

Confidence Interval Formation on Spatial Point Processes Using Thinning

by

Lucinda Khalil

2024

Declaration

This project report is submitted in partial fulfilment of the requirements for the degree of Imperial College London, MSci Mathematics. I declare that this thesis is my original work, except where stated otherwise. This thesis has never been submitted for any degree or examination to any other University.

Lucinda Khalil

This thesis was conducted under the supervision of Dr. Cohen.

Abstract

The abstract of the thesis (about 300 words long).

- Use thinning to obtain a distribution of a statistic calculated on spatial data
- This distribution can be used to find $(1 - \alpha)\%$ confidence intervals
- Derive result explaining the behaviour of this distribution as $n \rightarrow \infty$
- discuss where method works and does not work

Dedication

Your dedication. Try to keep it within one page.

Contents

Declaration	i
Abstract	ii
Dedication	iii
List of Figures	vii
List of Tables	xii
Abbreviations	xii
Nomenclature	xiii
1 Introduction	1
2 Background Theory	2
2.1 Spatial Point Processes	2
2.1.1 Definition of Point Processes	2
2.1.2 Properties	3
2.1.3 Poisson Point Process	3
2.1.4 Cluster Point Processes	4
2.1.5 Regular Point Processes	5
2.2 Ripley's K	7
2.2.1 Isotropic Estimator	7
2.2.2 Ripley's K Under CSR	8
2.2.3 Inhomogenous Ripley's K	9
2.3 Inhomogenous Intensity Estimates	9
2.3.1 Inhomogenous Poisson Point Process	9
2.3.2 Parametric Estimation	9
2.3.3 Kernel Density Estimation	11
3 Existing Bootstrapping Methods	12
3.1 Bootstrapping Methods for Distributions	12

3.1.1	Classical Bootstrapping	12
3.1.2	Jackknife	12
3.2	Bootstrapping Spatial Point Processes	12
3.2.1	Overview	12
3.2.2	Splitting Method	12
3.2.3	Tiling Method	13
3.2.4	Subsets Method	13
3.2.5	Marked Point Method	14
3.2.6	Bootstrapping inhomogenous processes	15
3.3	Bootstrapping Methods for Inhomogeneous Intensity Estimates	15
3.4	Proposed Method: Thinning	15
4	Analytic Results	16
4.1	Asymptotics of Thinning	16
4.1.1	Expansion of Statistical Functional	16
4.1.2	Convergence of Remainder Term	18
4.1.3	Derivation of Distribution of Thinned Estimates	19
4.1.4	Usefulness of Asymptotic Result	22
4.2	Link Between Thinning and Bootstrap Sampling	23
4.3	Asymptotics for Spatial Point Processes	25
4.3.1	Conditions Required	25
4.3.2	Conditions on Data	25
4.3.3	Conditions on Statistic	26
4.4	Numerical Demonstration of Results	26
4.5	Example: Intensity of Homogeneous Poisson Point Process	28
5	Confidence Interval Formation	33
5.1	Naive Confidence Interval	33
5.2	Studentized Confidence Interval	34
5.3	Example: Intensity of Homogeneous Poisson Point Process	36
6	Ripley's K	38
6.1	Results	38
6.1.1	Basic Bootstrap Confidence Intervals	38
6.1.2	Studentized Confidence Intervals	38
6.2	Distribution of Thinned Estimates	38
6.2.1	Normal Distribution of Thinned Estimates	38
6.2.2	Variance of Thinned Estimates	39
6.3	Finding Appropriate Scaling	41
7	Inhomogenous Intensity Estimates	42

References	46
A My First Appendix	51
A.1 Poisson as Limit of Binomial	51
A.2 Simulations Showing Deomonstrating Asymptotics	51

List of Figures

- 2.1 This figure shows examples of various point processes discussed in this section, all of which are shown on a unit square. The first subplot shows a homogenous Poisson point process with an intensity of $\lambda = 250$. The second subplot shows a Matérn cluster process which is a type of Neyman-Scott process. This has a parent and daughter intensities of $\lambda_p = 10$ and $\lambda_d = 25$ respectively, and a cluster radius of $R = 0.1$. The third subplot is a Thomas process, with the same parameters as the Matérn. The final subplot is a regular Soft Core point process with $\lambda_b = 500$, and the radii, ρ , are assigned with a probability density function of $f(\rho) = 800\rho$ where $0 < \rho < 0.05$ 6
- 2.2 This figure shows the flaw in the naive estimator of Ripley's K , 2.3. Assume we have observed the points in the unit square region A . We sum over each of the points inside A , counting how many other points are within a distance R . Consider the green point inside the region, which happens to be located at $(1.08, 1.45)$. If we set $h = 0.2$, then clearly, the circular region extends beyond the observed region, and the purple points which are outside of A but within the radius will not be included in the sum. Therefore, this estimator is biased for a bounded observation window. 7
- 3.1 Demonstration of bootstrapping via the tiling method with $N = 4$. The left subfigure shows the randomly selected subregions. Those outlined in red demonstrate toroidal wrapping. The right subfigure shows the new bootstrapped process. 14
- 4.1 Histograms of bootstrapped estimates, $\hat{\theta}^*$, and thinned estimates, $\hat{\theta}_s$, shown in green and purple respectively. The top row uses i.i.d data generated from $X \sim Poi(10)$ and $\lambda = 10$ is estimated using its MLE, which we know satisfies the requirements of theorems 3 and 6. The bottom row uses data distributed according to $X \sim Exp(10)$ and again, $\lambda = 10$ is estimated using its MLE. These results are produced by carrying out 500 simulations of 100 data points each. For each other these simulations, 1000 thinning samples are obtained and an estimate of the desired statistic is calculated. These estimates are then used to form the histograms. 24

- 4.2 Histograms showing distributions of normalised thinned estimates, compared to a standard normal distribution. Each row represents a thinning parameter of $p = 0.2, 0.5, 0.8$ respectively, and each column represents varying observed dataset size, in particular $n = 10, 20, 500$ respectively. Each subplot is generated using n i.i.d datapoints distributed according to a $Poi(10)$ distribution. For each dataset, 2000 thinned samples are obtained and used to calculate the estimate. These are then plotted alongside 2000 samples from a standard normal distribution. 27
- 4.3 Histograms showing distributions of normalised thinned estimates, compared to the normalised estimates calculated over the entire observed dataset. Each row represents a thinning parameter of $p = 0.2, 0.5, 0.8$ respectively, and each column represents varying observed dataset size, in particular $n = 10, 20, 500$ respectively. Each subplot is generated using n i.i.d datapoints distributed according to a $Poi(10)$ distribution. For each dataset, 2000 thinned samples are obtained and used to calculate the estimate. For each case, we also generate 2000 sets of observed data from the distribution, and calculate the estimator over all of these. These estimates are then normalised and plotted against the thinned normalised estimates. 28
- 5.1 Line plot showing the empirical coverage using the basic bootstrap confidence interval as the number of observed datapoints, n , increases from $n = 5$ to $n = 1005$ in steps of 10. The statistic being calculated here is the intensity of a homogenous Poisson distribution. 1000 simulations are carried out, each of which use 500 thinned subsamples. The purple, orange and green line represent thinning parameters of $p = 0.8, 0.5, 0.2$ respectively, and the red dashed line is the desired cover of 0.95. 34
- 5.2 Line plot showing the empirical coverage using the Studentized confidence interval as the number of observed datapoints, n , increases from $n = 5$ to $n = 1005$ in steps of 10. The statistic being calculated here is the intensity of a homogeneous Poisson distribution. 1000 simulations are carried out, each of which use 500 thinned subsamples. The purple, orange and green line represent thinning parameters of $p = 0.8, 0.5, 0.2$ respectively, and the red dashed line is the desired cover of 0.95. 35

5.3 Line plots showing the cover of confidence intervals calculated for the estimation of the intensity of a homogeneous Poisson point process with a true intensity of $\lambda = 250$. The green lines show the cover using the naive basic bootstrap method, and the purple lines show the results of the studentized confidence intervals. Each plot shows the covers using a thinning parameter of $p = 0.2, 0.5, 0.8$ respectively from left to right. These covers are shown for various sizes of observation window, A , with side lengths varying from 0.5 to 2 in steps of 0.25. For each calculated cover, 1000 simulations are carried out, and for each of these a realisation of a homogeneous Poisson process is generated, and 500 thinned subsamples are selected to form the confidence interval.	36
6.1 QQ-plots of Ripley's K estimates on thinned point processes after normalisation, $\sqrt{n}(\hat{K}_s(h) - \hat{K}(h))$. Values are calculated by simulating 100 realisations of a homogeneous Poisson point process with intensity $\lambda = 10000$ on a unit square, and generating 500 thinned processes for each of these. Thinned Ripley's K is then estimated on each of these, and normalised to obtain the desired value. This is carried out for thinning parameters $p = 0.2, 0.5, 0.8$ which are shown in the subfigures from left to right respectively.	39
6.2 Line plot showing estimates of the variance of Ripley's K over entire realisations of data, $\hat{\sigma}^2$, and the scaled variance over thinned subsamples of data, $\hat{\sigma}_{sp}/(1 - p)$. The former of these is shown in green and is generated using estimates $\hat{K}(h)$ for $h = 0.1$ on 10000 simulations of a homogeneous Poisson point process with $\lambda = 250$ on a unit square, and taking the sample variance of these. The latter is formed by simulating the same process 100 times, and calculating the thinned estimate, $\hat{K}_s(h)$, on 500 thinned subsamples of each of these. This is carried out for p taking values from 0.1 to 0.9 in steps of 0.1.	40
6.3 Line plot showing covers for estimation of Ripley's K , using a studentized confidence interval, and scaling the thinned variance using the empirical values. Covers are calculated using 1000 simulations of a homogenous Poisson process with $\lambda = 250$ on a unit square. For each of these, 500 thinned subsamples are generated, and $\hat{K}_s(0.1)$ is calculated on each of these. The plot shows the cover using thinning parameters of $p = 0.2, 0.5, 0.8$ in purple, orange and green respectively.	41
7.1 Caption	43

7.2 Line plots showing the empirical coverage using the basic bootstrap confidence interval as the size of the observed window, A , increases. This side length of the window varies from $W = 0.5$ to $W = 2.0$ in steps if 0.25. The process used is an inhomogenous Poisson point process with an intensity function of $\lambda(x, y) = \exp(a + bx)$, and the values being estimated are a and b . The left plot shows the covers for a , and the right plot shows the covers for b . 100 simulations are carried out, each of which use 500 thinned subsamples. The purple, orange and green line represent thinning parameters of $p = 0.8, 0.5, 0.2$ respectively, and the red dashed line is the desired cover of 0.95.	44
7.3 Line plots showing the empirical coverage using the studentized confidence interval as the size of the observed window, A , increases. This side length of the window varies from $W = 0.5$ to $W = 2.0$ in steps if 0.25. The process used is an inhomogenous Poisson point process with an intensity function of $\lambda(x, y) = \exp(a + bx)$, and the values being estimated are a and b . The left plot shows the covers for a , and the right plot shows the covers for b . 100 simulations are carried out, each of which use 500 thinned subsamples. The purple, orange and green line represent thinning parameters of $p = 0.8, 0.5, 0.2$ respectively, and the red dashed line is the desired cover of 0.95.	45
A.1 Caption	52
A.2 Caption	53
A.3 Caption	54
A.4 Caption	55

List of Tables

Abbreviations

CSR Complete Spatial Randomness.

Nomenclature

aiaostyle

C_D Discharge coefficient.

μ Mean.

σ Standard deviation.

nounits

Chapter 1

Introduction

- discuss issues with classical bootstrapping and spatial data, and the current spatial methods
- more difficult to resample 2nd order stats because this can produce new pairs of points which do not have same relationship as original distribution Loh and Stein (2004)
- there is an analytical expression for variance of K when process is homogenous poisson (in Ripley 1988), as well as a few other cases, but everything else is very difficult or impossible to obtain Loh (2010) Loh and Stein (2004)
- formalise thinning method here?
- discuss asymptotic result
- where it work and where it doesn't work
- explain each chapter

Chapter 2

Background Theory

Spatial statistics is a relatively young field aimed at summarising and analysing data which has a position assigned to each data point. Throughout this project, we consider data which is assumed to be generated from point processes in two dimensions, although there are extensions to this such as spatial processes indexed over continuous space, called spatial processes, or over lattices, called lattice processes. We can think of a spatial process as being ‘a (partial) realisation of a random process

$$\{Z(s) : s \in D\} \quad (2.1)$$

where the index set D allows s to vary continuously throughout a region of d -dimensional Euclidean space’, and $Z(\cdot)$ is random, whereas a point process takes the same form 2.1 but in this case, the index set D is itself a spatial process Cressie (1991). Note that throughout the project we also only consider *simple* point processes where the probability of more than one event occurring at a given position is almost surely zero.

Point processes emerged due to a need to model observed data in areas such as biology, physics, geography and geology among others (Ripley (1981)). Some examples which are important for this research are the Neyman-Scott process, as described in Neyman and Scott (1958); Stoyan et al. (1995), and the Soft Core process described in Stoyan (1987).

2.1 Spatial Point Processes

2.1.1 Definition of Point Processes

Point processes arise over some set X , in our case some subset of \mathbb{R}^2 . We can extend this idea to a marked point process, where each point $s \in \mathbb{R}^2$ is assigned a mark $z \in Y$, then we can think of this new marked point process simply as a point process on the product space $X = \mathbb{R}^2 \times Y$ (Cressie (1991)). In order to define a well-behaved point process, we first introduce *locally finite point configurations*:

Definition 1. (van Lieshout (2019)) A family $N^{lf}(\mathbb{R}^d)$ is a family of **locally finite point configurations** in \mathbb{R}^d if it consists entirely of subsets $x \subset \mathbb{R}^d$ which place finitely many

points in every bounded Borel set $A \subset \mathbb{R}^d$.

This ensures that the point processes do not contain accumulation points, and each Borel set contains at most a countably infinite number of points. We can now define a point process:

Definition 2. (*van Lieshout (2019)*) A **point process**, $X \in N^{lf}(\mathbb{R}^d)$, on \mathbb{R}^d is a random locally finite point configuration, such that for all bounded Borel sets $A \subset \mathbb{R}^d$, the number of points of X that fall in A is a finite random variable which is denoted as $N_X(A)$.

2.1.2 Properties

Definition 3. (*van Lieshout (2019); Ibe (2013)*) A **homogeneous** (or analogously stationary) spatial point process is one whose distribution is invariant under translation. Given the bounded Borel sets $B_1, B_2, \dots, B_k \in \mathbb{R}^d$ for some $k \in \mathbb{Z}_+$, the joint distribution of $N(B_1), N(B_2), \dots, N(B_k)$ is equal to the joint distribution of $N(B_1+y), N(B_2+y), \dots, N(B_k+y)$ for an arbitrary y :

$$P(N(B_1 + y) \leq n_1, \dots, N(B_k + y) \leq n_k) = P(N(B_1) \leq n_1, \dots, N(B_k) \leq n_k).$$

Definition 4. An **isotropic** spatial point process is one whose distribution is invariant under rotation.

Definition 5. (*Cressie (1991)*) A point process exhibits **complete spatial randomness** when, conditioned on the number of points in the bounded region A , $N(A)$, the points are independently and uniformly distributed over the region. This means that if $N(A) = n$, then the ordered tuple of points (s_1, s_2, \dots, s_n) satisfies

$$P(s_1 \in B_1, s_2 \in B_2, \dots, s_n \in B_n) = \prod_{i=1}^n \frac{|B_i|}{|A|}, \quad B_1, B_2, \dots, B_n \subset A$$

where $|B| = \int_B ds$.

Complete spatial randomness (CSR) is present in a process when points are equally likely to occur anywhere within a given bounded set, and the points do not interact repulsively or attractively. A homogeneous Poisson process exhibits complete spatial randomness. One statistic useful in measuring CSR is Ripley's K which is discussed further in section 2.2 Cressie (1991); Shekhar et al. (2005).

2.1.3 Poisson Point Process

Poisson point processes are very useful in modelling processes we see in the real world since they can be described as the limit of a binomial point processes (van Lieshout (2019)).

Further discussion of this based on van Lieshout (2019) can be found in Appendix A. Similarly, Berman (1987) and Ellis (1986) explore convergence of Markov point processes and transformed point processes respectively to Poisson. not done yet

Some examples of physical processes which are often modelled using a Poisson spatial process include random trials on a fine grid such as a lattice of radioactive atoms, and populations which are randomly thinned such as seeds which randomly germinate Baddeley et al. (2015). An example of a homogenous Poisson point process is displayed in figure 2.1.

A homogenous Poisson point process is defined in van Lieshout (2019) in the following way:

Definition 6. (*van Lieshout (2019)*) A homogenous Poisson point process, X , is a point process on \mathbb{R}^d with intensity $\lambda > 0$ if

- $N_X(A)$ is Poisson distributed with mean $\lambda|A|$ for every bounded Borel set $A \subset \mathbb{R}^d$;
- for any k disjoint bounded Borel sets A_1, A_2, \dots, A_k , $k \in \mathbb{N}$, the random variables $N_X(A_1), \dots, N_X(A_k)$ are independent.

This can be extended to an **inhomogenous Poisson point process** by allowing the intensity, λ , to vary over A , and therefore the mean will take the form

$$\int_A \lambda(x) dx$$

for some function $\lambda : \mathbb{R}^d \rightarrow \mathbb{R}^+$ which is integrable on bounded sets.

2.1.4 Cluster Point Processes

Cluster point processes have a wide range of practical applications such as modelling failures of complex systems such as computers Lewis (1964), ecology Bartlett (1964); Aryal and Jones (2021) and teletraffic to a server Faj et al. (2006). A cluster process, is itself a combination of two processes. The first of these is the parent process, N_p , which determines the positions of the cluster centres, and the second is the daughter process, N_d , which determines the positions of points in each cluster. Each point of N_p initiates a process N_d around it independently to generate the members of the cluster, therefore each of the clusters are independent and identically distributed point processes. The final process is a superposition of all these resulting clusters, and the parent points can either be included or discarded Westcott (1971).

Neyman-Scott processes are a class of point processes which result in clustered spatial data. They were first developed to model the distribution of galaxies on the universe, something which had been done deterministically beforehand Neyman and Scott (1958). They are a special case of cluster processes where the parent distribution is homogenous Poisson of intensity λ_p , and the number of daughter points of each cluster is found randomly according to some distribution and are scattered independently and with identical spatial probability density function f about the parent. Also, the parent points are discarded. The resulting process is stationary and isotropic, and if f is radially symmetric, then the process is isotropic. Let \bar{c} be the mean number of daughter points per parent, then the overall intensity, λ , of a

Neyman-Scott process is

$$\lambda = \lambda_p \bar{c}.$$

Cluster processes can be characterised by their probability generating functions. In the case of the Neymann-Scott process on \mathbb{R}^d , this is:

$$G_n(v) = \exp \left(-\lambda_p \int_{\mathbb{R}^d} \left(1 - G_n \left(\int_{\mathbb{R}^d} v(x+y) f(y) dy \right) \right) dx \right)$$

where G_n is the probability generating function the the distribution used to obtain the number of daughters of each parent point Stoyan et al. (2013).

Matérn Cluster Process

The Matérn cluster process, as described in Stoyan et al. (1995), is a specific case of the Neyman-Scott class of point processes. The positions of the parent points are of course generated with a Poisson point process of intensity λ_p . Each daughter cluster has a number of points determined by a Poisson distribution with intensity λ_d , which are independently and uniformly distributed on on the ball $B(\mathbf{x}, R)$ centered around the respective parent, \mathbf{x} . R is a radius parameter which is determined. Therefore, the intensity of the overall process is

$$\lambda = \lambda_p \lambda_d.$$

Thomas Process

A Thomas point process, as described in Stoyan et al. (1995), is similar to the Matérn cluster process, except the daughter points of each cluster are distributed according to a d -dimensional normal distribution with a determined variance parameter, σ^2 . Let μ be the mean number of points per cluster, then the overall intensity is

$$\lambda = \lambda_p \mu.$$

Examples of Matérn and Thomas processes are shown in figure 2.1.

2.1.5 Regular Point Processes

Regular point processes are those which deter points from being within some distance of each other, which in turn reduces clustering. They are useful in modelling distributed wireless networks ElSawy and Hossain (2013), and in forestry Stoyan and Penttinen (2000).

Soft Core Process

One example of these processes is the Soft Core process which is a more general form of Matern's second hard-core process, see Matern (1986) for more details on this. The Soft Core process, as introduced in Stoyan (1987), is simulated by first generating a stationary

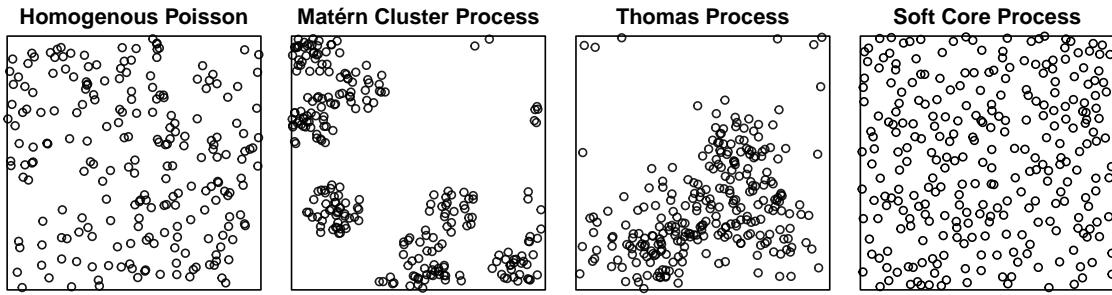


Figure 2.1: This figure shows examples of various point processes discussed in this section, all of which are shown on a unit square. The first subplot shows a homogenous Poisson point process with an intensity of $\lambda = 250$. The second subplot shows a Matérn cluster process which is a type of Neyman-Scott process. This has a parent and daughter intensities of $\lambda_p = 10$ and $\lambda_d = 25$ respectively, and a cluster radius of $R = 0.1$. The third subplot is a Thomas process, with the same parameters as the Matérn. The final subplot is a regular Soft Core point process with $\lambda_b = 500$, and the radii, ρ , are assigned with a probability density function of $f(\rho) = 800\rho$ where $0 < \rho < 0.05$.

Poisson process of intensity λ_b . Each point is then assigned a radius according to some continuous distribution function, F . Then, each point is then considered is either kept or removed according to some criteria. Consider an arbitrary point x with radius $r(x)$. This point survives if and only if there is no other point y with radius $r(y)$ such that x is in a circle of radius $r(y)$ centred at y , and $r(y) \geq r(x)$. Alternatively, a mark can also be randomly generated for each point alongside the radius according to a uniform distribution on $[0, 1]$. Then, the point x survives iff there is no point y with radius y and mark $m(y)$ such that x is in a circle of radius $r(y)$ centred at y , and $m(y) \geq m(x)$. Other slight variations to this process can be seen in Stoyan (1987).

In all cases, the resulting process is stationary and isotropic with an intensity, λ , of:

$$\lambda = \lambda_b p_s$$

where p_s is the probability of an arbitrary point from the initial Poisson process surviving. This probability is given by:

$$p_s = \int_0^1 \exp(-\lambda_b V(u)) du$$

where $V(u)$ is the volume of the set $M(o, u)$ which is three dimensional. For $x = (x_1, x_2)$ we have:

$$M(x, u) = \{(\xi, \eta, \zeta) : \sqrt{(\xi - x_1)^2 + (\eta - x_2)^2} \leq F^{-1}(\zeta), u \leq \zeta \leq 1\}.$$

explanation of this p_s in Stoyan et al. (2013) pg 177, come back to this!!

Mase et al. (2001) ball packing is an application of regular processes such as hard core gibbs? An example of a Soft Core process is shown in figure 2.1.

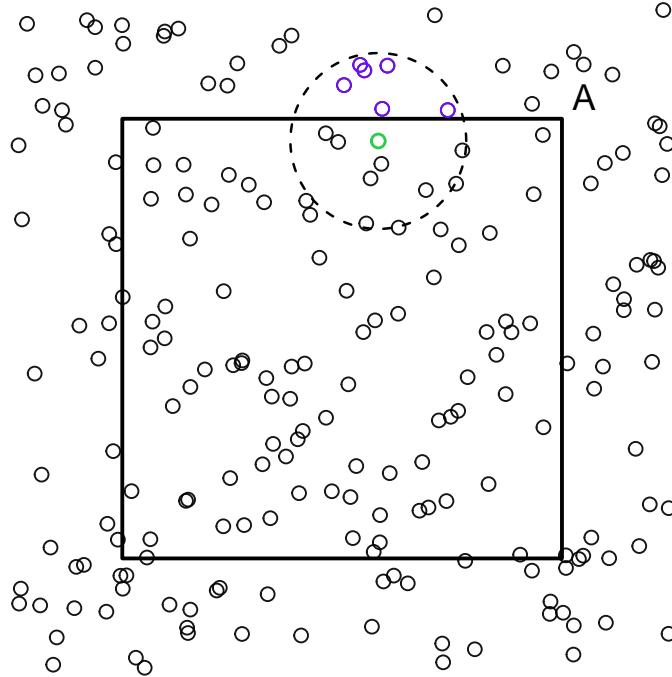


Figure 2.2: This figure shows the flaw in the naive estimator of Ripley's K , 2.3. Assume we have observed the points in the unit square region A . We sum over each of the points inside A , counting how many other points are within a distance R . Consider the green point inside the region, which happens to be located at $(1.08, 1.45)$. If we set $h = 0.2$, then clearly, the circular region extends beyond the observed region, and the purple points which are outside of A but within the radius will not be included in the sum. Therefore, this estimator is biased for a bounded observation window.

2.2 Ripley's K

Ripley's K , sometimes known as the reduced second moment measure, is a distance based statistic which ‘captures the spatial dependence between different regions of the point process’. Given some distance $h > 0$, Ripley's K is given as:

$$K(h) = \frac{E(C(\mathbf{x}, h)|\text{point at } \mathbf{x})}{\lambda} \quad (2.2)$$

where $C(\mathbf{x}, h)$ is the number of points within a distance h of a point \mathbf{x} , not counting \mathbf{x} itself. In other words, Ripley's K is the expected number of points within a distance h of any point in the process, normalised by the intensity, λ (Cressie (1991)). Note that we are assuming that the intensity, λ , is a constant, and therefore that the point process is homogenous.

2.2.1 Isotropic Estimator

Ripley's K of a point process can be estimated using an empirical average to replace the expectation. Let $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ be the locations of the n observed points over some region

$A \subset \mathbb{R}^d$. The simplest estimator then takes the following form:

$$\hat{K}(h) = \frac{1}{\hat{\lambda}} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{I(||\mathbf{x}_i - \mathbf{x}_j|| \leq h)}{n} \quad (2.3)$$

where $\hat{\lambda}$ is an estimate of the intensity. This estimator is unbiased if we assume all points which are within a distance h of the observed points are also observed. If this is not the case, then we have $E(\hat{K}(h)) < K(h)$, and therefore this estimator is biased (Ripley (1988)). This is illustrated in figure 2.2.

Denote the circular region of radius R centered at an arbitrary point \mathbf{x} as $B(\mathbf{x}, R)$, and its boundary as $\partial B(\mathbf{x}, R)$. The isotropic correction described in Ripley (1988) compensates for the issue presented for the estimator 2.3 by using the isotropy of the point process, and counting pairs more than once where required. Consider a pair of points $(\mathbf{x}_i, \mathbf{x}_j)$ for some $i, j \in \{1, \dots, n\}$, where the boundary $\partial B(\mathbf{x}_i, ||\mathbf{x}_i - \mathbf{x}_j||)$ is partially outside of the observation window. There could be unobserved points a distance of $||\mathbf{x}_i - \mathbf{x}_j||$ from \mathbf{x}_i , and so to compensate for this we count the pair $(\mathbf{x}_i, \mathbf{x}_j)$ more than once in the summation. We do this by multiplying each summand by a weight $w(\mathbf{x}_i, \mathbf{x}_j)$:

$$w(\mathbf{x}_i, \mathbf{x}_j) = \frac{|\partial B(\mathbf{x}_i, ||\mathbf{x}_i - \mathbf{x}_j||)|}{|\partial B(\mathbf{x}_i, ||\mathbf{x}_i - \mathbf{x}_j||) \cap A|}$$

which is essentially the inverse of the proportion of the boundary which lies within the observed region, A . Using these weights, an unbiased estimate can be formed

$$\hat{K}_{iso}(h) = \frac{|A|}{n(n-1)} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n I(||\mathbf{x}_i - \mathbf{x}_j|| \leq h) w(\mathbf{x}_i, \mathbf{x}_j) \quad (2.4)$$

where $|A|$ is the volume of the observed region.

The variance of Ripley's K with the isotropic correction is derived in closed form in Ripley (1988) as:

$$var(\hat{K}_{iso}(h)) = \frac{2|A|^2 \beta(h)}{n^2} (1 + 0.305\gamma(h) + \beta(h)(-1 + 0.0132m\gamma(h))) \quad (2.5)$$

where $\beta(h) = \frac{\pi h^2}{|A|}$ and $\gamma(h) = \frac{|\partial A|h}{|A|}$.

For the remainder of this report, we will discard the subscript and $\hat{K}(h)$ will refer to the Ripley's K estimate with an isotropic correction.

2.2.2 Ripley's K Under CSR

Assuming that a point process is completely spatially random in \mathbb{R}^2 , such as a homogenous Poisson point process, we know that $K(h) = \pi h^2$ (Cressie (1991)). Intuitively, this is because the numerator of 2.2 is simply the area of the circular region we are considering, multiplied

by the intensity:

$$K(h) = \frac{E(C(\mathbf{x}, h)|\text{point at } \mathbf{x})}{\lambda} = \frac{\lambda\pi h^2}{\lambda} = \pi h^2.$$

For clustered processes, Ripley's K is typically less than πh^2 , and for regular processes it is typically greater than πh^2 .

2.2.3 Inhomogenous Ripley's K

Now we consider the case where we no longer assume the point process to be homogenous, but we instead assume that the intensity at a point, \mathbf{x} , is determined by the function $\lambda(\mathbf{x})$. The isotropic estimate is extended to this case in [Baddeley et al. \(2000\)](#) in the following way:

$$\hat{K}_{inhom}(h) = \frac{1}{|A|} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \frac{I(||\mathbf{x}_i - \mathbf{x}_j|| \leq h) w(\mathbf{x}_i, \mathbf{x}_j)}{\lambda(\mathbf{x}_i)\lambda(\mathbf{x}_j)}.$$

2.3 Inhomogenous Intensity Estimates

Another property which will be estimated throughout this project is the intensity of an inhomogenous spatial point process. This can be done both parametrically and non-parametrically, both of which will be outlined here.

2.3.1 Inhomogenous Poisson Point Process

definition in pg 599 or section 8.5.1 of Cressie??

2.3.2 Parametric Estimation

The problem of parametric inhomogenous intensity estimation is called 'estimation with incomplete covariate data' in [Waagepetersen \(2008\)](#). When using this method, we will consider an inhomogenous Poisson point process, whose intensity function is dependent on some parameters β , and of course the spatial location/covariates.

Log-linear Model

A popular model discussed in [Waagepetersen \(2008\)](#) is the log-linear model for a Poisson point process, which is also sometimes called 'a modulated Poisson process'. Let $z(\mathbf{x})$ be a vector of the spatial covariates at some location $\mathbf{x} \in A \subset \mathbb{R}^d$. The intensity function following this model is then

$$\lambda(\mathbf{x}; \beta) = \exp(z(\mathbf{x})^T \beta), \quad (2.6)$$

where A is the region of interest. The reason this function is used is that it guarantees non negativity of the intensity at all positions [Rathbun et al. \(2007\)](#).

Maximum Likelihood Estimator

In order to estimate the parameters β , we will use a variance of the maximum likelihood estimator. The loglikelihood of an inhomogenous Poisson point process with a log-linear intensity, X , observed within the bounded observation window A , is

$$\mathcal{L}_A(\beta) = \beta^T z(\mathbf{x}) - \int_A \exp(\beta^T z(\mathbf{x})) d\mathbf{x}$$

as seen in Rathbun et al. (2007).

The issue which arises here is that in order to obtain the optimal β through maximising this likelihood, we require the values of $z(\cdot)$ to be known within the entire region, A , however, these are usually only observed at a finite subset of locations which are the events of the point process X . One way to overcome this is to approximate the integral above term above.

Firstly, note that the MLE we are aiming to obtain, $\hat{\beta}$ satisfies $\psi_A(\hat{\beta}) = 0$, where

$$\psi_A(\beta) = \frac{1}{|A|} \left(\sum_{x \in X} z(\mathbf{x}) - \int_A z(\mathbf{x}) \exp(\beta^T z(\mathbf{x})) d\mathbf{x} \right) \quad (2.7)$$

where the extra $1/|A|$ multiplier is added here for consistency with Rathbun et al. (2007), in which it is a technical requirement for the derivation of large-sample properties of $\hat{\beta}$. This quantity is also sometimes called the score function, such as in Waagepetersen (2008).

We now want to approximate the integral

$$Y(A; \beta) = \int_A z(\mathbf{x}) \exp(\beta^T z(\mathbf{x})) d\mathbf{x}.$$

Since we do not have the values of the covariates for all $\mathbf{x} \in A$, we use a 'design-unbiased' estimator for $Y(A; \beta)$ using the observed covariates which are obtained from a probability based sample of sites $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m \in A$. These are called 'covariate sample sites', and are sampled from a known joint probability distribution. Note that these are different to the event of the point process.

We then define the inclusion density,

$$\pi(\mathbf{x}) = \sum_{i=1}^m f_i(\mathbf{x}),$$

where $f_i(\cdot)$ is the known marginal probability density function for $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$. Assuming that $\pi(\mathbf{x}) > 0$ for all $\mathbf{x} \in A$, then an unbiased estimator for $Y(A; \beta)$ using the Horvitz-Thomson estimator is

$$\hat{Y}(A; \beta) = \sum_{i=1}^m \frac{z(\mathbf{u}_i) \exp(\beta^T z(\mathbf{u}_i))}{\pi(\mathbf{u}_i)},$$

(Rathbun et al. (2007)). Using this estimator, 2.7 becomes

$$\hat{\psi}_A(\beta) = \frac{1}{|A|} \left(\sum_{x \in A} z(x_i) - \sum_{i=1}^m \frac{z(u_i) \exp(\beta^T z(u_i))}{\pi(u_i)} \right), \quad (2.8)$$

and we can solve for $\tilde{\beta}$ which satisfies $\hat{\psi}_A(\tilde{\beta}) = 0$.

Asymptotic Properties of $\hat{\beta}$

Rathbun and Cressie (1994) show that the estimator $\hat{\beta}$ satisfying $\hat{\psi}_A(\hat{\beta}) = 0$ as in 2.7, has various asymptotic properties.

It is shown that $\hat{\beta}$ is consistent, and asymptotically Gaussian with a covariance matrix equal to the inverse Fisher information, $\text{var}(\hat{\beta}) = I(\beta)^{-1}$ where

$$I(\beta) = \int_A z(x) z^T(x) \exp(\beta^T z(x)) dx.$$

As well as this, properties of $\tilde{\beta}$ which is derived using 2.8 in Rathbun et al. (2007), including property that it is also asymptotically Gaussian.

2.3.3 Kernel Density Estimation

Chapter 3

Existing Bootstrapping Methods

3.1 Bootstrapping Methods for Distributions

3.1.1 Classical Bootstrapping

3.1.2 Jackknife

discuss how delete d is much better, which is similar to thinning

3.2 Bootstrapping Spatial Point Processes

3.2.1 Overview

Loh and Stein (2004) explores the effectiveness of four bootstrapping methods to find confidence intervals for estimates of Ripley's K statistic for various spatial point processes. This is more difficult task than resampling a point process to estimate a first-order statistic, since it is important to avoid producing pairs of points which have a different clustering structure to the original process. [first page of Loh and Stein \(2004\) has a list of a few papers talking about bootstrapping for first moment statistics](#)

[paragraph on basic bootstrap Davison and Hinkley \(1997\)](#)

3.2.2 Splitting Method

The splitting method is the most simple of the four. It entails splitting the region over which the point process has been observed into N congruent subregions. We assume that the statistics calculated on each subregion are independent with a Gaussian distribution, which becomes a more sensible assumption at N increases (as a result of the central limit theorem, Billingsley (1986)). We also assume that the variance of the statistic in a subregion is N times greater than the statistic observed over the entire observed region. To obtain a confidence interval the variance of the statistic is estimated by the sample variance of the set of N statistics estimated from each of the subregions, more precisely, for an arbitrarily

statistic θ , we obtain an $(1 - \alpha)\%$ confidence interval as:

$$\left(\hat{\theta} - t_{N-1,\alpha/2} \sqrt{\frac{\widehat{Var}(\hat{\theta}_i)}{N}}, \hat{\theta} + t_{N-1,\alpha/2} \sqrt{\frac{\widehat{Var}(\hat{\theta}_i)}{N}} \right), \quad \widehat{Var}(\hat{\theta}) = \frac{1}{N-1} \sum_{i=1}^N (\hat{\theta}_i - \hat{\theta})^2$$

where $\hat{\theta}_i$ is the statistic estimated from the i th subregion, $\hat{\theta}$ is the statistic estimated from the entire observed region, and $t_{N-1,\alpha/2}$ is the $(1 - \alpha/2)$ th percentile of the t -distribution with $N - 1$ degrees of freedom.

3.2.3 Tiling Method

This method is carried out by randomly sampling N subregions of area a/N where a is the area of the entire observed window. The patterns within each of these subregions are then rearranged in a predetermined way into a new point process. This new process has \hat{n} points and an estimate of the statistic is found from it. This process is repeated B times, and using the basic bootstrap method, a $(1 - \alpha)\%$ confidence interval is given by

$$[2\hat{\theta} - \tilde{\theta}_{(B+1)(1-\alpha/2)}, 2\hat{\theta} - \tilde{\theta}_{(B+1)\alpha/2}] \quad (3.1)$$

where $\tilde{\theta}_{(B+1)(1-\alpha/2)}$ and $\tilde{\theta}_{(B+1)\alpha/2}$ are the $(B + 1)(1 - \alpha/2)$ th and $(B + 1)\alpha/2$ th ordered values of the B bootstrapped estimates. For this method, we consider a toroidal wrapping as described in [Statistics et al. \(1991\)](#); [Loh and Stein \(2008\)](#), where we treat the observed window as being wrapped around a torus. Figure 3.1 demonstrates how this method is carried out when $N = 4$. The left subfigure is the original point process, in the case a Poisson process with intensity of $\lambda = 50$. The overlaid boxes are the 4 randomly selected subregions. The boxes shown in red are ones where toroidal wrapping is applied since the subregions extend beyond one observed region. The right subfigure shows the selected subregions arranged in a pre-determined way into the new bootstrapped process.

This also shows a key limitation of the tiling method, in that when artificially placing resampled tiles next to each other can introduce new point pairs which effect the second order characteristics. This is most prominent in hard core processes where points cannot be within a distance of r_0 , but this may not be true for the bootstrapped process. It is shown in [Lahiri \(1993\)](#) that this method of bootstrapping does not provide an accurate distribution of the desired statistic when the data exhibits long range dependence, meaning that the covariance between points persists over long distances [Cressie \(1991\)](#).

3.2.4 Subsets Method

The subsets method is similar to the tiling method in that it requires selecting subregions of the entire observed region, but in this case they are not rearranged into a new process, but the statistic is calculated by only considering the effects between points in the same subregion. This method therefore does not have the problem of introducing correlation which did not

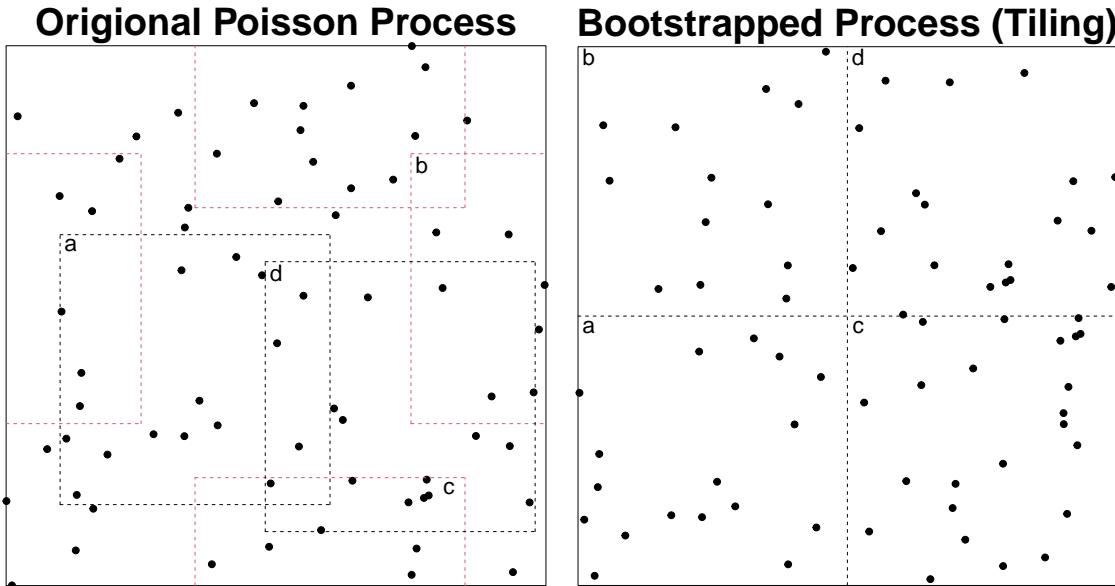


Figure 3.1: Demonstration of bootstrapping via the tiling method with $N = 4$. The left subfigure shows the randomly selected subregions. Those outlined in red demonstrate toroidal wrapping. The right subfigure shows the new bootstrapped process.

exist in the original process. Note however that use of toroidal wrapping results in some artificially produced pairs, but much fewer than the tiling method. This method is similar to subsampling introduced in Politis and Romano (1994).

3.2.5 Marked Point Method

The marked point method was also developed in Brani and Kulperger (1998) to estimate the second order intensity of a point process in \mathbb{R} and is adapted in Loh and Stein (2004) to spatial point processes, in particular when the desired statistic is Ripley's K with an isotropic edge correction. This is done by assigning each observed point, x , a mark, m_x which is equal to the sum of the weights of all points within a distance r of it. Note that this is the weight applied during the isotropic edge correction, as discussed in section ???. More precisely,

$$m_x = \sum_{y:y \neq x} 1\{|y - x| \leq r\} w_A(x, y)$$

where $w_A(x, y)$ is the weight arising from the isotropic edge correction. We now again resample the points, along with their marks, by selecting randomly placed subregions. Note that no new marks are calculated. Then, the bootstrapped Ripley's K is calculated by finding the sum of these marks, and normalising them:

$$\tilde{K}(r) = \frac{a}{\sum n_i (\sum n_i - 1)} \sum_{i=1}^N \sum_{j=1}^{n_i} m_{ij}$$

where a is the area of the entire observed region, and m_{ij} is the mark corresponding to the j th point of the i th sampled subregion.

Again, this method does not artificially introduce point pairs, and it is useful since the mark assigned to each point encodes information of the point pairs beyond the subregion window.

3.2.6 Bootstrapping inhomogenous processes

use Loh (2010) (this also uses marked point method i think) (note that this is for second-order intensity reweighted stationary point processes)

3.3 Bootstrapping Methods for Inhomogeneous Intensity Estimates

3.4 Proposed Method: Thinning

Chapter 4

Analytic Results

4.1 Asymptotics of Thinning

4.1.1 Expansion of Statistical Functional

change thetas to T so match literature notations? We now consider processes more generally in an arbitrary dimension, d . We aim to obtain the distribution of a statistical functional, $\theta : \mathcal{F} \rightarrow \mathbb{R}$, when estimates are obtained from thinned samples and as the number of observations increases $n \rightarrow \infty$. \mathcal{F} is the space of distributions, including the true distribution F and the empirical distribution F_n . Also, we require \mathcal{F} to be convex, such that for any $F, G \in \mathcal{F}$, the linear combination $(1 - \lambda)F + \lambda G$ is also in \mathcal{F} .

Let $\mathbf{x} = (x_1, x_2, \dots, x_n)$ be the observed data, which are independent and identically distributed according to some unknown distribution, F , with finite variance. Here we allow for spatial data, i.e., $\mathbf{x}_i \in \mathbb{R}^d$, $\forall i \in \{1, 2, \dots, n\}$, $d \in \mathbb{N}$. Let $\theta = \theta(F)$ be the true statistical functional which is a function of the true distribution, and $\hat{\theta} = \hat{\theta}(F_n) = \hat{\theta}(\mathbf{x})$ be an estimator of θ which is a function of the empirical distribution.

Firstly, we aim to understand the distribution of $\hat{\theta} - \theta$ for an arbitrary statistical functional. v. Mises (1947) describes the Taylor expansion for a statistical functional:

$$\hat{\theta}(F_n) = \theta(F) + d_1\theta(F; F_n - F) + \frac{1}{2!}d_2\theta(F; F_n - F) + \dots$$

There are two possible formulations of this notion of a derivative of a statistical functional, the Gâteaux derivative, and the Fréchet derivative. The first Gâteaux derivative of θ at F in the direction of G is defined as

$$d_1\theta(F; G - F) = \lim_{\lambda \rightarrow 0^+} \frac{\theta((1 - \lambda)F + \lambda G) - \theta(F)}{\lambda}$$

where $G \in \mathcal{F}$. More generally, the k th derivative is given by

$$d_k\theta(F; F_n - F) = \left. \frac{d^k}{d\lambda^k} \theta((1 - \lambda)F + \lambda G) \right|_{\lambda=0^+}$$

given that the limit exists. We can then set $G = F_n$ in order to formulate the remainder term as

$$R_n = \hat{\theta}(F_n) - \theta(F) - \sum_{k=1}^m \frac{1}{k!} d_k \theta(F; F_n - F)$$

for some $m \in \mathbb{N}$. For the asymptotic result which will be derived in 4.1.3, we will take $m = 1$.

In general, we are considering estimators which can be expanded in the following form:

$$\hat{\theta}(F_n) = \theta(F) + \frac{1}{n} \sum_{i=1}^n \phi_F(x_i) + R_n, \quad (4.1)$$

where ϕ_F is measurable in x with $E_F(\phi_F(X)) = 0$, $0 < E_F(\phi_F^2(X)) = \sigma^2 < \infty$. The condition that ϕ_F should be measurable ensures that the resulting $\phi_F(x_i)$ terms are all also random variables, since random variables are defined to be measurable functions (Fristedt and Gray (1997)), and the composition of measurable functions is also measurable (Strichartz (2010)). We then see that this form of the estimator splits it into terms which arise from the true distribution, and the remainder which captures the error arising from having limited observations. We can therefore describe estimators satisfying (4.1) as having a **linear part which is $O(n^{-1/2})$ (or higher?????????????)** and a remainder term with a lower order.

Note that the summation in 4.1 matches up to the first derivative in the Taylor expansion:

$$d_1 \theta(F; F_n - F) = \frac{1}{n} \sum_{i=1}^n \phi_F(x_i).$$

This happens via the von Mises derivative (Fernholz (1983)). θ is von Mises differentiable if its Gâteaux derivative, $d_1 \theta(F; G - F)$, and can be expressed as an integral of a measurable function ϕ_F , over the difference of the distributions:

$$d_1 \theta(F; F_n - F) = \frac{d}{d\lambda} \theta((1 - \lambda)F + \lambda G) \Big|_{\lambda=0^+} = \int \phi_F(x) d(G - F)(x).$$

If this holds, we call ϕ_F the influence function of θ at F . As explained in Fernholz (1983), this function applied to some point x can be written as

$$\phi_F(x) = \frac{d}{d\lambda} \theta((1 - \lambda)F + \lambda \delta_x) \Big|_{\lambda=0^+}$$

where δ_x is the Dirac measure at x , which assigns a measure of 1 to any set containing x and a measure of 0 otherwise. Also, note that the influence function is only unique up to an additive constant, so we can choose it such that **(CAN ADD MORE DETAIL ON WHY THIS IS TRUE)**

$$\int \phi_F(x) dF(x) = 0.$$

Therefore, given that θ is von Mises differentiable, the expansion 4.1 can be obtained by

truncating the Taylor series of the statistical functional in the following way:

$$\begin{aligned}
\hat{\theta}(F_n) &= \theta(F) + d_1\theta(F; F_n - F) + \frac{1}{2!}d_2\theta(F; F_n - F) + \dots \\
&= \theta(F) + d_1\theta(F; F_n - F) + R_n \\
&= \theta(F) + \int \phi_F(x)d(F_n - F)(x) + R_n \\
&= \theta(F) + \int \phi_F(x)dF_n(x) - 0 + R_n \\
&= \theta(F) + \frac{1}{n} \sum_{i=1}^n \phi_F(x_i) + R_n
\end{aligned}$$

4.1.2 Convergence of Remainder Term

In the case of $m = 1$, it is desirable for the remainder term to satisfy

$$\sqrt{n}R_n \xrightarrow{P} 0. \quad (4.2)$$

which also shows that the σ^2 seen in 4.1 is simply the limiting variance of $\sqrt{n}(\hat{\theta}(F_n) - \theta(F))$. We can now also interpret this expansion of $\hat{\theta}(F_n)$ as allowing us to write it as a [linear statistic](#), with a lower order approximation error.

The convergence seen in 4.2 can be guaranteed with the condition that

$$\sqrt{n} \sup_{0 \leq \lambda \leq 1} \left| \frac{d^2}{d\lambda^2} \theta((1-\lambda)F + \lambda F_n) \right| \xrightarrow{P} 0,$$

but this is difficult to analyse as it requires a derivative of higher order than necessary. Numerical explorations of this are less complicated and produce useful results. Another way to ensure 4.2 holds is via the Fréchet derivative.

Let $F, G \in \mathcal{F}$, and \mathcal{D} be the linear space generated by differences $F - G$. Let $\|\cdot\|$ be a norm on \mathcal{D} , then the functional $\theta : \mathcal{F} \rightarrow \mathbb{R}$ is said to have a Fréchet derivative at $F \in \mathcal{F}$ with respect to norm $\|\cdot\|$ if there exists a functional $\theta(F; \Delta) : \mathcal{D} \rightarrow \mathbb{R}$ and linear in the argument Δ such that

$$\theta(G) - \theta(F) - \theta(F; G - F) = O(\|G - F\|)$$

as $\|G - F\| \rightarrow 0$. $\theta(F; \Delta)$ is called the Fréchet derivative ([Serfling \(1980\)](#)).

Although Fréchet differentiability is a stronger condition than Gâteaux differentiability, they both provide approximations to the desired

$$\hat{\theta}(F_n) - \theta(F)$$

as seen in equation 4.1. This term is approximated using the random variable $\theta(F; F_n - F)$ and $d_1\theta(F; F_n - F)$ respectively. The following lemma, shows that these approaches are equivalent, and that the Fréchet derivative is stronger than the Gâteaux derivative.

Lemma 1. *If θ has a Fréchet derivative at F with respect to $\|\cdot\|$, then for any $G \in \mathcal{F}$,*

$d_1\theta(F; G - F)$ exists and

$$d_1\theta(F; G - F) = \theta(F; G - F)$$

This lemma is proven on pg. 218 of Serfling (1980). The following lemma shows that Fréchet differentiability is sufficient to satisfy the condition in equation 4.2.

Lemma 2. Assume θ has a Fréchet derivative at F with respect to $\|\cdot\|$. Let $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ be observations from F such that $\sqrt{n}\|F_n - F\| = O(1)$. Then

$$\sqrt{n}R_n \xrightarrow{P} 0.$$

Again, this proof can be found on pg. 218 of Serfling (1980). The requirement $\sqrt{n}\|F_n - F\| = O(1)$ is established in \mathbb{R} by the Dvoretzky–Kiefer–Wolfowitz inequality (Dvoretzky et al. (1956)), which states

$$P\left(\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| > \epsilon\right) < C \exp(-2n\epsilon^2)$$

where $\epsilon > 0$ and C is some constant.

Many statistics satisfy the condition 4.2, including L statistics, maximum likelihood estimates, least-squares estimates, and U statistics (Serfling (1980); Wu (1990)).

4.1.3 Derivation of Distribution of Thinned Estimates

Consider \mathbf{x}_s to be a subsample of \mathbf{x} which is obtained by thinning, (sometimes described as Poisson subsampling as in Hájek (1960)). This is done by selecting each data point in \mathbf{x} with probability p . Here we will consider the case where the probability assigned to each x_i is the same, this method is sometimes called Bernoulli subsampling (Wang and Zou (2021)). Thinning is equivalent to simple random sampling without replacement where the size of the subsample, r , is a random variable with a binomial distribution. In order to obtain an estimate of the distribution of θ , the statistic is calculated on B thinned samples where B is large, so this size of each thinned subsample is conditionally distributed as $r_s|N = n \sim Bin(n, p)$ for $s = 1, 2, \dots, B$. Also set $d_s = n - r_s$ to be the number of deleted samples. Note that if r_s is sampled to be 0 or n , we either have an empty subsample, or one equivalent to the original observed sample, so we discard this and resample r_s . Both of these cases are unlikely when p is not close to 0 or 1, and the probability of this happening converges to 0 as $n \rightarrow \infty$. This result is proven on page 151 of Feller (1968) and will be discussed in more detail later in this proof.

We will assume that $\hat{\theta}$ obeys the central limit theorem and is asymptotically normal with limiting variance σ^2 ,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \sigma^2) \tag{4.3}$$

so we need to show that the distribution of the thinned $\hat{\theta}_s$ is also asymptotically normal.

Note that thinning can also be reframed as a delete- d Jackknife where d , the number of x_i deleted for each subsample is instead a variable. The following proof is based on the

asymptotic normality of the distribution of the Jackknife estimates presented in Theorem 2 (iii) of [Wu \(1990\)](#).

Here we define the Thinning histogram analogously to the 'Jackknife histogram' in [Wu \(1990\)](#)

$$H(t) = P \left\{ \left(\frac{nr_s}{d_s} \right)^{1/2} \frac{\hat{\theta}_s - \hat{\theta}}{\hat{\sigma}} \leq t \right\}, \quad (4.4)$$

where $\hat{\sigma}^2$ is a consistent estimator of the limiting variance σ^2 . Since $\hat{\theta}$ is assumed to be normal, the aim is to prove that $\hat{\theta}_s$ is also $\forall s \in 1, 2, \dots, B$.

Theorem 3. *If $d_s/n \geq \lambda$ for some $\lambda > 0$ and $r_s \rightarrow \infty$ for all $s = 1, 2, \dots, B$, then*

$$\sup_t |H(t) - \Phi(t)| \xrightarrow{P} 0 \quad (4.5)$$

is this convergence in distribution or probability?

For this proof, we will use the assumed form of the estimator seen in (4.1) to obtain:

$$\begin{aligned} \hat{\theta}_s &= \theta + \frac{1}{r_s} \sum_{i \in x_s} \phi_F(x_i) + R_{n,s}, \\ \hat{\theta}_s - \hat{\theta} &= \underbrace{\frac{1}{r_s} \sum_{i \in x_s} \phi_F(x_i) - \frac{1}{n} \sum_{i=1}^n \phi_F(x_i)}_A + \underbrace{R_{n,s} - R_n}_B. \end{aligned}$$

Firstly, we will show in theorem 4 that the term linear in ϕ_F , A, converges asymptotically to a normal distribution, and then in theorem 5 that the remainder term, B, is asymptotically negligible.

To show the convergence of the linear part, define $\hat{\theta}_s^L$ and $\hat{\theta}^L$ to be the truncated versions of $\hat{\theta}_s$ and $\hat{\theta}$ respectively by removing the remainders:

$$\begin{aligned} \hat{\theta}_s^L &= \theta + \frac{1}{r_s} \sum_{i \in x_s} \phi_F(x_i), \\ \hat{\theta}^L &= \theta + \frac{1}{n} \sum_{i=1}^n \phi_F(x_i). \end{aligned}$$

Let the corresponding Thinning histogram be:

$$H_L(t) = P \left\{ \left(\frac{nr_s}{d_s} \right)^{1/2} \frac{\hat{\theta}_s^L - \hat{\theta}}{\hat{\sigma}} \leq t \right\}.$$

Theorem 4. *If $0 < E[(X - \mu)^2] < \infty$, $r_s \rightarrow \infty$ and $d_s \rightarrow \infty$ for all $s = 1, 2, \dots, B$, then*

$$\sup_t |H_L(t) - \Phi(t)| \xrightarrow{a.s.} 0,$$

where Φ is the standard normal cdf.

Proof. We will begin by showing that the following is satisfied for a thinned set x_s of size r_s :

$$\lim_{n \rightarrow \infty} \frac{1}{(n-1)\hat{\sigma}^2} \sum_{i=1}^n (x_i - \hat{\theta})^2 \mathbf{1}\left(|x_i - \hat{\theta}| \geq \tau \sqrt{\frac{r_s d_s}{n}} \hat{\sigma}\right) = 0, \quad \tau > 0. \quad (4.6)$$

Note that $\hat{\sigma}^2 r_s d_s / n$ is the variance of $\hat{\theta}_s$ as seen in Hájek (1960).

Since we have assumed $\hat{\theta} \rightarrow \theta$ and $\hat{\sigma} \rightarrow \sigma < \infty$ almost surely, (4.6) becomes:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i - \theta)^2 \mathbf{1}\left(|x_i - \theta| \geq \tau \sqrt{\frac{r_s d_s}{n}}\right), \quad (4.7)$$

where the coefficient of the sum has been simplified but remains $O(1/n)$ and the σ inside the indicator is taken into the constant τ .

Now we bound $r_s d_s / n$ above and below. Without loss of generality, let $\min(r_s, d_s) = d_s$, such that $\hat{p} = r_s/n \geq \frac{1}{2}$, then

$$\begin{aligned} \frac{r_s d_s}{n} &= n\hat{p}(1-\hat{p}) = d_s \hat{p} > \frac{1}{2} \min(r_s, d_s) \\ \frac{r_s d_s}{n} &= \frac{(n-d_s)d_s}{n} = d_s - \frac{d_s^2}{n} < d = \min(r_s, d_s) \\ \implies \frac{1}{2} \min(r_s, d_s) &< \frac{r_s d_s}{n} < \min(r_s, d_s) \end{aligned}$$

which shows $\min(r_s, d_s) \rightarrow \infty \iff r_s d_s / n \rightarrow \infty$ as $n \rightarrow \infty$. This means that given any $k > 0$, we can choose an m large enough such that for $n \geq m$, $r_s d_s / n > k^2$, then (4.7) can be bounded above:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\theta})^2 \mathbf{1}(|x_i - \theta| \geq \tau k)$$

which is arbitrarily small since k is arbitrarily large. So we have shown that (4.6) holds, then clearly

$$\lim_{n \rightarrow \infty} \frac{1}{(n-1)\hat{\sigma}^2} \sum_{i \in x_s} (x_i - \hat{\theta})^2 \mathbf{1}\left(|x_i - \hat{\theta}| \geq \tau \sqrt{\frac{r_s d_s}{n}} \hat{\sigma}\right) = 0 \quad \tau > 0$$

since we have only removed d_s terms from the sum here. By Theorem 3.1 in Hájek (1960), this implies that $\hat{\theta}_s^L$ has an asymptotically normal distribution with mean θ and variance $\hat{\sigma}^2 r_s d_s / n$. Then Pólya (1920) implies the initial claim:

$$\sup_t |H_L(t) - \Phi(t)| \xrightarrow{P} 0.$$

Therefore the claim holds. \square

We now show that the remainder term, B , is asymptotically negligible.

Theorem 5. If $\sqrt{n}R_n \xrightarrow{P} 0$, $\frac{d_s}{n} > \lambda$ for some $\lambda > 0$ for all $s = 1, 2, \dots, B$ and $r_s \rightarrow \infty$, then

$$Q(\epsilon) = P \left\{ \left(\frac{nr_s}{d_s} \right)^{1/2} |R_{n,s} - R_n| \leq \epsilon \right\} \xrightarrow{P} 0.$$

Proof. $Q(\epsilon)$ is clearly non-negative, and we can use the constant λ to bound it above:

$$\left(\frac{nr_s}{d_s} \right)^{1/2} |R_{n,s} - R_n| \leq \left(\frac{r_s}{\lambda} \right)^{1/2} |R_{n,s} - R_n|.$$

Then using the assumption that $\sqrt{n}R_n \xrightarrow{P} 0$:

$$\lim_{n \rightarrow \infty} \left(\frac{r_s}{\lambda} \right)^{1/2} |R_{n,s} - R_n| = \lim_{n \rightarrow \infty} |\sqrt{r_s}R_{n,s} - \sqrt{r_s}R_n|,$$

where the first term clearly converges to 0 since $R_{n,s}$ is of size r_s , and the second term is bounded above by $\sqrt{n}R_n$ which also converges to 0. Therefore, the claim is proven. \square

Note that the requirement that $\frac{d_s}{n} > \lambda$ for some $\lambda > 0$ may not always be met since d_s is a random variable, however as n becomes large we can see using Feller (1968) that for $p < 0.9$, the probability of d_s less than $0.1n$ for example converges to 0:

$$P(d_s \leq 0.1n) \leq \frac{(n - 0.1n)(1 - p)}{(n(1 - p) - 0.1n)^2} = \frac{0.9(1 - p)}{n(0.9 - p)^2} \rightarrow 0.$$

This means that asymptotically, given that we have chose $p < 0.9$, we can assume $\frac{d_s}{n} > 0.1$ and more generally that this term is $O(n)$.

Finally, combining Theorems 2 and 3 prove the desired result stated in Theorem 1.

4.1.4 Usefulness of Asymptotic Result

Having now proven theorem 3, we have shown that the 'root' converges to a standard normal distribution:

$$\left(\frac{nr_s}{d_s} \right)^{1/2} \frac{\hat{\theta}_s - \hat{\theta}}{\hat{\sigma}} \xrightarrow{P} N(0, 1).$$

In order for this result to be useful, we want to show that the distribution of the standardised thinned estimate $\hat{\theta}_s$, is close to that of the standardised estimate over all the observed data, $\hat{\theta}$. To do this, we require a condition on $\hat{\theta}$ stronger than that imposed earlier.

Firstly, analogously to 4.4, let

$$H_0(t) = P \left\{ \frac{\sqrt{n}(\hat{\theta} - \theta)}{\sigma} \leq t \right\},$$

and assume now that

$$\sup_t |H_0(t) - \Phi(t)| \xrightarrow{P} 0. \quad (4.8)$$

The following theorem then provides us with the desired result.

Theorem 6. *If $d_s/n \geq \lambda$ for some $\lambda > 0$ and $r_s \rightarrow \infty$ for as $s = 1, 2, \dots, B$, then*

$$\sup_t |H(t) - H_0(t)| \xrightarrow{P} 0$$

Proof.

$$\begin{aligned} \sup_t |H(t) - H_0(t)| &= \sup_t |H(t) - \Phi(t) + \Phi(t) - H_0(t)| \\ &\leq \sup_t (|H(t) - \Phi(t)| + |H_0(t) - \Phi(t)|) \\ &\leq \sup_t |H(t) - \Phi(t)| + \sup_t |H_0(t) - \Phi(t)| \end{aligned}$$

By using theorem 3, and assuming 4.8, we know that

$$\begin{aligned} \sup_t |H(t) - \Phi(t)| &\xrightarrow{P} 0, \\ \sup_t |H_0(t) - \Phi(t)| &\xrightarrow{P} 0. \end{aligned}$$

Since all terms are positive, we have that also

$$\sup_t |H(t) - \Phi(t)| + \sup_t |H_0(t) - \Phi(t)| \xrightarrow{P} 0,$$

and so we can use the Squeeze theorem (Sohrab (2014)) to prove the result:

$$\begin{aligned} 0 \leq \sup_t |H(t) - H_0(t)| &\leq \sup_t |H(t) - \Phi(t)| + \sup_t |H_0(t) - \Phi(t)| \\ \implies \sup_t |H(t) - H_0(t)| &\xrightarrow{P} 0 \end{aligned}$$

□

We have now shown that the standardised distributions of the thinned estimate and the observed estimate asymptotically approach each other. Intuitively, since we know that for large n , the distribution of $\hat{\theta}$ is close to the distribution of θ , and the distribution of $\hat{\theta}_s$ is close to the distribution of $\hat{\theta}$, then $\hat{\theta}_s$ can be used to approximate the distribution of the desired θ .

4.2 Link Between Thinning and Bootstrap Sampling

In this section, we will show that when the thinning parameter takes a value of 0.5, there is an equivalence between the distribution that arises from the thinned estimates, and the distribution arising from the classical bootstrap.

Firstly, we carry out a numerical exploration. Figure 4.1 shows histograms of the estimators calculated from thinned samples and from bootstrapped samples for varying thinning

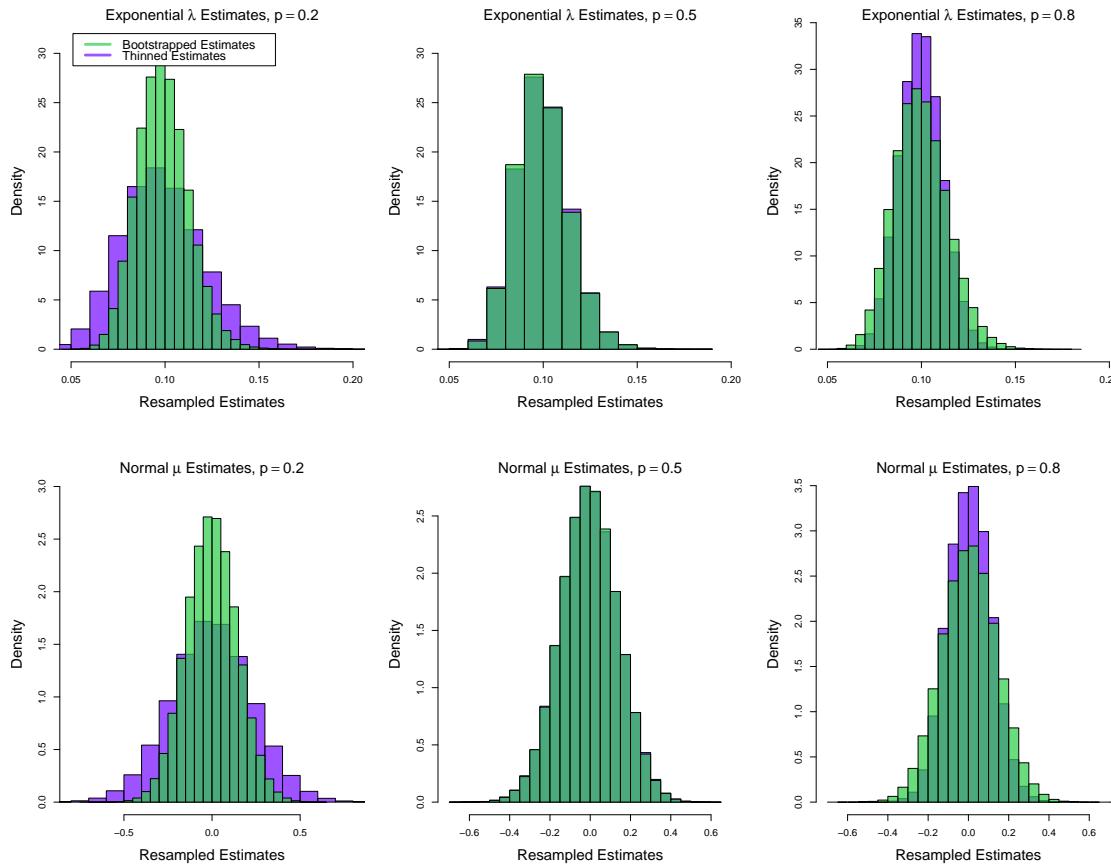


Figure 4.1: Histograms of bootstrapped estimates, $\hat{\theta}^*$, and thinned estimates, $\hat{\theta}_s$, shown in green and purple respectively. The top row uses i.i.d data generated from $X \sim Poi(10)$ and $\lambda = 10$ is estimated using its MLE, which we know satisfies the requirements of theorems 3 and 6. The bottom row uses data distributed according to $X \sim Exp(10)$ and again, $\lambda = 10$ is estimated using its MLE. These results are produced by carrying out 500 simulations of 100 data points each. For each other these simulations, 1000 thinning samples are obtained and an estimate of the desired statistic is calculated. These estimates are then used to form the histograms.

parameters, p . The plots for a poisson distribution with $\lambda = 10$, and for an exponential distribution with $\lambda = 10$ are shown here.

We see that a thinning parameter of $p = 0.2$, results in a distribution with a higher variance than the bootstrapped distribution because fewer observations are sampled, and therefore, we do not capture enough information. On the other hand, for $p = 0.8$, the variance of the thinned distribution is smaller than that of the bootstrapped distribution. In this case, we overfit to the observed data, and do not capture enough of the variance in the distribution. For a value of $p = 0.5$, the histograms are very similar, and seem to represent the same distribution up to randomness.

We can make this observation more precise by considering the asymptotic distributions which arise from thinning and bootstrapping. Note that we have assumed the estimator $\hat{\theta}$

obeys 4.3.

Let $\hat{\theta}^*$ be the estimate of the desired statistic based on the bootstrapped sample. Bickel and Freedman (1981) show that, given, 4.3, we have that

$$\frac{\sqrt{m}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}} \rightarrow N(0, 1)$$

where m is the number of bootstrap samples selecting. Here, we will consider the case $m = n$, such that for large n

$$\frac{\sqrt{n}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}} \sim N(0, 1). \quad (4.9)$$

The result from theorem 3 can be written as

$$\left(\frac{nr_s}{d_s} \right)^{1/2} \frac{\hat{\theta}_s - \hat{\theta}}{\hat{\sigma}} \rightarrow N(0, 1).$$

Clearly, when $p = 0.5$, $r_s \approx 0.5n$ and $d_s \approx 0.5n$, and therefore for large n

$$\frac{\sqrt{n}(\hat{\theta}_s - \hat{\theta})}{\hat{\sigma}} \sim N(0, 1),$$

which is approximately the same distribution to the bootstrapped distribution, 4.9.

Therefore, we conclude that thinning with a parameter of $p = 0.5$ performs very similarly to bootstrapping for statistics which obey the CLT.

4.3 Asymptotics for Spatial Point Processes

4.3.1 Conditions Required

We now turn our attention to the spatial point process setting, and whether theorem 4.5 applies in this case. There are various conditions required on both the statistic of interest, θ , and the underlying process. These will be explored in this section. These conditions are summarised here:

1. The observed data is independent and identically distributed according to some unknown process with finite covariance (analogous to finite variance? add moment measures to background theory maybe)
2. θ needs to be a statistical functional which maps some convex space of distributions, \mathcal{F} , to \mathbb{R}
3. θ can be approximated as a linear statistic with an error of order lower than $O(\sqrt{n})$

4.3.2 Conditions on Data

An important assumption we place on the sampled data is that they arise independently and are identically distributed. This assumption is much less restrictive when sampling from

probability distributions. For spatial point processes, the introduction of any clustering or regularity structure also introduces dependence into the data, and these are often the features we want to understand better via thinning. This issue has been explored thoroughly for the classical bootstrap and jackknife, and as stated in Cressie (1991), 'blind application of the classical jackknife and bootstrap techniques on dependent data can give incorrect answers'. This is because these resampling techniques do not generally preserve the underlying dependence structure according to Loh (2008), and it is possible that the same holds for thinning.

4.3.3 Conditions on Statistic

Needs updating

The two statistics of interest which we will focus on here are the parametric intensity estimate, $\hat{\lambda}$, and Ripley's K . Both of these can be considered as statistical functionals applied to the empirical functions, F_n .

The parametric intensity estimate we use here is a maximum likelihood estimate, and it is shown in Serfling (1980) that MLEs can be approximated by a linear statistic with the error being of lower order.

For Ripley's K this is more difficult. Sufficient conditions required to satisfy requirement 3 above have been discussed in section 4.1.2. It is difficult to generalise these analytically for a statistical functional in the spatial setting since we are no longer working with a space of distribution functions, \mathcal{F} . Instead, we will investigate this further through simulations.

4.4 Numerical Demonstration of Results

This section shows simulations which demonstrate the theorems previously proven. Firstly, we consider theorem 3. Figure 4.2 shows the distribution of the thinned estimates, $\hat{\theta}_s$, after normalisation compared with a standard normal distribution:

$$\left(\frac{nr_s}{d_s} \right)^{1/2} \frac{\hat{\theta}_s - \hat{\theta}}{\hat{\sigma}}.$$

The statistic being estimated in this case is the intensity, λ , of a poisson distribution which is estimated using its MLE. We choose this setting since we know it satisfies the desired conditions summarised in the previous section. As $n \rightarrow \infty$, we expect these to converge to the same distribution. This aligns with the results in figure 4.2, since for all thinning parameters considered, the histograms become more similar as n increases. We see also that the convergence occurs quite quickly, such that for $n = 20$, the method produces a reasonable distribution, and the improvements thereafter are less significant. Furthermore, there are no large differences in the distribution when varying the thinning parameter between $p = 0.2, 0.5, 0.8$. This is because this effect has been eliminated in the normalisation through the r_s/d_s term.

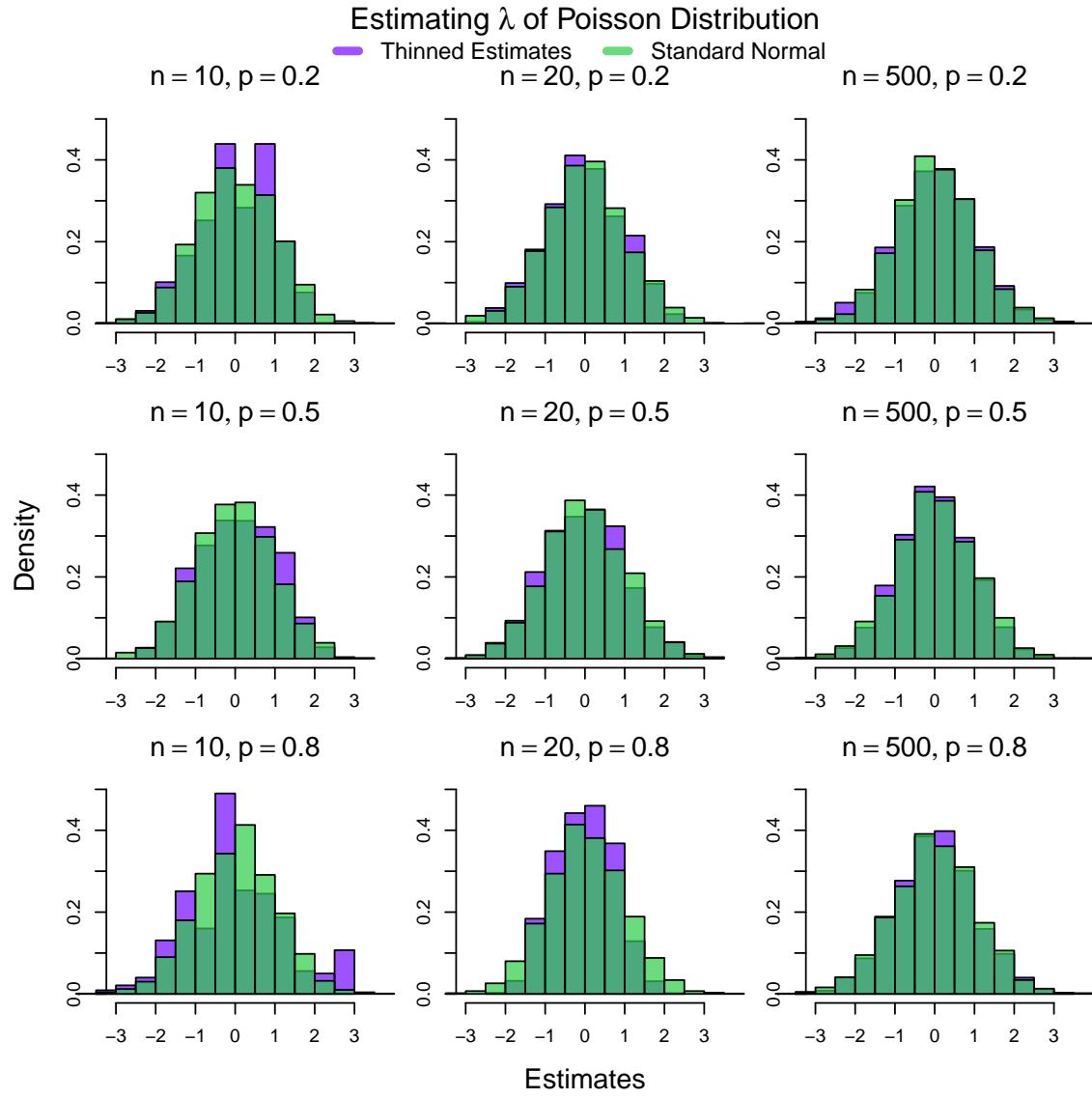


Figure 4.2: Histograms showing distributions of normalised thinned estimates, compared to a standard normal distribution. Each row represents a thinning parameter of $p = 0.2, 0.5, 0.8$ respectively, and each column represents varying observed dataset size, in particular $n = 10, 20, 500$ respectively. Each subplot is generated using n i.i.d datapoints distributed according to a $Poi(10)$ distribution. For each dataset, 2000 thinned samples are obtained and used to calculate the estimate. These are then plotted alongside 2000 samples from a standard normal distribution.

Next we demonstrate the result of theorem 6 in figure 4.3. we use the same statistic and estimate as before, except in this case, we compare against the normalised estimates over the entire observed data rather than a standard normal:

$$\frac{\sqrt{n}(\hat{\theta}_s - \theta)}{\sigma}.$$

Again, we see that as n increases, the the normalised $\hat{\theta}_s$ and $\hat{\theta}$ become closer in distribution, as expected.

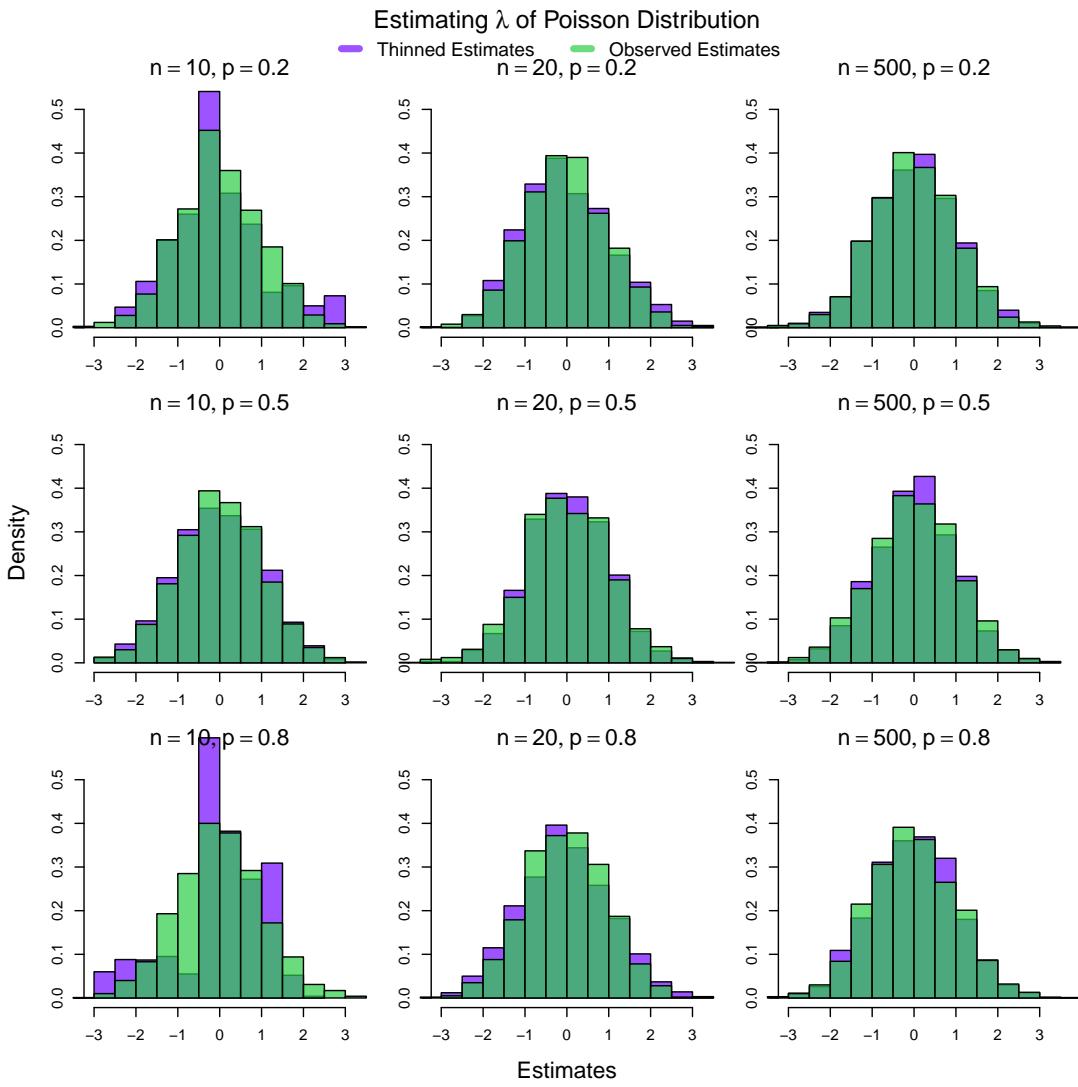


Figure 4.3: Histograms showing distributions of normalised thinned estimates, compared to the normalised estimates calculated over the entire observed dataset. Each row represents a thinning parameter of $p = 0.2, 0.5, 0.8$ respectively, and each column represents varying observed dataset size, in particular $n = 10, 20, 500$ respectively. Each subplot is generated using n i.i.d datapoints distributed according to a $Poi(10)$ distribution. For each dataset, 2000 thinned samples are obtained and used to calculate the estimate. For each case, we also generate 2000 sets of observed data from the distribution, and calculate the estimator over all of these. These estimates are then normalised and plotted against the thinned normalised estimates.

4.5 Example: Intensity of Homogeneous Poisson Point Process

In this section, we will explicitly show the result proven in theorem 3 for the intensity of a homogeneous Poisson spatial point process.

Firstly, we define $\hat{\lambda}$. We know that for a homogeneous Poisson process, the number of points within a bounded Borel set, A , is

$$N(A) \sim Poi(\lambda|A|), \quad (4.10)$$

where λ is the true intensity of the point process, and $|A|$ is the area of A . Let r be a random variable corresponding to the number of points which are retained after thinning. We know this is conditionally distributed as

$$r|N(A) = n \sim Bin(n, p). \quad (4.11)$$

By definition, this intensity of a homogeneous Poisson process over a set A with volume $|A|$ is

$$\lambda = \frac{E[N(A)]}{|A|}$$

and thus, given a realisation $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$, the intensity can be estimated in the following way

$$\hat{\lambda} = \frac{N(A)}{|A|},$$

van Lieshout (2019). Now, to define $\hat{\lambda}_s$ using the thinned subsample, we again take the ratio of the number of points and the area of the set. This time, since the mean of the number of points retained is np , we need to also normalise by the thinning probability, p :

$$\hat{\lambda}_s = \frac{r}{|A|p}.$$

The aim here is to show that the distribution of $\sqrt{n}(\hat{\lambda} - \lambda)$ is asymptotically the same as that of

$$\sqrt{\frac{nr_s}{d_s}} (\hat{\theta}_s - \hat{\theta}).$$

Since we know $\hat{\lambda}$ has a Poisson distribution with finite variance, we can apply the central limit theorem:

$$\sqrt{n}(\hat{\lambda} - \lambda) \rightarrow N(0, \sigma^2)$$

for large n . Now, before showing the desired result, we need to find the variances of $\hat{\lambda}$ and $\hat{\lambda}_s$.

Theorem 7. *The intensity estimate has a variance of*

$$var(\hat{\lambda}) = \frac{\lambda}{|A|}$$

and the intensity estimate of a thinned homogenous Poisson point process has a variance of

$$var(\hat{\lambda}_s) = \frac{\lambda}{|A|p}.$$

Proof. The first of these is straight forward, using the known distribution of $N(A)$:

$$\text{var}(\hat{\lambda}) = \text{var}\left(\frac{N(A)}{|A|}\right) = \frac{1}{|A|^2} \text{var}(N(A)) = \frac{\lambda|A|}{|A|^2} = \frac{\lambda}{|A|}.$$

Finding the variance of the thinned estimate is trickier, since we only know the conditional probability of $r|N = n$, whereas the distribution of r is required. To find this, we use the law of total probability:

$$P(r = x) = \sum_{m=0}^{\infty} P(r = x|N = m)P(N = m).$$

Since the terms where $m < x$ are zero, the index of the sum can begin at x . Using the known distributions 4.10 and 4.11:

$$\begin{aligned} P(r = x) &= \sum_{m=x}^{\infty} P(r = x|N = m)P(N = m) \\ &= \sum_{m=x}^{\infty} \binom{m}{x} p^r (1-p)^{m-r} \frac{(\lambda|A|)^m \exp(-\lambda|A|)}{m!} \\ &= \sum_{m=x}^{\infty} \frac{m!}{r!(m-r)!} p^r (1-p)^{m-r} \frac{(\lambda|A|)^m \exp(-\lambda|A|)}{m!} \\ &= \sum_{m=x}^{\infty} \frac{1}{r!(m-r)!} p^r (1-p)^{m-r} (\lambda|A|)^m \exp(-\lambda|A|) \\ &= \frac{p^r \exp(-\lambda|A|)}{r!} \sum_{m=x}^{\infty} \frac{1}{(m-r)!} (1-p)^{m-r} (\lambda|A|)^m \\ &= \frac{(\lambda|A|p)^r \exp(-\lambda|A|)}{r!} \sum_{m=x}^{\infty} \frac{1}{(m-r)!} (1-p)^{m-r} (\lambda|A|)^{m-r} \end{aligned}$$

Applying a change of variables $k = m - r$, and recognising the Taylor expansion of an exponential:

$$\begin{aligned} P(r = x) &= \frac{(\lambda|A|p)^r \exp(-\lambda|A|)}{r!} \sum_{k=0}^{\infty} \frac{(\lambda|A|(1-p))^k}{k!} \\ &= \frac{(\lambda|A|p)^r \exp(-\lambda|A|)}{r!} \exp(\lambda|A|(1-p)) \\ &= \frac{(\lambda|A|p)^r \exp(-\lambda|A|p)}{r!} \end{aligned}$$

which we recognise as a Poisson probability distribution, such that

$$r \sim \text{Poi}(\lambda|A|p). \quad (4.12)$$

Using this, we can now find the variance of the thinned estimator:

$$\text{var}(\hat{\lambda}_s) = \text{var}\left(\frac{r}{|A|p}\right) = \frac{1}{(|A|p)^2} \text{var}(r) = \frac{\lambda|A|p}{(|A|p)^2} = \frac{\lambda}{|A|p}.$$

The resulting relationship between the variances of the two estimators $\hat{\lambda}$ and $\hat{\lambda}_s$ is

$$\text{var}(\hat{\lambda}_s) = \frac{\text{var}(\hat{\lambda})}{p}. \quad (4.13)$$

□

Theorem 8. *The variance of $\hat{\lambda}_s - \hat{\lambda}$ is given by*

$$\text{var}(\hat{\lambda}_s - \hat{\lambda}) = \frac{(1-p)\lambda}{|A|p}.$$

Proof.

$$\begin{aligned} \text{var}(\hat{\lambda}_s - \hat{\lambda}) &= \text{var}\left(\frac{r}{|A|p} - \frac{N(A)}{|A|}\right) \\ &= \frac{1}{|A|^2} \text{var}\left(\frac{r}{p} - N(A)\right) \\ &= \frac{1}{|A|^2} \left[\text{var}\left(\frac{r}{p}\right) + \text{var}(N(A)) - 2\text{cov}\left(\frac{r}{p}, N(A)\right) \right] \end{aligned}$$

The first two variances are known using 4.10 and 4.12, so we need to find the covariance term:

$$\text{cov}\left(\frac{r}{p}, N(A)\right) = E\left(\frac{rN(A)}{p}\right) - E\left(\frac{r}{p}\right)E(N(A))$$

DasGupta (2011) shows that

$$\begin{aligned} E(rN(A)) &= p(\lambda|A| + \lambda^2|A|^2) \\ \Rightarrow \text{cov}\left(\frac{r}{p}, N(A)\right) &= \lambda|A| + \lambda^2|A|^2 - \lambda^2|A|^2 = \lambda|A| \end{aligned}$$

Now, combining these results

$$\begin{aligned} \text{var}(\hat{\lambda}_s - \hat{\lambda}) &= \frac{1}{|A|^2} \left[\frac{\lambda|A|}{p} + \lambda|A| - 2\lambda|A| \right] \\ &= \frac{\lambda(1-p)}{|A|p} \end{aligned}$$

which is the desired result. □

In summary, we have

$$\left(\frac{n|A|}{\lambda} \right)^{1/2} (\hat{\lambda} - \lambda) \rightarrow N(0, 1),$$

$$\left(\frac{np|A|}{\lambda(1-p)} \right)^{1/2} (\hat{\lambda}_s - \hat{\lambda}) \rightarrow N(0, 1),$$

but for large n , r_s/d_s is approximately $p/(1-p)$. So since $\hat{\sigma}^2 = \lambda/|A|$, we have shown that

$$\left(\frac{nr_s}{d_s} \right)^{1/2} \frac{\hat{\lambda}_s - \hat{\lambda}}{\hat{\sigma}} \rightarrow N(0, 1)$$

as in theorem 3.

Chapter 5

Confidence Interval Formation

In this chapter, we will investigate how to best form a confidence interval for a desired statistic, given the results from the previous chapter. We will begin by using the basic bootstrap confidence interval, and adapt this appropriately.

5.1 Naive Confidence Interval

We first attempt to form confidence intervals using the popular method of the basic bootstrap, developed by [Davison and Hinkley \(1997\)](#). This method is also used in [Loh and Stein \(2004\)](#) for various spatial settings. It is designed for independently and identically distributed samples, which is something we have also assumed earlier for theorem 3, so is not restrictive.

Let R be the number of thinned samples we obtain and calculate the estimator from, and α be the confidence level. Let $\hat{\theta}_s^{(R+1)(1-\alpha/2)}$ and $\hat{\theta}_s^{(R+1)\alpha/2}$ be the $(R+1)(1-\alpha/2)$ th and $(R+1)\alpha/2$ th ordered values of $\hat{\theta}_s$ respectively. Then, using the distribution of $\hat{\theta}_s - \hat{\theta}$ to estimate that of $\hat{\theta} - \theta$, we can then derive a confidence interval in the following way:

$$\begin{aligned} 1 - \alpha &= P\left(\hat{\theta}_s^{(R+1)\alpha/2} \leq \hat{\theta}_s \leq \hat{\theta}_s^{(R+1)(1-\alpha/2)}\right) \\ &= P\left(\hat{\theta}_s^{(R+1)\alpha/2} - \hat{\theta} \leq \hat{\theta}_s - \hat{\theta} \leq \hat{\theta}_s^{(R+1)(1-\alpha/2)} - \hat{\theta}\right) \\ &\approx P\left(\hat{\theta}_s^{(R+1)\alpha/2} - \hat{\theta} \leq \hat{\theta} - \theta \leq \hat{\theta}_s^{(R+1)(1-\alpha/2)} - \hat{\theta}\right) \\ &= P\left(\hat{\theta}_s^{(R+1)\alpha/2} - 2\hat{\theta} \leq -\theta \leq \hat{\theta}_s^{(R+1)(1-\alpha/2)} - 2\hat{\theta}\right) \\ &= P\left(2\hat{\theta} - \hat{\theta}_s^{(R+1)(1-\alpha/2)} \leq \theta \leq 2\hat{\theta} - \hat{\theta}_s^{(R+1)\alpha/2}\right). \end{aligned}$$

The approximate confidence interval is then

$$(2\hat{\theta} - \hat{\theta}_s^{(R+1)(1-\alpha/2)}, 2\hat{\theta} - \hat{\theta}_s^{(R+1)\alpha/2}).$$

This method is implemented in figure 5.1 for $\alpha = 0.05$ on order to obtain 95% confidence intervals. This is done for the simple case of estimating the intensity, λ of a homogenous Poisson distribution, as the number of observed points, n , increases. The red dotted line

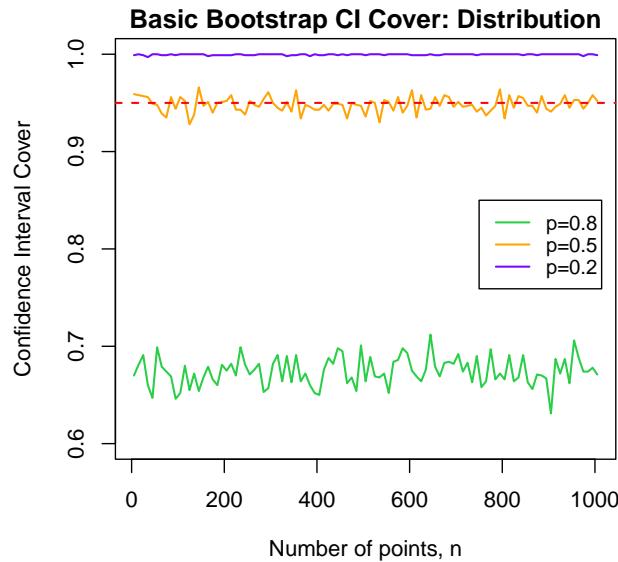


Figure 5.1: Line plot showing the empirical coverage using the basic bootstrap confidence interval as the number of observed datapoints, n , increases from $n = 5$ to $n = 1005$ in steps of 10. The statistic being calculated here is the intensity of a homogenous Poisson distribution. 1000 simulations are carried out, each of which use 500 thinned subsamples. The purple, orange and green line represent thinning parameters of $p = 0.8, 0.5, 0.2$ respectively, and the red dashed line is the desired cover of 0.95.

represents a cover of 0.95 which is the desired behaviour, given the choice of α here. We see that only a thinning probability of $p = 0.5$ results in the desired behaviour, while $p = 0.2$ results in thinned samples with too much variance, and $p = 0.8$ does not capture enough of the variance.

The benefit of this confidence interval formation is that the distribution of the true statistic, or its estimator do not need to be known. The assumption made which does not align with thinning is that the distribution of $\hat{\theta}_s - \hat{\theta}$ approximates $\hat{\theta} - \theta$. We have instead that the asymptotic distribution of

$$\left(\frac{nr_s}{d_s} \right)^{1/2} \frac{\hat{\theta}_s - \hat{\theta}}{\hat{\sigma}}$$

approximates $\sqrt{n}(\hat{\theta} - \theta)$. It is now clear why a thinning probability of $p = 0.5$ produces a good confidence interval cover using the basic bootstrap method. Since $E(r_s) = np$ and $E(d_s) = 1 - np$, when $p = 0.5$, the extra term $\sqrt{r_s/d_s}$ is approximately 1, and the two distributions coincide.

5.2 Studentized Confidence Interval

We now formulate an alternative confidence interval which should adjust for the thinning probability used. To do this, we use a studentized confidence interval. This typically provides a better confidence interval when the distributions of $\sqrt{n}(\hat{\theta} - \theta)$ and $\sqrt{n}(\hat{\theta}_s - \hat{\theta})$ are not close

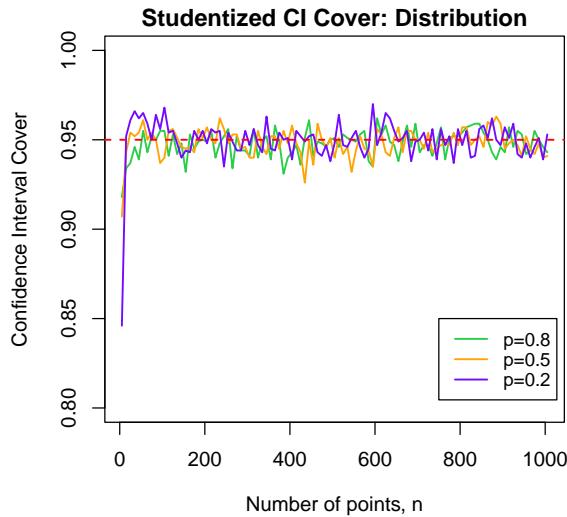


Figure 5.2: Line plot showing the empirical coverage using the Studentized confidence interval as the number of observed datapoints, n , increases from $n = 5$ to $n = 1005$ in steps of 10. The statistic being calculated here is the intensity of a homogeneous Poisson distribution. 1000 simulations are carried out, each of which use 500 thinned subsamples. The purple, orange and green line represent thinning parameters of $p = 0.8, 0.5, 0.2$ respectively, and the red dashed line is the desired cover of 0.95.

but the distributions of $\sqrt{n}(\hat{\theta} - \theta)/\hat{\sigma}$ and $\sqrt{n}(\hat{\theta}_s - \hat{\theta})/\hat{\sigma}_s$ are. Here, we have the case that the asymptotic variance of $\hat{\theta}_s$ is dependent on the thinning probability used, which is not taken into account with the basic bootstrap confidence intervals.

The interval we are trying to approximate is an interval around $\hat{\theta}$, given the true θ . We are working with statistics with the assumed property that

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, \hat{\sigma}^2)$$

which leads to a confidence interval of

$$\left(\hat{\theta} - t_{R-1,1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}}, \hat{\theta} + t_{R-1,1-\alpha/2} \frac{\hat{\sigma}}{\sqrt{n}} \right),$$

where $t_{R-1,1-\alpha/2}$ is the $(1 - \alpha/2)$ th percentile of a Student's t statistic with $R - 1$ degrees of freedom. Note that R is the number of thinned sabsamples we obtain in each simulation.

Since we only have $\hat{\theta}_s$, $\hat{\theta}$, we can instead form the interval in the following way

$$\hat{\theta}_s \sim N(\hat{\theta}, \hat{\sigma}_s^2) \implies (\hat{\theta} - t_{R-1,1-\alpha/2} \hat{\sigma}_s, \hat{\theta} + t_{R-1,1-\alpha/2} \hat{\sigma}_s).$$

where $\hat{\sigma}_s$ is the variance of the thinned samples. Using the distribution derived in theorem 3, this is the same as

$$\left(\hat{\theta} - t_{R-1,1-\alpha/2} \hat{\sigma} \sqrt{\frac{d_s}{nr_s}}, \hat{\theta} + t_{R-1,1-\alpha/2} \hat{\sigma} \sqrt{\frac{d_s}{nr_s}} \right)$$

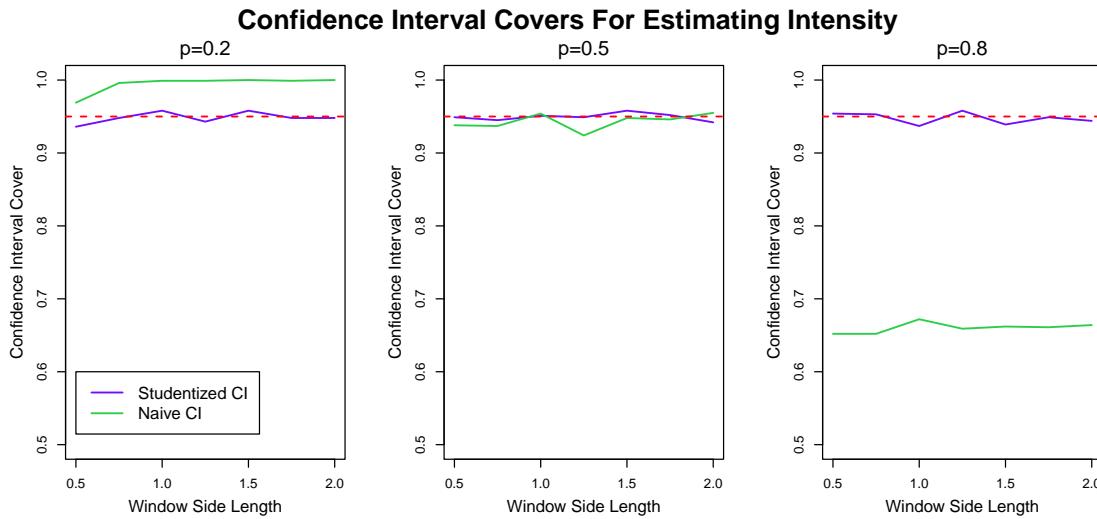


Figure 5.3: Line plots showing the cover of confidence intervals calculated for the estimation of the intensity of a homogeneous Poisson point process with a true intensity of $\lambda = 250$. The green lines show the cover using the naive basic bootstrap method, and the purple lines show the results of the studentized confidence intervals. Each plot shows the covers using a thinning parameter of $p = 0.2, 0.5, 0.8$ respectively from left to right. These covers are shown for various sizes of observation window, A , with side lengths varying from 0.5 to 2 in steps of 0.25. For each calculated cover, 1000 simulations are carried out, and for each of these a realisation of a homogeneous Poisson process is generated, and 500 thinned subsamples are selected to form the confidence interval.

which has an extra $\sqrt{\frac{d_s}{r_s}}$ multiplying the desired variance, $\hat{\sigma}$. We want to use $\hat{\sigma}_s$ to obtain an interval, so we can just cancel out this extra term to obtain

$$\left(\hat{\theta} - t_{R-1,1-\alpha/2} \hat{\sigma}_s \sqrt{\frac{r_s}{d_s}}, \hat{\theta} + t_{R-1,1-\alpha/2} \hat{\sigma}_s \sqrt{\frac{r_s}{d_s}} \right). \quad (5.1)$$

This formation of the confidence interval clearly eliminates the effect of the thinning parameter, p , which is encoded into r_s and d_s , the number of points retained and discarded respectively.

Similarly to figure 5.1, figure 5.2 shows the covers calculated for the estimating of the intensity of a Poisson distribution with intensity, $\lambda = 10$. However in this case, we now used the Studentized confidence interval, 5.1. We can see that this method performs much better, with the covers for $p = 0.2, 0.5, 0.8$ all sitting around the desired value of 0.95 shown by the red dashed line. The coverage is no longer dependent on the thinning probability, and the convergence to the desired value occurs quite quickly.

5.3 Example: Intensity of Homogeneous Poisson Point Process

The same theory also applies for various statistics of spatial point processes. Here we will present the simple example of estimating the intensity of a homogeneous Poisson point

process. We have seen already in section 4.5 that this intensity estimator satisfies theorem 3, and so we can expect to form accurate confidence intervals as in the case of the Poisson distribution.

Figure 5.3 shows how both the basic bootstrap and studentized methods of forming confidence intervals perform. Very similarly to the case of a Poisson distribution, the basic bootstrap covers, shown in green, are dependent on the thinning probability which is not desirable. The studentized confidence intervals on the other hand perform well for all values shown here, $p = 0.2, 0.5, 0.8$.

This shows that the use of thinning to obtain confidence intervals is applicable to spatial point processes. In the following chapters we will explore where it is particularly useful, and where it fails.

Chapter 6

Ripley's K

In this chapter, we will explore how well thinning performs for finding confidence intervals of Ripley's K estimates, and adapt these accordingly.

6.1 Results

Show plots using basic bootstrap and studentised for CSR, clustered and regular

6.1.1 Basic Bootstrap Confidence Intervals

plot/table with bootstrapped CIs

6.1.2 Studentized Confidence Intervals

plot/table with studentized CIs, with scaling of $p/(1-p)$

6.2 Distribution of Thinned Estimates

The reason that the studentized interval shown above does not produce the desired results is that the result shown in theorem 3 does not hold for Ripley's K . More precisely, the variance of the thinned estimates is not $\sqrt{r_s/d_s}$ times that of the estimates of the entire observed data. This is most likely because Ripley's K cannot be written as a linear statistical functional, as required. This is difficult to show analytically, but we will show that the estimates do not have the desired distributions through simulation.

6.2.1 Normal Distribution of Thinned Estimates

Firstly, we can check whether $\sqrt{n}(\hat{K}_s(h) - \hat{K}(h))$ has an asymptotically normal distribution. This is done by simulating a homogeneous Poisson point process, estimating $\hat{K}(h)$ from this realisation, then generating 500 thinned samples and estimating $\hat{K}_s(h)$ from each of these. The value of $\sqrt{n}(\hat{K}_s(h) - \hat{K}(h))$ is then stored. This process is repeated 100 times

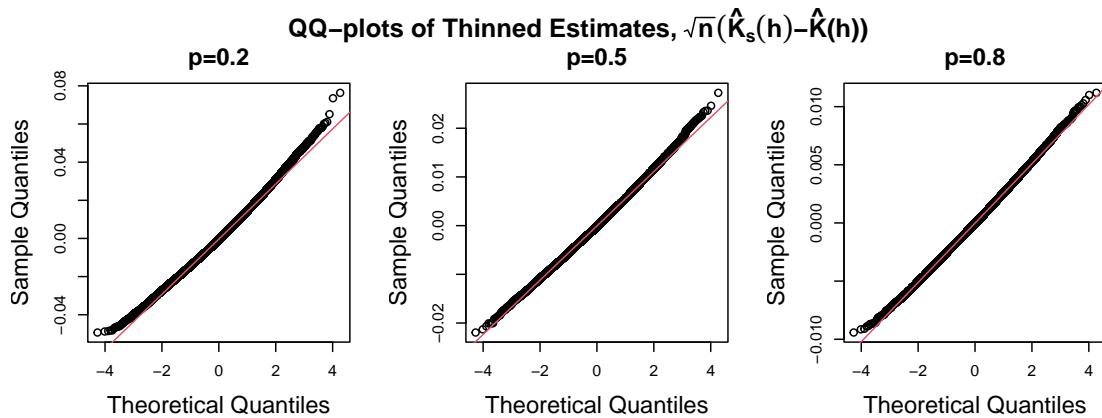


Figure 6.1: QQ-plots of Ripley's K estimates on thinned point processes after normalisation, $\sqrt{n}(\hat{K}_s(h) - \hat{K}(h))$. Values are calculated by simulating 100 realisations of a homogeneous Poisson point process with intensity $\lambda = 10000$ on a unit square, and generating 500 thinned processes for each of these. Thinned Ripley's K is then estimated on each of these, and normalised to obtain the desired value. This is carried out for thinning parameters $p = 0.2, 0.5, 0.8$ which are shown in the subfigures from left to right respectively.

to generate 50000 thinned estimates. The Poisson process used is one with an intensity of $\lambda = 10000$, so that n is large enough to observe the asymptotic effects. The QQ-plots thinning parameters of $p = 0.2, 0.5, 0.8$ for shown in figure 6.1.

We see that for a large n , the distributions are roughly normal, however it is clear that is a slight curvature to the plots. This shape indicates that the samples are skewed to the right, meaning that more than half of the points are greater than 0. This effect is more prominent for lower values of the thinning probability, p , possibly because the thinned samples contain fewer points.

6.2.2 Variance of Thinned Estimates

We now consider the variance of these quantities, and more specifically their relationship. Let

$$\text{var}(\sqrt{n}(\hat{K}(h) - K(h))) = \hat{\sigma}^2,$$

$$\text{var}(\sqrt{n}(\hat{K}_s(h) - \hat{K}(h))) = \hat{\sigma}_s^2.$$

Then if theorem 3 holds, it suggests that for large n ,

$$\frac{r_s \sigma_s^2}{d_s} \approx \frac{p \hat{\sigma}_s^2}{1-p} \approx \hat{\sigma}^2.$$

We can determine whether this relationship holds through simulation, as in figure 6.2. This is done by simulating a homogeneous Poisson point process of intensity $\lambda = 250$ on a unit square 10000 times, and calculating $\hat{K}(h)$ on each of these with $h = 0.1$. The variance of all of these samples is then taken to be $\hat{\sigma}^2$. To find approximate $\hat{\sigma}^2$, we simulate the same process 100 times, and for each of these, generate 500 thinned subsamples and calculate

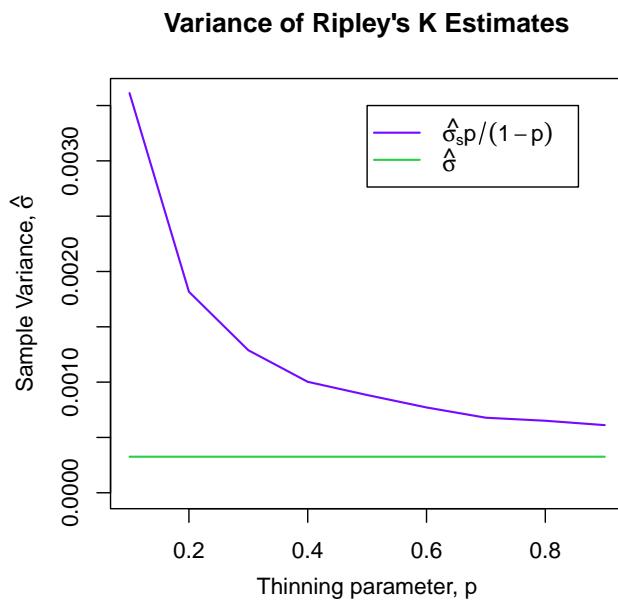


Figure 6.2: Line plot showing estimates of the variance of Ripley's K over entire realisations of data, $\hat{\sigma}^2$, and the scaled variance over thinned subsamples of data, $\hat{\sigma}_s p / (1 - p)$. The former of these is shown in green and is generated using estimates $\hat{K}(h)$ for $h = 0.1$ on 10000 simulations of a homogeneous Poisson point process with $\lambda = 250$ on a unit square, and taking the sample variance of these. The latter is formed by simulating the same process 100 times, and calculating the thinned estimate, $\hat{K}_s(h)$, on 500 thinned subsamples of each of these. This is carried out for p taking values from 0.1 to 0.9 in steps of 0.1.

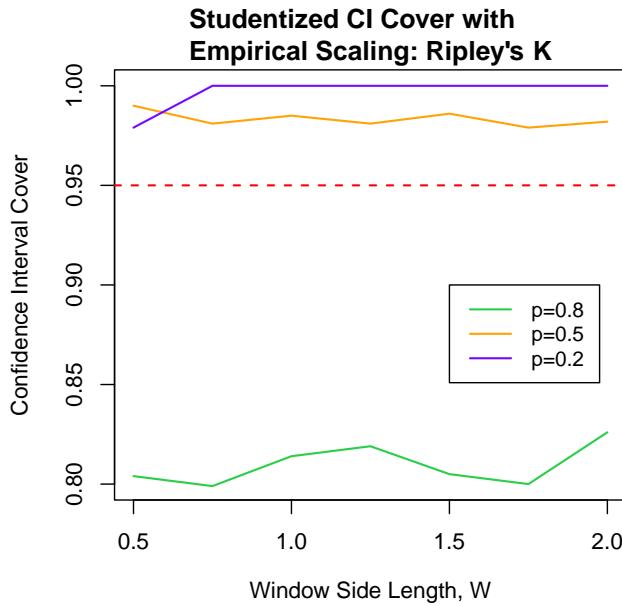


Figure 6.3: Line plot showing covers for estimation of Ripley's K , using a studentized confidence interval, and scaling the thinned variance using the empirical values. Covers are calculated using 1000 simulations of a homogenous Poisson process with $\lambda = 250$ on a unit square. For each of these, 500 thinned subsamples are generated, and $\hat{K}_s(0.1)$ is calculated on each of these. The plot shows the cover using thinning parameters of $p = 0.2, 0.5, 0.8$ in purple, orange and green respectively.

$\hat{K}_s(h)$ on each of these, then we take the variance. Figure 6.2 shows the variance of the thinned estimates scaled by $p/(1-p)$ for a range of p from 0.1 to 0.9, which would be close to $\hat{\sigma}^2$ if the result holds. We see that this relationship does not hold for Ripley's K , and in fact the scaled thinned variance lies consistently above the desired variance, is consistent with the fact that we see covers which are too high when using this scaling with a studentized confidence interval [reference to table/figure here](#).

6.3 Finding Appropriate Scaling

Using empirical scalings

Using these scalings on homogeneous poisson is shown in figure 6.3

Chapter 7

Inhomogenous Intensity Estimates

This is also demonstrated on an inhomogenous Poisson spatial point process in 7.2. The intensity function takes the form $\lambda(x, y) = \exp(a + bx)$, and the values being estimated here are the coefficients a and b . This is shown as the size of the observed window, A , increases, and therefore the number of observed points increases. Out of the thinning probabilities shown here, $p = 0.2, 0.5, 0.8$, only a value of $p = 0.5$ produces the desired behavior for estimating b . It is clear that the cover is dependent on the thinning probability.

The same is also shown for an inhomogeneous Poisson point process where the coefficients a and b of the intensity function $\lambda(x, y) = \exp(a + bx)$ are being estimated in figure 7.3. Again, we see a significant improvement in the performance for all thinning probabilities. The coverage is no longer dependent on p since we have cancelled this term out.

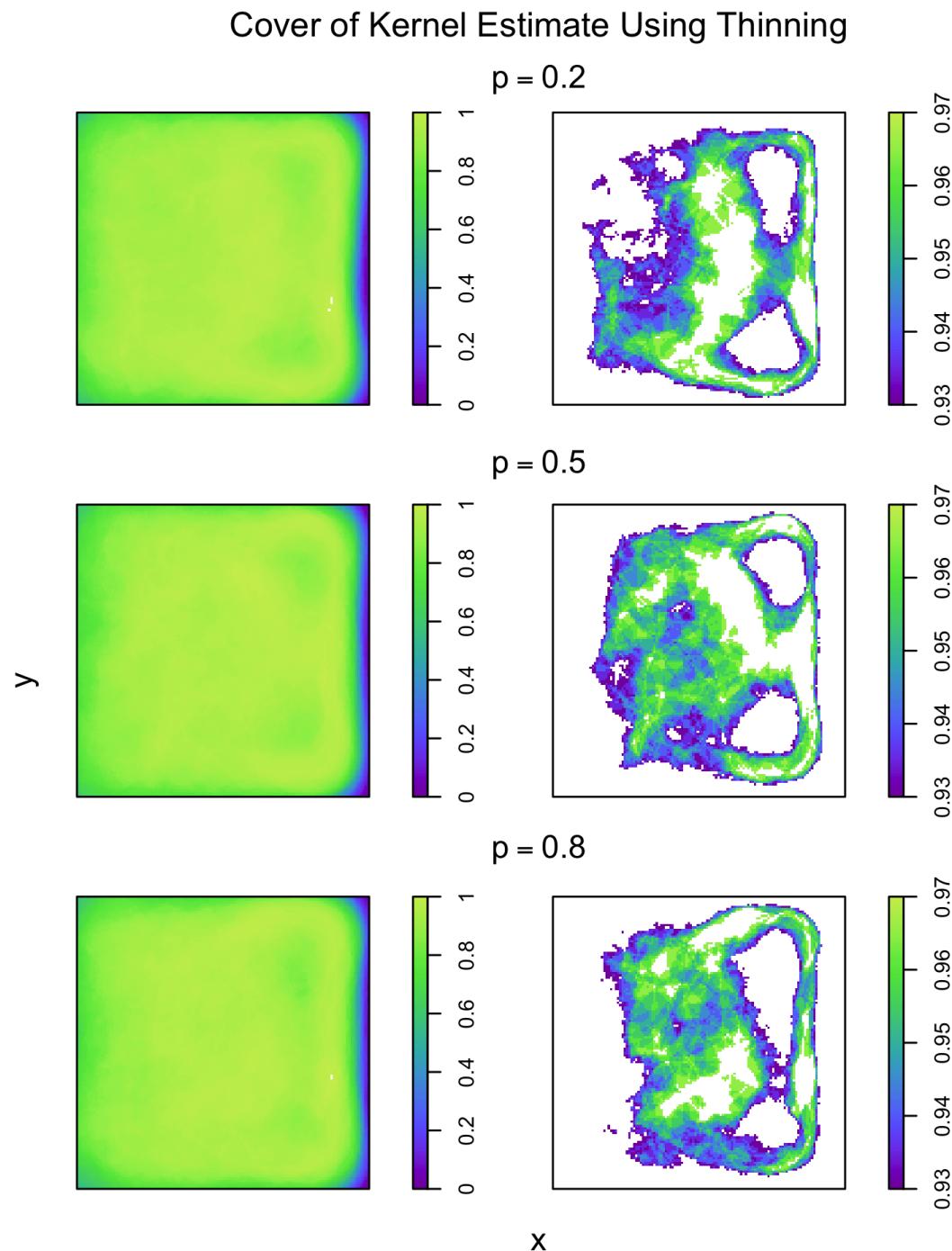


Figure 7.1: Caption

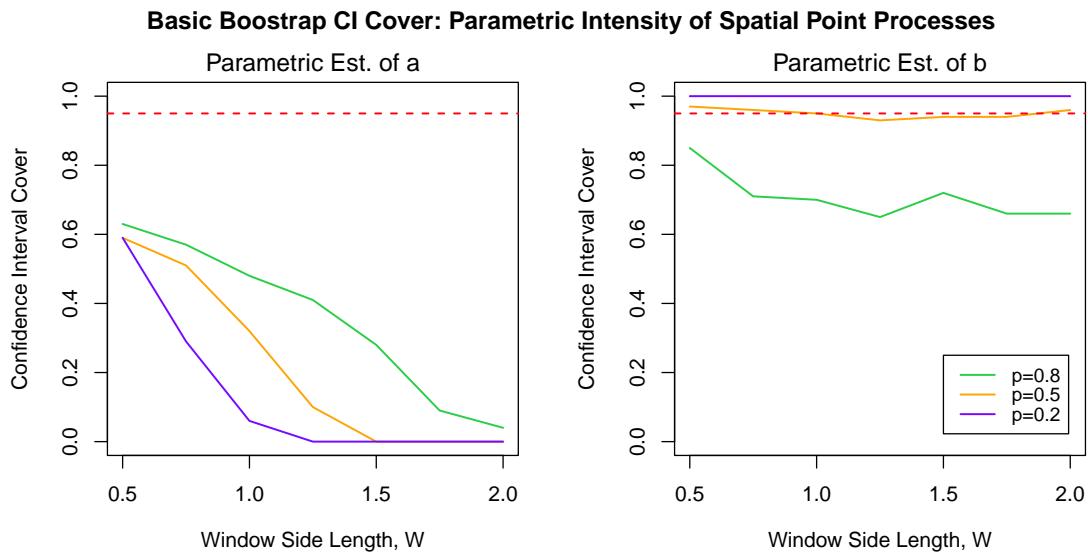


Figure 7.2: Line plots showing the empirical coverage using the basic bootstrap confidence interval as the size of the observed window, A , increases. This side length of the window varies from $W = 0.5$ to $W = 2.0$ in steps of 0.25. The process used is an inhomogenous Poisson point process with an intensity function of $\lambda(x, y) = \exp(a + bx)$, and the values being estimated are a and b . The left plot shows the covers for a , and the right plot shows the covers for b . 100 simulations are carried out, each of which use 500 thinned subsamples. The purple, orange and green line represent thinning parameters of $p = 0.8, 0.5, 0.2$ respectively, and the red dashed line is the desired cover of 0.95.

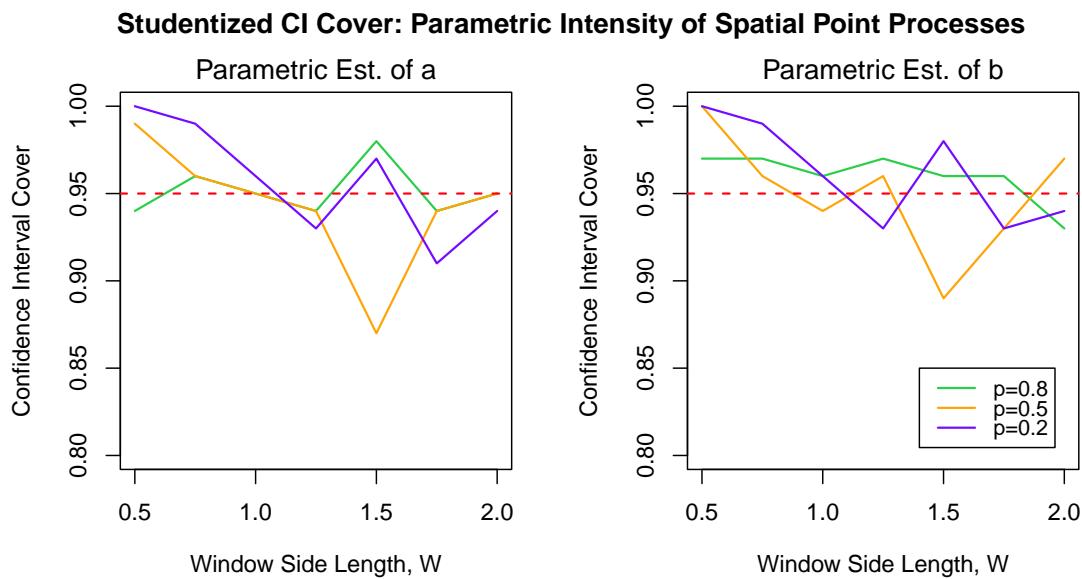


Figure 7.3: Line plots showing the empirical coverage using the studentized confidence interval as the size of the observed window, A , increases. This side length of the window varies from $W = 0.5$ to $W = 2.0$ in steps of 0.25. The process used is an inhomogenous Poisson point process with an intensity function of $\lambda(x, y) = \exp(a + bx)$, and the values being estimated are a and b . The left plot shows the covers for a , and the right plot shows the covers for b . 100 simulations are carried out, each of which use 500 thinned subsamples. The purple, orange and green line represent thinning parameters of $p = 0.8, 0.5, 0.2$ respectively, and the red dashed line is the desired cover of 0.95.

References

- J. M. Loh and M. L. Stein. Bootstrapping a spatial point process. *Statistica Sinica*, 14(1):69–101, 2004. ISSN 10170405, 19968507. URL <http://www.jstor.org/stable/24307180>.
- Ji Meng Loh. Bootstrapping an inhomogeneous point process. *Journal of Statistical Planning and Inference*, 140(3):734–749, 2010. ISSN 0378-3758. doi: <https://doi.org/10.1016/j.jspi.2009.09.003>. URL <https://www.sciencedirect.com/science/article/pii/S0378375809002845>.
- N.A.C. Cressie. *Statistics for Spatial Data*. A Wiley-interscience publication. J. Wiley, 1991. ISBN 978-0-471-84336-8. URL https://books.google.co.uk/books?id=k_JQAAAAMAAJ.
- B.D. Ripley. *Spatial Statistics*. Wiley Series in Probability and Statistics. Wiley, 1981. ISBN 978-0-471-08367-2. URL <https://books.google.co.uk/books?id=dxvvAAAAMAAJ>.
- Jerzy Neyman and Elizabeth L. Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(1):1–43, 1958. ISSN 00359246. URL <http://www.jstor.org/stable/2983905>.
- D. Stoyan, W.S. Kendall, and J. Mecke. *Stochastic Geometry and Its Applications*. Inorganic Chemistry. Wiley, 1995. ISBN 978-0-471-95099-8. URL <https://books.google.co.uk/books?id=oFDvAAAAMAAJ>.
- D. Stoyan. Statistical analysis of spatial point processes: A soft-core model and cross-correlations of marks. *Biometrical Journal*, 29(8):971–980, 1987. doi: <https://doi.org/10.1002/bimj.4710290811>. URL <https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.4710290811>.
- M.N.M. van Lieshout. *Theory of Spatial Statistics: A Concise Introduction*. Chapman & Hall/CRC Texts in Statistical Science. CRC Press, 2019. ISBN 978-0-429-62703-3. URL <https://books.google.co.uk/books?id=BNWNDwAAQBAJ>.
- Oliver C. Ibe. 15 - Markov Point Processes. In Oliver C. Ibe, editor, *Markov Processes for Stochastic Modeling (Second Edition)*, pages 453–480. Elsevier, Oxford, second edition edition, 2013. ISBN 978-0-12-407795-9. doi: <https://doi.org/10.1002/bimj.4710290811>.

- 1016/B978-0-12-407795-9.00015-3. URL <https://www.sciencedirect.com/science/article/pii/B9780124077959000153>.
- Shashi Shekhar, Pusheng Zhang, and Yan Huang. Spatial Data Mining. In Oded Maimon and Lior Rokach, editors, *Data Mining and Knowledge Discovery Handbook*, pages 848–8549. Springer US, Boston, MA, 2005. ISBN 978-0-387-25465-4. doi: 10.1007/0-387-25465-X_39. URL https://doi.org/10.1007/0-387-25465-X_39.
- Simeon M. Berman. Poisson and extreme value limit theorems for markov random fields. *Advances in Applied Probability*, 19(1):106–122, 1987. ISSN 00018678. URL <http://www.jstor.org/stable/1427375>.
- Steven P. Ellis. A limit theorem for spatial point processes. *Advances in Applied Probability*, 18(3):646–659, 1986. ISSN 00018678. URL <http://www.jstor.org/stable/1427181>.
- Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press, London, 2015. ISBN 9781482210200. URL <https://www.routledge.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/p/book/9781482210200/>.
- Peter A. W. Lewis. A branching poisson process model for the analysis of computer failure patterns. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(3):398–456, 1964. ISSN 00359246. URL <http://www.jstor.org/stable/2984497>.
- M. S. Bartlett. The spectral analysis of two-dimensional point processes. *Biometrika*, 51(3/4):299–311, 1964. ISSN 00063444. URL <http://www.jstor.org/stable/2334136>.
- Nanda R. Aryal and Owen D. Jones. Spatial-temporal rainfall models based on poisson cluster processes. *Stochastic Environmental Research and Risk Assessment*, 35(12):2629–2643, December 2021. ISSN 1436-3259. doi: 10.1007/s00477-021-02046-5. URL <https://doi.org/10.1007/s00477-021-02046-5>.
- Gilles Faÿ, Bárbara González-Arévalo, Thomas Mikosch, and Gennady Samorodnitsky. Modeling teletraffic arrivals by a Poisson cluster process. *Queueing Systems*, 54(2):121–140, October 2006. ISSN 1572-9443. doi: 10.1007/s11134-006-9348-z. URL <https://doi.org/10.1007/s11134-006-9348-z>.
- M. Westcott. On existence and mixing results for cluster point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 33(2):290–300, 1971. ISSN 00359246. URL <http://www.jstor.org/stable/2985009>.
- Dietrich Stoyan, Sung Nok Chiu, W. S. Kendall, and Joseph Mecke. *Stochastic geometry and its applications*. Wiley series in probability and statistics. John Wiley & Sons Inc, Chichester, West Sussex, United Kingdom, third edition edition, 2013. ISBN 978-1-118-65823-9 978-1-118-65825-3 978-1-118-65824-6.

- Hesham ElSawy and Ekram Hossain. A modified hard core point process for analysis of random csma wireless networks in general fading environments. *IEEE Transactions on Communications*, 61(4):1520–1534, 2013. doi: 10.1109/TCOMM.2013.020813.120594.
- Dietrich Stoyan and Antti Penttinen. Recent applications of point process methods in forestry statistics. *Statistical Science*, 15(1):61–78, 2000. ISSN 08834237. URL <http://www.jstor.org/stable/2676677>.
- B. Matern. *Spatial Variation*. Lecture Notes in Statistics. Springer New York, 1986. ISBN 978-3-540-96365-3. URL <https://books.google.co.uk/books?id=s-xczaXRpt0C>.
- S. Mase, J. Møller, D. Stoyan, R. P. Waagepetersen, and G. Döge. Packing Densities and Simulated Tempering for Hard Core Gibbs Point Processes. *Annals of the Institute of Statistical Mathematics*, 53(4):661–680, December 2001. ISSN 1572-9052. doi: 10.1023/A:1014662415827. URL <https://doi.org/10.1023/A:1014662415827>.
- B. D. Ripley. *Statistical Inference for Spatial Processes*. Cambridge University Press, 1988.
- Adrian J Baddeley, Jesper Møller, and Rasmus Waagepetersen. Non-and semi-parametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350, 2000.
- Rasmus Waagepetersen. Estimating functions for inhomogeneous spatial point processes with incomplete covariate data. *Biometrika*, 95(2):351–363, 2008. ISSN 00063444, 14643510. URL <http://www.jstor.org/stable/20441469>.
- Stephen L. Rathbun, Saul Shiffman, and Chad J. Gwaltney. Modelling the effects of partially observed covariates on poisson process intensity. *Biometrika*, 94(1):153–165, 2007. ISSN 00063444, 14643510. URL <http://www.jstor.org/stable/20441360>.
- Stephen L. Rathbun and Noel Cressie. A space-time survival point process for a longleaf pine forest in southern georgia. *Journal of the American Statistical Association*, 89(428):1164–1174, 1994. ISSN 01621459. URL <http://www.jstor.org/stable/2290979>.
- A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.
- P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Statistics. Wiley, 1986. ISBN 978-0-471-80478-9. URL <https://books.google.co.uk/books?id=Q2IPQAQAMAAJ>.
- Stanford University Department of Statistics, D.N. Politis, J.P. Romano, and National Science Foundation (U.S.). *A Circular Block-resampling Procedure for Stationary Data*. Purdue University. Department of Statistics, 1991. URL <https://books.google.co.uk/books?id=bR5UHwAACAAJ>.

- Ji Meng Loh and Michael Stein. Spatial bootstrap with increasing observations in a fixed domain. *Statistica Sinica*, 18:667–688, 04 2008.
- S.N. Lahiri. On the moving block bootstrap under long range dependence. *Statistics Probability Letters*, 18(5):405–413, 1993. ISSN 0167-7152. doi: [https://doi.org/10.1016/0167-7152\(93\)90035-H](https://doi.org/10.1016/0167-7152(93)90035-H). URL <https://www.sciencedirect.com/science/article/pii/016771529390035H>.
- Dimitris N. Politis and Joseph P. Romano. Large Sample Confidence Regions Based on Sub-samples under Minimal Assumptions. *The Annals of Statistics*, 22(4):2031 – 2050, 1994. doi: 10.1214/aos/1176325770. URL <https://doi.org/10.1214/aos/1176325770>.
- WJ Branu and RJ Kulperger. A bootstrap for point processes. *Journal of Statistical Computation and Simulation*, 60(2):129–155, 1998.
- R. v. Mises. On the asymptotic distribution of differentiable statistical functions. *The Annals of Mathematical Statistics*, 18(3):309–348, 1947. ISSN 00034851. URL <http://www.jstor.org/stable/2235734>.
- B. Fristedt and L. Gray. *A Modern Approach to Probability Theory*. Birkhauser, Boston, MA, 1997.
- Robert S. Strichartz. *The way of analysis*. World Publishing Corporation, 2010.
- Luisa Turrin Fernholz. *Von Mises calculus for statistical functionals*. Number 19 in Lecture notes in statistics. Springer, New York [u.a.], 1983. ISBN 0387908994. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+027318796&sourceid=fbw_bibsonomy.
- Robert J. Serfling. *Approximation theorems of mathematical statistics*. Wiley series in probability and mathematical statistics : Probability and mathematical statistics. Wiley, New York, NY [u.a.], [nachdr.] edition, 1980. ISBN 0471024031. URL http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+024353353&sourceid=fbw_bibsonomy.
- A. Dvoretzky, J. Kiefer, and J. Wolfowitz. Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator. *The Annals of Mathematical Statistics*, 27(3):642–669, 1956. ISSN 00034851. URL <http://www.jstor.org/stable/2237374>.
- C. F. J. Wu. On the asymptotic properties of the jackknife histogram. *The Annals of Statistics*, 18(3):1438–1452, 1990. ISSN 00905364. URL <http://www.jstor.org/stable/2242062>.
- Jaroslav Hájek. Limiting distributions in simple random sampling from a finite population. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 5:361–374, 1960.

- HaiYing Wang and Jiahui Zou. A comparative study on sampling with replacement vs poisson sampling in optimal subsampling. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 289–297. PMLR, 13–15 Apr 2021. URL <https://proceedings.mlr.press/v130/wang21a.html>.
- William Feller. *An introduction to probability theory and its applications*, page 150–152. John Wiley, 1968.
- G. Pólya. Über den zentralen grenzwertsatz der wahrscheinlichkeitsrechnung und das momentenproblem. *Mathematische Zeitschrift*, 8:171–181, 1920. URL <http://eudml.org/doc/167598>.
- H.H. Sohrab. *Basic Real Analysis*. Springer New York, 2014. ISBN 978-1-4939-1841-6. URL <https://books.google.co.uk/books?id=QnpqBQAAQBAJ>.
- Peter J. Bickel and David A. Freedman. Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 9(6):1196–1217, 1981. ISSN 00905364. URL <http://www.jstor.org/stable/2240410>.
- Ji Meng Loh. A valid and fast spatial bootstrap for correlation functions. *The astrophysical journal*, 681(1):726, 2008.
- A. DasGupta. *Probability for Statistics and Machine Learning: Fundamentals and Advanced Topics*. Springer Texts in Statistics. Springer New York, 2011. ISBN 9781441996343. URL <https://books.google.co.in/books?id=nAa7PCTIBLkC>.

Appendix A

My First Appendix

The context of Appendix A.

A.1 Poisson as Limit of Binomial

pg 96 of [van Lieshout \(2019\)](#)

A.2 Simulations Showing Deomonstrating Asymptotics

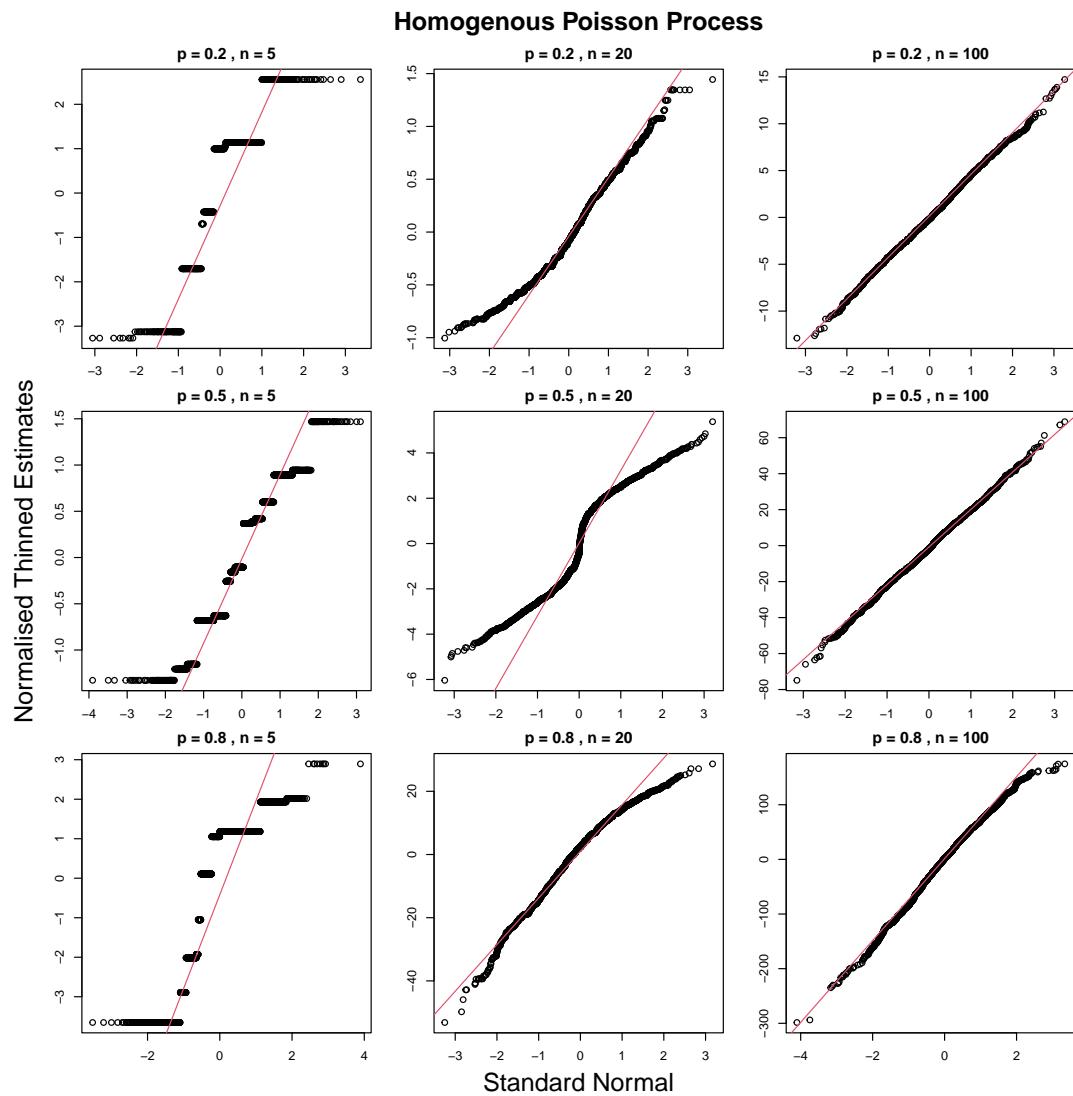


Figure A.1: Caption

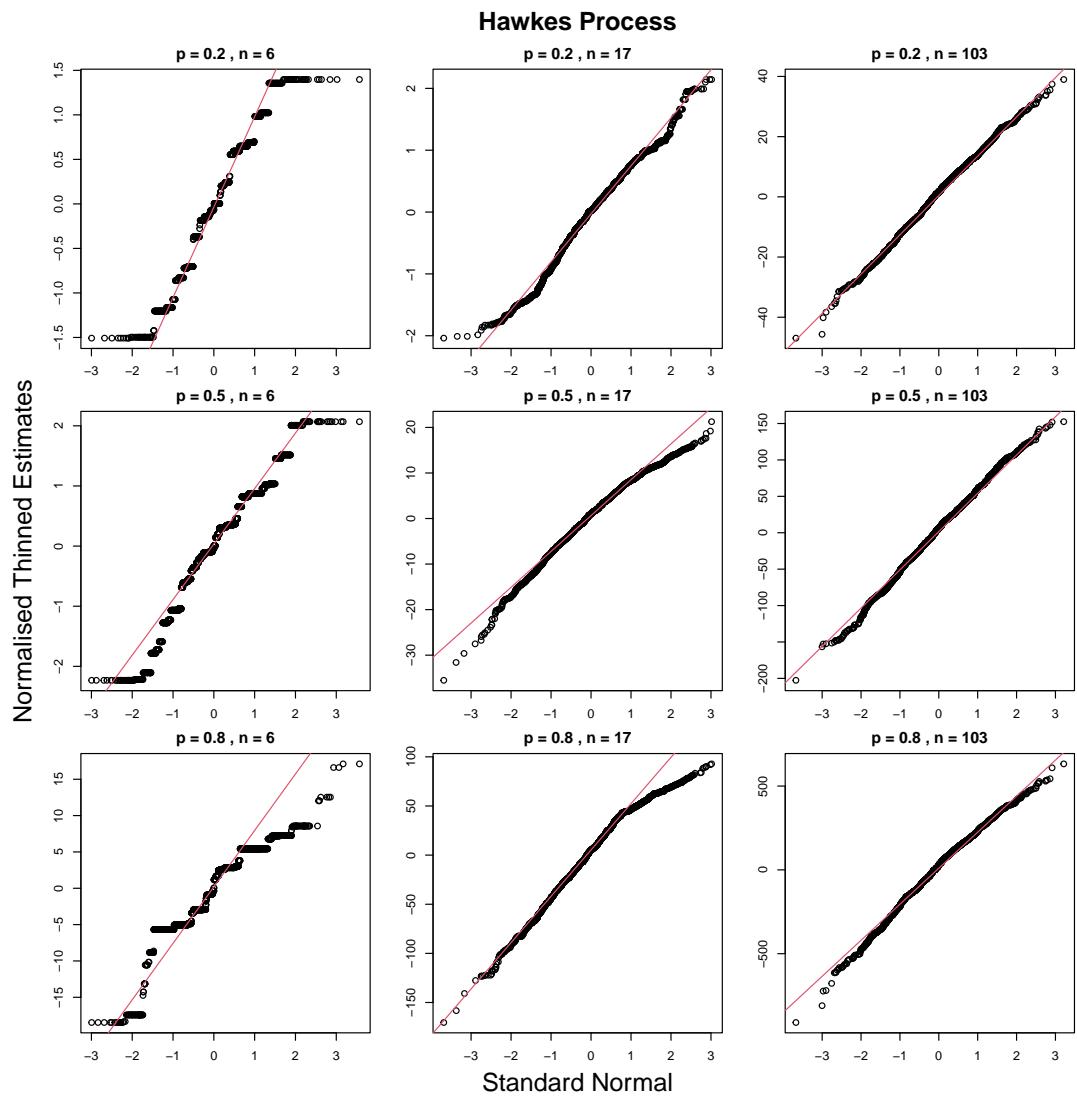


Figure A.2: Caption

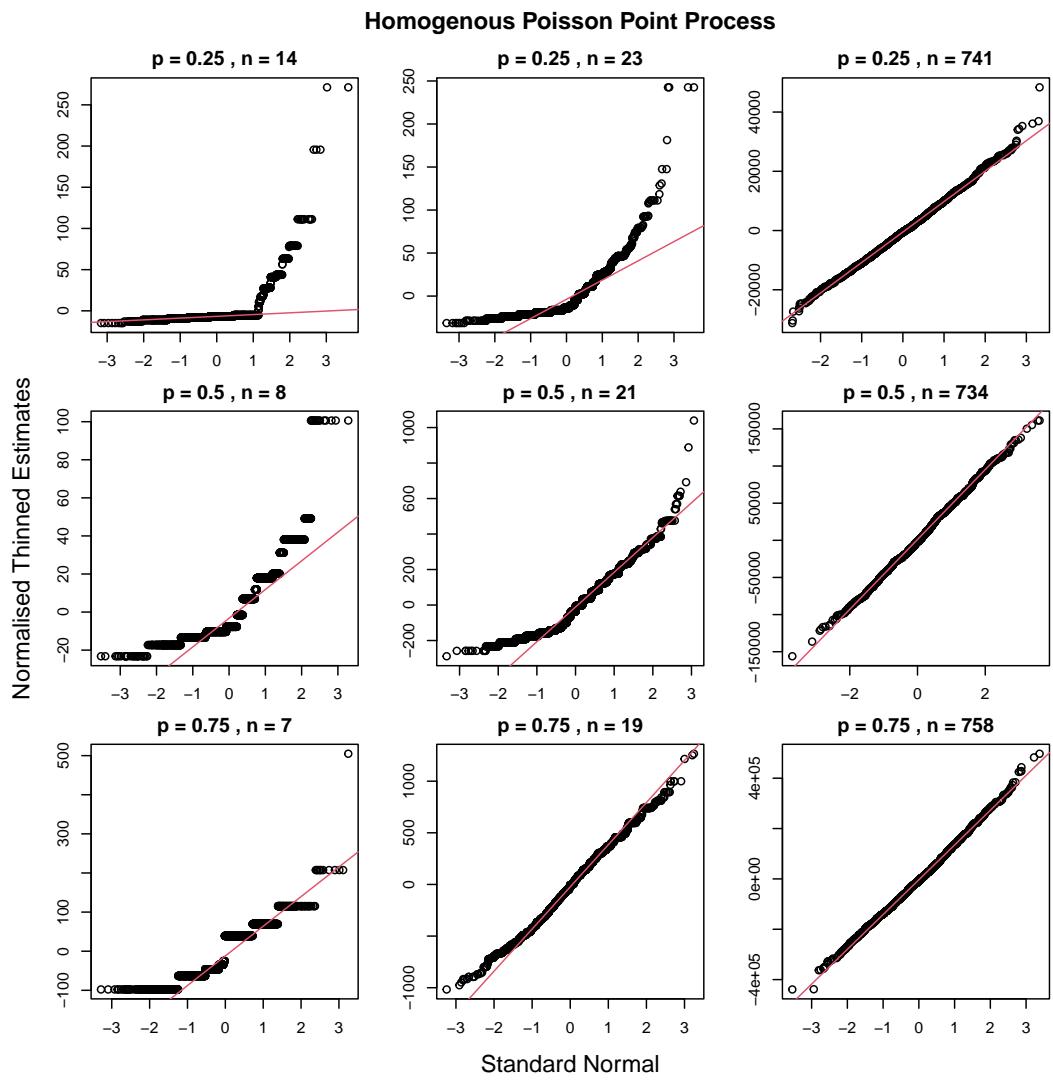


Figure A.3: Caption

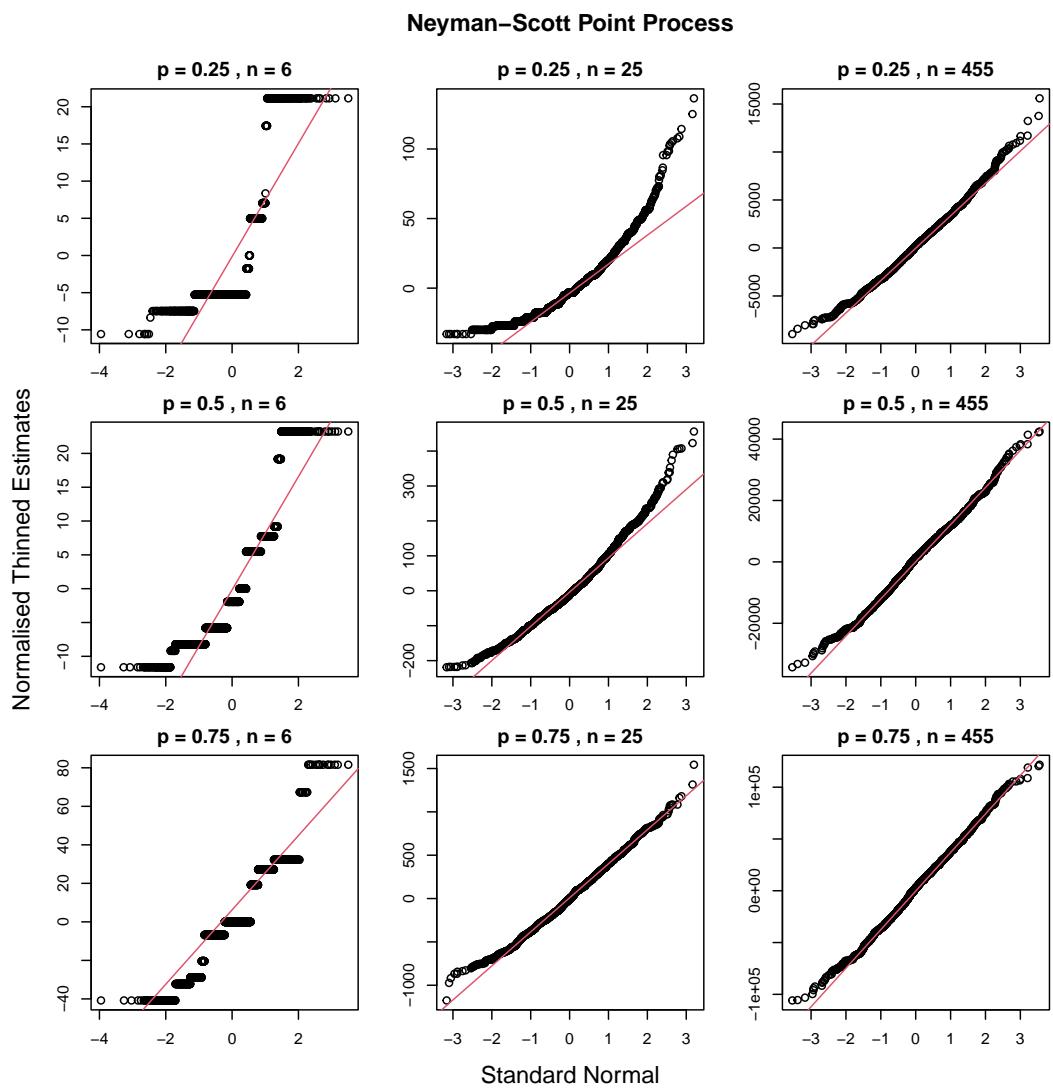


Figure A.4: Caption