b_2 - (R_{k2} - R_{i2}))}{4h_1h_2}$ as that constitutes the subset of the purple block that is captured by these blocks. (The maximum width is $b_1 - (R_{i1} - R_{k1})$ while the maximum height is $b_2 - (R_{k2} - R_{i2})$. Here, the width is only $b_1 - R_{i1}$.)

Next, we have the right blocks. Notice that there are a total of $2\left\lceil b_2^{\alpha-1}\right\rceil - 1$ such blocks highlighted in Figure 7.18. Each of these blocks contains the complete height, b_2, of the grey blocks, but the widths are not the full b_1.

Indeed, we have that $p_{ij} = \dfrac{R_{i1}b_2}{4h_1h_2}$ and $p_{kj} = \dfrac{R_{k1}b_2}{4h_1h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{R_{k1}(b_2 - (R_{k2} - R_{i2}))}{4h_1h_2}$ as that constitutes the subset of the purple block that is captured by these blocks.

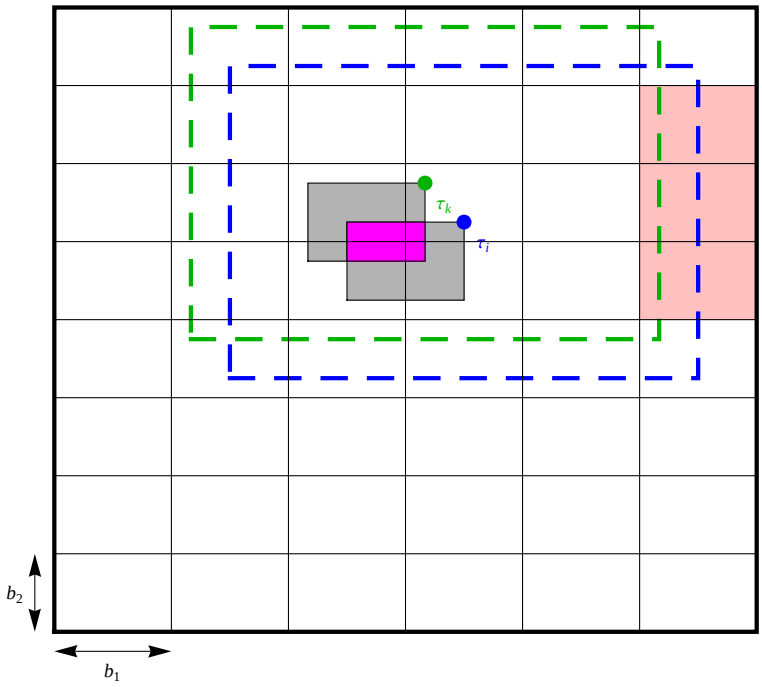

Figure 7.18: Covariance formula for two dimensions, Case (ii), right blocks

Finally, there are four corners to deal with as highlighted in Figure 7.19. We shall compute those probabilities working clockwise from the bottom-left.

In the bottom-left corner, we have that

$$p_{ij} = \frac{(b_1 - R_{i1})(b_2 - R_{i2})}{4h_1h_2}, \; p_{kj} = \frac{(b_1 - R_{k1})(b_2 - R_{k2})}{4h_1h_2},$$

$$\text{and } p_{ij} \cap p_{kj} = \frac{(b_1 - R_{i1})(b_2 - R_{k2})}{4h_1h_2}.$$

In the top-left corner, we have that

$$p_{ij} = \frac{(b_1 - R_{i1})R_{i2}}{4h_1h_2}, \; p_{kj} = \frac{(b_1 - R_{k1})R_{k2}}{4h_1h_2},$$

$$\text{and } p_{ij} \cap p_{kj} = \frac{(b_1 - R_{i1})R_{i2}}{4h_1h_2}.$$

In the top-right corner, we have that

$$p_{ij} = \frac{R_{i1}R_{i2}}{4h_1h_2}, \; p_{kj} = \frac{R_{k1}R_{k2}}{4h_1h_2},$$

$$\text{and } p_{ij} \cap p_{kj} = \frac{R_{k1}R_{i2}}{4h_1h_2}.$$

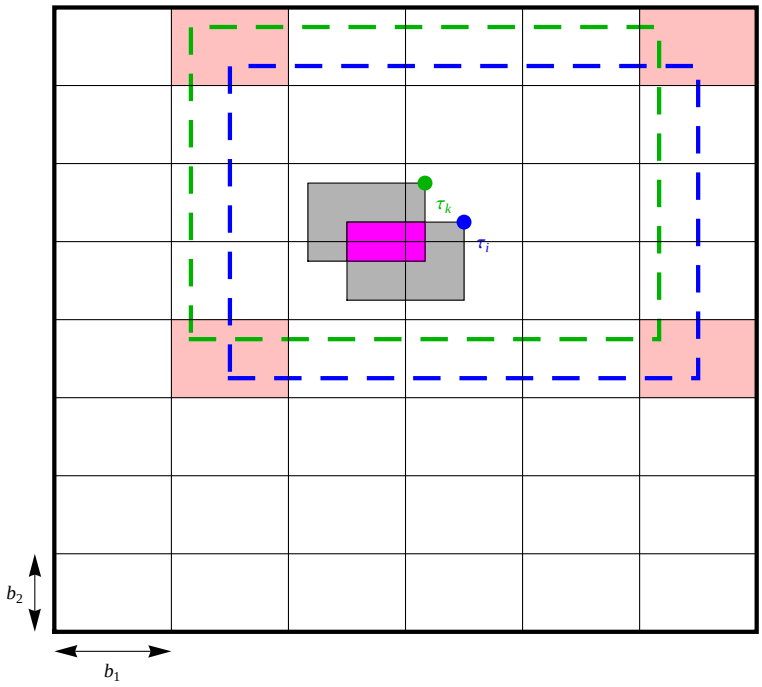

Figure 7.19: Covariance formula for two dimensions, Case (ii), corner blocks

In the bottom-right corner, we have that

$$p_{ij} = \frac{R_{i1}(b_2 - R_{i2})}{4h_1h_2},\ p_{kj} = \frac{R_{k1}(b_2 - R_{k2})}{4h_1h_2},$$

and $p_{ij} \cap p_{kj} = \dfrac{R_{k1}(b_2 - R_{k2})}{4h_1h_2}$.

Putting this all together, we have the following covariance formula:

$$
\begin{aligned}
&Cov^*(W_i, W_k) \\
&= \sum_{j=1}^{l} Cov^*(Y_{ij}, Y_{kj}) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 1\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)\left[\frac{(b_1 - (R_{i1} - R_{k1}))(b_2 - (R_{k2} - R_{i2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 1\right)\left[\frac{(b_1 - (R_{i1} - R_{k1}))(b_2 - R_{k2})}{4h_1h_2} - \left(\frac{b_1(b_2 - R_{i2})}{4h_1h_2}\right)\left(\frac{b_1(b_2 - R_{k2})}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 1\right)\left[\frac{(b_1 - (R_{i1} - R_{k1}))R_{i2}}{4h_1h_2} - \left(\frac{b_1R_{i2}}{4h_1h_2}\right)\left(\frac{b_1R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)\left[\frac{(b_1 - R_{i1})(b_2 - (R_{k2} - R_{i2}))}{4h_1h_2} - \left(\frac{(b_1 - R_{i1})b_2}{4h_1h_2}\right)\left(\frac{(b_1 - R_{k1})b_2}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)\left[\frac{R_{k1}(b_2 - (R_{k2} - R_{i2}))}{4h_1h_2} - \left(\frac{R_{i1}b_2}{4h_1h_2}\right)\left(\frac{R_{k1}b_2}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(b_1 - R_{i1})(b_2 - R_{k2})}{4h_1h_2} - \left(\frac{(b_1 - R_{i1})(b_2 - R_{i2})}{4h_1h_2}\right)\left(\frac{(b_1 - R_{k1})(b_2 - R_{k2})}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(b_1 - R_{i1})R_{i2}}{4h_1h_2} - \left(\frac{(b_1 - R_{i1})R_{i2}}{4h_1h_2}\right)\left(\frac{(b_1 - R_{k1})R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{R_{k1}R_{i2}}{4h_1h_2} - \left(\frac{R_{i1}R_{i2}}{4h_1h_2}\right)\left(\frac{R_{k1}R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{R_{k1}(b_2 - R_{k2})}{4h_1h_2} - \left(\frac{R_{i1}(b_2 - R_{i2})}{4h_1h_2}\right)\left(\frac{R_{k1}(b_2 - R_{k2})}{4h_1h_2}\right)\right] \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(b_1 - (R_{i1} - R_{k1}))(b_2 - (R_{k2} - R_{i2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + O\left(\frac{b_1b_2}{h_1h_2}\right)
\end{aligned}
$$

Notice, we have that $R_{i1} - R_{k1} = |\tau_{i1} - \tau_{k1}|$ and $R_{k2} - R_{i2} = |\tau_{i2} - \tau_{k2}|$, so we can reduce the above formula to the following lemma:

Lemma 7.4.5. *For τ_i and τ_k sufficiently close and in the same block (with $\tau_{k1} < \tau_{i1}$ and $\tau_{i2} < \tau_{k2}$), τ_i and τ_k removed from the boundary, we have that*

$$
\begin{aligned}
&Cov^*(W_i, W_k) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(b_1 - |\tau_{i1} - \tau_{k1}|)(b_2 - |\tau_{i2} - \tau_{k2}|)}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + O\left(\frac{b_1b_2}{h_1h_2}\right)
\end{aligned}
$$

While technically there would be four cases to consider with τ_i and τ_k in the same block,

since our covariance formula only involves the points via $|\tau_{i1} - \tau_{k1}|$ and $|\tau_{i2} - \tau_{k2}|$, we can assume without loss of generality that the two cases above cover all four possibilities.

Alternatively, one can notice that in the derivation above, the only thing that changed were some instances of i's and k's.

For this reason (and in the interest of conserving space), the next three cases will be illustrated without loss of generality using $\tau_{k1} < \tau_{i1}$ and $\tau_{k2} < \tau_{i2}$.

Case (iii): τ_i and τ_k occur in different blocks, horizontally.

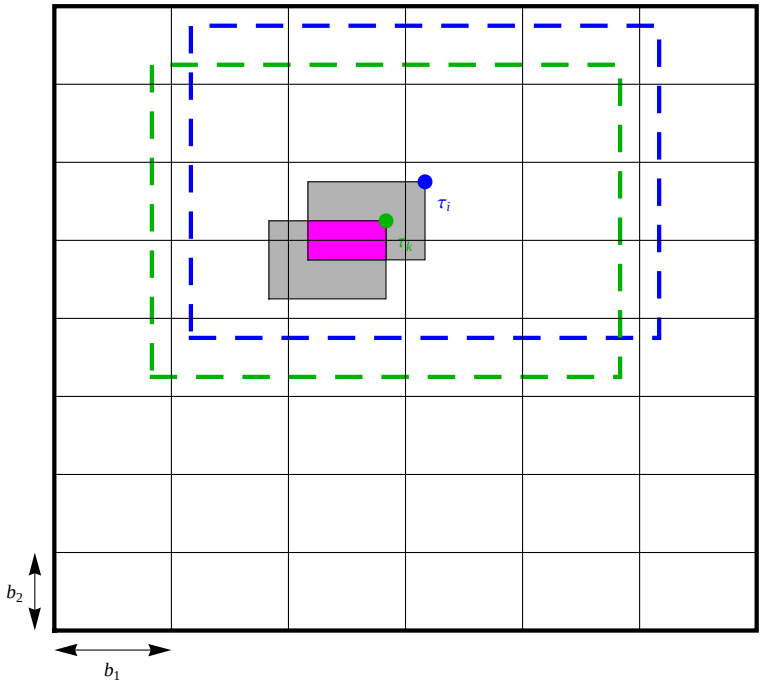

Figure 7.20: Covariance formula for two dimensions, Case (iii)

Notice that there are a total of $\left(2\left\lceil b_1^{\alpha-1}\right\rceil + 2\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil + 1\right)$ blocks contained within the boundaries of the dashed rectangles in Figure 7.20.

However, the left-most and right-most blocks do not contain both sets of dashed rectangles, and hence the covariance will be zero in those blocks. So, in actuality, we only need to consider $\left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil + 1\right)$ blocks.

And before we continue, it will be helpful to write out explicitly the dimensions of the purple region. Zooming in on the appropriate 2×3 blocks, we have Figure 7.21.

As in the previous cases, we shall break down the covariance calculations into groups of blocks. We begin with the middle blocks.

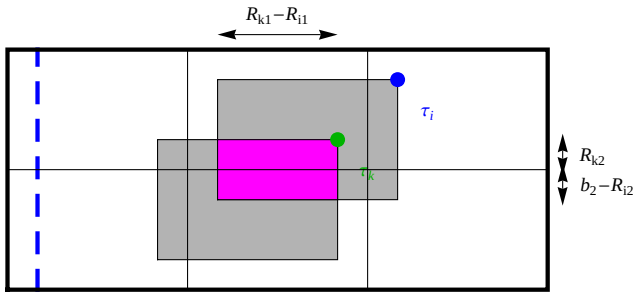

Figure 7.21: Covariance formula for two dimensions, Case (iii), enhanced

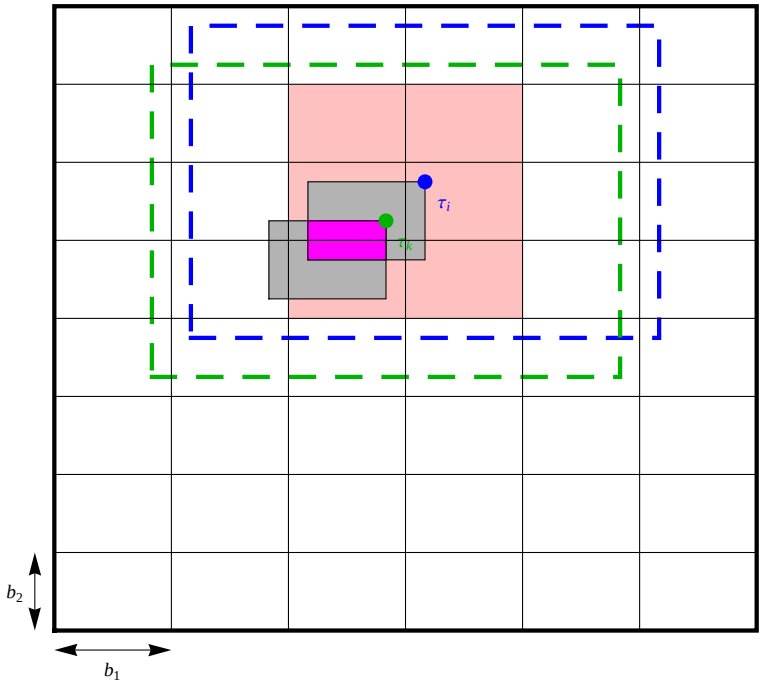

Figure 7.22: Covariance formula for two dimensions, Case (iii), middle blocks

Notice that there are a total of $\left(2\left\lceil b_1^{\alpha-1}\right\rceil - 2\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)$ such blocks highlighted in Figure 7.22. Each of these blocks completely contains the grey rectangles that correspond to τ_i and τ_k, as well as their intersection (the purple area).

Thus, we have that $p_{ij} = \dfrac{b_1 b_2}{4h_1 h_2} = p_{kj}$

and $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(b_2 - (R_{i2} - R_{k2}))}{4h_1 h_2}$.

Next, we have the bottom blocks. Notice that there are a total of $2\left\lceil b_1^{\alpha-1}\right\rceil - 2$ such blocks highlighted in Figure 7.23. Each of these blocks contains the complete width, b_1, of the grey blocks, but the heights are not the full b_2.

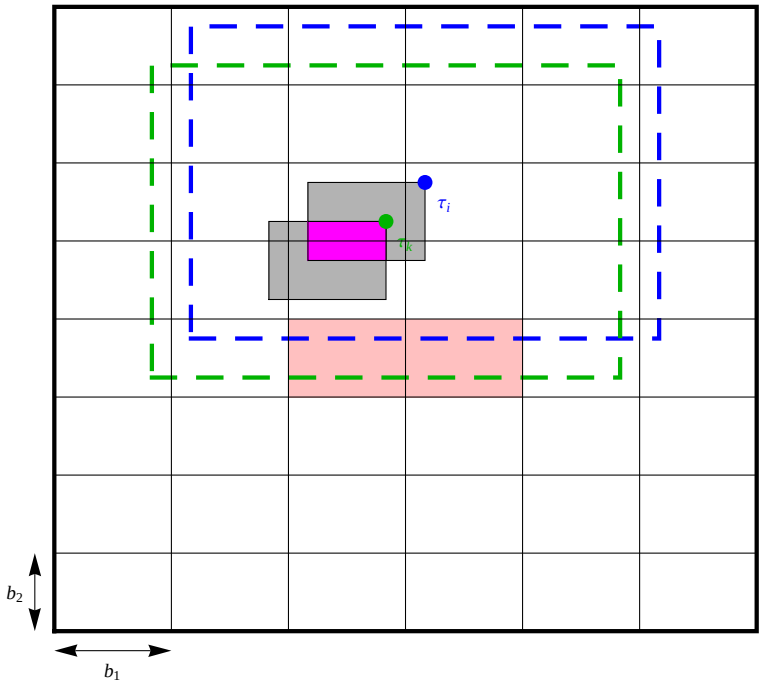

Figure 7.23: Covariance formula for two dimensions, Case (iii), bottom blocks

Indeed, we have that $p_{ij} = \dfrac{b_1(b_2 - R_{i2})}{4h_1 h_2}$ and $p_{kj} = \dfrac{b_1(b_2 - R_{k2})}{4h_1 h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(b_2 - R_{i2})}{4h_1 h_2}$ as that constitutes the subset of the purple block that is captured by these blocks.

Next, we have the top blocks. Notice that there are a total of $2\left\lceil b_1^{\alpha-1}\right\rceil - 2$ such blocks highlighted in Figure 7.24. Each of these blocks contains the complete width, b_1, of the grey

blocks, but the heights are not the full b_2.

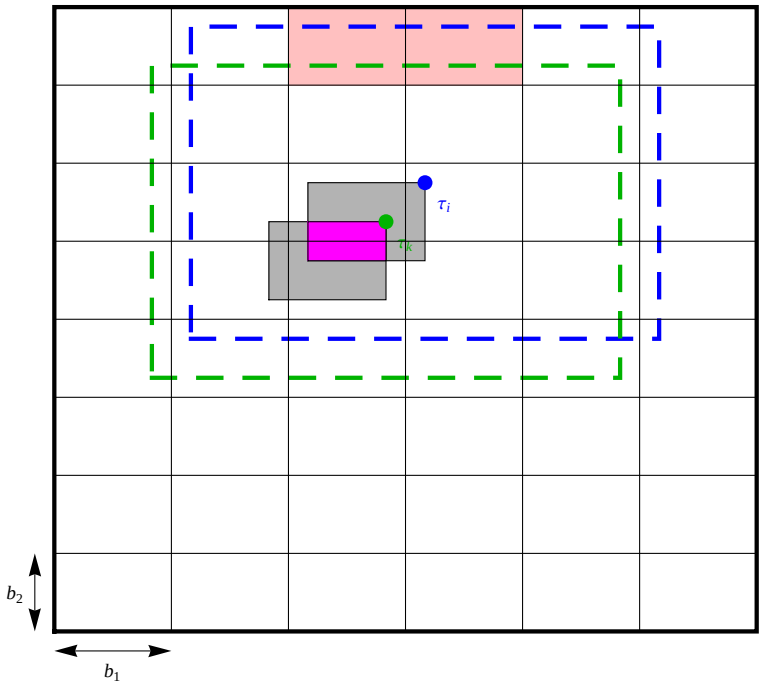

Figure 7.24: Covariance formula for two dimensions, Case (iii), top blocks

Indeed, we have that $p_{ij} = \dfrac{b_1 R_{i2}}{4h_1h_2}$ and $p_{kj} = \dfrac{b_1 R_{k2}}{4h_1h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})R_{k2}}{4h_1h_2}$ as that constitutes the subset of the purple block that is captured by these blocks.

Next, we have the left blocks. Notice that there are a total of $2\left\lceil b_2^{\alpha-1}\right\rceil - 1$ such blocks highlighted in Figure 7.25. Each of these blocks contains the complete height, b_2, of the grey blocks, but the widths vary.

Indeed, we have that $p_{ij} = \dfrac{(b_1 - R_{i1})b_2}{4h_1h_2}$ and $p_{kj} = \dfrac{b_1 b_2}{4h_1h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(b_2 - (R_{i2} - R_{k2}))}{4h_1h_2}$ as the entire purple block is captured by these blocks.

Next, we have the right blocks. Notice that there are a total of $2\left\lceil b_2^{\alpha-1}\right\rceil - 1$ such blocks highlighted in Figure 7.26. Each of these blocks contains the complete height, b_2, of the grey blocks, but the widths vary.

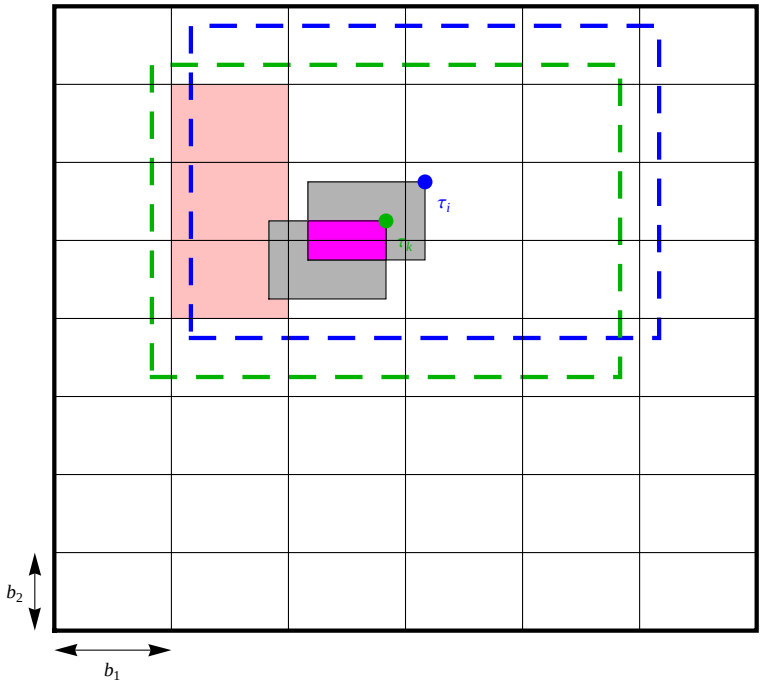

Figure 7.25: Covariance formula for two dimensions, Case (iii), left blocks

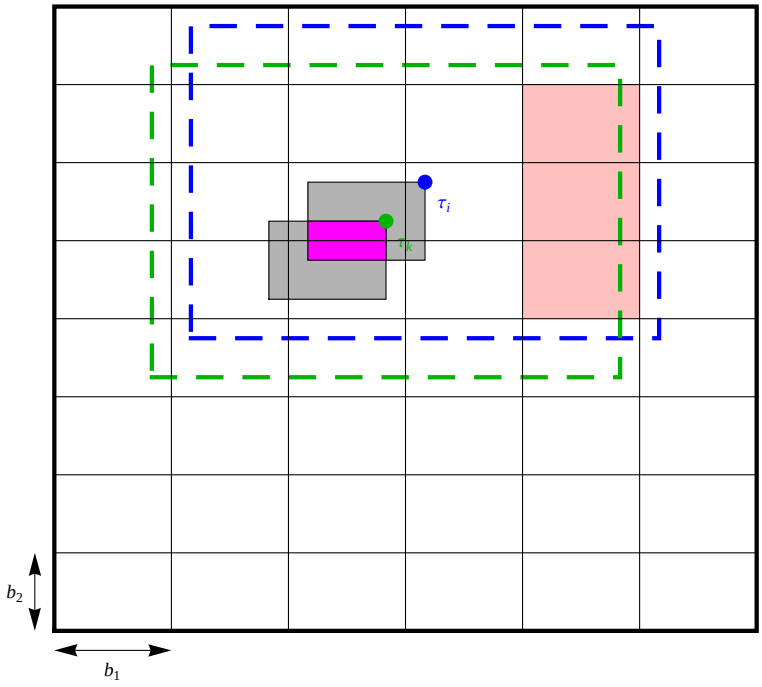

Figure 7.26: Covariance formula for two dimensions, Case (iii), right blocks

Indeed, we have that $p_{ij} = \dfrac{b_1 b_2}{4h_1 h_2}$ and $p_{kj} = \dfrac{R_{k1} b_2}{4h_1 h_2}$.

Here, we have that $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(b_2 - (R_{i2} - R_{k2}))}{4h_1 h_2}$ as the entire purple block is captured by these blocks.

Finally, there are four corners to deal with as highlighted in Figure 7.27. We shall compute those probabilities working clockwise from the bottom-left.

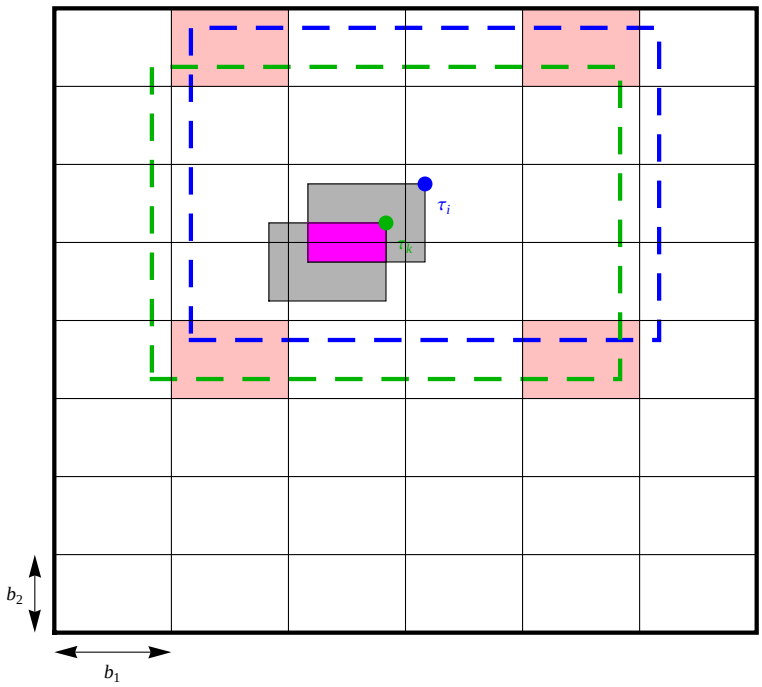

Figure 7.27: Covariance formula for two dimensions, Case (iii), corner blocks

In the bottom-left corner, we have that

$$p_{ij} = \frac{(b_1 - R_{i1})(b_2 - R_{i2})}{4h_1 h_2},\ p_{kj} = \frac{b_1(b_2 - R_{k2})}{4h_1 h_2},$$

and $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(b_2 - R_{i2})}{4h_1 h_2}$.

In the top-left corner, we have that

$$p_{ij} = \frac{(b_1 - R_{i1})R_{i2}}{4h_1 h_2},\ p_{kj} = \frac{b_1 R_{k2}}{4h_1 h_2},$$

and $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})R_{k2}}{4h_1 h_2}$.

In the top-right corner, we have that

$$p_{ij} = \frac{b_1 R_{i2}}{4h_1 h_2},\ p_{kj} = \frac{R_{k1} R_{k2}}{4h_1 h_2},$$

$$\text{and } p_{ij} \cap p_{kj} = \frac{(R_{k1} - R_{i1}) R_{k2}}{4h_1 h_2}.$$

In the bottom-right corner, we have that

$$p_{ij} = \frac{b_1 (b_2 - R_{i2})}{4h_1 h_2},\ p_{kj} = \frac{R_{k1} (b_2 - R_{k2})}{4h_1 h_2},$$

$$\text{and } p_{ij} \cap p_{kj} = \frac{(R_{k1} - R_{i1})(b_2 - R_{i2})}{4h_1 h_2}.$$

Putting this all together, we have the following covariance formula:

$$
\begin{aligned}
&Cov^*(W_i, W_k) \\
&= \sum_{j=1}^{l} Cov^*(Y_{ij}, Y_{kj}) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 2\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)\left[\frac{(R_{k1}-R_{i1})(b_2-(R_{i2}-R_{k2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 2\right)\left[\frac{(R_{k1}-R_{i1})(b_2-R_{i2})}{4h_1h_2} - \left(\frac{b_1(b_2-R_{i2})}{4h_1h_2}\right)\left(\frac{b_1(b_2-R_{k2})}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 2\right)\left[\frac{(R_{k1}-R_{i1})R_{k2}}{4h_1h_2} - \left(\frac{b_1R_{i2}}{4h_1h_2}\right)\left(\frac{b_1R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)\left[\frac{(R_{k1}-R_{i1})(b_2-(R_{i2}-R_{k2}))}{4h_1h_2} - \left(\frac{(b_1-R_{i1})b_2}{4h_1h_2}\right)\left(\frac{b_1b_2}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_2^{\alpha-1}\right\rceil - 1\right)\left[\frac{(R_{k1}-R_{i1})(b_2-(R_{i2}-R_{k2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)\left(\frac{R_{k1}b_2}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(R_{k1}-R_{i1})(b_2-R_{i2})}{4h_1h_2} - \left(\frac{(b_1-R_{i1})(b_2-R_{i2})}{4h_1h_2}\right)\left(\frac{b_1(b_2-R_{k2})}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(R_{k1}-R_{i1})R_{k2}}{4h_1h_2} - \left(\frac{(b_1-R_{i1})R_{i2}}{4h_1h_2}\right)\left(\frac{b_1R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(R_{k1}-R_{i1})R_{k2}}{4h_1h_2} - \left(\frac{b_1R_{i2}}{4h_1h_2}\right)\left(\frac{R_{k1}R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(R_{k1}-R_{i1})(b_2-R_{i2})}{4h_1h_2} - \left(\frac{b_1(b_2-R_{i2})}{4h_1h_2}\right)\left(\frac{R_{k1}(b_2-R_{k2})}{4h_1h_2}\right)\right] \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(R_{k1}-R_{i1})(b_2-(R_{i2}-R_{k2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + O\left(\frac{b_1^2b_2}{h_1^2h_2}\right) + O\left(\frac{b_1^2b_2^2}{h_1^2h_2^2}\right) + O\left(\frac{b_1b_2}{h_1h_2}\right) + O\left(\frac{b_1b_2^2}{h_1h_2^2}\right) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(R_{k1}-R_{i1})(b_2-(R_{i2}-R_{k2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] + O\left(\frac{b_1b_2}{h_1h_2}\right)
\end{aligned}
$$

Referring back to Figure 7.21, notice that $R_{i1} - R_{k1} = b_1 - (\tau_{i1} - \tau_{k1}) = b_1 - |\tau_{i1} - \tau_{k1}|$ and that $R_{i2} - R_{k2} = \tau_{i2} - \tau_{k2} = |\tau_{i2} - \tau_{k2}|$.

With this in mind, we can reduce the above formula to the following lemma:

Lemma 7.4.6. *For τ_i and τ_k sufficiently close and in the adjacent horizontal blocks, with τ_i and τ_k removed from the boundary, we have that*

$$
\begin{aligned}
&Cov^*(W_i, W_k) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(b_1 - |\tau_{i1} - \tau_{k1}|)(b_2 - |\tau_{i2} - \tau_{k2}|)}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + O\left(\frac{b_1b_2}{h_1h_2}\right)
\end{aligned}
$$

This is precisely the same formula that we found in cases (i) and (ii).

The only difference is in the explicit formula for the product of p_{ij} and p_{kj}. The formula was simpler in the first two cases, whereas it is more involved in this setting. Still, that only contributes a negligible amount to the overall covariance. Indeed, it is of the same order as the error terms from the first two cases.

Case (iv): τ_i and τ_k occur in different blocks, vertically.

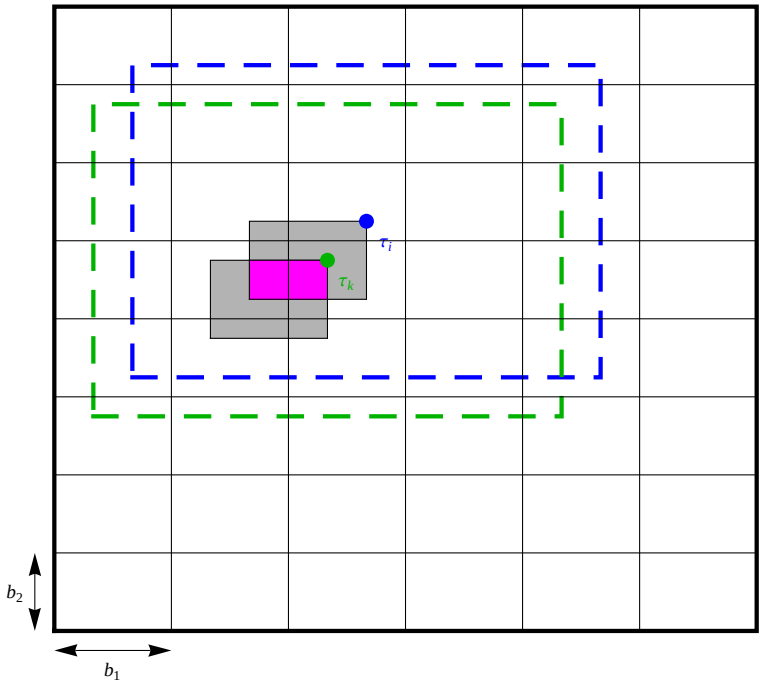

Figure 7.28: Covariance formula for two dimensions, Case (iv)

Notice that there are a total of $\left(2\left\lceil b_1^{\alpha-1}\right\rceil + 1\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil + 2\right)$ blocks contained within the boundaries of the dashed rectangles in Figure 7.28.

However, the top row and bottom row of blocks do not contain both sets of dashed rectangles, and hence the covariance will be zero in those blocks. So, in actuality, we only need to consider $\left(2\left\lceil b_1^{\alpha-1}\right\rceil + 1\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)$ blocks.

And before we continue, it will be helpful to write out explicitly the dimensions of the purple region. Zooming in on the appropriate 3×2 blocks, we have Figure 7.29.

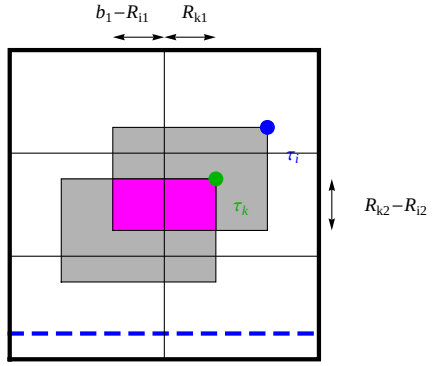

Figure 7.29: Covariance formula for two dimensions, Case (iv), enhanced

We could go through the six sets of blocks as we have done in the previous cases, but that is not needed here. Indeed, notice that the only difference between case (iii) and case (iv) is that all of the subscripts that were 1's are now 2's.

As a result, the covariance calculations will be exactly the same as in case (iii). Thus, we proceed directly to the next lemma.

Lemma 7.4.7. *For τ_i and τ_k sufficiently close and in the adjacent vertical blocks, with τ_i and τ_k removed from the boundary, we have that*

$$\begin{aligned} &Cov^*(W_i, W_k) \\ &= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(b_1 - |\tau_{i1} - \tau_{k1}|)(b_2 - |\tau_{i2} - \tau_{k2}|)}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\ &+ O\left(\frac{b_1b_2}{h_1h_2}\right) \end{aligned}$$

Case (v): τ_i and τ_k occur in neighboring diagonal blocks.

Notice that there are a total of $\left(2\left\lceil b_1^{\alpha-1}\right\rceil + 2\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil + 2\right)$ blocks contained within the boundaries of the dashed rectangles in Figure 7.30.

However, the top and bottom rows as well as the left-most and right-most blocks do not contain both sets of dashed rectangles, and hence the covariance will be zero in those blocks. So, in actuality, we only need to consider $\left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)$ blocks.

Also, the purple area is not split across multiple blocks. Thus, the covariance will be the same in all instances, greatly reducing the complexity of our computations.

In fact, the purple area is just

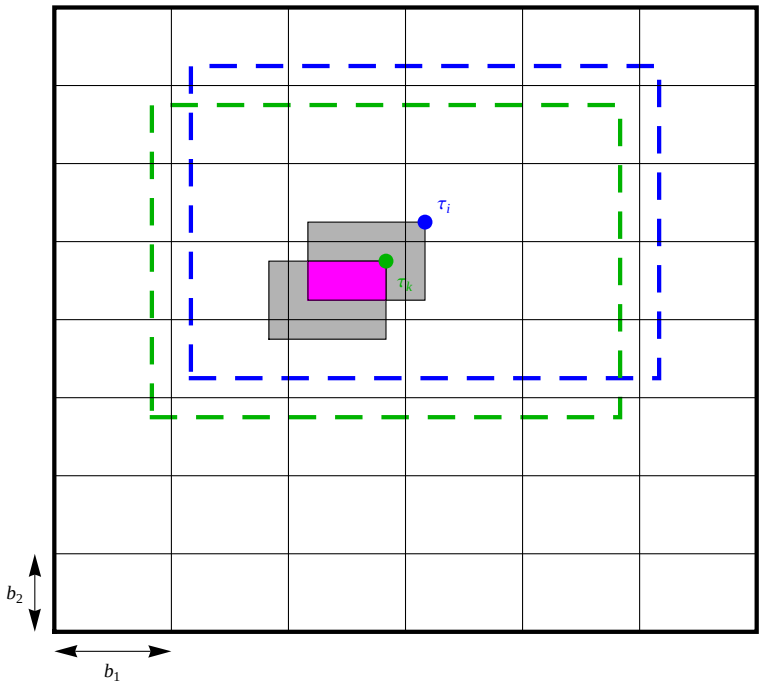

Figure 7.30: Covariance formula for two dimensions, Case (v)

$$\begin{aligned}(R_{k1} - R_{i1})(R_{k2} - R_{i2}) &= (b_1 - (\tau_{i1} - \tau_{k1}))(b_2 - (\tau_{i2} - \tau_{k2})) \\ &= (b_1 - |\tau_{i1} - \tau_{k1}|)(b_2 - |\tau_{i2} - \tau_{k2}|)\end{aligned}$$

As in the previous cases, we shall break down the covariance calculations into groups of blocks. We begin with the middle blocks.

Notice that there are a total of $\left(2\left\lceil b_1^{\alpha-1}\right\rceil - 2\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil - 2\right)$ such blocks highlighted in Figure 7.31. Each of these blocks completely contains the grey rectangles that correspond to τ_i and τ_k, as well as their intersection (the purple area).

Thus, we have that $p_{ij} = \dfrac{b_1 b_2}{4h_1 h_2} = p_{kj}$

and $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(R_{k2} - R_{i2})}{4h_1 h_2}$.

Next, we have the bottom blocks. Notice that there are a total of $2\left\lceil b_1^{\alpha-1}\right\rceil - 2$ such blocks highlighted in Figure 7.32. Each of these blocks contains the complete width, b_1, of the grey blocks, but the heights vary.

Indeed, we have that $p_{ij} = \dfrac{b_1(b_2 - R_{i2})}{4h_1 h_2}$ and $p_{kj} = \dfrac{b_1 b_2}{4h_1 h_2}$.

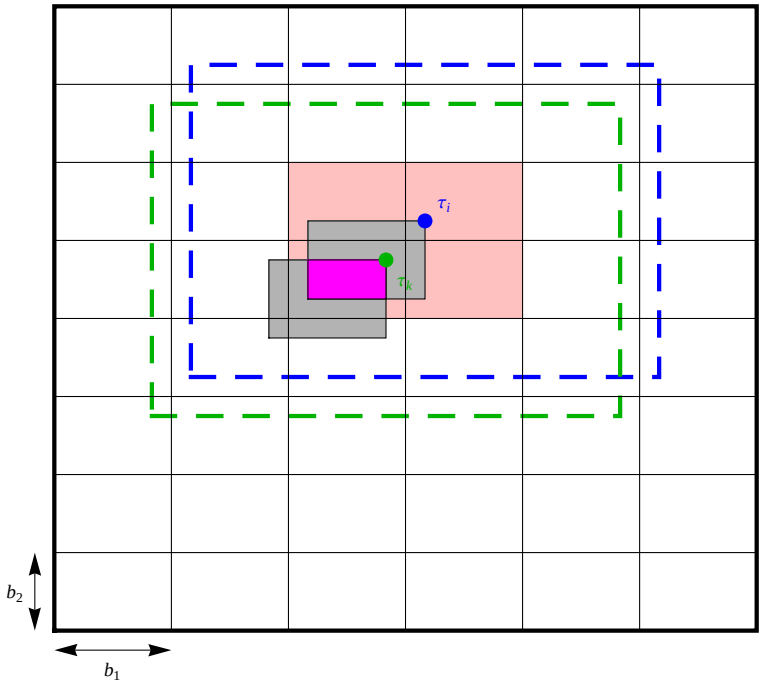

Figure 7.31: Covariance formula for two dimensions, Case (v), middle blocks

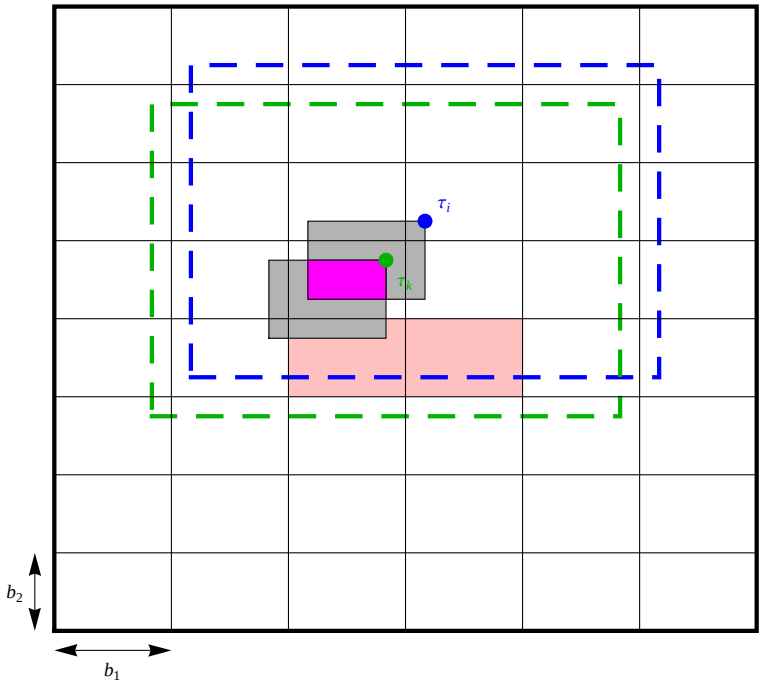

Figure 7.32: Covariance formula for two dimensions, Case (v), bottom blocks

Again, $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(R_{k2} - R_{i2})}{4h_1h_2}$.

Next, we have the top blocks. Notice that there are a total of $2\left\lceil b_1^{\alpha-1}\right\rceil - 2$ such blocks highlighted in Figure 7.33. Each of these blocks contains the complete width, b_1, of the grey blocks, but the heights vary.

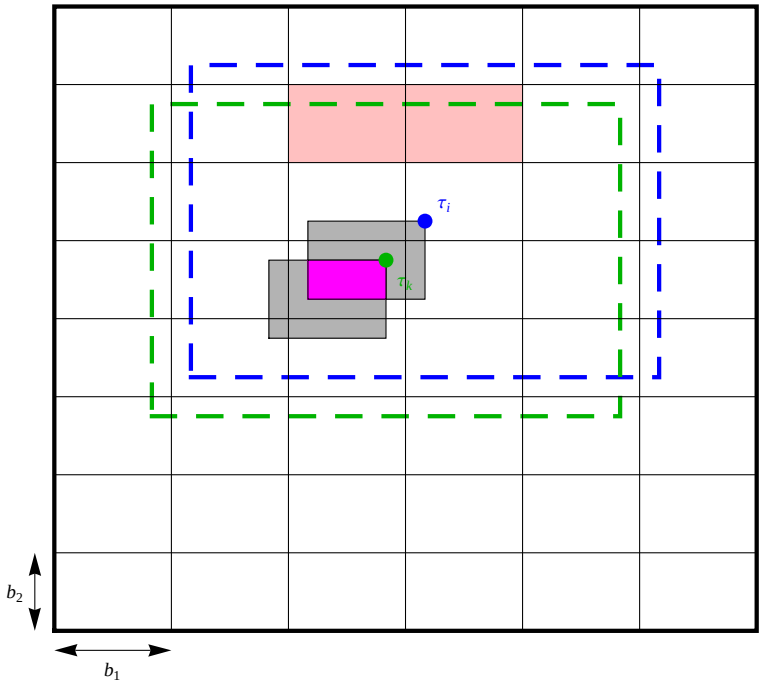

Figure 7.33: Covariance formula for two dimensions, Case (v), top blocks

Indeed, we have that $p_{ij} = \dfrac{b_1b_2}{4h_1h_2}$ and $p_{kj} = \dfrac{b_1R_{k2}}{4h_1h_2}$.

Again, $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(R_{k2} - R_{i2})}{4h_1h_2}$.

Next, we have the left blocks. Notice that there are a total of $2\left\lceil b_2^{\alpha-1}\right\rceil - 2$ such blocks highlighted in Figure 7.34. Each of these blocks contains the complete height, b_2, of the grey blocks, but the widths vary.

Indeed, we have that $p_{ij} = \dfrac{(b_1 - R_{i1})b_2}{4h_1h_2}$ and $p_{kj} = \dfrac{b_1b_2}{4h_1h_2}$.

Again, $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(R_{k2} - R_{i2})}{4h_1h_2}$.

Next, we have the right blocks. Notice that there are a total of $2\left\lceil b_2^{\alpha-1}\right\rceil - 2$ such blocks

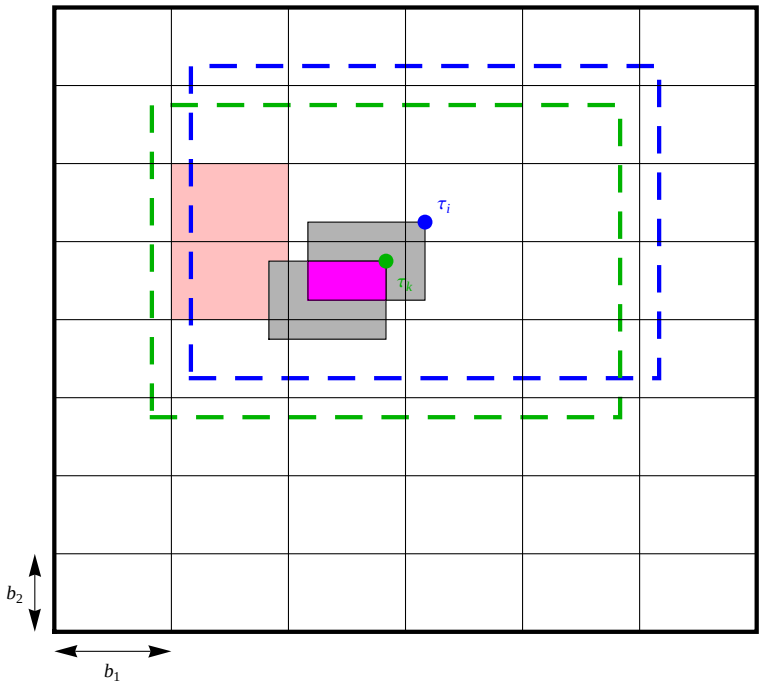

Figure 7.34: Covariance formula for two dimensions, Case (v), left blocks

highlighted in Figure 7.35. Each of these blocks contains the complete height, b_2, of the grey blocks, but the widths vary.

Indeed, we have that $p_{ij} = \dfrac{b_1 b_2}{4h_1 h_2}$ and $p_{kj} = \dfrac{R_{k1} b_2}{4h_1 h_2}$.

Again, $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(R_{k2} - R_{i2})}{4h_1 h_2}$.

Finally, there are four corners to deal with as highlighted in Figure 7.36. We shall compute those probabilities working clockwise from the bottom-left.

In the bottom-left corner, we have that

$$p_{ij} = \frac{(b_1 - R_{i1})(b_2 - R_{i2})}{4h_1 h_2},\ p_{kj} = \frac{b_1 b_2}{4h_1 h_2},$$

and $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(R_{k2} - R_{i2})}{4h_1 h_2}$.

In the top-left corner, we have that

$$p_{ij} = \frac{(b_1 - R_{i1}) b_2}{4h_1 h_2},\ p_{kj} = \frac{b_1 R_{k2}}{4h_1 h_2},$$

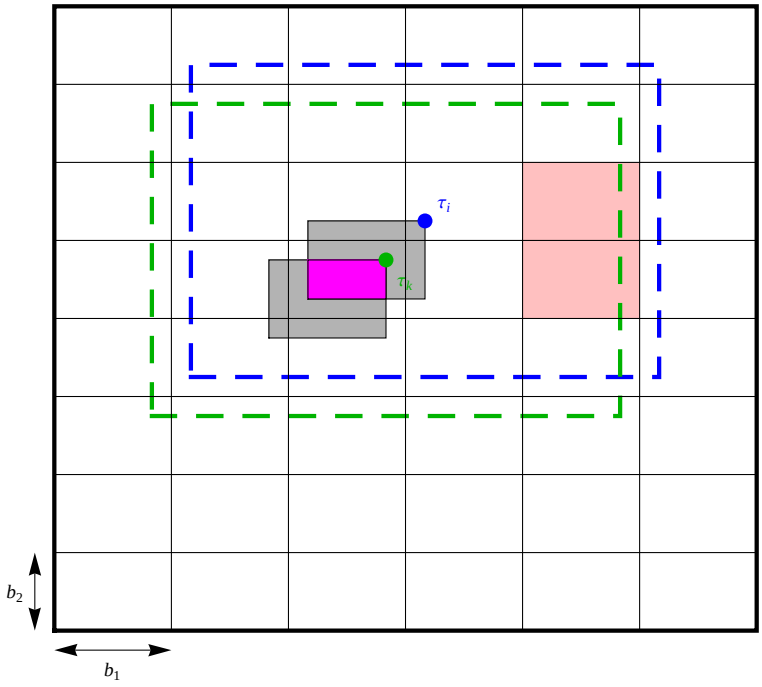

Figure 7.35: Covariance formula for two dimensions, Case (v), right blocks

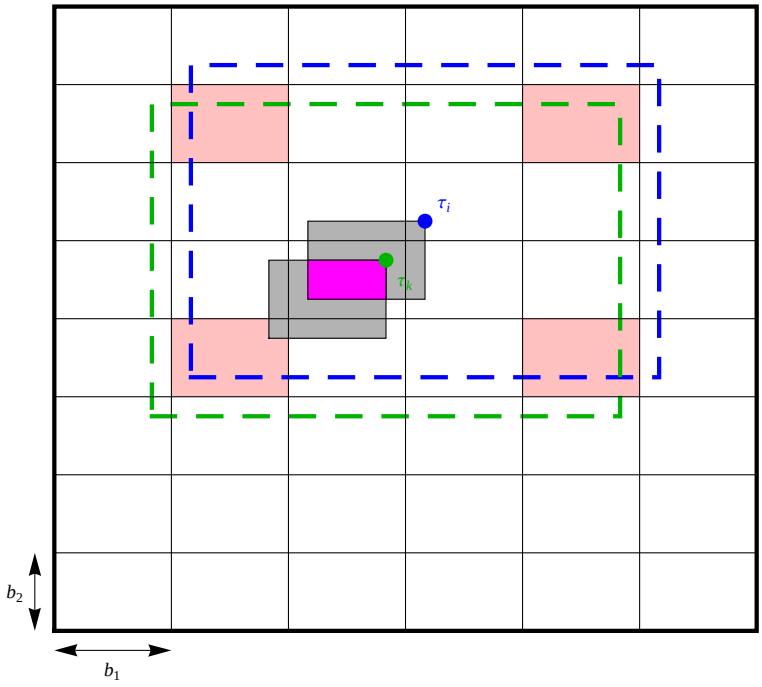

Figure 7.36: Covariance formula for two dimensions, Case (v), corner blocks

and $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(R_{k2} - R_{i2})}{4h_1h_2}$.

In the top-right corner, we have that

$$p_{ij} = \frac{b_1b_2}{4h_1h_2},\ p_{kj} = \frac{R_{k1}R_{k2}}{4h_1h_2},$$

and $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(R_{k2} - R_{i2})}{4h_1h_2}$.

In the bottom-right corner, we have that

$$p_{ij} = \frac{b_1(b_2 - R_{i2})}{4h_1h_2},\ p_{kj} = \frac{R_{k1}b_2}{4h_1h_2},$$

and $p_{ij} \cap p_{kj} = \dfrac{(R_{k1} - R_{i1})(R_{k2} - R_{i2})}{4h_1h_2}$.

Putting this all together, we have the following covariance formula:

$$
\begin{aligned}
&Cov^*(W_i, W_k) \\
&= \sum_{j=1}^{l} Cov^*(Y_{ij}, Y_{kj}) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 2\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil - 2\right)\left[\frac{(R_{k1}-R_{i1})(R_{k2}-R_{i2})}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 2\right)\left[\frac{(R_{k1}-R_{i1})(R_{k2}-R_{i2})}{4h_1h_2} - \left(\frac{b_1(b_2-R_{i2})}{4h_1h_2}\right)\left(\frac{b_1b_2}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_1^{\alpha-1}\right\rceil - 2\right)\left[\frac{(R_{k1}-R_{i1})(R_{k2}-R_{i2})}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)\left(\frac{b_1R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_2^{\alpha-1}\right\rceil - 2\right)\left[\frac{(R_{k1}-R_{i1})(R_{k2}-R_{i2})}{4h_1h_2} - \left(\frac{(b_1-R_{i1})b_2}{4h_1h_2}\right)\left(\frac{b_1b_2}{4h_1h_2}\right)\right] \\
&\quad + \left(2\left\lceil b_2^{\alpha-1}\right\rceil - 2\right)\left[\frac{(R_{k1}-R_{i1})(R_{k2}-R_{i2})}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)\left(\frac{R_{k1}b_2}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(R_{k1}-R_{i1})(R_{k2}-R_{i2})}{4h_1h_2} - \left(\frac{(b_1-R_{i1})(b_2-R_{i2})}{4h_1h_2}\right)\left(\frac{b_1b_2}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(R_{k1}-R_{i1})(R_{k2}-R_{i2})}{4h_1h_2} - \left(\frac{(b_1-R_{i1})b_2}{4h_1h_2}\right)\left(\frac{b_1R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(R_{k1}-R_{i1})(R_{k2}-R_{i2})}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)\left(\frac{R_{k1}R_{k2}}{4h_1h_2}\right)\right] \\
&\quad + \left[\frac{(R_{k1}-R_{i1})(R_{k2}-R_{i2})}{4h_1h_2} - \left(\frac{b_1(b_2-R_{i2})}{4h_1h_2}\right)\left(\frac{R_{k1}b_2}{4h_1h_2}\right)\right] \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(R_{k1}-R_{i1})(b_2-(R_{i2}-R_{k2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + O\left(\frac{b_1^2b_2}{h_1^2h_2}\right) + O\left(\frac{b_1^2b_2^2}{h_1^2h_2^2}\right) + O\left(\frac{b_1b_2}{h_1h_2}\right) + O\left(\frac{b_1b_2^2}{h_1h_2^2}\right) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(R_{k1}-R_{i1})(b_2-(R_{i2}-R_{k2}))}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] + O\left(\frac{b_1b_2}{h_1h_2}\right)
\end{aligned}
$$

Making use of the observation $(R_{k1}-R_{i1})(R_{k2}-R_{i2}) = (b_1-|\tau_{i1}-\tau_{k1}|)(b_2-|\tau_{i2}-\tau k2|)$, we can reduce the above formula to the following lemma:

Lemma 7.4.8. *For τ_i and τ_k sufficiently close and in neighboring diagonal blocks, with τ_i and τ_k removed from the boundary, we have that*

$$
\begin{aligned}
&Cov^*(W_i, W_k) \\
&= \left(2\left\lceil b_1^{\alpha-1}\right\rceil\right)\left(2\left\lceil b_2^{\alpha-1}\right\rceil\right)\left[\frac{(b_1-|\tau_{i1}-\tau_{k1}|)(b_2-|\tau_{i2}-\tau_{k2}|)}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] \\
&\quad + O\left(\frac{b_1b_2}{h_1h_2}\right)
\end{aligned}
$$

Notice that in all of the lemmas, the formulas are identical. As was the case in one-dimension, the formulas themselves are slightly different, but are equal up to an error term of the same order. As we shall see later, this is asymptotically negligible.

Moreover, we can rewrite the above formula as:

$$\begin{aligned}
&Cov^*(W_i, W_k) \\
&= \frac{4h_1h_2}{b_1b_2}\left[\frac{(b_1 - |\tau_{i1} - \tau_{k1}|)(b_2 - |\tau_{i2} - \tau_{k2}|)}{4h_1h_2} - \left(\frac{b_1b_2}{4h_1h_2}\right)^2\right] + O\left(\frac{b_1b_2}{h_1h_2}\right) \\
&= \frac{(b_1 - |\tau_{i1} - \tau_{k1}|)(b_2 - |\tau_{i2} - \tau_{k2}|)}{b_1b_2} + O\left(\frac{b_1b_2}{h_1h_2}\right) \\
&= \left(1 - \frac{|\tau_{i1} - \tau_{k1}|}{b_1}\right)\left(1 - \frac{|\tau_{i2} - \tau_{k2}|}{b_2}\right) + O\left(\frac{b_1b_2}{h_1h_2}\right)
\end{aligned}$$

This same approach can be readily extended to three dimensions. The calculations are identical to those above, but verification will require nearly twice as many cases to consider. To help visualize this, consider Figure 7.37.

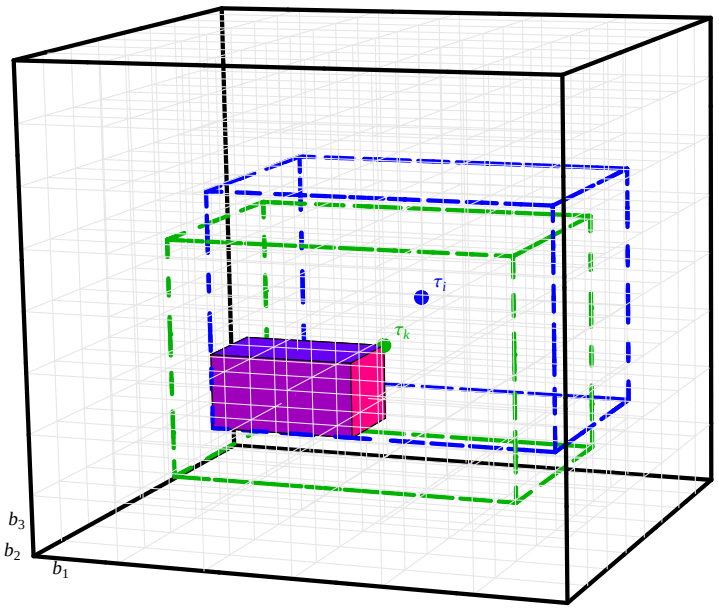

Figure 7.37: Covariance formula for three dimensions

For higher dimensions, we again need to modify our definitions. Again, this is a completely natural extension of the definition from lower dimensions.

Definition 7.4.9. Suppose $\tau_i, \tau_k \in \mathbb{R}^d$. If $|\tau_{ij} - \tau_{kj}| < b_j$ for $j = 1, 2, \ldots, d$, then we say that τ_i and τ_k are **sufficiently close**.

Definition 7.4.10. Given a d-dimensional "box" specified by $[a_{11}, a_{12}] \times \ldots \times [a_{d1}, a_{d2}]$, we shall say that a point $\tau_i \in \mathbb{R}^d$ is **removed from the boundary** provided that $|\tau_{ij} - a_{j1}| > h_j$ and $|\tau_{ij} - a_{j2}| > h_j$ for $j = 1, 2, \ldots, d$.

Now, we are finally ready to present our main result for the covariance of W_i and W_k in d-dimensions.

Theorem 7.4.11. *For τ_i and τ_k sufficiently close with τ_i and τ_k removed from the boundary, we have that*

$$Cov^*(W_i, W_k) = \prod_{j=1}^{d} \left(1 - \frac{|\tau_{ij} - \tau_{kj}|}{b_j}\right) + O\left(\prod_{j=1}^{d} \frac{b_j}{h_j}\right)$$

We remark that the formula has exactly the same structure as the one-dimensional formula developed in Chapter 4, but there are products in each dimension. Moreover, the projection of a d-dimensional Poisson process onto one dimension is also a Poisson process. (See [63] for a more detailed discussion.) Because of this, the results that we developed above can be adapted to our more general setting.

We have the following formula for the variance:

$$\begin{aligned}
&Var^*\left[\sqrt{\Lambda(K)}\,\tilde{\tilde{X}}^*\right] \\
&= \frac{1}{\Lambda(K)} \sum_{i \neq k} \left[\prod_{j=1}^{d} \left(1 - \frac{|\tau_{ij} - \tau_{kj}|}{b_j}\right) 1_{\{|\tau_{ij} - \tau_{kj}| < b\}}\right] X(\tau_i) X(\tau_k) \\
&+ \frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} X^2(\tau_i) \\
&+ \frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} \left[O\left(\prod_{j=1}^{d} \left(\frac{b_j}{h_j}\right) 1_{\{|\tau_{ij} - \tau_{kj}| < b\}}\right)\right] X(\tau_i) X(\tau_k)
\end{aligned}$$

Similar to the one-dimensional case, we have the following results:

Lemma 7.4.12. *Assuming A.7.1, A.7.5, and A.7.6, we have that*

$$\frac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} \sum_{k=1}^{N(K)} \left[O\left(\prod_{j=1}^{d} \left(\frac{b_j}{h_j}\right) 1_{\{|\tau_{ij} - \tau_{kj}| < b\}}\right)\right] X(\tau_i) X(\tau_k) \xrightarrow{p} 0.$$

Thus, the last term is asymptotically negligible and can be safely ignored.

Lemma 7.4.13. *Let* $\hat{R}(\boldsymbol{0}) = \dfrac{1}{\Lambda(K)} \sum_{i=1}^{N(K)} X^2(\tau_i)$. *Then, under Assumptions A.4.1-A.4.3,*

i. $E[\hat{R}(\boldsymbol{0})] = R(\boldsymbol{0}) + \mu^2$

ii. $Var[\hat{R}(\boldsymbol{0})] \to 0$ *as* $diam(K) \to \infty$

Again, as was the case with the one-dimensional case, to establish Lemma 7.4.13(ii), we shall appeal to Lemma 4.4.12. (Note that the lemma, as stated, applies to $\mathbb{R}^d$, so no additional work is needed on our part.)

We have the following result pertaining to the unequal terms in the variance formula.

Lemma 7.4.14. *Suppose Assumptions A.7.1-A.7.4, and A.7.6 hold.*

Let $\hat{\theta}^2 = \dfrac{1}{\Lambda(K)} \sum_{i \neq k} \left[\prod_{j=1}^{d} \left(1 - \dfrac{|\tau_{ij} - \tau_{kj}|}{b_j} \right) 1_{\{|\tau_{ij} - \tau_{kj}| < b\}} \right] X(\tau_i) X(\tau_k)$. *Then*

i. $E[\hat{\theta}^2] \xrightarrow{p} \theta^2$

ii. $Var[\hat{\theta}^2] \to 0$ *as* $diam(K) \to \infty$

where $\theta^2 = \lim_{diam(K) \to \infty} \dfrac{1}{\Lambda(K)} \int_K \int_K R(\boldsymbol{s} - \boldsymbol{t}) \lambda(\boldsymbol{s}) \lambda(\boldsymbol{t}) d\boldsymbol{s} d\boldsymbol{t}$

Lemmas 7.4.12 - 7.4.14 give us the following result.

Lemma 7.4.15. *Suppose Assumptions A.7.1-A.7.6 hold. As* $diam(K) \to \infty$, *we have that* $Var^*[\sqrt{\Lambda(K)} \tilde{\tilde{X}}^*] \xrightarrow{p} \theta^2 + R(\boldsymbol{0}) + \mu^2$.

Now that we have established that the bootstrap variance tends (asymptotically) to the true variance, we are ready to state our main theorem.

Theorem 7.4.16. *Suppose Assumptions A.7.1-A.7.6 hold. Then we have the following:*

i. $E^* \left[\tilde{\tilde{X}}^* \right] = \tilde{X}_K$

ii. $\dfrac{Var^*[\tilde{\tilde{X}}^*]}{Var[\tilde{X}_K]} \xrightarrow{p} 1$

iii. $\sup_x \left| P^* \left(\sqrt{\Lambda(K)} (\tilde{\tilde{X}}^* - \tilde{X}_K) \leq x \right) - P \left(\sqrt{\Lambda(K)} (\tilde{X}_K - \mu) \leq x \right) \right| \xrightarrow{p} 0$

As in the one-dimensional setting, instead of using torodial wrapping, we could simply reduce the window that our shift can move each block near the boundary. This approach may be preferrable in cases where the intensity at the edges differs significantly.

Again, the drawback for this approach, though, is that not all points will be resampled once (on average). Indeed, those points within h units of the boundary will appear (on average) less than once. The solution for this is to modify our theorems not to subtract $\tilde{X}_K$, but rather subtract $E^*[\tilde{X}^*]$. This recenters the data around the bootstrap mean.

In the theory above, it was assumed that K was a "box" in d dimensions. However, our methods can be extended to the case where K is an arbitrary convex set in $\mathbb{R}^d$. The most straight-forward approach is to decompose K into a collection of non-overlapping boxes and apply our resampling method to each segment. While this would introduce additional edge effects, if the region can be covered (or at least well-approximated) by a few rectangles, then the impact would be minimal.

7.5 Simulations

While the following algorithm can be adapted to dimensions greater than 2, the compuational complexity of the problem leads us to only consider the case of $d = 2$. Unlike the one-dimensional setting in Section 4.5, we need to restrict our region to something much more modest, $[0, 20] \times [0, 20]$, for example. (Note: The region need not be a square. In fact, we can use *any* convex region in $\mathbb{R}^d$.)

The first step is to generate data from a two-dimensional marked point process. Again, since we have assumed independence between the point process and the associated marks, the first step is to generate an inhomogeneous Poisson process.

We do this using the Accept-Reject Method as specified by Lewis and Shedler ([43]). The idea is to generate a two-dimensional homogeneous Poisson process on the rectangle $[0, x_{max}] \times [0, y_{max}]$, and then use thinning to create an inhomogeneous Poisson process.

Formally, we have the following algorithm:

1. Set $i = 0$, $j = 0$ and $X(0) = 0$
2. Generate a random number $U_1 \sim U(0, 1)$
3. While $X(i) < x_{max}$, set $i = i + 1$ and $X(i) = X(i-1) - \ln(U_1)/(\lambda y_{max})$
4. Generate another random number $Y(i) \sim U(0, y_{max})$
5. Generate another random number $U_2 \sim U(0, 1)$
6. If $U_2 \leq \lambda(X(i), Y(i))/\lambda$, set $j = j + 1$ and $P(k) = (X(i), Y(i))$
7. Go to step 2.

The output will consist of the counter j, which is the number of events that occur on the region K and $P(1), \ldots, P(j)$ will constitute the location of the events.

Next, we need to generate the corresponding marks. (We use the same code as in the one-dimensional case.)

We assume that $X(\cdot)$ is a stationary, Gaussian process with mean 0 and covariance function $R(t)$. Our algorithm for generating the marks is as follows:

1. Initialize a matrix of covariances, Σ with dimensions given by $N(K)$, the number of points found in the algorithm above
2. Populate Σ with values $R(P(i) - P(j))$ for all pairs i, j
3. Consider the eigenvalue decomposition of $\Sigma = VDV^T$ and express $\Sigma^{1/2} = VD^{1/2}V^T$
4. Generate $N(K)$ i.i.d. $N(0, 1)$ random variables and place them into the vector Z
5. Set the marks equal to $X = \Sigma^{1/2}Z$

Finally, we record the pair $\{P(j), X(P(j))\}$. This is our randomly generated marked point process.

We run simulations using using the statistical software `R` [61]. We use two resampling methods (torodial wrapping and no wrapping), two different covariance functions, five different models, and four combinations of b_1, b_2, h_1 and h_2 for a total of 80 models. Each model was simulated 1,000 times and each time, a 95% confidence interval was constructed.

The five different intensity functions considered are as follows:

First, we consider $\lambda_1(\mathbf{t}) = 1 + 9\exp\left(-\frac{1}{10}\left((t_1 - 10)^2 + (t_2 - 10)^2\right)\right)$ for $0 \leq t_1 \leq 20$, $0 \leq t_2 \leq 20$. This is shown in Figure 7.38.

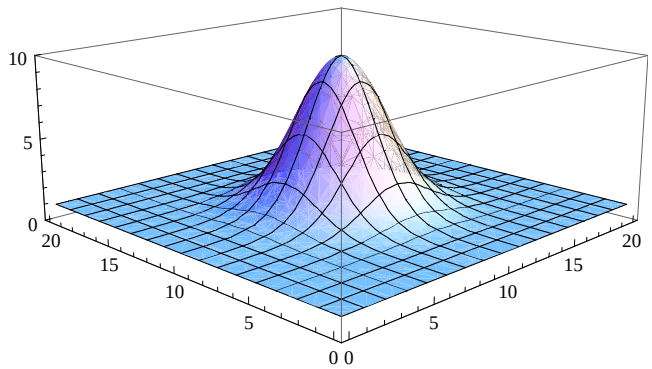

Figure 7.38: Two-Dimensional Inhomogeneous Simulation Intensity Function 1

Second, we consider $\lambda_1(\mathbf{t}) = 1 + 9\exp\left(-\frac{1}{100}\left((t_1-10)^2+(t_2-10)^2\right)\right)$ for $0 \leq t_1 \leq 20$, $0 \leq t_2 \leq 20$. This is shown in Figure 7.39.

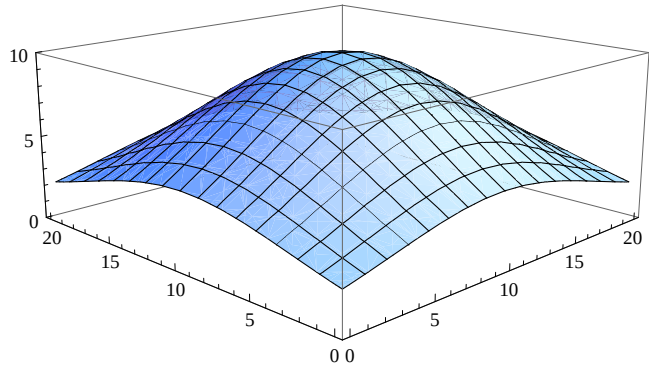

Figure 7.39: Two-Dimensional Inhomogeneous Simulation Intensity Function 2

Third, we consider $\lambda_3(\mathbf{t}) = 3 + 2\sin\left(\frac{2\pi}{15}t_1\right)$ for $0 \leq t_1 \leq 30$, $0 \leq t_2 \leq 20$. This is shown in Figure 7.40.

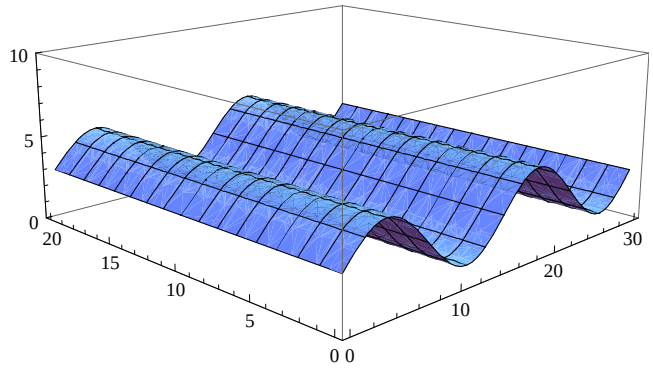

Figure 7.40: Two-Dimensional Inhomogeneous Simulation Intensity Function 3

Fourth, we consider $\lambda_4(\mathbf{t}) = \frac{1}{2} + \frac{1}{4}t_1 + \frac{1}{32}t_2$ for $0 \leq t_1 \leq 16$, $0 \leq t_2 \leq 64$. This is shown in Figure 7.41.

Fifth, we consider $\lambda_5(\mathbf{t}) = 2 + 1.5\cos\left(\frac{\pi}{4}\left((t_1-25)^2+(t_2-5)^2\right)\right)$ for $0 \leq t_1 \leq 50$, $0 \leq t_2 \leq 15$. This is shown in Figure 7.42.

In the tables that follow, we present the mean and standard deviation of the constructed confidence intervals as well as the percentage of confidence intervals containing the true mean, 0 (the coverage probability).

Looking at all of the tables, we see that torodial wrapping performs better than non-

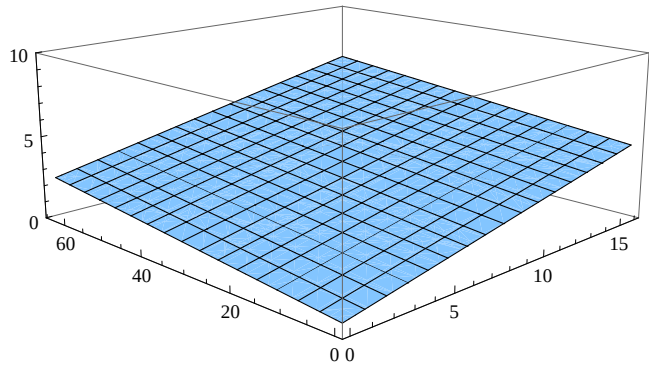

Figure 7.41: Two-Dimensional Inhomogeneous Simulation Intensity Function 4

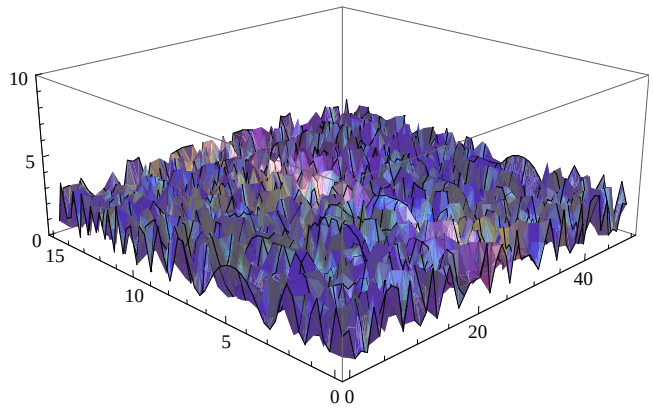

Figure 7.42: Two-Dimensional Inhomogeneous Simulation Intensity Function 5

wrapping by approximately 10%, though noticeable exceptions occur in Table 7.4 and Table 7.5, where wrapping outperforms non-wrapping by 20-40%. Usually, one would expect the edge effects to hinder performance and thus *reduce* coverage, as was seen in Chapter 4.

The most likely explanation is the issues that arise with using relatively small regions (and relatively large blocks). Due to the computational complexities that arise when increasing dimensions (i.e., "the curse of dimensionality"), a fair amount of the regions under consideration have blocks near the edge. Still, one would not expect such a drastic difference and this will be a source of further study.

We also see that the mean and standard deviation of the lengths of the confidence intervals are comparable to the one-dimensional case (0.1 and 0.025, respectively), so our algorithm appears to be fairly stable.

When comparing Tables 7.1 - 7.5 with Tables 7.6 - 7.10, we see that the increased dependence leads to a lower coverage probability. That is to be expected.

In Table 7.1 and Table 7.2, we see that the choices of parameters does not make a significant difference with regards to the coverage probabilities, though again a smaller choice for b_i seems to produce higher coverage. The effect becomes noticeable when considering Table 7.3 and Table 7.4. There, the difference between the smallest choice of b_i and the largest choice is around 30%.

This suggests that care must be taken when choosing the parameters. As was discussed in Chapter 2 and Chapter 4, methods exist for choosing the local parameter and methods will need to be developed for selecting a block size for higher dimensions. Such a data-driven approach will be the focus of future work.

Similar results holds for Table 7.6 - 7.10.

Table 7.1: Two-Dimensional LBB with Intensity function $\lambda_1(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.25	1.25	2	2	658.813	0.11819	0.00575	89.9%
	2.00	1.25	8	2	668.912	0.11144	0.00528	82.2%
	1.00	1.25	2	3	659.653	0.11771	0.00596	89.3%
	1.00	1.00	2	2	660.186	0.11807	0.00531	94.7%
Yes	1.25	1.25	2	2	663.840	0.11739	0.00567	93.5%
	2.00	1.25	8	2	666.714	0.11716	0.00573	84.7%
	1.00	1.25	2	3	665.070	0.11750	0.00571	90.7%
	1.00	1.00	2	2	664.460	0.11747	0.00528	97.1%

Table 7.2: Two-Dimensional LBB with Intensity function $\lambda_2(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.25	1.25	2	2	2431.567	0.06656	0.00177	72.3%
	2.00	1.25	8	2	2565.542	0.05881	0.00224	76.4%
	1.00	1.25	2	3	2443.673	0.06681	0.00181	68.8%
	1.00	1.00	2	2	2424.828	0.06708	0.00177	73.1%
Yes	1.25	1.25	2	2	2399.877	0.06863	0.00195	83.2%
	2.00	1.25	8	2	2393.463	0.06872	0.00300	81.0%
	1.00	1.25	2	3	2400.606	0.06861	0.00187	87.7%
	1.00	1.00	2	2	2399.061	0.06858	0.00179	92.2%

Table 7.3: Two-Dimensional LBB with Intensity function $\lambda_3(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.25	1.25	2	2	1874.802	0.11289	0.00352	84.3%
	2.00	1.25	8	2	1875.239	0.11354	0.00393	58.3%
	1.00	1.25	2	3	1877.288	0.11299	0.00356	89.3%
	2.00	2.00	8	8	1846.922	0.11459	0.00464	52.2%
Yes	1.25	1.25	2	2	1856.771	0.11488	0.00359	86.7%
	2.00	1.25	8	2	1853.601	0.11481	0.00393	70.9%
	1.00	1.25	2	3	1856.222	0.11507	0.00358	89.1%
	2.00	2.00	8	8	1859.261	0.11447	0.00501	58.3%

Table 7.4: Two-Dimensional LBB with Intensity function $\lambda_4(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.0	1.6	2	5	3559.368	0.05407	0.00115	72.6%
	1.0	4.0	2	16	3753.267	0.05412	0.00172	56.1%
	1.6	2.0	3	8	3573.271	0.05414	0.00120	38.4%
	1.6	4.0	3	16	3578.571	0.05414	0.00162	38.7%
Yes	1.0	1.6	2	5	3572.319	0.05386	0.00133	93.2%
	1.0	4.0	2	16	3572.690	0.05381	0.00182	69.3%
	1.6	2.0	3	8	3571.011	0.05387	0.00140	83.3%
	1.6	4.0	3	16	3572.810	0.05383	0.00196	64.6%

Table 7.5: Two-Dimensional LBB with Intensity function $\lambda_5(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.25	1.00	3	2	1587.845	0.09447	0.00263	58.3%
	1.25	1.25	3	3	1583.023	0.09454	0.00272	55.8%
	2.00	1.00	8	2	1582.945	0.09305	0.00259	40.1%
	2.00	1.25	8	3	1579.912	0.09325	0.00291	37.5%
Yes	1.25	1.00	3	2	1573.039	0.09900	0.00270	98.2%
	1.25	1.25	3	3	1572.671	0.09905	0.00292	92.7%
	2.00	1.00	8	2	1571.596	0.09921	0.00274	93.7%
	2.00	1.25	8	3	1574.156	0.09875	0.00311	78.5%

Table 7.6: Two-Dimensional LBB with Intensity function $\lambda_1(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|/3)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.25	1.25	2	2	668.056	0.10484	0.00467	91.9%
	2.00	1.25	8	2	669.935	0.10512	0.00438	74.2%
	1.00	1.25	2	3	670.256	0.10544	0.00472	95.9%
	1.00	1.00	2	2	670.278	0.10504	0.00428	96.1%
Yes	1.25	1.25	2	2	679.286	0.10573	0.00485	93.6%
	2.00	1.25	8	2	679.770	0.10575	0.00440	76.2%
	1.00	1.25	2	3	677.376	0.10593	0.00485	94.9%
	1.00	1.00	2	2	678.235	0.10581	0.00447	98.2%

Table 7.7: Two-Dimensional LBB with Intensity function $\lambda_2(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|/3)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.25	1.25	2	2	2542.989	0.05309	0.00145	80.5%
	2.00	1.25	8	2	2516.383	0.05046	0.00159	55.5%
	1.00	1.25	2	3	2551.763	0.05314	0.00146	84.1%
	1.00	1.00	2	2	2530.640	0.05353	0.00135	87.2%
Yes	1.25	1.25	2	2	2491.238	0.05493	0.00150	85.8%
	2.00	1.25	8	2	2498.064	0.05473	0.00183	57.3%
	1.00	1.25	2	3	2492.586	0.05482	0.00154	86.5%
	1.00	1.00	2	2	2494.139	0.05480	0.00138	92.3%

Table 7.8: Two-Dimensional LBB with Intensity function $\lambda_3(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|/3)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.25	1.25	2	2	1819.024	0.10271	0.00267	61.3%
	2.00	1.25	8	2	1820.306	0.09246	0.00461	43.2%
	1.00	1.25	2	3	1819.114	0.10265	0.00304	68.4%
	2.00	2.00	8	8	1815.099	0.09336	0.00535	47.7%
Yes	1.25	1.25	2	2	1820.761	0.09950	0.00309	87.2%
	2.00	1.25	8	2	1815.847	0.09943	0.00417	64.7%
	1.00	1.25	2	3	1821.131	0.09951	0.00302	90.5%
	2.00	2.00	8	8	1816.536	0.09917	0.00517	56.0%

Table 7.9: Two-Dimensional LBB with Intensity function $\lambda_4(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|/3)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.0	1.6	2	5	3482.352	0.05699	0.00112	81.5%
	1.0	4.0	2	16	3488.159	0.05831	0.00161	64.2%
	1.6	2.0	3	8	3486.516	0.05876	0.00130	49.4%
	1.6	4.0	3	16	3482.737	0.05888	0.00171	47.7%
Yes	1.0	1.6	2	5	3491.297	0.05624	0.00115	91.5%
	1.0	4.0	2	16	3486.862	0.05619	0.00182	70.1%
	1.6	2.0	3	8	3489.723	0.05624	0.00149	78.7%
	1.6	4.0	3	16	3493.034	0.05619	0.00212	62.5%

Table 7.10: Two-Dimensional LBB with Intensity function $\lambda_5(\mathbf{t})$ and $R(\mathbf{t}) = \exp(-|\mathbf{t}|/3)$

Wrap?	b_1	b_2	h_1	h_2	N Obs.	CI Len. Mean	CI Len. SD	Cover %
No	1.25	1.00	3	2	1521.448	0.09553	0.00253	80.4%
	1.25	1.25	3	3	1524.086	0.09559	0.00274	77.2%
	2.00	1.00	8	2	1519.078	0.09588	0.00270	72.3%
	2.00	1.25	8	3	1514.525	0.09618	0.00304	74.6%
Yes	1.25	1.00	3	2	1528.350	0.09401	0.00257	93.2%
	1.25	1.25	3	3	1525.531	0.09434	0.00268	93.1%
	2.00	1.00	8	2	1528.590	0.09387	0.00257	92.5%
	2.00	1.25	8	3	1527.644	0.09380	0.00290	79.2%

7.6 Proofs

Proof of Lemma 7.4.12: The proof follows almost immediately from the proof of Lemma 4.4.10.

Again, we shall focus on the points since $X(\tau_i) = O_p(1)$. Thus, we are left with an integral over K with a product of terms in each dimension.

But since $\int_K$ refers to $\int_0^{K_1}\int_0^{K_2}\cdots\int_0^{K_d}$, we can split the integrand into d single integrals and use the results of Lemma 4.4.10 for each coordinate.

Thus, it follows that

$$\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\left[O\left(\prod_{j=1}^{d}\left(\frac{b_j}{h_j}\right)1_{\{|\tau_{ij}-\tau_{kj}|<b\}}\right)\right] = O_p\left(\prod_{j=1}^{d} b_j K_j\right),$$

since dividing through by $\prod_{j=1}^{d} b_j K_j$ would give us constant values for the mean and the variance.

But by Assumption A.7.1, we have that $\Lambda(K) = O\left(\prod_{j=1}^{d} K_j\right)$. Thus,

$$\frac{1}{\Lambda(K)}\sum_{i=1}^{N(K)}\sum_{k=1}^{N(K)}\left[O\left(\prod_{j=1}^{d}\left(\frac{b_j}{h_j}\right)1_{\{|\tau_{ij}-\tau_{kj}|<b\}}\right)\right]X(\tau_i)X(\tau_k) = O_p\left(\prod_{j=1}^{d}\frac{b_j^2}{h_j}\right)$$

By Assumptions A.7.5 and A.7.6, we have that this tends to 0, completing the proof.

□

Proof of Lemma 7.4.13: The proof of (i) follows be exactly the same reasoning as in the proof of Lemma 4.4.11(i) and the details are omitted.

Many of the steps of (ii) can be copied from the one-dimensional case. For example, using Assumption A.7.1, we bound $\lambda(\mathbf{t})$ by λ_{max} to reduce the variance calculation to a double integral over K of the covariance of $X^2(\mathbf{t})$ and $X^2(\mathbf{s})$.

One notable exception with the change to higher dimensions is the transformation to reduce the double integral to a single integral. Namely, we have that

$$\int_K\int_K \left|Cov(X^2(\mathbf{t}), X^2(\mathbf{s}))\right| d\mathbf{t}d\mathbf{s} \leq |K|\int |R_2(\mathbf{u})|\frac{|K\cap(K-y)|}{|K|}d\mathbf{u}$$

where $K - y = \{x - y : x \in K\}$ and $|\cdot|$ denotes the Lebesgue measure (volume).

Lemma 3.2 in [30] gives us that

$$\lim_{diam(K)\to\infty} \frac{|K \cap (K-y)|}{|K|} = 1$$

However, Lemma 4.4.12 (which we can apply because of Assumption A.7.2) gives us the same bound on the covariance, $R_2(\mathbf{u})$, that we obtained in the one-dimensional case. Using Assumption A.7.3, we have that the integral is finite, so we see that

$$Var\left[\hat{R}(\mathbf{0})\right] = O\left(\frac{1}{\prod_{j=1}^{d} K_j}\right),$$

which tends to 0 as $diam(K) \to \infty$.

□

Proof of Lemma 7.4.14: Again, we appeal to the results from the proof of Lemma 4.4.14. Just as in the proof of Lemma 7.4.12, we recognize that $\int_K$ refers to $\int_0^{K_1}\int_0^{K_2}\cdots\int_0^{K_d}$ and so we can use Lemma 4.4.14 on each of the d components of the formula.

Thus, we have that

$$E\left[\hat{\theta}^2\right] = \theta^2 + O\left(\frac{1}{b_1 b_2 ... b_d}\right)$$

and

$$Var\left[\hat{\theta}^2\right] = O\left(\prod_{j=1}^{d} \frac{b_j^2}{K_j}\right)$$

□

Proof of Lemma 7.4.15: This follows immediately since Lemma 7.4.14 gives us that the first term tends to θ^2, Lemma 7.4.13 gives us that the second term tends to $R(\mathbf{0}) + \mu^2$, and Lemma 7.4.12 gives us that the third term tends to 0.

□

Proof of Theorem 7.4.16: Similar to above, we appeal to the results from the one-dimensional setting. The proof is nearly identical to the proof of Theorem 4.4.16 and so we omit the details here. The key differences are that b is replaced with $\prod_{j=1}^{d} b_j$ and K is replaced with $\prod_{j=1}^{d} K_j$.

□

Chapter 8

Conclusions

8.1 Overview

The objective of this work was to develop methods to resample inhomogeneous marked point processes to construct confidence intervals for the mean. We limited our focus to the subset of Poisson processes and considered how we could adapt existing methods.

In the case of a one-dimensional Poisson process, we can estimate its intensity function and transform the data to a homogeneous Poisson process with rate 1. From there, we can resample the data using a block bootstrap technique to obtained the desired confidence intervals for simulated data.

We saw that a small choice for block size ($b = 2$) and a moderate choice for local window size ($w = 10$) produced the best results (with around 90% coverage). As we noted, though, future work will involve the determination of b and w from the data itself. As the block size increased, there was a notable difference in coverage when the dependence of the marks was increased.

We also considered using a local block bootstrap approach as an alternative method which does not require us to estimate the intensity function. However, we are required to provide a second parameter, h, which is similar in nature to the local window size w used to estimate the intensity. Similar to the homogeneous setting in the existing literature, we establish the asymptotically normality of our estimators and construct confidence intervals for the mean of the simulated data.

Again, we observed that a small block size led to the best results. Also, increasing h led to a reduction in coverage. This is to be expected, as a large h reduces our local block bootstrap to a regular block bootstrap, which is likely to lose information regarding changes in intensities. When considering torodial wrapping and no wrapping, we see that there is a slight improvement when not wrapping the data. This is likely due to edge effects.

When comparing the two approaches, we see that the local block bootstrap consistently

performs as well, if not better than the transformation methods. In fact, for a few models, the coverage using the transformation method was upwards of 60% worse!

We note that the local block bootstrap had mean confidence interval lengths similar to those of the transformed methods, but the standard deviation was ten times smaller. This suggests that our algorithm is more stable than the transformation method. This could explain some of the observed differences, but will be a source of future study.

Finally, we applied the local block bootstrap method to higher dimensional data. In this setting, a transformation to a homogeneous Poisson process would not be feasible unless we make unrealistically strong assumptions on the measure. For that reason, our method is the only method currently available to resample such data.

Similar to the one-dimensional case, we observed that smaller block sizes (in each dimension) led to greater coverage, in general. However, unlike the one-dimensional case, torodial wrapping produced significantly higher coverage than non-wrapping. Part of the difference is likely due to the small windows under consideration, but the difference was greater than anticipated. Also, the coverage seemed to be quite sensitive to changes in parameters. This suggests a need for good estimates for b_i and h_i so that our methods can perform as well as possible.

8.2 Future Work

The subtle differences surrounding $\tilde{X}_K$ and $\bar{X}_K$ need to be explored further so that a satisfactory resolution may be found. This will bring closure to the contradictory results in the existing literature.

There are many possible extensions for this work. First and foremost, a data-driven method for selecting b needs to be well-defined. It should be possible to modify the methods used in [28]. Second, simulations should be carried out with increasing intervals to assess how well the finite sample variance compares with the asymptotic variance. This would allow us to estimate the error involved with using our sample variance in calculations.

Moreover, we may drop the requirement of stationarity of the marks. Instead, we might consider local stationarity (see [11] and [12]) of the marks. This would permit a slowly changing mean over time, for example. Since we fill blocks "close" to their original block, our method should be robust enough to preserve the local stationarity.

Implicit in our proofs was the independence of the point process and the associated marks. There are many situations in nature where this is not the case. For example, in a forest the distribution of trees and their associated heights are likely to be correlated. The independence assumption allowed us to establish asymptotic normality of our estimators. Again, since the resampled data preserves local structure, it is quite plausible that our method would accurately resample such data.

We may consider relaxing the Poisson assumption. In its place we could include an assumption on maximal correlation mixing as was considered in [60].

At the defense of this thesis, it was brought up that Brillinger [5] considered the problem of estimating the variance of $\bar{X}_K$. His results have been included in Chapter 3 to strengthen the arguments for the asymptotic normality of $\bar{X}_K$. His work is based upon assumptions of cumulants while we make assumptions on α-mixing. Work remains to verify that our assumptions are enough to establish the theorems.

Brillinger also includes an approximation to the variance that can serve as an alternative method for computing coverage probabilities for the models considered in Chapter 2. Early results are inconclusive and will be a source of further study.

Appendix A

Code for Simulations

The following R code was used to generate the numerical results contained in the text. Code for Chapter 5 is not included, as it is precisely the same code used in Chapter 2 and Chapter 4, with the outputs combined to allow for easy comparisons between the methods.

Also, the function calls are not provided, but should be fairly clear from the context on the individual tables as to what parameters were used.

A.1 Chapter 2 Code

```
######################################################################
####### Generate One-Dimensional Inhomogeneous Poisson Data #######
######################################################################

generate_one_dim_inhom_pp = function(lambda, lambda_max = lambda, T = 1000) {
  # initialize
  t = 0;
  i = 0;
  N = NULL;
  while(t < T) {
    U1 = runif(1);
    t = t - log(U1)/lambda_max;
    if(t > T){break};

    U2 = runif(1);
    if(U2 <= lambda(t)/lambda_max) {
      i = i + 1;
      N[i] = t;
    }
  }
  return(N);
}
```

```
########################################
####### Generate Stationary Marks #######
########################################


cov_func = function(t) {
  R = exp(-abs(t))
  return(R);
}

stationary_process = function(pts) {

  # A stationary process can be generated by
  # creating a multivariate normal distribution with mean 0
  # and covariance struction given by R(t) = exp(-abs(t))

  # We begin by initializing a covariance matrix of 0s

  len = length(pts);
  cov = matrix(rep(0,len^2),len);

  # Next, we populate the entries based upon R(t)

  for(i in 1:len) {
    for(j in i:len) {
      cov[i,j] = cov_func(pts[i] - pts[j]);
      cov[j,i] = cov[i, j];
    }
  }

  # We find the square root of the covariance matrix

  e = eigen(cov);
  V = e$vectors;
  B = V %*% diag(sqrt(e$values)) %*% t(V);

  # Next, we generate a vector of standard normal R.V.s

  Z = rnorm(len);

  # Finally, we create our marks, X(t)

  X = B %*% Z;

  return(X);
}


############################################
####### Resample Using Transformation #######
############################################
```

```
# Next, we compare the above results with those from using a block
# bootstrap on transformed data

# That is, we estimate the intensity function (or use the actual intensity)
# and transform the points from an inhomogeneous Poisson process to a
# homogeneous Poisson process with rate 1


# Estimate the intensity function by using piecewise constant functions

estimate_intensity = function(points, T, win) {

  Counts = NULL;

  for(j in 1:(T/win)) {

    start_loc = (j-1)*win + 0.00000001;
    end_loc = start_loc + win;

    # Determine the indices in the original data that will be captured in
    # this block
    start_index = which(points == min(points[(points - start_loc) >= 0]));
    end_index = which(points == max(points[(points - end_loc) <= 0]));

    if(length(start_index) == 0 || length(end_index) == 0) {
      Counts[j] = 0;
    }  else {
      Counts[j] = (end_index - start_index +1)/win;
    }
  }
  return(Counts);
}


# Transform our inhomogeneous data to homogeneous data by using the fact
# that by plugging in our points into \Lambda(t) = int_0^t {\lambda(s)ds}
# we will have a Poisson(1) process

# The output is a list of original and transformed points


transform_data_via_piecewise_linear = function(points, T, win) {

  intensities = estimate_intensity(points, T, win);
  Mapped = NULL;

  for(i in 1:length(points)) {
    value_int = floor(points[i]/win);
    value_rem = points[i] - win*value_int;
```

```
    if (value_int == 0) {
      Mapped[i] = intensities[1]*value_rem;
    } else {
      Mapped[i] = win*sum(intensities[1:value_int])
                  + intensities[(value_int+1)]*value_rem;
    }
  }
  correspondence = matrix(c(points, Mapped), nrow=length(points),ncol=2,
                          dimnames=list(NULL,c("Orig","Trans")));
  return(correspondence);
}


# If there is a need to "invert" \Lambda(t), we can use the follow to restore
# the original points

restore_original = function(corresp, list) {

  Original_List = NULL;

  for(i in 1:length(list)) {
    Original_List[i] = corresp[which(corresp[,2]==list[i]),1];
  }
  return(Original_List);
}


# We can estimate \lambda(t) a second way
# First, we make a fine grid, cast a "net" of size h on either side
# (or until we hit a boundary) to estimate the intensity on that interval
# This gives us point estimates for \lambda(t) at each point along the grid
# We can then sum the areas under the curve to estimate \Lambda(t) for each
# of our datapoints

estimate_intensity2 = function(points, T, win, step) {

  # Make a grid of points
  seq = seq(from = 0, to = T, by = step);
  lambda_est = NULL;

  # Next, cast a net of length h in both directions, but be mindful of
  # boundaries (which shrinks the net's size)
  for(i in 1:length(seq)) {
    l_bd = max(0, seq[i] - win/2);
    u_bd = min(T, seq[i] + win/2);
    count = 0;
    for(j in 1:length(points)) {
      if(points[j] > l_bd && points[j] < u_bd) {
        count = count + 1;
      }
```

```
    }
  lambda_est[i] = count/(u_bd - l_bd);
  }
  output = matrix(c(seq, lambda_est), nrow = length(seq), ncol = 2,
                  dimnames = list(NULL,c("seq","lamba_est")));
  return(output);
}


# Transform our inhomogeneous data to homogeneous data by using the fact
# that by plugging in our points into \Lambda(t) = int_0^t {\lambda(s)ds}
# we will have a Poisson(1) process

# The output is a list of original and transformed points

transform_data_via_estimation = function(points, T, win, step) {

  intensities = estimate_intensity2(points, T, win, step)[,2];
  Mapped = NULL;

  for(i in 1:length(points)) {
    value_int = floor(points[i]/step);
    value_rem = points[i] - step*value_int;
    if (value_int == 0) {
      Mapped[i] = intensities[1]*value_rem;
    } else {
      Mapped[i] = step*sum(intensities[1:value_int])
                  + intensities[(value_int+1)]*value_rem;
    }
  }
  correspondence = matrix(c(points, Mapped),
                          nrow=length(points),
                          ncol=2,dimnames=list(NULL,c("Orig","Trans")));
  return(correspondence);
}


##############################################
####### Resample Homogeneous MPP Code #######
##############################################


# Resample the mpp for a particular block (called in run_circular_bootstrap())

resample_homogeneous_block = function(mpp, j, T, b) {

  # Determine how much to shift the block from the starting point, j*b
  shift = runif(1, 0, T);

  # Determine starting and ending block locations (on the real line)
  block_start_loc = (j*b + shift) %% T;
```

```
  block_end_loc = (block_start_loc + b) %% T;

  points = mpp[,2];
  block_data = NULL;

  for(i in 1:length(points)) {
    if (block_start_loc < block_end_loc) {
      if (points[i] >= block_start_loc && points[i] <= block_end_loc) {
        block_data = rbind(block_data,mpp[i,]);
      }
    } else {
      if (points[i] >= block_start_loc || points[i] <= block_end_loc) {
        block_data = rbind(block_data,mpp[i,]);
      }
    }
  }
  # Return the points and marks for the particular resampled block
  return(block_data);
}


# Call the "circular" bootstrap function to generate one replicant of data

run_circular_bootstrap = function(mpp, T, b) {
  # initialize the resampled data vectors
  resample_orig = NULL;
  resample_trans = NULL;
  resample_marks = NULL;

  # Define the number of blocks to resample over
  L = T/b;

  # run a loop over all of the L blocks
  for(j in 0:(L-1)) {
    # resample the jth block
    resampled_block = resample_homogeneous_block(mpp, j, T, b);

    # determine the points and marks corresponding to that resampling
    resample_orig = c(resample_orig, resampled_block[,1]);
    resample_trans = c(resample_trans, resampled_block[,2]);
    resample_marks = c(resample_marks, resampled_block[,3]);
  }

  # combine the points and marks to make a new mpp matrix
  resampled_mpp = matrix(c(resample_orig, resample_trans, resample_marks),
                  nrow=length(resample_orig),
                  ncol=3,dimnames=list(NULL,c("Orig", "Trans", "Marks")))

  # sort the data in ascending order and return the value
  return(resampled_mpp[sort.list(resampled_mpp[,1]), ])
}
```

```
gen_95_CI_em = function(mpp, model, num_resamp, T, b, win) {
  # Set the count equal to zero to begin with
  count = 0;

  # initialize a vector for CI length
  CI_length = NULL;
  Num = NULL;

  # Run a loop over all of the times we wish to resample
  for(j in 1:num_resamp) {
    values = run_circular_bootstrap(mpp, T, b)[,ncol(mpp)];
    m = mean(values);
    sd = sd(values);
    n = length(values);
    error = qt(0.975, df = n - 1)*sd/sqrt(n);
    lower = m - error;
    upper = m + error;

    CI_length[j] = upper - lower;
    Num[j] = n;

    if((lower < 0)  && (upper > 0)) {
      count = count + 1;
    }
  }

  outputs = matrix(c(model, b, win, mean(Num), mean(CI_length), sd(CI_length),
                     100*count/num_resamp), nrow=1,
                     dimnames=list(NULL,c("Model", "b", "win", "N Obs (avg)",
                     "CI Len Mean", "CI Len SD", "Cover %")));


  # Return the percentage of confidence intervals that contain 0,
  # the mean CI length, and the standard deviation of CI lengths
  return(outputs);
}
```

A.2 Chapter 4 Code

The code used to generate the inhomogeneous Poisson process is the same as in Chapter 2. The main difference here is with the use of the local block bootstrap to resample the data.

```
#####################################################################
####### Generate One-Dimensional Inhomogeneous Poisson Data #######
#####################################################################
```

```
generate_one_dim_inhom_pp = function(lambda, lambda_max = lambda, T = 1000) {
  # initialize
  t = 0;
  i = 0;
  N = NULL;
  while(t < T) {
    U1 = runif(1);
    t = t - log(U1)/lambda_max;
    if(t > T){break};

    U2 = runif(1);
    if(U2 <= lambda(t)/lambda_max) {
      i = i + 1;
      N[i] = t;
    }
  }
  return(N);
}


########################################
####### Generate Stationary Marks #######
########################################


cov_func = function(t) {
  R = exp(-abs(t))
  return(R);
}

stationary_process = function(pts) {

  # A stationary process can be generated by
  # creating a multivariate normal distribution with mean 0
  # and covariance struction given by R(t) = exp(-abs(t))

  # We begin by initializing a covariance matrix of 0s

  len = length(pts);
  cov = matrix(rep(0,len^2),len);

  # Next, we populate the entries based upon R(t)

  for(i in 1:len) {
    for(j in i:len) {
      cov[i,j] = cov_func(pts[i] - pts[j]);
      cov[j,i] = cov[i, j];
    }
  }

  # We find the square root of the covariance matrix
```

```
  e = eigen(cov);
  V = e$vectors;
  B = V %*% diag(sqrt(e$values)) %*% t(V);

  # Next, we generate a vector of standard normal R.V.s

  Z = rnorm(len);

  # Finally, we create our marks, X(t)

  X = B %*% Z;

  return(X);
}


#########################################################
####### Local Block Bootstrap Resampling Code #######
#########################################################


# Resample the mpp for a particular block (used as part of run_lbb())
# Note: Wrap = 1 uses torodial wrapping while Wrap = 0 does not wrap
# and instead shrinks the bandwidth near the boundaries

resample_block = function(mpp, j, T, b, h, wrap) {

  # Determine how much to shift the block from the starting point, j*b
  if (wrap == 0) {
    shift_left = max(-j*b, -h);
    shift_right = min(T - (j+1)*b, h);
    shift = runif(1, shift_left, shift_right);
  } else {
    shift = runif(1, -h, h);
  }

  # Determine starting and ending block locations (on the real line)
  block_start_loc = (j*b + shift) %% T;
  block_end_loc = (block_start_loc + b) %% T;

  points = mpp[,1];
  block_data = NULL;

  for(i in 1:length(points)) {
    if (block_start_loc < block_end_loc) {
      if (points[i] >= block_start_loc && points[i] <= block_end_loc) {
        block_data = rbind(block_data,mpp[i,]);
      }
    } else {
      if (points[i] >= block_start_loc || points[i] <= block_end_loc) {
```

```
        block_data = rbind(block_data,mpp[i,]);
      }
    }
  }

  # Return the points and marks for the particular resampled block
  return(block_data);
}


# call the local block bootstrap function to generate one replicant of data

run_lbb = function(mpp, T, b, h, wrap) {
  # initialize the resampled data vectors
  resample_points = NULL;
  resample_marks = NULL;

  # Define the number of blocks to resample over
  L = T/b;

  # run a loop over all of the L blocks
  for(j in 0:(L-1)) {
    # resample the jth block
    resampled_block = resample_block(mpp, j, T, b, h, wrap);

    # determine the points and marks corresponding to that resampling
    resample_points = c(resample_points, resampled_block[,1]);
    resample_marks = c(resample_marks, resampled_block[,2]);
  }

  # combine the points and marks to make a new mpp matrix
  resampled_mpp = matrix(c(resample_points, resample_marks),
                  nrow=length(resample_points), ncol=2,
                  dimnames=list(NULL,c("Points","Marks")))

  # sort the data in ascending order (based on the observed points) and
  # return the value
  return(resampled_mpp[sort.list(resampled_mpp[,1]), ])
}


# Generate 95% confidence intervals and return the percentage that contain 0
gen_95_CI = function(mpp, model, num_resamp, T, b, h, wrap) {
  # Set the count equal to zero to begin with
  count = 0;

  # initialize a vector for CI length
  CI_length = NULL;
  Num = NULL;

  # Run a loop over all of the times we wish to resample
```

```
  for(j in 1:num_resamp) {
    values = run_lbb(mpp, T, b, h, wrap)[,ncol(mpp)];
    m = mean(values);
    sd = sd(values);
    n = length(values);
    error = qt(0.975, df = n - 1)*sd/sqrt(n);
    lower = m - error;
    upper = m + error;

    CI_length[j] = upper - lower;
    Num[j] = n;

    if((lower < 0)  && (upper > 0)) {
      count = count + 1;
    }
  }

  outputs = matrix(c(model, wrap, b, h, mean(Num), mean(CI_length),
                     sd(CI_length), 100*count/num_resamp), nrow=1,
                     dimnames=list(NULL,c("Model", "Wrapped?", "b", "h",
                     "N Obs (avg)", "CI Len Mean", "CI Len SD", "Cover %")));

  # Return the percentage of confidence intervals that contain 0, the mean CI
  # length, and the standard deviation of CI lengths
  return(outputs);
}
```

A.3 Chapter 7 Code

We modify the algorithm from Chapter 4 to allow for two-dimensional data.

```
#####################################################################
####### Generate Two-Dimensional Inhomogeneous Poisson Data #######
#####################################################################

generate_one_dim_hom_pp = function(lambda, T) {
  # initialize
  t = 0;
  i = 0;
  N = NULL;
  while(t < T) {
    U1 = runif(1);
    t = t - log(U1)/lambda;
    if(t > T){break};
    i = i + 1;
    N[i] = t;
```

```
  }
  return(N);
}

generate_two_dim_inhom_pp = function(lambda, lambda_max = lambda, T1, T2) {

  # First, we generate points for a one-dimensional homogeneous Poisson
  # process with rate lambda_max*T2 on [0, T1]

  X = generate_one_dim_hom_pp(lambda_max*T2, T1);

  # Second, we generate Y_1, Y_2, ..., as independent, uniformly distributed
  # random variables on [0, T2]

  Y = NULL;
  for (i in 1:length(X)) {
    Y[i] = runif(1, 0, T2);
  }

  # Putting these together, we have a two-dimensional homogeneous Poisson
  # process with rate lambda_max

  two_dim_hom_pp = matrix(c(X, Y), nrow=length(X),ncol=2,
                          dimnames=list(NULL,c("X","Y")));

  # Next, we thin the process to arrive at our two-dimensional inhomogeneous
  # Poisson process

  k = 0;
  X_final = NULL;
  Y_final = NULL;

  for (i in 1:length(X)) {
    U1 = runif(1);
    if (U1 <= lambda(X[i], Y[i])/lambda_max) {
      k = k + 1;
      X_final[k] = X[i];
      Y_final[k] = Y[i];
    }
  }
  return(matrix(c(X_final, Y_final), nrow=length(X_final),ncol=2,
                dimnames=list(NULL,c("X","Y"))));
}


########################################
####### Generate Stationary Marks #######
########################################

cov_func_two_dim = function(t) {
  val = 0;
```

```
  for(i in 1:length(t)) {
    val = val - abs(t[i]);
  }
  R = exp(val);
  return(R);
}

stationary_process_two_dim = function(pts) {

  # A stationary process can be generated by
  # creating a multivariate normal distribution with mean 0
  # and covariance struction given by R(t) = exp(-abs(t))

  # We begin by initializing a covariance matrix of 0s

  len = nrow(pts);
  cov = matrix(rep(0,len^2),len);

  # Next, we populate the entries based upon R(t)

  for(i in 1:len) {
    for(j in i:len) {
      cov[i,j] = cov_func_two_dim(pts[i] - pts[j]);
      cov[j,i] = cov[i, j];
    }
  }

  # We find the square root of the covariance matrix

  e = eigen(cov);
  V = e$vectors;
  B = V %*% diag(sqrt(e$values)) %*% t(V);

  # Next, we generate a vector of standard normal R.V.s

  Z = rnorm(len);

  # Finally, we create our marks, X(t)

  X = B %*% Z;

  return(X);
}




#######################################################
####### Begin Two-Dimensional Resampling Code #######
#######################################################
```

```
# Resample the mpp for a particular block (used as part of run_lbb())
# Note: Wrap = 1 uses torodial wrapping while Wrap = 0 does not wrap
# and instead shrinks the bandwidth near the boundaries

resample_block_two_dim = function(mpp, j, T1, T2, b1, b2, h1, h2, wrap) {

  # First, we determine the coordinates of the start of the block

  L1 = T1/b1;
  L2 = T2/b2;

  d = floor(j/L1);
  r = j - d*L1;

  j_block_start_loc1 = r*b1;
  j_block_start_loc2 = d*b2;

  # Determine how much to shift the block from the starting point (the
  # lower left corner of the block)
  if (wrap == 0) {
    shift_left = max(-r*b1, -h1);
    shift_right = min(T1 - (r+1)*b1, h1);
    shift_down = max(-d*b2, -h2);
    shift_up = min(T2 - (d+1)*b2, h2);
    shift1 = runif(1, shift_left, shift_right);
    shift2 = runif(1, shift_down, shift_up);
  } else {
    shift1 = runif(1, -h1, h1);
    shift2 = runif(1, -h2, h2);
  }

  # Determine starting and ending block locations
  block_start_loc1 = (j_block_start_loc1 + shift1) %% T1;
  block_start_loc2 = (j_block_start_loc2 + shift2) %% T2;
  block_end_loc1 = (block_start_loc1 + b1) %% T1;
  block_end_loc2 = (block_start_loc2 + b2) %% T2;

  points = mpp[,1:2];
  block_data = NULL;

  for(i in 1:nrow(points)) {
    if(block_start_loc1 > block_end_loc1) {
      if(block_start_loc2 > block_end_loc2) {  # This is the upper right corner
        if((points[i,1] >= block_start_loc1 || points[i,1] <= block_end_loc1)
        && (points[i,2] >= block_start_loc2 || points[i,2] <= block_end_loc2)){
          block_data = rbind(block_data,mpp[i,]);
        }
      } else {                                 # This is along the right-side
        if((points[i,1] >= block_start_loc1 || points[i,1] <= block_end_loc1)
        && (points[i,2] >= block_start_loc2 && points[i,2] <= block_end_loc2)){
```

```
        block_data = rbind(block_data,mpp[i,]);
      }
    }
  } else {
    if(block_start_loc2 > block_end_loc2) {  # This is along the top
      if((points[i,1] >= block_start_loc1 && points[i,1] <= block_end_loc1)
      && (points[i,2] >= block_start_loc2 || points[i,2] <= block_end_loc2)){
        block_data = rbind(block_data,mpp[i,]);
      }
    } else {                                 # Middle blocks
      if((points[i,1] >= block_start_loc1 && points[i,1] <= block_end_loc1)
      && (points[i,2] >= block_start_loc2 && points[i,2] <= block_end_loc2)){
        block_data = rbind(block_data,mpp[i,]);
      }
    }
  }
}

  # Return the points and marks for the particular resampled block
  return(block_data);
}



# call the local block bootstrap function to generate one replicant of data

run_lbb_two_dim = function(mpp, T1, T2, b1, b2, h1, h2, wrap) {
  # initialize the resampled data vectors
  resample_points1 = NULL;
  resample_points2 = NULL;
  resample_marks = NULL;

  # Define the number of blocks to resample over
  L = (T1/b1)*(T2/b2);

  # run a loop over all of the L blocks
  for(j in 0:(L-1)) {
    # resample the jth block
    r_block = resample_block_two_dim(mpp, j, T1, T2, b1, b2, h1, h2, wrap);

    # determine the points and marks corresponding to that resampling
    resample_points1 = c(resample_points1, r_block[,1]);
    resample_points2 = c(resample_points2, r_block[,2]);
    resample_marks = c(resample_marks, r_block[,3]);
  }

  # combine the points and marks to make a new mpp matrix
  resampled_mpp = matrix(c(resample_points1, resample_points2, resample_marks),
                         nrow=length(resample_points1),ncol=3,
                         dimnames=list(NULL,c("Points1", "Points2", "Marks")));
```

```
  # sort the data in ascending order (based on the observed points)
  # and return the value
  return(resampled_mpp[sort.list(resampled_mpp[,1]), ])
}


# Generate 95% confidence intervals and return the percentage that contain 0
gen_95_CI_two_dim = function(mpp, model, num_resamp, T1, T2, b1, b2, h1, h2,
                             wrap) {
  # Set the count equal to zero to begin with
  count = 0;

  # initialize a vector for CI length
  CI_length = NULL;
  Num = NULL;

  # Run a loop over all of the times we wish to resample
  for(j in 1:num_resamp) {
    values = run_lbb_two_dim(mpp, T1, T2, b1, b2, h1, h2, wrap)[,ncol(mpp)];
    m = mean(values);
    sd = sd(values);
    n = length(values);
    error = qt(0.975, df = n - 1)*sd/sqrt(n);
    lower = m - error;
    upper = m + error;

    CI_length[j] = upper - lower;
    Num[j] = n;

    if((lower < 0)  && (upper > 0)) {
      count = count + 1;
    }
  }

  outputs = matrix(c(model, wrap, b1, b2, h1, h2, mean(Num), mean(CI_length),
                     sd(CI_length), 100*count/num_resamp), nrow=1,
                     dimnames=list(NULL,c("Model", "Wrapped?", "b1", "b2",
                     "h1", "h2", "N Obs (avg)", "CI Len Mean", "CI Len SD",
                     "Cover %")));

  # Return the percentage of confidence intervals that contain 0,
  # the mean CI length, and the standard deviation of CI lengths
  return(outputs);
}
```

Bibliography

[1] A. J. Baddeley, J. Møller, and R. Waagepetersen, *Non- and semi-parametric estimation of interaction in inhomogeneous point patterns*, Statist. Neerlandica **54** (2000), no. 3, 329–350. MR 1804002 (2001i:62103)

[2] Adrian Baddeley, *Spatial point processes and their applications*, Stochastic geometry, Lecture Notes in Math., vol. 1892, Springer, Berlin, 2007, pp. 1–75. MR 2327290 (2008c:60045)

[3] Peter J. Bickel and Anat Sakov, *On the choice of m in the m out of n bootstrap and confidence bounds for extrema*, Statist. Sinica **18** (2008), no. 3, 967–985. MR 2440400 (2009k:62081)

[4] W. J. Braun and R. J. Kulperger, *A bootstrap for point processes*, J. Statist. Comput. Simulation **60** (1998), no. 2, 129–155. MR 1628927

[5] David R. Brillinger, *Estimation of the mean of a stationary time series by sampling*, J. Appl. Probability **10** (1973), 419–431. MR 0373202 (51 #9403)

[6] Peter J. Brockwell and Richard A. Davis, *Time series: theory and methods*, second ed., Springer Series in Statistics, Springer-Verlag, New York, 1991. MR 1093459 (92d:62001)

[7] Edward Carlstein, *The use of subseries values for estimating the variance of a general statistic from a stationary sequence*, Ann. Statist. **14** (1986), no. 3, 1171–1179. MR 856813 (88a:62095)

[8] Shaojie Chen, *Statistical analysis of point processes and associated marks*, ProQuest LLC, Ann Arbor, MI, 2006, Thesis (Ph.D.)–University of California, Riverside. MR 2709686

[9] Ann Cowling, Peter Hall, and Michael J. Phillips, *Bootstrap confidence regions for the intensity of a Poisson point process*, J. Amer. Statist. Assoc. **91** (1996), no. 436, 1516–1524. MR 1439091 (98b:62150)

[10] Noel A. C. Cressie, *Statistics for spatial data*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, 1993, Revised reprint of the 1991 edition, A Wiley-Interscience Publication. MR 1239641 (94h:62155)

[11] R. Dahlhaus, *On the Kullback-Leibler information divergence of locally stationary processes*, Stochastic Process. Appl. **62** (1996), no. 1, 139–168. MR 1388767 (98b:60070)

[12] ———, *Fitting time series models to nonstationary processes*, Ann. Statist. **25** (1997), no. 1, 1–37. MR 1429916 (98b:62168)

[13] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes. Vol. I*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2003, Elementary theory and methods. MR 1950431 (2004c:60001)

[14] A. C. Davison and D. V. Hinkley, *Bootstrap methods and their application*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 1, Cambridge University Press, Cambridge, 1997, With 1 IBM-PC floppy disk (3.5 inch; HD). MR 1478673 (98g:62080)

[15] Peter J. Diggle, *Statistical analysis of spatial point patterns*, Mathematics in Biology, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1983. MR 743593 (85m:62205)

[16] P.J. Diggle, N. Lange, and F.M. Beneš, *Analysis of variance for replicated spatial point patterns in clinical neuroanatomy.*, J. Amer. Statist. Assoc. **86** (1991), 618–625.

[17] Paul Doukhan, *Mixing*, Lecture Notes in Statistics, vol. 85, Springer-Verlag, New York, 1994, Properties and examples. MR 1312160 (96b:60090)

[18] Arif Dowla, *Local block bootstrap based inference for nonstationary time series*, ProQuest LLC, Ann Arbor, MI, 2002, Thesis (Ph.D.)–University of California, San Diego. MR 2703695

[19] Arif Dowla, Efstathios Paparoditis, and Dimitris N. Politis, *Locally stationary processes and the local block bootstrap*, Recent advances and trends in nonparametric statistics, Elsevier B. V., Amsterdam, 2003, pp. 437–444. MR 2498256 (2010i:62248)

[20] Friedrich Götze and Alfredas Račkauskas, *Adaptive choice of bootstrap sample sizes*, State of the art in probability and statistics (Leiden, 1999), IMS Lecture Notes Monogr. Ser., vol. 36, Inst. Math. Statist., Beachwood, OH, 2001, pp. 286–309. MR 1836566 (2002f:62047)

[21] Yongtao Guan, *Tests for independence between marks and points of a marked point process*, Biometrics **62** (2006), no. 1, 126–134, 317. MR 2226565

[22] Yongtao Guan and David R. Afshartous, *Test for independence between marks and points of marked point processes: a subsampling approach*, Environ. Ecol. Stat. **14** (2007), no. 2, 101–111. MR 2370957

[23] Peter Hall, *Resampling a coverage pattern*, Stochastic Process. Appl. **20** (1985), no. 2, 231–246. MR 808159 (87d:62071)

[24] Roelof Helmers and I Wayan Mangku, *On estimating the period of a cyclic Poisson process*, Mathematical statistics and applications: Festschrift for Constance van Eeden, IMS Lecture Notes Monogr. Ser., vol. 42, Inst. Math. Statist., Beachwood, OH, 2003, pp. 345–356. MR 2138302 (2006d:62081)

[25] Roelof Helmers, I. Wayan Mangku, and Ričardas Zitikis, *Consistent estimation of the intensity function of a cyclic Poisson process*, J. Multivariate Anal. **84** (2003), no. 1, 19–39. MR 1965821 (2004a:60095)

[26] I. A. Ibragimov and Yu. V. Linnik, *Independent and stationary sequences of random variables*, Wolters-Noordhoff Publishing, Groningen, 1971, With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman. MR 0322926 (48 #1287)

[27] Janine Illian, Antti Penttinen, Helga Stoyan, and Dietrich Stoyan, *Statistical analysis and modelling of spatial point patterns*, Statistics in Practice, John Wiley & Sons Ltd., Chichester, 2008. MR 2384630 (2008k:62004)

[28] Agnieszka Jach, Tucker McElroy, and Dimitris N. Politis, *Subsampling inference for the mean of heavy-tailed long-memory time series*, Journal of Time Series Analysis (2011), no–no.

[29] Nazgul Jenish and Ingmar R. Prucha, *Central limit theorems and uniform laws of large numbers for arrays of random fields*, J. Econometrics **150** (2009), no. 1, 86–98. MR 2525996 (2010h:60071)

[30] Alan F. Karr, *Inference for stationary random fields given Poisson samples*, Adv. in Appl. Probab. **18** (1986), no. 2, 406–422. MR 840101 (87k:62142)

[31] ———, *Point processes and their statistical inference*, second ed., Probability: Pure and Applied, vol. 7, Marcel Dekker Inc., New York, 1991. MR 1113698 (92f:62116)

[32] J. F. C. Kingman, *Poisson processes*, Oxford Studies in Probability, vol. 3, The Clarendon Press Oxford University Press, New York, 1993, Oxford Science Publications. MR 1207584 (94a:60052)

[33] K. Krickeberg, *Processus ponctuels en statistique*, Tenth Saint Flour Probability Summer School—1980 (Saint Flour, 1980), Lecture Notes in Math., vol. 929, Springer, Berlin, 1982, pp. 205–313. MR 665597 (84h:60095)

[34] M.E. Kuhl, J.R. Wilson, and M.A. Johnson, *Estimation and simulation of nonhomogeneous poisson processes having multiple periodicities*, Simulation Conference Proceedings, 1995. Winter, dec 1995, pp. 374 –383.

[35] Hans R. Künsch, *The jackknife and the bootstrap for general stationary observations*, Ann. Statist. **17** (1989), no. 3, 1217–1241. MR 1015147 (90k:62089)

[36] Yu. A. Kutoyants, *Statistical inference for spatial Poisson processes*, Lecture Notes in Statistics, vol. 134, Springer-Verlag, New York, 1998. MR 1644620 (99k:62149)

[37] Soumendra Nath Lahiri, *Second order optimality of stationary bootstrap*, Statist. Probab. Lett. **11** (1991), no. 4, 335–341. MR 1108356 (92k:62088)

[38] Averill M. Law and W. David Kelton, *Simulation modeling and analysis*, McGraw-Hill Book Co., New York, 1982, McGraw-Hill Series in Industrial Engineering and Management Science. MR 630193 (82i:68058)

[39] Sanghoon Lee, James R. Wilson, and Melba M. Crawford, *Modeling and simulation of a nonhomogeneous poisson process having cyclic behavior*, Communications in Statistics - Simulation and Computation **20** (1991), no. 2-3, 777–809.

[40] L. M. Leemis, *Nonparametric estimation of the cumulative intensity function for a nonhomogeneous poisson process*, Management Sci. **37** (1991), no. 7, 886–900.

[41] N. N. Leonenko and A. V. Ivanov, *Statisticheskii analiz sluchainykh polei*, "Vishcha Shkola", Kiev, 1986, With a preface by A. V. Skorokhod. MR 917486 (89e:62125)

[42] P. A. W. Lewis and G. S. Shedler, *Statistical analysis of non-stationary series of events in a data base system*, IBM J. Res. Develop. **20** (1976), no. 5, 465–482. MR 0443292 (56 #1662)

[43] ———, *Simulation of nonhomogeneous Poisson processes by thinning*, Naval Res. Logist. Quart. **26** (1979), no. 3, 403–413. MR 546120 (80k:68089)

[44] Keh-Shin Lii and Elias Masry, *Model fitting for continuous-time stationary processes from discrete-time data*, J. Multivariate Anal. **41** (1992), no. 1, 56–79. MR 1156681 (93e:62235)

[45] ———, *Spectral estimation of continuous-time stationary processes from random sampling*, Stochastic Process. Appl. **52** (1994), no. 1, 39–64. MR 1289167 (95m:62215)

[46] Regina Y. Liu and Kesar Singh, *Moving blocks jackknife and bootstrap capture weak dependence*, Exploring the limits of bootstrap (East Lansing, MI, 1990), Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., Wiley, New York, 1992, pp. 225–248. MR 1197787 (94f:62079)

[47] Shigeru Mase, *The threshold method for estimating total rainfall*, Ann. Inst. Statist. Math. **48** (1996), no. 2, 201–213. MR 1405927 (98d:62179)

[48] Elias Masry, *Alias-free sampling: an alternative conceptualization and its applications*, IEEE Trans. Information Theory **IT-24** (1978), no. 3, 317–324. MR 0504299 (58 #20763)

[49] ———, *Poisson sampling and spectral estimation of continuous-time processes*, IEEE Trans. Information Theory **IT-24** (1978), no. 2, 173–183. MR 0483238 (58 #3256)

[50] ———, *Nonparametric covariance estimation from irregularly-spaced data*, Adv. in Appl. Probab. **15** (1983), no. 1, 113–132. MR 688009 (84e:62141)

[51] Efstathios Paparoditis and Dimitris N. Politis, *Local block bootstrap*, C. R. Math. Acad. Sci. Paris **335** (2002), no. 11, 959–962. MR 1952557 (2003k:62237)

[52] Raghu Pasupathy, *Generating nonhomogeneous poisson processes*, Wiley Encyclopedia of Operations Research and Management Science, pp. 1–11, John Wiley & Sons, Inc., 2010.

[53] Dimitris N. Politis, Efstathios Paparoditis, and Joseph P. Romano, *Resampling marked point processes*, Multivariate analysis, design of experiments, and survey sampling, Statist. Textbooks Monogr., vol. 159, Dekker, New York, 1999, pp. 163–185. MR 1719074 (2000k:62087)

[54] Dimitris N. Politis and Joseph P. Romano, *A circular block-resampling procedure for stationary data*, Exploring the limits of bootstrap (East Lansing, MI, 1990), Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., Wiley, New York, 1992, pp. 263–270. MR 1197789 (93j:62123)

[55] ———, *A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation*, Ann. Statist. **20** (1992), no. 4, 1985–2007. MR 1193322 (94d:62113)

[56] ———, *A nonparametric resampling procedure for stationary data*, Computing Science and Statistics, Proceedings of the 22nd Symposium on the, 1990), Springer, New York, 1992, pp. 98–103.

[57] ———, *Nonparametric resampling for homogeneous strong mixing random fields*, J. Multivariate Anal. **47** (1993), no. 2, 301–328. MR 1247380 (94k:62158)

[58] ———, *The stationary bootstrap*, J. Amer. Statist. Assoc. **89** (1994), no. 428, 1303–1313. MR 1310224 (96a:62051)

[59] Dimitris N. Politis, Joseph P. Romano, and Michael Wolf, *Subsampling*, Springer Series in Statistics, Springer-Verlag, New York, 1999. MR 1707286 (2001d:62047)

[60] Dimitris N. Politis and Michael Sherman, *Moment estimation for statistics from marked point processes*, J. R. Stat. Soc. Ser. B Stat. Methodol. **63** (2001), no. 2, 261–275. MR 1841414 (2002c:62144)

[61] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0.

[62] Arif Raïs, *Méthodes de reéchantillonage et de sous échantillonage dans le contexte spatial et pour des données dépendantes*, ProQuest LLC, Ann Arbor, MI, 1992, Thesis (Ph.D.)–University of Montreal, Montreal.

[63] Sidney Resnick, *Adventures in stochastic processes*, Birkhäuser Boston Inc., Boston, MA, 1992. MR 1181423 (93m:60004)

[64] _____, *A probability path*, Birkhäuser Boston Inc., Boston, MA, 1999. MR 1664717 (2000b:60002)

[65] Brian D. Ripley, *The second-order analysis of stationary point processes*, J. Appl. Probability **13** (1976), no. 2, 255–266. MR 0402918 (53 #6732)

[66] _____, *Spatial statistics*, John Wiley & Sons Inc., New York, 1981, Wiley Series in Probability and Mathematical Statistics. MR 624436 (83a:62217)

[67] Murray Rosenblatt, *Stationary sequences and random fields*, Birkhäuser Boston Inc., Boston, MA, 1985. MR 885090 (88c:60077)

[68] Sheldon M. Ross, *Introduction to probability models*, seventh ed., Harcourt/Academic Press, Burlington, MA, 2000. MR 1766683 (2001b:60002)

[69] George G. Roussas and D. Ioannides, *Moment inequalities for mixing sequences of random variables*, Stochastic Anal. Appl. **5** (1987), no. 1, 61–120. MR 882698 (88f:60033)

[70] Ju. A. Rozanov, *On the density of Gaussian distributions and Wiener-Hopf integral equations*, Dokl. Akad. Nauk SSSR **165** (1965), 1000–1002. MR 0189131 (32 #6558)

[71] Mats Rudemo, *Empirical choice of histograms and kernel density estimators*, Scand. J. Statist. **9** (1982), no. 2, 65–78. MR 668683 (83j:62059)

[72] H. Schäbe, *Nonparametric estimation of intensities of nonhomogeneous Poisson processes*, Statist. Papers **34** (1993), no. 2, 113–131. MR 1232649 (94i:62066)

[73] Martin Schlather, Paulo J. Ribeiro, Jr., and Peter J. Diggle, *Detecting dependence between marks and locations of marked point processes*, J. R. Stat. Soc. Ser. B Stat. Methodol. **66** (2004), no. 1, 79–93. MR 2035760

[74] Robert J. Serfling, *Approximation theorems of mathematical statistics*, John Wiley & Sons Inc., New York, 1980, Wiley Series in Probability and Mathematical Statistics. MR 595165 (82a:62003)

[75] Nan Shao and Keh-Shin Lii, *Modelling non-homogeneous Poisson processes with almost periodic intensity functions*, J. R. Stat. Soc. Ser. B Stat. Methodol. **73** (2011), no. 1, 99–122. MR 2797738

[76] Michael Sherman and Edward Carlstein, *Nonparametric estimation of the moments of a general statistic computed from spatial data*, J. Amer. Statist. Assoc. **89** (1994), no. 426, 496–500. MR 1294075 (95f:62081)

[77] _____, *Replicate histograms*, J. Amer. Statist. Assoc. **91** (1996), no. 434, 566–576. MR 1395726 (97a:62092)

[78] Martin Snethlage, *Is bootstrap really helpful in point process statistics?*, Metrika **49** (1999), no. 3, 245–255. MR 1731770 (2001f:62060)

[79] V. Ventura, A.C. Davison, and S.J. Boniface, *Statistical inference for the effect of magnetic brain stimulation on a motoneurone*, J. Appl. Stat. **46** (1997), 77–94.

[80] William W. S. Wei, *Time series analysis*, second ed., Addison Wesley/Pearson, Boston, MA, 2006, Univariate and multivariate methods. MR 2517831 (2010i:62262)

[81] Wolfram Research, Inc., *Mathematica, Version 8.0*, Wolfram Research, Inc., Champaign, Illinois, 2010.

[82] M. Ĭ. Yadrenko, *Spectral theory of random fields*, Translation Series in Mathematics and Engineering, Optimization Software Inc. Publications Division, New York, 1983, Translated from the Russian. MR 697386 (84f:60003)

[83] G. Yakovlev, J.B. Rundle, R. Shcherbakov, and D.L. Turcotte, *Inter-arrival time distribution for the non-homogeneous poisson process*, eprint arXiv:cond-mat/0507657 (2005), 1–6.