Statistical methods for spatial screening

Zhigang Zhou
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/1790
NOTE TO USERS

This reproduction is the best copy available.

UMI®
Statistical methods for spatial screening

by

Zhigang Zhou

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:
Max Morris, Major Professor
Philip Dixon
Mark Kaiser
Kirk Moloney
Huaiqing Wu

Iowa State University
Ames, Iowa
2005
Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Zhigang Zhou

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor
Signature was redacted for privacy.

For the Major Program
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Physical characteristics of UXO deposition</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Statistical model</td>
<td>6</td>
</tr>
<tr>
<td>1.4 Sampling methodology</td>
<td>8</td>
</tr>
<tr>
<td>1.5 Literature review</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER 2 GLOBAL SAMPLING</td>
<td>14</td>
</tr>
<tr>
<td>2.1 Model assumptions</td>
<td>14</td>
</tr>
<tr>
<td>2.2 The Neyman-Scott line transect distribution</td>
<td>16</td>
</tr>
<tr>
<td>2.3 Estimation of parameters and standard errors</td>
<td>21</td>
</tr>
<tr>
<td>2.3.1 Estimation of $\theta$</td>
<td>22</td>
</tr>
<tr>
<td>2.3.2 Estimation of $\rho^2$</td>
<td>23</td>
</tr>
<tr>
<td>2.3.3 Estimation of $\mu$</td>
<td>24</td>
</tr>
<tr>
<td>2.3.4 Estimation of $\lambda$</td>
<td>25</td>
</tr>
<tr>
<td>2.3.5 Estimation of object intensity $\tau$</td>
<td>26</td>
</tr>
<tr>
<td>2.4 More about estimating $\theta$</td>
<td>27</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Figure 1.1 Some examples of UXO ........................................ 2
Figure 1.2 Minex 2FD 4.500 Compact Metal Detector .................. 3
Figure 1.3 Minex Compact Metal Detector in use ........................ 4
Figure 1.4 An Oak Ridge National Laboratory UXO-survey aircraft .... 5

Figure 2.1 A realization of Neyman-Scott process with $\lambda=0.002$, $\mu=20$, and $\rho^2=4$ .................................................. 15

Figure 3.1 Center line transect position. The dot is the location of the cluster center $(\mu_x, \mu_y)$ ............................................. 71
Figure 3.2 Line transect positions 2-5. The dot is the location of the cluster center $(\mu_x, \mu_y)$ ............................................. 73
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Comparison of two methods of estimating $\theta$</td>
<td>29</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Values of ratios of $T(\theta)$ over $T(\theta/2)$</td>
<td>32</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Estimation of large $\theta$</td>
<td>33</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Estimate of $\theta$</td>
<td>34</td>
</tr>
<tr>
<td>Table 2.5</td>
<td>Estimate of $\rho^2$</td>
<td>36</td>
</tr>
<tr>
<td>Table 2.6</td>
<td>Estimate of $\mu$ ($\mu = 50$)</td>
<td>37</td>
</tr>
<tr>
<td>Table 2.7</td>
<td>Estimate of $1000\lambda$ ($\lambda = 0.005$)</td>
<td>37</td>
</tr>
<tr>
<td>Table 2.8</td>
<td>Correlations among estimators of $\theta$, $\rho^2$ and $k$</td>
<td>38</td>
</tr>
<tr>
<td>Table 2.9</td>
<td>Average Number of Undetected Clusters</td>
<td>39</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Some values of $Q(N_u, P_u)$</td>
<td>70</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Standard deviations of $\hat{\mu}_x$ and $\hat{\mu}_y$</td>
<td>74</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Average estimate and standard deviation of $\hat{N}$</td>
<td>75</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Standard deviation of $\hat{\rho^2}$</td>
<td>75</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>Estimate and standard deviation of $n_{rec}$</td>
<td>76</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>Coverage of 95% IR</td>
<td>77</td>
</tr>
<tr>
<td>Table 3.7</td>
<td>Coverage of 99% IR</td>
<td>77</td>
</tr>
<tr>
<td>Table 3.8</td>
<td>Coverage of 95% IR using $\hat{n}<em>{rec,1}$ and $\hat{n}</em>{rec,3}$</td>
<td>78</td>
</tr>
<tr>
<td>Table 3.9</td>
<td>Coverage of 99% IR using $\hat{n}<em>{rec,1}$ and $\hat{n}</em>{rec,3}$</td>
<td>78</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my major professor, Dr. Morris, for your patient guidance, constant encouragement and creative research ideas. What I learned from your deep understanding of applied statistics and great personalities during these years will benefit me forever.

I would like to thank my POS committee members, Dr. Dixon, Dr. Kaiser, Dr. Moloney and Dr. Wu, for your help and services.

This research was supported in part by Oak Ridge National Laboratory through SERDP Project CU-1201.

Many thanks go to Han Wu for helpful discussions and to Chunbai Wang for the LaTeX help.

Finally, I am indebted to my parents and my wife for their patient waiting, never-ending love and support.
Military bases that have been used for weapon-testing and training usually are contaminated with unexploded ordnance (UXO). These sites can be returned to public use only after UXO remediation. The cleaning-up procedure is usually very expensive and time-consuming. This demands statistical tools to provide more effective sampling strategy and to characterize the UXO distribution. Based on the physical characteristics of UXO deposition, we adopt a simplified Neyman-Scott process to model the UXO distribution. A line transect survey is used to collect data on one coordinate of individual object locations. Two-stage (global and local) sampling strategy is applied to screen the contaminated site. In the global sampling, the estimators of the cluster intensity, mean cluster size and cluster dispersion are provided. The theoretical variance estimators of all the cluster parameters are also given. Simulation studies show that all the parameter estimates perform well and their theoretical variance estimates are reasonably close to their corresponding sample variances. In the local sampling, an inclusion region for covering all the unobserved objects in a cluster is proposed. Its asymptotic coverage property is given and proved. Simulation studies show the actual coverage of the inclusion region is very close to the nominal level.
CHAPTER 1 INTRODUCTION

1.1 Background

Over the span of history, military forces have trained and fought on this continent, resulting in millions of rounds of unexploded ordnance left on and in the land and under water. Unexploded ordnance (UXO) is technically defined as: explosive ordnance which has been primed, fused, armed or otherwise prepared for action, and which has been fired, dropped, launched, projected, or placed in such a manner as to constitute a hazard to operations, installations, personnel, or material, and remains unexploded either by malfunction or design, or for any other cause [1] [2]. The presence of UXO is inevitable on any land that the military used for training or weapons development and testing. No types of munition explodes 100 percent of the time when fired [3]. Surveys in Laos and Cambodia after the Vietnam War indicated that 10-30 percent of bombs dropped on these countries failed to denotate [4]. The types of UXO vary widely depending on the types of military activities that took place at the site. UXO can range from small-arms ammunition to bombs weighing up to a ton. Other types of UXO include rounds, mortars, aircraft cannon, tank-fired projectiles, summunitions (which are designed to scatter over a large area), rockets, guided missiles, grenades, torpedoes, mines, chemical munitions, bulk explosives, pyrotechnics, etc [3]. Fig. 1.1 [5] gives some examples of UXO. According to a report of the Defence Science Board Task Force [6], it is estimated that over 15 million acres of land in the United States may contain some level of UXO contamination at about 1500 different sites. Taken together, all the contaminated sites
comprise an area the size of Florida.

Figure 1.1 Some examples of UXO

Generally speaking, there are two types of risks associated with the UXO sites. The first is the obvious risk of explosion. Exploding UXO causes serious injuries, dismemberment, or even deaths. The second type of risk is environmental contamination due to the infiltration of munitions-related chemicals such as explosives and perchlorate (a component of pyrotechnics and rocket fuel) into soil and groundwater. A prominent example exists at the Massachusetts Military Reservation on Cape Cod, Massachusetts, where decades of artillery training has contaminated the only drinking water for thousands of surrounding residents.

While locating UXO by sight is sometimes possible, most UXO is extremely difficult to locate by the raw eye without the aid of detection equipment, because UXO may be buried under the ground or UXO is deteriorated and camouflaged by soil, grasses, and leaves. The tools and technology available for UXO detection have not changed very much from those that the military has employed for cleaning mines and explosive
ordnance since World War II. The typical tools are magnetometers, electromagnetic induction (EMI) systems, infrared detectors, etc. Magnetometers measure distortion in the earth’s magnetic field caused by the presence of metal objects. EMI systems generate a magnetic field in the ground that induces current to flow in buried metal; this current in turn induces a secondary magnetic field with a voltage that is detected by the EMI instrument. Infrared detectors measure the thermal energy signatures; UXO on or near the soil surface may possess a different heat capacity or heat transfer properties than the surrounding soil, and this temperature difference can theoretically be detected and used to identify UXO. Fig. 1.2 exhibits Minex 2FD 4.500 Compact Metal Detector manufactured by Foerster Instruments Inc. [7].

Figure 1.2  Minex 2FD 4.500 Compact Metal Detector

The typical UXO detection process is a so-called “Mag and Flag” process. The UXO clearance crew is equipped with a metal detector and a shovel. The crew first clears vegetation from the suspicious contaminated site (using mechanical method or controlled burning). The site is divided into grids and the grids are splitted into lanes
(usually one meter wide). The crew member slowly advances along each lane, swinging the metal detector close to the ground. When the detector beeps, the crew member either plants a flag to indicate that the excavation will need to occur or starts digging with a shovel until the metal object is found. Fig. 1.3 shows a UXO clearance worker is swinging a Minex compact metal detector for UXO detection [7]. This “Mag and Flag” process is very time-consuming, expensive and labor-intensive, so it can only be applied to a small site. For a large contaminated site, usually tens of thousands of acres,
future land use determines the cleanup standards with respect to UXO. If a base being closed is going to be fenced off and left as a wildlife refuge which is inaccessible to the public, minimal cleanup may be applied. However, if the base is going to be transferred to public use, such as farm land, housing, park, etc, then extensive cleanup may be required to meet a 100% clearance objective. So far, the Army has made substantial progress in transferring land without UXO and very little progress in transferring land with UXO. About 61% of the acreage of BRAC land without UXO has been transferred to organizations outside the Army, while only 10% of the land with the presence of UXO has been transferred. The leading factors of delay of the transfer process are lack of information about UXO locations, types and quantities before land-reuse decisions were made; inability of detection technology to ensure that all UXO items have been located and removed; inability to meet regulators' requirement for reducing risk from UXO [9]. The BRAC program and the environment cleanup of other DOD sites have
generated a huge demand for UXO remediation which challenges Army engineers, outside scientists, and private UXO-clearance corporations to provide more accurate equipment, more economical sampling mechanism, safer removal procedure, etc.

1.2 Physical characteristics of UXO deposition

Before trying to develop a statistical model to characterize UXO over a contaminated site, we should know the physical process leading to the distribution of UXO. In a military-training or weapon-testing site, UXO typically is not deposited uniformly over the site, but in spatial clusters around some centers of activity. Generally speaking, UXO is deposited through a series of activities that occurred over the entire history of this site. Two stages apply to each ordnance activity. First, a decision is made on the location or locations of the activity. Then the activity is performed near the locations and leaves ordnance in a cluster pattern that is determined by the activity [10]. For example, a commanding officer may set up targets scattered over the site to train the new soldiers or to test the new weapons, then soldiers fire the weapons aiming at those targets. Under this scenario, UXO is not distributed uniformly over the entire site, but rather in clusters around those targets. As a result, much of the area is UXO-free while sub-areas around the targets may contain UXO. UXO within a cluster tends to be closer to each other, relative to the distance between clusters.

The task of locating all the UXO over the site will be fulfilled if we can find all the clusters and all the UXO within each cluster.

1.3 Statistical model

The distribution of locations of events (UXO in our case) is known as the spatial point process in the terminology of spatial statistics [11–13]. A spatial point process which mimics the physical characteristics of UXO deposition very closely is the Neyman-Scott
process [14]. Neyman and Scott applied this process to problems in cosmology [14] in 1958, and they also used this process to model the geometry of bombing in 1972 [15]. A spatial point process is a Neyman-Scott process if it satisfies the following conditions [11]:

1. Parent events are realized from an inhomogeneous Poisson process.

2. Each parent produces a random number of offspring, realized independently and identically for each parent according to a discrete probability distribution.

3. The positions of the offspring relative to their parents are independently and identically distributed.

4. The final process is composed of the superposition of offspring only.

In our UXO deposition setting, targets correspond to parents and UXO to the offspring. For condition 1, knowledge about where the commanding officer set up the targets may be needed to model some kinds of inhomogeneity in the parent process. In cases where such information is lacking or where all of the affected area may have been used, a homogeneous model of the parents may be appropriate. For condition 2, the number of ordnance fired at one target can be reasonably assumed as independent of that of a different target. For condition 3, we will assume that UXO is scattered around the targets randomly and independently of each other. For condition 4, we only consider the location of UXO, not the location of the targets, i.e., we assume that the targets have been removed and are not part of the material to be located. Therefore, for the deposition of UXO, the four conditions of the Neyman-Scott process are satisfied, or at least not severely violated. So the Neyman-Scott process is a reasonable model to use.

The specific model we will use in this research is a simplified Neyman-Scott process, in which

1. Parent events are realized from a homogeneous Poisson process with parameter \( \lambda \).
2. Each parent independently produces a Poisson number of offspring with mean $\mu$.

3. The offspring are independently and identically located around their parent with a radially symmetric normal density with variance $\rho^2$.

4. The final process is composed of the offspring only.

Note that in this simplified Neyman-Scott process, we have specified the parent distribution as a homogeneous Poisson process, which means the parents are uniformly scattered around the area. And the two coordinates of the positions of the UXO are independently and identically distributed as normal around their targets with the same dispersion.

1.4 Sampling methodology

We use a line transect survey to observe the data, partial location information on each detected item of UXO. The ultimate goal is to locate all UXO for removal. Line transect survey and point transact survey are two primary methods of distance sampling. Distance sampling is widely used in the biological community to estimate the density or abundance of biological populations [16–18]. Distance sampling provides a practical, cost-effective way to obtain reliable estimates of density of objects of interest. For objects distributed sparsely across large geographic areas, there are no competing methods against distance sampling [16]. In a line transect survey, an observer walks slowly along the pre-determined line transects (paths) to detect the objects (with or without instruments) and records the distance of any object which is detected to the transect. The sample data are the set of distances of detected objects and any relevant covariants. One assumption of line transact survey is that not all objects are detected, in fact, the probability of any object being detected is assumed to be $g(r)$, where $r$ is the perpendicular distance of this object to the line transect. Here, $g(r)$ is called the detection function. Any object which happens to be located on the transect is assumed to be
certainly detected, namely, $g(0) = 1$. Detectability decreases with increasing distance of the object to the transect, i.e., the detection function $g(r)$ is a decreasing function. Besides the certain detection for objects on the transect, it’s also desirable that detection remains almost certain for objects at small distance from the transect line; this is called a “shoulder” property of the detection function. When $g(r)$ possesses a “shoulder” property, it requires $g'(0)$ to be zero and $g(r)$ can’t decrease sharply near 0. From both theoretical and empirical points of view, the assumption of the “shoulder” property of the detection function is very important in getting reliable estimates of object intensity [16]. In a line transect survey, there might be many objects which remain undetected during the course of survey. In this sense, it’s an extension of plot sampling method in which all objects are assumed to be observed.

In most applications, UXO sampling involves inspection of a large acreage of land. Economically speaking, it’s not practical to screen all the area. Moreover, due to the limit of the current detection technology (magnetometers, EMI systems or other detection instruments can not guarantee to detect all the UXO in any situation), it’s reasonable to assume only a portion of UXO at the site can be detected. Since line transect survey is an efficient method for objects distributed sparsely over a large area, it is a natural selection as a method of UXO sampling. In fact, it indeed is the current UXO sampling methodology. Our sampling mechanism should take into consideration the generic two-stage point process from the model assumption of the Neyman-Scott process. In the model assumption, the distance between objects within the same cluster is generally smaller than the distance between clusters. Much of the area is UXO-free due to this cluster structure. So we adopt two-stage sampling. The objective of the stage-I sampling (or global sampling) is to detect the clusters and to estimate parameters associated with the Neyman-Scott process. The objective of the stage-II sampling (or local sampling) is to locate the individual UXO within a detected cluster and to provide some probability statements about UXO distribution within this detected cluster. Without detailed prior
information about where UXO is located, a single-stage sampling plan will usually detect few UXO per unit of sampling path since most of the area is UXO-free. The two-stage sampling is an adaptive strategy; intuitively it is more effective than single-stage sampling in our UXO setting.

Differences between a line transect survey for UXO and for some species in biology need to be emphasized here. First, the detection processes are different. In a UXO survey, the detection is through the aid of the instrument, which has limited reliability in the current technology. In a biological survey, the detection is mainly by the observer's visual sight and effectiveness may partly depend on the species' unique characteristic. Birds, for example, are very often conspicuous because of their bright coloration, distinctive song or call. So electromagnetic observation of UXO may be much more difficult than visual identification of a species. Secondly, information about resulting survey data are different. In a biological survey, the perpendicular distance from the transect to the detected object can be recorded, which means the exact positions of the detected objects in the study area are available. In a UXO survey where a magnetometer or an EMI system is applied, when UXO is detected, we don't know how far it is from the transect. The perpendicular distance of the detected UXO to the path is not available. So only one coordinate of the detected UXO (along the path) is recorded.

In the global sampling, we put some evenly spaced parallel transect lines in the large contaminated site and may let the aircraft equipped with an EMI system travel along those pre-determined transects to detect UXO. In the local sampling, for each of the clusters detected in the global sampling, more dense evenly spaced parallel transect line survey will be conducted in order to get more detailed information.
1.5 Literature review

For estimating the cluster intensity, mean object size of the cluster and the cluster dispersion of a Neyman-Scott process, the conventional approach is based on the Ripley’s K-function [11, 12, 19]. K-function captures the second-order spatial dependence between objects. The parameters of interest are estimated by minimizing the difference between theoretical and estimated values of the K-function. Using K-function generally requires the completely mapped data, which means the locations of all the objects of interest in the predefined study region should be observed. Stoyan [20] used K-function to estimate the parameters of planar Neyman-Scott processes, where the distances between the parent and the offspring follow a specified distribution and the orientations of the directed lines between the parent and the offspring are independently and uniformly on \([0, 2\pi]\). Taylor, Dryden and Farnoosh [21] modeled a generalized Neyman-Scott process, where the offspring are Gaussian perturbations from a regular mean configuration, through the K-function method.

For line transect surveys, Buckland et al. [16] provided comprehensive techniques to deal with this type of data. They include parametric modeling of the detection function, a nonparametric method for estimation of probability density functions by trigonometric Fourier series, etc. Borchers et al. [17] and Barry et al. [18] also described methods of distance sampling in detail. Quang [22] applied the bootstrap method to construct the confidence intervals for the density of objects or animals. A Bayesian approach was used by Karunamuni and Quinn [23] for estimating the density of a closed animal population, where the detection function was assumed to be Gaussian. Kernel estimation of population density was studied in a series of paper by Chen [24–27] and Mack et al. [28]. Hiby and Krishna [29] proposed estimating methods for curved transect survey when straight line transects are difficult to execute, such as in a dense forest where cutting straight line transects is too time-consuming and expensive.
For modeling Neyman-Scott processes with line transect survey data, Brown and Cowling [30] proposed an approach based on the likelihood theory of the simplified Neyman-Scott model with observations taken from line transect survey with a Gaussian detection function, where the distances of the detected objects to the line transects are available. First, they defined the distribution of the size of the detected cluster (i.e., number of detected objects) as the Neyman-Scott line transect distribution and showed that this distribution depends only on a single parameter which is a function of the mean cluster size, cluster dispersion and the detection function parameter. Through this Neyman-Scott line transect distribution and the likelihood theory, they provided the estimators of the cluster intensity, mean number of objects in each cluster, and the Gaussian detection function parameter. The standard errors of those parameter estimates were also given. They argued that the estimators are more efficient than those derived from the existing K-function method because they are based on the likelihood function. The standard error expressions are more tractable and easier to compute.

Cowling [31] provided a method of estimating the clustering parameters of the simplified Neyman-Scott process, for use when the observations are also assumed to be collected from line transect survey with a Gaussian detection function with a known detection parameter. As in our UXO sampling, the distances of the detected objects to the line transect are assumed to be unavailable in this paper; the author only used one coordinate of the detected objects. The author derived the estimators from a traditional K-function approach, which has the drawback of not allowing the parameter estimators to be explicitly expressed in terms of data. Therefore, the explicit expression of bias and standard errors of the parameter estimators are not available, which is not as appealing as the approach in Brown and Cowling [30]. Since there was only one-dimension information of location of the detected objects, the author was forced to assume the Gaussian detection function is absolutely known. In this paper, the author also suggested transect-based Hopkins statistic and Kolmogorov-Smirnoff statistic to test the
clustering assumption and compared the power of two tests based on the above two test statistics.

Although the data form of Cowling [31] is exactly as ours, three concerns make us to try Brown and Cowling's approach of estimation instead of Cowling's. The first concern is that the K-function method usually requires completely mapped data, or at least a large number of observed objects in the detected clusters; this requirement is often not satisfied in our case because UXO are sparsely distributed. The second concern is that the estimators from Cowling's approach are not as reliable as those from Brown and Cowling's approach because the latter procedure is based on the likelihood approach. The third concern is that there are no explicit expressions of standard errors of parameter estimators for Cowling's approach. In Chapter 2 (Global Sampling), we extend Brown and Cowling's approach of combination of the Neyman-Scott line transect distribution and the likelihood theory from the full information of location of detected objects, to only one-dimension coordinates along the transect lines. We provide estimators of all the clustering parameters and their asymptotic standard errors with explicit expressions, and simulation studies are conducted to evaluate the performance of the given estimators and their stand errors. In Chapter 3 (Local Sampling), we deal with follow-up sampling to further document individual clusters. We propose an inclusion region to cover all the undetected objects with a controlled probability. The asymptotic property of the inclusion region is given and proved. Simulation studies are also conducted to compare the actual coverage of the inclusion region with its nominal coverage.
CHAPTER 2 GLOabal Sampling

2.1 Model assumptions

We assume UXO are spatially distributed over the studied region as a simplified Neyman-Scott process [30]. In this process, there are Pois(\lambda A) number of "parents", where \( A \) is the area of the study region; for each parent, there are Pois(\mu) number of "offspring" in a cluster and the cluster sizes are independent. The parents are uniformly distributed over the study region, that is, the process generating parent locations is assumed to be homogeneous. The offspring are independently and identically located with a radially symmetric normal density with variance \( \rho^2 \) around the location of the parent. The final process consists of just the offspring, i.e., the parents serve to locate the clusters, but do not correspond to individual UXO objects. Fig. 2.1 is a realization of simplified Neyman-Scott process with \( \lambda=0.002, \mu=20 \), and \( \rho^2=4 \) in a \((0,100) \times (0,100)\) region. In biological applications, it is important, and of primary interest in preliminary global sampling, to estimate the cluster intensity \( \lambda \), the mean number of objects in each cluster \( \mu \) and the cluster dispersion \( \rho^2 \). In this chapter, besides these objectives, we also want to estimate the variances of these estimates. The traditional approach for estimation is based on Ripley's K-function [19] through a combination of a survey area in which we detect all the objects and know their locations, and a parametric model such as a Neyman-Scott process. The huge economic cost associated with surveying the whole region prevents us from implementing this approach. So we adopt a line transect survey to observe the data. The second goal in the global sampling is to detect at least
one UXO in each cluster, to serve as a starting point in the local sampling (Chapter 3) to characterize each cluster.

![Figure 2.1 A realization of Neyman-Scott process with $\lambda=0.002$, $\mu=20$, and $\rho^2=4$](image)

Figure 2.1 A realization of Neyman-Scott process with $\lambda=0.002$, $\mu=20$, and $\rho^2=4$

We assume a Gaussian detection function $g(r) = e^{-\frac{r^2}{2\sigma^2}}$ in the line transect survey; here $r$ is the perpendicular distance between the object and the line transect, and $\sigma$ is the detection parameter. $g(r)$ is the probability that any object located distance $r$ from a transect will be detected in the sampling along that transect. The "effective strip half-width" [17] denoted by $\omega$ associated with this detection function is defined to be $\frac{1}{2}\sqrt{2\pi}\sigma$. The meaning of the effective strip half-width $\omega$ is that we expect to observe the same number of objects as would be the case for a detection function $g(r) = 1$, $0 \leq r < \omega$ and $g(r) = 0$, otherwise. The effective area covered is $2\omega$ per unit length of
the line transect ignoring any "overlap" associated with intersecting transects. There is another interesting property of \( \omega \); the expected number of undetected objects within the distance \( \omega \) of the line transect is equal to the expected number of detected objects outside the distance \( \omega \) from the line. Note that for this Gaussian detection function, we implicitly assume any objects which by chance are located on a transect will be detected surely. This detection function also has the "shoulder" property (Chapter 1) because \( g'(0) = 0 \).

A magnetometer or an EMI system is used to detect the UXO by traveling along each of the transects. We can record the position of each detected object along the transect line, but the perpendicular distance of the detected objects to the transect is unobservable. For example, if the transect runs north-to-south (sometimes described as "vertical", since it would appear as a vertical line segment on a map), we only observe the \( y \)-coordinates of the detected objects. \( \sigma \) is assumed to be a known parameter (also assumed by Cowling [31]); this assumption is reasonable in this application based on the fact that we know the physical detection capability of the magnetometer or EMI system.

2.2 The Neyman-Scott line transect distribution

This section summarizes the results given by Brown and Cowling [30] regarding to the Neyman-Scott line transect distribution.

In the simplified Neyman-Scott Poisson process, consider a parent event at \((c, d)\) and suppose a line transect survey is conducted along the line \( X = 0 \) with the detection function \( g(x) = e^{-\frac{x^2}{2\sigma^2}} \). The detection function specifies the probability that an object will be detected if it is (signed) distance \( x \) from the transect line. This detection function implies that objects located on the transect line will surely be detected, while objects more than a few multiples of \( \sigma \) away from the line will seldom be detected. Conditional
upon \((c, d)\), the probability that any one unspecified object in the cluster is detected is

\[
P_c = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{2\pi \rho^2} e^{-\frac{(x-c)^2 + (y-d)^2}{2\rho^2}} \, dx \, dy
\]

\[
= \int_{-\infty}^{\infty} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi \rho}} e^{-\frac{(x-c)^2}{2\rho^2}} \, dx
\]

\[
= \frac{1}{\sqrt{2\pi \rho}} e^{-\frac{x^2}{2(\sigma^2 + \rho^2)}} \int_{-\infty}^{\infty} e^{-\frac{(x-c)^2}{2\rho^2}} \, dx
\]

\[
= \frac{1}{\sqrt{2\pi \rho}} e^{-\frac{x^2}{2(\sigma^2 + \rho^2)}} \sqrt{2\pi \rho} \frac{\sigma}{\sqrt{\sigma^2 + \rho^2}}
\]

\[
= \frac{\sigma}{\sqrt{\sigma^2 + \rho^2}} e^{-\frac{x^2}{2(\sigma^2 + \rho^2)}}
\]

(2.1)

Note that the above expression of \(P_c\) has a Gaussian form in \(c\); this leads to

\[
\int_{-\infty}^{\infty} P_c^j \, dc = \int_{-\infty}^{\infty} \left( \frac{\sigma}{\sqrt{\sigma^2 + \rho^2}} \right)^j e^{-\frac{j^2}{2(\sigma^2 + \rho^2)}} \, dc
\]

\[
= \left( \frac{\sigma}{\sqrt{\sigma^2 + \rho^2}} \right)^j \sqrt{2\pi} \sqrt{\frac{\sigma^2 + \rho^2}{j}}
\]

\[
= \sqrt{\frac{2\pi}{j}} \frac{\sigma^j}{(\sigma^2 + \rho^2)^{\frac{j}{2} - \frac{1}{2}}}
\]

(2.2)

for any positive integer \(j\).

Let \((x, y)\) be the location of a detected object, then its joint probability density function is

\[
f(x, y) = \frac{1}{P_c} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{2\pi \rho^2} e^{-\frac{(x-c)^2 + (y-d)^2}{2\rho^2}}.
\]

(2.3)

Before the detection process, \(x\) and \(y\) are independent with \(x \sim N(c, \rho^2)\) and \(y \sim N(d, \rho^2)\). From Eq. (2.3), we know the detection process doesn’t change the independence of \(x\) and \(y\), or the marginal distribution of \(y\), but the marginal distribution of \(x\)
conditional on the object being detected is

\[ x \sim N \left( \frac{\sigma x^2}{\sigma^2 + \rho^2}, \frac{\sigma^2 \rho^2}{\sigma^2 + \rho^2} \right), \tag{2.4} \]

which can be obtained from the development of the expression of \( P_c \) in Eq. (2.1).

Now let \( M \) be the number of detected objects in this cluster and let \( N \) be the true size of the cluster; then \( M \leq N \) and the cluster is not detected if \( M = 0 \). Note that \( N \sim \text{Pois}(\mu) \) from the model assumption. Conditional on \( c \), for any nonnegative integer \( m \), we have

\[
\Pr(M = m \mid c) = \Pr(M = m, N \geq m \mid c) = \sum_{k=m}^{\infty} \Pr(M = m, N = k \mid c) = \sum_{k=m}^{\infty} \Pr(M = m \mid N = k, c) \Pr(N = k \mid c) = \sum_{k=m}^{\infty} \binom{k}{m} P_c^m (1 - P_c)^{k-m} \left[ \frac{\mu^k e^{-\mu}}{k!} \right] = \frac{(\mu P_c)^m e^{-\mu P_c}}{m!}. \tag{2.5} \]

The above fourth equation holds because \( M \mid N, c \) is a binomial distribution and \( N \) is independent of \( c \). Therefore, conditional on \( c \), \( M \) follows a \( \text{Pois}(\mu P_c) \) distribution, i.e., the detected objects form a "thinned" Poisson process. The probability that a cluster is detected (i.e., that at least one object in the cluster is detected) is

\[
\Pr(M > 0 \mid c) = 1 - e^{-\mu P_c}. \tag{2.6} \]

Denote by \( M^* \) the size of a detected cluster, then \( M^* = M \) when \( M > 0 \). The only difference between \( M^* \) and \( M \) is whether the cluster is detected or not. When the
cluster is not detected, i.e. \( M = 0 \), \( M^* \) is not defined. Then, conditional on \( c \), \( M^* \) has a \( \text{Pois}(\mu P_c) \) distribution truncated below 1, i.e.,

\[
\Pr(M^* = m \mid c) = \frac{1}{1 - e^{-\mu P_c}} \frac{m e^{-\mu P_c}}{m!}
\]

(2.7)

for any positive integer \( m \).

Next we want to develop an unconditional distribution of \( M^* \). For any unspecified cluster, the parent position \((c, d)\) is uniformly distributed from the model assumptions, so the density of \( c \), denoted by \( f(c) \), is proportional to 1. For any positive integer \( m \),

\[
\Pr(M^* = m) = \frac{\Pr(M = m)}{\Pr(M > 0)} = \frac{\int_{-\infty}^{\infty} \Pr(M = m \mid c) f(c) dc}{\int_{-\infty}^{\infty} \Pr(M > 0 \mid c) f(c) dc}
\]

(2.8)

Expanding \( e^{-\mu P_c} \) into a Taylor series and applying Eq. (2.2), we have

\[
\Pr(M^* = m) = \frac{\mu^m}{m!} \sum_{j=0}^{\infty} \frac{(-\mu)^j}{j!} \int_{-\infty}^{\infty} P_c^{m+j} dc
\]

\[
= \frac{\mu^m}{m!} \sum_{j=1}^{\infty} \frac{(-\mu)^j}{j!} \int_{-\infty}^{\infty} P_c^{m+j} dc
\]

\[
= \frac{\mu^m}{m!} \sum_{j=1}^{\infty} \frac{(-\mu)^j}{j!} \frac{\sqrt{2\pi} \sigma^m}{\sqrt{m+j(\sigma^2+p^2)(m+j-1)/2}}
\]

\[
= \left( \frac{\mu \sigma}{\sqrt{\sigma^2+p^2}} \right)^m \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \frac{\mu \sigma}{\sqrt{\sigma^2+p^2}} \left( \frac{\mu \sigma}{\sqrt{\sigma^2+p^2}} \right)^j
\]

(2.9)
Now define
\[
\theta = \frac{\mu \sigma}{\sqrt{\sigma^2 + \rho^2}} \quad (2.10)
\]
and a function \( T(\theta) \) as
\[
T(\theta) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1} \theta^j}{j! \sqrt{j}}. \quad (2.11)
\]

After some algebraic work, the first and \( m^{th} \) derivatives of \( T(\theta) \) can be derived as
\[
T'(\theta) = \sum_{j=0}^{\infty} \frac{(-1)^j \theta^j}{j! \sqrt{j} + 1}, \quad (2.12)
\]
and
\[
T^{(m)}(\theta) = (-1)^{m+1} \sum_{j=0}^{\infty} \frac{(-1)^j \theta^j}{j! \sqrt{m+j}}, \quad (2.13)
\]
respectively. Using the definition of \( \theta \) and \( T(\theta) \) in Eq. (2.9), the unconditional distribution of \( M^* \) can be derived as
\[
\Pr(M^* = m) = \frac{\theta^m}{m!} \frac{(-1)^{m+1} T^{(m)}(\theta)}{T(\theta)}, \quad (2.14)
\]
for any positive integer \( m \).

The above distribution of \( M^* \) is called the Neyman-Scott line transect distribution [30], which depends only on the parameter \( \theta \) through the function \( T \). We can find its moment generating function \( h(s) \) after a lengthy calculation as
\[
h(s) = 1 - \frac{T((1 - \log(s))\theta)}{T(\theta)} \quad (s > 0). \quad (2.15)
\]

Through this moment generating function, the mean and variance of \( M^* \) can be easily derived as:
\[
E(M^*) = \frac{\theta}{T(\theta)} \quad (2.16)
\]
and

\[ \text{Var}(M^*) = \frac{\theta}{T(\theta)} \left( 1 + \frac{\theta}{\sqrt{2}} - \frac{\theta}{T(\theta)} \right). \]  

(2.17)

At the beginning of this section, we assume a line transect positioned at \( X = 0 \). If this transect is put at \( X = x_0 \) instead of 0, the probability that an object is detected is generalized to

\[ P_{c,x_0} = \frac{\sigma}{\sqrt{\sigma^2 + \rho^2}} e^{-\frac{(x_0-c)^2}{2(\sigma^2+\rho^2)}} \]  

(2.18)

and the marginal density of \( x \) in an observed object \((x, y)\) is generalized to

\[ x \sim N \left( \frac{c\sigma^2 + x_0\rho^2}{\sigma^2 + \rho^2}, \frac{\sigma^2\rho^2}{\sigma^2 + \rho^2} \right), \]  

(2.19)

while the remaining results are unchanged.

2.3 Estimation of parameters and standard errors

In this section, we extend the approach of parameter estimation in Brown and Cowl­ling [30] from the case that two coordinates \((x, y)\) of any detected object are available to the case that only \( y \)-coordinate is available.

Suppose \( m \) vertical line transects are evenly positioned through the study region. Let \( L \) be the total length of all these \( m \) transects. We will assume that the horizontal distance between two neighboring transects is so large that the probability any object will be detected along more than one transect is essentially zero. Hence, for purposes of process parameter estimation, data from these \( m \) transects can be analyzed as if they were from one transect with length \( L \).

We assume the clusters in the region are identifiable. That is, we know which pairs of detected objects are from a common cluster and which are not. Suppose we detect \( k \) clusters and \( n_i, i = 1, 2, \ldots, k \), is the detected size of the \( i^{th} \) detected cluster; clearly, \( n_i \geq 1 \), otherwise the cluster is undetected. Denote by \( n \) the total number of detected
objects within the region, i.e., \( n = \sum_{i=1}^{k} n_i \). Let \( y_{ij}, i = 1, 2, \ldots, k, j = 1, 2, \ldots, n_i \), be the \( y \)-coordinate of the \( j^{th} \) detected object in the \( i^{th} \) detected cluster. Define \( S_i^2 = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \), the corrected sum of squares of \( y \) within the \( i^{th} \) cluster, where \( \bar{y}_i = \frac{\sum_j y_{ij}}{n_i} \), the average detected \( y \) value within the \( i^{th} \) cluster. Let \( S^2 = \sum_{i=1}^{k} S_i^2 \) be the pooled version of sums of squares of \( y \) across all detected clusters.

### 2.3.1 Estimation of \( \theta \)

The first-order moment of \( M^* \) in Eq. (2.16) suggests a method of moments estimator of \( \theta \) by inverting

\[
\frac{\hat{\theta}}{T(\theta)} = \overline{M}^* ,
\]

where \( \overline{M}^* = \frac{n}{k} \), the sample average of detected cluster size. Denote this estimator by \( \hat{\theta} \). Define \( g(\theta) = \frac{\theta}{T(\theta)} \), then \( g'(\theta) = \frac{T(\theta) - \theta T'(\theta)}{T^2(\theta)} \). Applying the delta method to Eq. (2.20), we have

\[
\text{Var}(\overline{M}^*) \approx g'^2(\theta) \text{Var}(\hat{\theta}).
\]

On the other hand, \( \text{Var}(\overline{M}^*) = \frac{1}{k} \text{Var}(M^*) \) since cluster sizes are independent of each other. Thus, equating these two expressions and substituting Eq. (2.17) give the asymptotic variance of \( \hat{\theta} \) as

\[
\text{Var}(\hat{\theta}) \approx \frac{1}{k} \frac{\theta T^3(\theta) \left( 1 + \frac{\theta}{\sqrt{2}} - \frac{\theta}{T(\theta)} \right)}{(T(\theta) - \theta T'(\theta))^2}.
\]

Replacing \( \theta \) with its estimator \( \hat{\theta} \) in this variance formula produces a variance estimator of \( \hat{\theta} \) as

\[
\hat{\text{Var}}(\hat{\theta}) = \frac{1}{k} \frac{\hat{\theta} T^3(\hat{\theta}) \left( 1 + \frac{\hat{\theta}}{\sqrt{2}} - \frac{\hat{\theta}}{T(\hat{\theta})} \right)}{(T(\hat{\theta}) - \hat{\theta} T'(\hat{\theta}))^2}.
\]
2.3.2 Estimation of $\rho^2$

Since the marginal distribution of $y$, the vertical coordinate of any detected object, is normal with variance $\rho^2$ after the detection, and the location and detection of any two objects are independent, then $S_i^2 \sim \rho^2 \chi_{n_i-1}^2$ for given $n_i$. Thus, for the pooled sums of squares of $y$, $S^2$, we have $S^2 \sim \rho^2 \chi_{n-k}^2$ for given $n$ and $k$. This suggests an estimator of $\rho^2$ as

$$\hat{\rho}^2 = \frac{1}{n-k} S^2.$$  \hspace{1cm} (2.24)

Note that conditional only on $n - k$ being positive $\hat{\rho}^2$ is an unbiased estimator of $\rho^2$ because

$$E(\hat{\rho}^2) = E(E(\rho^2 \mid n, k))$$
$$= E \left( \frac{1}{n-k} E(S^2 \mid n, k) \right)$$
$$= E \left( \frac{1}{n-k} (n-k)\rho^2 \right)$$
$$= \rho^2.$$  \hspace{1cm} (2.25)

Further,

$$Var(\hat{\rho}^2) = Var \left( E \left( \frac{1}{n-k} S^2 \mid n, k \right) \right) + E \left( Var \left( \frac{1}{n-k} S^2 \mid n, k \right) \right)$$
$$= Var(\rho^2) + E \left( \frac{1}{(n-k)^2} 2(n-k)\rho^4 \right)$$
$$= 0 + 2\rho^4 E \left( \frac{1}{n-k} \right)$$
$$\approx \frac{2\rho^4}{n-k}.$$  \hspace{1cm} (2.26)
which suggests a variance estimator of \( \hat{\rho}^2 \) given by

\[
\text{Var}(\hat{\rho}^2) = \frac{2\Sigma^4}{(n-k)^3}.
\] (2.27)

### 2.3.3 Estimation of \( \mu \)

The relationship between \( \theta \) and \((\mu, \rho^2)\) from Eq. (2.10) can yield an estimator of \( \mu \) because we already have the estimators of \( \theta \) and \( \rho^2 \). Denote this estimator by \( \hat{\mu} \), then

\[
\hat{\mu} = \frac{\hat{\theta} \sqrt{\sigma^2 + \hat{\rho}^2}}{\sigma}.
\] (2.28)

Recall that we assume \( \sigma \) is a known parameter since, in practice, it represents sensitivity of the physical detection system. If the correlation between \( \hat{\theta} \) and \( \hat{\rho}^2 \) is very small (this appears to be the case as demonstrated in the later simulation studies), then we can construct an approximate variance estimator of \( \hat{\mu} \) by using the delta method as

\[
\text{Var}(\hat{\mu}) = \left( \frac{\partial \hat{\mu}}{\partial \hat{\theta}} \right)^2 \text{Var}(\hat{\theta}) + \left( \frac{\partial \hat{\mu}}{\partial \hat{\rho}^2} \right)^2 \text{Var}(\hat{\rho}^2),
\] (2.29)

where

\[
\frac{\partial \hat{\mu}}{\partial \hat{\theta}} = \frac{\sqrt{\sigma^2 + \hat{\rho}^2}}{\sigma} = \frac{\hat{\mu}}{\hat{\theta}}
\] (2.30)

and

\[
\frac{\partial \hat{\mu}}{\partial \hat{\rho}^2} = \frac{\hat{\theta}}{2\sigma \sqrt{\sigma^2 + \hat{\rho}^2}} = \frac{\hat{\theta}^2}{2\sigma^2 \hat{\mu}}.
\] (2.31)

Therefore,

\[
\text{Var}(\hat{\mu}) = \left( \frac{\hat{\mu}}{\hat{\theta}} \right)^2 \text{Var}(\hat{\theta}) + \frac{1}{4\hat{\mu}^2} \left( \frac{\hat{\theta}}{\sigma} \right)^4 \text{Var}(\hat{\rho}^2),
\] (2.32)

where \( \text{Var}(\hat{\theta}) \) and \( \text{Var}(\hat{\rho}^2) \) are as defined in Eq. (2.23) and (2.27), respectively.
2.3.4 Estimation of $\lambda$

Before we estimate $\lambda$, let's first examine the distribution of $k$, the number of detected clusters. Suppose $K_t$ is the true number of clusters, then $K_t \sim \text{Pois}(\lambda A)$ from the model assumptions. As in the case of $M$ in the above section, the number of detected clusters is also related to a thinned Poisson process, $\text{Pois}(\lambda a)$, where $a$ is the integrated cluster detection probability, given by

$$a = L \int_{-\infty}^{\infty} \Pr(M > 0|c) dc,$$

and where $L$ is the length of all the line transects. Using the expression of $\Pr(M > 0|c)$ in Eq. (2.6) and expanding $e^{-\mu P_x}$ into a Taylor series, we obtain

$$a = L\sqrt{2\pi(\sigma^2 + \rho^2)} T(\theta).$$

The $\text{Pois}(\lambda a)$ distribution of $k$ suggests estimating $\lambda$ by $\hat{\lambda} = \frac{k}{\hat{a}}$, where $\hat{a}$ is the estimator of $a$ formed by replacing the unknown parameters with their corresponding estimators in Eq. (2.34). Thus, we have

$$\hat{\lambda} = \frac{k}{L\sqrt{2\pi(\sigma^2 + \rho^2)} T(\hat{\theta})}.$$

The above $\hat{\lambda}$ involves three random quantities, $k$, $\hat{\theta}$ and $\hat{\rho^2}$. If the pairwise correlations among them are small (as in the case of $\text{Var}(\hat{\mu})$, the simulation studies offer evidence supporting this), an estimator of the variance of $\hat{\lambda}$ is derived through the delta method as

$$\text{Var}(\hat{\lambda}) = \left(\frac{\partial \hat{\lambda}}{\partial k}\right)^2 \text{Var}(k) + \left(\frac{\partial \hat{\lambda}}{\partial \hat{\theta}}\right)^2 \text{Var}(\hat{\theta}) + \left(\frac{\partial \hat{\lambda}}{\partial \hat{\rho^2}}\right)^2 \text{Var}(\hat{\rho^2}),$$

(2.36)
where

\[
\frac{\partial \hat{\lambda}}{\partial k} = \frac{1}{L \sqrt{2\pi(\sigma^2 + \rho^2)T(\hat{\theta})}} = \frac{\hat{\lambda}}{k},
\]

\[
\frac{\partial \hat{\lambda}}{\partial \theta} = \frac{k}{L \sqrt{2\pi(\sigma^2 + \rho^2)T^2(\theta)}} \frac{-T'(\hat{\theta})}{T(\theta)} = -\hat{\lambda} \frac{T'(\hat{\theta})}{T(\theta)},
\]

\[
\frac{\partial \hat{\lambda}}{\partial \rho^2} = \frac{1}{L \sqrt{2\pi T(\hat{\theta})}} \frac{-1}{2(\sigma^2 + \rho^2)^{3/2}} = \frac{\hat{\lambda}}{2(\sigma^2 + \rho^2)},
\]

and

\[
\text{Var}(k) = \hat{\lambda} \hat{\alpha} = k.
\]

Substituting the above four expressions into Eq. (2.36) results in

\[
\text{Var}(\lambda) = \hat{\lambda} \left( \frac{1}{k} + \left( \frac{T'(\hat{\theta})}{T(\hat{\theta})} \right)^2 \text{Var}(\hat{\theta}) + \frac{1}{4(\sigma^2 + \rho^2)^2} \text{Var}(\hat{\rho}^2) \right).
\]

\[
\text{Var}(\lambda) = \hat{\lambda} \left( \frac{1}{k} + \left( \frac{T'(\hat{\theta})}{T(\hat{\theta})} \right)^2 \text{Var}(\hat{\theta}) + \frac{1}{4(\sigma^2 + \rho^2)^2} \text{Var}(\hat{\rho}^2) \right).
\]

### 2.3.5 Estimation of object intensity \( \tau \)

Often times the estimation of the mean object intensity per unit area \( \tau = \lambda \mu \) is of interest in clustered biological populations. Although estimation of \( \tau \) is not our interest, we give its estimator here to demonstrate that it takes the form one would expect, based on estimates derived to this point. If we let \( \hat{\tau} = \hat{\lambda} \hat{\mu} \), then

\[
\hat{\tau} = \frac{k}{L \sqrt{2\pi(\sigma^2 + \rho^2)T(\hat{\theta})}} \frac{\hat{\theta} \sqrt{\sigma^2 + \rho^2}}{\sigma} = \frac{k}{L \sqrt{2\pi \sigma T(\hat{\theta})}} = \frac{n}{L \sqrt{2\pi \sigma}}.
\]

The last equation holds because \( \frac{\hat{\theta}}{T(\hat{\theta})} = \frac{n}{k} \).
Denote by $\omega$ the effective strip half-width, then

$$\omega = \frac{1}{2} \sqrt{2\pi \sigma}$$  \hspace{1cm} (2.43)

and the effective area surveyed is

$$L \int_{-\infty}^{\infty} g(x)dx = L\sqrt{2\pi \sigma} = 2L\omega.$$  \hspace{1cm} (2.44)

Therefore, $\tau$ in Eq. (2.42) takes the familiar form of

\[
\frac{\text{total number of objects observed}}{\text{effective area surveyed}}
\]

2.4 More about estimating $\theta$

$\theta$ is the only parameter of the Neyman-Scott line transect distribution. In Section 2.3, we derived a method of moments estimator. Brown and Cowling [30] found that this MOM estimator is efficient when the true $\theta$ is very small (less than 5), and the efficiency decreases when $\theta$ becomes moderately large. They concluded it is an open question whether or not other sample functions of $M^*$ can yield more efficient estimators for moderate values of $\theta$. Although $\theta$ is not of primary interest in our application, the estimators of $\mu$ and $\lambda$ are based on the estimator of $\theta$. Thus, an efficient estimator of $\theta$ is desirable.

Since all information about $\theta$ lies in $\Pr(M^* = m)$, the distribution of the size of detected clusters, the maximum likelihood estimator is an alternative to the MOM estimator. The likelihood function of $\theta$ after observing $k$ clusters where the $i^{th}$ cluster size is $n_i$ ($i = 1, 2, \cdots, k$) is

$$L(\theta) = \prod_{i=1}^{k} \frac{\theta^{n_i} (-1)^{n_i+1} T^{(n_i)}(\theta)}{n_i! T(\theta)}$$  \hspace{1cm} (2.45)
and the log-likelihood function is, apart from additive constants,

\[ l(\theta) = n \log(\theta) - k \log(T(\theta)) + \sum_{i=1}^{k} \log \left( (-1)^{n_i+1} T^{(n_i)}(\theta) \right). \]  

(2.46)

Apparently, there is no closed form for the MLE of \( \theta \), denoted by \( \hat{\theta}_{ml} \). Its variance estimator can be obtained from the inverse of the observed Fisher information \[32\], i.e.,

\[ \text{Var} \left( \hat{\theta}_{ml} \right) = \frac{1}{l'' \left( \hat{\theta}_{ml} \right)}, \]  

(2.47)

where

\[ l''(\theta) = -\frac{n}{\theta^2} - k \frac{T''(\theta)}{T(\theta)} + k \left( \frac{T'(\theta)}{T(\theta)} \right)^2 + \sum_{i=1}^{k} \frac{T^{(n_i+2)}(\theta)}{T^{(n_i)}(\theta)} - \sum_{i=1}^{k} \left( \frac{T^{(n_i+1)}(\theta)}{T^{(n_i)}(\theta)} \right)^2 \]  

(2.48)

is the second derivative of the log-likelihood function \( l(\theta) \).

Evaluation of \( T(\theta) \) and its derivatives \( T^{(m)}(\theta) \) is the most difficult aspect of numerical evaluation of the MLE. Note that \( T(\theta) = \sum_{j=1}^{\infty} \frac{(-1)^{j+1}\theta^j}{j!} \) is similar to \( e^{-\theta} = \sum_{j=0}^{\infty} \frac{(-1)^j\theta^j}{j!} \) except that each term in \( T(\theta) \) has an additional \( \sqrt{j} \) in its denominator. When evaluating the alternating series of \( e^{-\theta} \), we can evaluate \( 1/e^\theta \) instead of directly computing that series to avoid cancellation error \[33\]. But \( T(\theta) \) doesn’t possess this nice property because of the additional \( \sqrt{j} \) term. Since we don’t know the true value of \( T(\theta) \) for any given \( \theta \), but \( e^{-\theta} \) is available in any programming language, we can test the accuracy of direct computing of the alternating series of \( e^{-\theta} \). This accuracy can be related to that of \( T(\theta) \) because of the similarity of those two alternating series. We found that the accuracy is extremely good when \( \theta \) is less than 15 and perhaps acceptable for \( \theta \) between 15 and 20, but that the results are essentially useless when \( \theta \) reaches 40. This suggests we may only have a reliable evaluation of \( T(\theta) \) for small \( \theta \), say less than 20. For \( \theta \) larger than 20, if we could find a relationship between \( T(\theta) \) and \( T(c\theta) \), where \( c \) is a constant, then
we could evaluate $T(\theta)$ for small $\theta$ and avoid evaluation for large values. But we failed to find any useful relationship. The evaluation of $T^{(m)}(\theta)$ has similar problems because of the similarity of these alternating series.

A simulation study was conducted to compare the two estimators, the maximum likelihood estimator $\hat{\theta}_{ml}$ and method of moments estimator $\hat{\theta}$, for true $\theta$ values less than 20. In the simulation, $\mu$ was set to be 50, $\lambda$ to be 0.005, and we assume the known parameter $\sigma$ of the detection function to be 0.8. $\rho^2$ was chosen as four values 4, 9, 16 and 25 to yield different values of $\theta$. For each value of $\theta$, 10,000 realizations of the simplified Neyman-Scott process were simulated, MLE and MOM of $\theta$ were computed for each realization. We used two numerical methods, Newton-Ralphson iteration method and grid searching method, to evaluate $\hat{\theta}_{ml}$. There was no substantial difference in results for these two numerical evaluation methods from the simulation. Table 2.1 is the summary of the simulation results. In this table, we recorded the means, standard deviations and mean squared errors (MSE) of the 10,000 computed estimates of each method for four different true values of $\theta$. Here, MSE is the sum of the sample variance and the squared deviance of mean from the true value. Overall, both MOM and MLE estimate $\theta$ reliably except where $\theta = 7.899$, in which the mean of the MOM estimator is a little farther from the true value. MSE for MLE is smaller than that for MOM across all four values of $\theta$, which suggests MLE performs much better than MOM.

<table>
<thead>
<tr>
<th>true $\theta$</th>
<th>method</th>
<th>mean</th>
<th>std</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.899</td>
<td>MOM</td>
<td>9.099</td>
<td>0.964</td>
<td>2.367</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>8.015</td>
<td>0.755</td>
<td>0.583</td>
</tr>
<tr>
<td>9.806</td>
<td>MOM</td>
<td>10.027</td>
<td>1.262</td>
<td>1.642</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>9.579</td>
<td>1.016</td>
<td>1.085</td>
</tr>
<tr>
<td>12.883</td>
<td>MOM</td>
<td>12.638</td>
<td>1.852</td>
<td>3.488</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>12.549</td>
<td>1.408</td>
<td>2.095</td>
</tr>
<tr>
<td>18.570</td>
<td>MOM</td>
<td>18.303</td>
<td>3.212</td>
<td>10.386</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
<td>18.061</td>
<td>1.840</td>
<td>3.645</td>
</tr>
</tbody>
</table>

Table 2.1  Comparison of two methods of estimating $\theta$
For $\theta$ larger than 20, because the numerical evaluation of $T(\theta)$ and $T^{(m)}(\theta)$ is not accurate, neither MOM nor MLE is practical. In those situations, we used an ad hoc modification to develop an estimator of $\theta$ in the following way. We count two observed objects in any detected cluster as one observed object, that is to say, if $n_i$ objects are detected in one cluster, we pretend to observe just $\frac{n_i}{2}$ objects. By doing this, we cut the mean size of the detected clusters by half. Data such as this would be typical of what would be seen if the detection parameter $\sigma$ were actually $\frac{\sigma}{2}$, provided $\rho$ is large relative to $\sigma$. We can approximate the MOM by using the half mean size of detected clusters, then "reverse" this MOM back to the estimator of the original $\theta$ by approximately doubling it. More precisely, suppose $\theta^*$ is the Neyman-Scott line distribution parameter associated with the detection parameter $\frac{\sigma}{2}$, then

$$\theta^* = \frac{\mu(\sigma/2)}{\sqrt{(\sigma/2)^2 + \rho^2}}. \quad (2.49)$$

The relationship between $\theta$ and $\theta^*$ is

$$\theta = 2\theta^* \sqrt{(\sigma/2)^2 + \rho^2} \Big/ \sqrt{\sigma^2 + \rho^2}; \quad (2.50)$$

when $\sigma$ is much smaller than $\rho$ (this is our case), then $\theta \approx 2\theta^*$. Let $n$ and $k$ be the number of total detected objects and the number of detected clusters, and let $n^*$ and $k^*$ be the corresponding quantities associated with $\frac{\sigma}{2}$. Note that $n^*$ and $k^*$ are not real data; we don't observe them. Since the effective area surveyed with $\frac{\sigma}{2}$ is half that surveyed with $\sigma$, then $n^*$ is about half of $n$. Generally speaking, $k^*$ will be less than $k$, but we ignore this point for the moment. Define

$$M_1 = \frac{n}{2k} \quad (2.51)$$

and pretend $M_1$ is the observed mean detected cluster size from half the detection.
parameter $\frac{\theta}{2}$ to find $\hat{\theta}_1^*$, a MOM estimator of $\theta^*$, then apply Eq. (2.50) to get an estimator of $\theta$ as

$$
\hat{\theta}_{L,1} = 2\hat{\theta}_1^* \sqrt{\frac{(\sigma/2)^2 + \rho^2}{\sigma^2 + \rho^2}}. \tag{2.52}
$$

When we estimate $\theta$ using the MOM method, the larger the observed mean detected cluster size the larger the estimator of $\theta$. Since $k^*$ is usually less than $k$, $\bar{M}_1$ will be less than the true observed mean detected cluster size associated with $\frac{\theta}{2}$. Hence, $\hat{\theta}_1^*$ will systematically underestimate $\theta^*$. Therefore, we expect that $\hat{\theta}_{L,1}$ will underestimate $\theta$ too. Thus, more details of the relationship between $k$ and $k^*$ are needed. First, note that $k \sim \text{Pois}(\lambda a)$ and $k^* \sim \text{Pois}(\lambda a^*)$, where $a$ is defined in Eq. (2.34) and

$$
a^* = L \sqrt{2\pi((\sigma/2)^2 + \rho^2)T(\theta^*)}. \tag{2.53}
$$

Now we can say

$$
\frac{k}{k^*} \approx \frac{a}{a^*} = \frac{\sqrt{\sigma^2 + \rho^2} T(\theta)}{\sqrt{(\sigma/2)^2 + \rho^2} T(\theta^*)}. \tag{2.54}
$$

Since we do not know the value of $\frac{T(\theta^*)}{T(\theta^*)}$, a modification is necessary to provide a basis for estimation. Apply this relationship between $k$ and $k^*$, ignoring the factor $\frac{T(\theta)}{T(\theta^*)}$, to adjust $\bar{M}_1$ to

$$
\bar{M}_2 = \bar{M}_1 \frac{\sqrt{\sigma^2 + \rho^2}}{\sqrt{(\sigma/2)^2 + \rho^2}}. \tag{2.55}
$$

Treat $\bar{M}_2$ as the new observed mean detected cluster size to get $\hat{\theta}_2^*$, another approximate MOM estimator of $\theta^*$. We can construct another estimator of $\theta$, $\hat{\theta}_{L,2}$, from $\hat{\theta}_2^*$. The relationship between $\hat{\theta}_{L,2}$ and $\hat{\theta}_2^*$ is the same as that of $\hat{\theta}_{L,1}$ and $\hat{\theta}_1^*$ in Eq. (2.52).

Since $\frac{T(\theta)}{T(\theta^*)}$ is greater than 1 from the fact that $T(\theta)$ is an increasing function, then $\bar{M}_2$
obtained from ignoring $\frac{T(\theta)}{T(\theta^*)}$ is still less than the true observed value of mean detected cluster size for $\frac{\theta}{2}$. Hence, $\hat{\theta}_{L,2}$ may also underestimate $\theta$. From the relationship between $\theta^*$ and $\theta$ in Eq. (2.50), $\theta^*$ is slightly larger than half of $\theta$. We constructed a table of values of $\frac{T(\theta)}{T(\theta/2)}$ for $\theta$ between 20 and 40 in table 2.2. The values in this table may not be very accurate as we mentioned before that the evaluation of $T(\theta)$ for $\theta$ between 20 and 40 is not numerically accurate, but they at least give us some idea of the range of $\frac{T(\theta)}{T(\theta^*)}$. Since $T(\theta^*)$ is slightly larger than $T(\theta/2)$, and now from the values in the table which run between 1.12 and 1.22, we approximate $\frac{T(\theta)}{T(\theta^*)}$ as 1.1 to further adjust $\overline{M}_2$ as

$$\overline{M}_3 = 1.1 \overline{M}_2.$$  

As with $\overline{M}_1$ and $\overline{M}_2$, $\overline{M}_3$ produces the third estimator of $\theta$, $\hat{\theta}_{L,3}$.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>20</th>
<th>22</th>
<th>24</th>
<th>26</th>
<th>28</th>
<th>30</th>
<th>32</th>
<th>34</th>
<th>36</th>
<th>38</th>
<th>40</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{T(\theta)}{T(\theta/2)}$</td>
<td>1.12</td>
<td>1.14</td>
<td>1.15</td>
<td>1.17</td>
<td>1.18</td>
<td>1.19</td>
<td>1.20</td>
<td>1.21</td>
<td>1.22</td>
<td>1.20</td>
<td>1.14</td>
</tr>
</tbody>
</table>

Table 2.2 values of ratios of $T(\theta)$ over $T(\theta/2)$

A simulation study was conducted to explore the performance of the above three estimators of large $\theta$. We set $\mu$ to be 100, $\lambda$ to be 0.005, and pick five $\rho^2$ values as 4, 6, 9, 12 and 16, to result in five true $\theta$ values which run between 19.612 and 37.139. For each value of $\theta$, 100,000 realizations of the simplified Neyman-Scott process were simulated and the three estimators of $\theta$ were computed. Table 2.3 summarizes the simulation results. From the simulation, both $\hat{\theta}_{L,1}$ and $\hat{\theta}_{L,3}$ underestimate $\theta$ as we expected. The mean of $\hat{\theta}_{L,3}$ is very close to the true value of $\theta$ and its bias is much smaller than those of $\hat{\theta}_{L,1}$ and $\hat{\theta}_{L,2}$, especially when the true $\theta$ is larger. $\hat{\theta}_{L,3}$ has the largest standard deviations among these three estimators, while $\hat{\theta}_{L,1}$ has the smallest. $\hat{\theta}_{L,2}$ outperforms $\hat{\theta}_{L,1}$ across all values of $\theta$ in terms of MSE or bias, and $\hat{\theta}_{L,3}$ is the best estimator except that when $\theta = 19.612$. But for values of $\theta$ below 20, we can use unmodified estimates of $\theta$. In
conclusion, $\hat{\theta}_{L,3}$ is a competitive choice for estimating large $\theta$ (>$20$), where we cannot accurately evaluate $T(\theta)$ and $T^{(m)}(\theta)$.

<table>
<thead>
<tr>
<th>true $\theta$</th>
<th>estimator</th>
<th>mean</th>
<th>std</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>19.612</td>
<td>$\hat{\theta}_{L,1}$</td>
<td>18.085</td>
<td>2.223</td>
<td>7.274</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,2}$</td>
<td>18.418</td>
<td>2.259</td>
<td>6.530</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,3}$</td>
<td>20.726</td>
<td>2.511</td>
<td>7.548</td>
</tr>
<tr>
<td>22.502</td>
<td>$\hat{\theta}_{L,1}$</td>
<td>19.419</td>
<td>2.683</td>
<td>16.700</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,2}$</td>
<td>19.890</td>
<td>2.741</td>
<td>14.335</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,3}$</td>
<td>22.362</td>
<td>3.047</td>
<td>9.304</td>
</tr>
<tr>
<td>25.766</td>
<td>$\hat{\theta}_{L,1}$</td>
<td>21.671</td>
<td>3.274</td>
<td>27.487</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,2}$</td>
<td>22.359</td>
<td>3.368</td>
<td>22.954</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,3}$</td>
<td>25.105</td>
<td>3.745</td>
<td>14.461</td>
</tr>
<tr>
<td>31.046</td>
<td>$\hat{\theta}_{L,1}$</td>
<td>25.958</td>
<td>4.280</td>
<td>44.204</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,2}$</td>
<td>27.157</td>
<td>4.462</td>
<td>35.038</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,3}$</td>
<td>30.436</td>
<td>4.961</td>
<td>24.984</td>
</tr>
<tr>
<td>37.139</td>
<td>$\hat{\theta}_{L,1}$</td>
<td>30.967</td>
<td>5.579</td>
<td>69.222</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,2}$</td>
<td>33.039</td>
<td>5.928</td>
<td>51.951</td>
</tr>
<tr>
<td></td>
<td>$\hat{\theta}_{L,3}$</td>
<td>36.972</td>
<td>6.590</td>
<td>43.452</td>
</tr>
</tbody>
</table>

Table 2.3  Estimation of large $\theta$

For $\theta$ larger than 40, we could similarly use the above *ad hoc* method to count more than two objects as one to make $\theta^*$ less than 20, estimate $\theta^*$, then transform it back to estimate the original $\theta$. These modified estimators for $\theta$ larger than 20 are not required in cases covered in the simulations to follow.

2.5 Simulation

A simulation study was conducted to investigate the performance of the estimators of $\mu$, $\lambda$ and $\rho^2$, and their theoretical variance estimators, and to compare MOM and MLE estimators of $\theta$. Since the assumption that the pairwise correlations among the estimators of $\rho^2$, $\theta$ and $k$ are negligible is crucial in developing the variance estimators, these
pairwise correlations were also studied. Moreover, the number of undetected clusters was recorded.

In the simulation, the study region was fixed as a square of (0,100) x (0,100). \( \lambda \) was set at 0.005 and \( \mu \) at 50. We considered four values of \( \rho^2 \), 4, 9, 16 and 25. Five vertical line transects were evenly spaced on the study region, i.e., the transects were positioned at \( X = 10, 30, 50, 70 \) and 90. We set \( \sigma \), the parameter of the detection function, to the value 0.8, which makes the effective strip width \( 2w \) about 2. Since the total length of transects is 500, then the effective survey area is about 1000, which is about 10% of the total area of the study region.

For each value of \( \rho^2 \), a simplified Neyman-Scott process was generated according to the parameters. The detection process was executed and the observations were recorded. Based on the data, MOM and MLE of \( \theta \) were computed separately, and then estimators and variance estimators of the other parameters were obtained using the MOM or MLE of \( \theta \), respectively. We repeated the above process 10,000 times.

<table>
<thead>
<tr>
<th>( \rho^2 )</th>
<th>4</th>
<th>9</th>
<th>16</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Value</td>
<td>18.570</td>
<td>12.883</td>
<td>9.806</td>
<td>7.899</td>
</tr>
<tr>
<td>Bias</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOM</td>
<td>-0.267</td>
<td>-0.245</td>
<td>0.223</td>
<td>1.200</td>
</tr>
<tr>
<td>MLE</td>
<td>-0.509</td>
<td>-0.335</td>
<td>-0.227</td>
<td>0.115</td>
</tr>
<tr>
<td>Sample Std</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOM</td>
<td>3.212</td>
<td>1.852</td>
<td>1.262</td>
<td>0.964</td>
</tr>
<tr>
<td>MLE</td>
<td>1.501</td>
<td>1.408</td>
<td>1.016</td>
<td>0.775</td>
</tr>
<tr>
<td>Avg. Est. Std</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOM</td>
<td>3.160</td>
<td>1.842</td>
<td>1.330</td>
<td>1.169</td>
</tr>
<tr>
<td>MLE</td>
<td>1.380</td>
<td>1.343</td>
<td>1.000</td>
<td>0.808</td>
</tr>
<tr>
<td>Avg. Coverage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOM</td>
<td>0.9186</td>
<td>0.9250</td>
<td>0.9601</td>
<td>0.9252</td>
</tr>
<tr>
<td>MLE</td>
<td>0.6965</td>
<td>0.9151</td>
<td>0.9243</td>
<td>0.9655</td>
</tr>
</tbody>
</table>

Table 2.4 Estimate of \( \theta \)

Table 2.4 is the summary of the estimator and variance estimator of \( \theta \) across all four values of \( \rho^2 \). The row "Bias" represents the average deviation of estimators from the true value over the 10,000 iterations, the row "Avg. Est. Std" is the average
estimated standard deviation for the 10,000 iterations, the row “Sample Std” is the sample standard deviation of the 10,000 estimates, and the row “Avg. Coverage” is the average coverage of 95% confidence interval for the true value $\theta$, where the confidence interval is constructed by the estimate plus/minus 1.96 times of the estimated standard deviation in each realization. The rows with “MOM” are associated with the method of moments estimator of $\theta$ and the rows with “MLE” are for the maximum likelihood estimator. In “MLE”, its variance estimator is the inverse of observed Fisher information, which is defined in Eq. (2.47). From this table, we can see MLE estimates $\theta$ very well and MOM is reasonable except when $\theta = 7.899 (p^2 = 25)$. For MOM, the average estimated standard deviation is very close to the sample standard deviation, which suggests the variance estimator of $\theta$ defined in Eq. (2.23) asymptotically approaches the standard deviation of MOM of $\theta$. For MLE, the average estimated standard deviation is also close to the sample standard deviation. It appears that using the inverse of the observed Fisher information as the variance estimator of MLE works well in this context. The sample standard deviation for MLE of $\theta$ is smaller than that for MOM across all four values of $\theta$. Overall, MLE is a more efficient estimator than MOM, especially when $\theta$ is large. For the average coverage of the approximate 95% confidence interval, they are larger than 0.90 except for “MLE” when $p^2 = 4$, but a little off the target level 0.95. Perhaps, using the normal approximation to construct the confidence interval here is not appropriate, or the variance estimators have large variances. When $p^2 = 4$, we used the grid searching method to compute the MLE (Newton-Ralphson method doesn’t converge in this case) by pre-determining the scope of $\theta$. So both the sample and the estimated standard deviations of MLE of $\theta$ may be smaller than the true value of the standard deviation of MLE of $\theta$; this may contribute to the low average coverage (only 0.6965) for this case.

Table 2.5 summarizes the estimator and variance estimator of $p^2$. Because estimation of $p^2$ does not involve the estimator of $\theta$, there are no rows labeled “MOM” and “MLE” in
This table as in Table 2.4. For all cases of $\rho^2$, $\hat{\rho^2}$ and $Var(\hat{\rho^2})$ perform very well. It appears that $\rho^2$ slightly underestimates $\rho^2$ because all four biases are negative. Theoretically, $\hat{\rho^2}$ should be an unbiased estimator. This systematic underestimation may result from the edge effect. During the simulation, when a generated object is located outside of the $(0,100) \times (0,100)$ square, it is ignored. This will cause the true $\rho^2$ to effectively "shrink" a little. Therefore, underestimation is not very surprising when we consider the edge effect. When $\rho^2$ increases, the bias and average standard error also increase.

Generally speaking, the more spread the cluster (i.e. larger $\rho^2$), the harder it is to estimate the dispersion parameter. All average coverages of the 95% confidence interval (still a normal approximation as in the case of $\theta$) are smaller than the target level 0.95. The average coverage decreases when $\rho^2$ increases. Here $\hat{\rho^2}$ is $\chi^2$ distributed, using normal approximation clearly is not a good way to construct the confidence interval. Systematic underestimation resulting from the edge effect may also play a role in the undercoverage.

The results for the estimator of $\mu$ are summarized in Table 2.6. The biases are all very small except the scenario in which $\rho^2 = 25$ and MOM of $\theta$ was used, perhaps because the bias of the MOM estimator of $\theta$ is very large for this scenario. As in estimation of $\theta$, the variance estimators in "MOM" and "MLE" are reasonably close to the sample variances for each value of $\rho^2$. Unlike what was seen in estimating $\rho^2$, the sample standard deviation of $\mu$ in "MOM" decreases when $\rho^2$ increases, but doesn't
Table 2.6 Estimate of $\mu (\mu = 50)$

<table>
<thead>
<tr>
<th></th>
<th>$\rho^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Bias</td>
<td>MOM</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
</tr>
<tr>
<td>Sample Std</td>
<td>MOM</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
</tr>
<tr>
<td>Avg. Est. Std</td>
<td>MOM</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
</tr>
<tr>
<td>Avg. Coverage</td>
<td>MOM</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
</tr>
</tbody>
</table>

change very much in “MLE”. Overall, the “MLE” estimator outperforms the “MOM” estimator, following the pattern observed in the estimates of $\theta$ themselves. As in the cases of $\theta$ and $\rho^2$, the average coverages of normality-based 95% confidence interval are not realistic. The average coverages for “MOM” are greater than those for “MLE” across all four values of $\rho^2$.

Table 2.7 Estimate of $1000\lambda (\lambda = 0.005)$

<table>
<thead>
<tr>
<th></th>
<th>$\rho^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Bias</td>
<td>MOM</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
</tr>
<tr>
<td>Sample Std</td>
<td>MOM</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
</tr>
<tr>
<td>Avg. Est. Std</td>
<td>MOM</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
</tr>
<tr>
<td>Avg. Coverage</td>
<td>MOM</td>
</tr>
<tr>
<td></td>
<td>MLE</td>
</tr>
</tbody>
</table>

We tabulated the results of the estimate of $\lambda$ in Table 2.7. Since the magnitude of $\lambda$ is small, we recorded the results for $1000\lambda$. As with other parameters, the biases are very small relative to the parameter value, the average estimated standard deviations in “MOM” or “MLE” are very close to their corresponding sample standard deviations for
each value of $\rho^2$, and the sample standard deviation decreases when $\rho^2$ increases. Unlike
was seen in the estimation of $\theta$ and $\mu$, the sample standard deviations in “MLE” are larger
than their corresponding values for “MOM” estimators. The performances of “MLE”
and “MOM” here are comparable. For the average coverage of the normality-based
95% confidence interval, most of them are not realistic. The extremely low coverage for
“MOM” (0.7703) when $\rho^2 = 25$ might result from the large negative bias of $\lambda$.

$\begin{array}{cccc}
4 & 9 & 16 & 25 \\
\begin{array}{cc}
(\hat{\theta}, \hat{\rho^2}) & 0.0246 & 0.0195 & 0.0307 & 0.0360 \\
0.0352 & 0.0361 & 0.0338 & 0.0358 \\
(\hat{\theta}, k) & 0.0067 & -0.0127 & 0.0233 & 0.0077 \\
0.0643 & 0.0159 & 0.0345 & 0.0104 \\
(\hat{\rho^2}, k) & 0.0052 & 0.0159 & 0.0136 & -0.0172 \\
\end{array}
\end{array}$

Table 2.8 Correlations among estimators of $\theta$, $\rho^2$ and $k$

The pairwise correlations among the estimators of $\theta$, $\rho^2$ and $k$ are tabulated in Table 2.8. In the rows of the pairs of $(\hat{\theta}, \hat{\rho^2})$ and $(\hat{\theta}, k)$, each cell has two numbers; the top one is for estimates based on the MOM of $\theta$, and the bottom one is for estimates based on the MLE of $\theta$. Since $\hat{\rho^2}$ and $k$ do not involve the estimator of $\theta$, each cell of the row $(\hat{\rho^2}, k)$ has only one number. From this table, it is apparent that all the correlations are relatively small over the range of parameter values considered. This result lends some justification to our assumption that these pairwise correlations are negligible in the development of the variance estimators of $\hat{\mu}$ and $\hat{\lambda}$ in Eq. (2.32) and Eq. (2.41).

In conclusion, the estimators of parameters proposed in Section 2.3 perform well in most cases. The MLE of $\theta$ is more efficient than the MOM of $\theta$. This also holds true for the estimate of $\mu$. The pairwise correlations among $k$ and the estimators of $\theta$ and $\rho^2$ are indeed relatively small, and all the variance estimators match very well with their corresponding sample variances, but most of the normality-based 95% confidence intervals do not have idealistic coverages.
Table 2.9 Average Number of Undetected Clusters

<table>
<thead>
<tr>
<th>$\rho^2$</th>
<th>Avg. Num. of Undetected Clusters</th>
<th>Sample Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>22.01</td>
<td>4.69</td>
</tr>
<tr>
<td>9</td>
<td>12.22</td>
<td>3.51</td>
</tr>
<tr>
<td>16</td>
<td>4.89</td>
<td>2.21</td>
</tr>
<tr>
<td>25</td>
<td>1.92</td>
<td>1.37</td>
</tr>
</tbody>
</table>

In the next local sampling step, we want to survey around the detected clusters. So it is very important to detect as many clusters as possible. Table 2.9 records the average number of undetected clusters and its sample standard deviation for each value of $\rho^2$. Note that the average number of clusters in the study region is around 50, since the parameter $\lambda$ is 0.005 and the area of the study region is 10,000. When $\rho^2$ is small, the clusters do not spread widely, thus there will be a large number of undetected clusters. The more spread the cluster, the smaller the number of undetected clusters. In the case $\rho^2=25$, clusters are widely spread, there are on average only 1.92 undetected clusters out of about 50 total.

As to the problem of minimizing the number of undetected clusters, there is lengthy discussion in [10]. An algorithm for constructing an optimal sample path which minimizes the probability of missing all objects associated with any cluster was proposed there. Three types of information are required as input for this algorithm. The first type is the information characterizing the intensity of targets across the area to be screened, along with the likely size and number of objects included in a potential cluster located at any point. The second type is the information on the detection methodology to be used, in particular about the sensitivity of the method and the width of the path to be scanned along each transect. The third type is the information of an explicit list of linear transect segments that would be acceptable for inclusion in a sample path. Given the above information, the algorithm will provide an optimal sample path for any fixed length.
If the number of undetected clusters is a critical concern (as would obviously be the case in UXO sampling), estimates of $\rho^2$ based on initial global sampling might be used for calculations of the sort displayed in Table 2.9. If these suggest that an unacceptable number of undetected clusters may remain, additional transects might be added to the initial set of transects (e.g., evenly spaced between the initial set) to increase sampling coverage and reduce the likely number of undiscovered clusters.
CHAPTER 3 LOCAL SAMPLING

3.1 Introduction

Using the data collected in the global sampling, we detect clusters over the study region, and estimate intensity, mean cluster size and cluster dispersion. This gives us an overall picture of how the UXO are distributed. But our ultimate objective of defining an appropriate area to dig out all the UXO remains unsolved. We approach this goal by defining an inclusive region corresponding to a probability statement for UXO in each individual cluster. Even under the assumption that a bivariate normal distribution is an adequate model for the location of UXO within a cluster, we still need information about the center and spread of the cluster to obtain any probability statement. From estimates in the global sampling, we have a rough idea where the clusters are located, at least for one coordinate of the center. But due to the spatial sparsity of data in the global sampling, the estimate of the center is usually very inaccurate. Moreover, although we assume all the clusters have radially symmetric bivariate normal distributions, their dispersions may actually vary substantially. The global sampling estimation is based on a simpler model in which the dispersion characteristics of each cluster are assumed to be the same, resulting in only the pooled estimate of the dispersion. All the above concerns point to a need for more observations from each of the identified clusters, i.e., we need to perform additional local sampling.

The main objective of the local sampling is to infer a region for the cluster such that the probability that all the unobserved objects in the cluster are inside this region is $1 - \alpha$. 
We call this kind of region a 100 (1 — α) % inclusion region. The concept of the inclusion region is different from the traditional confidence region, prediction region or tolerance region [34]. When referring to 100 (1 — α) % confidence region, we mean the repeated sampling frequentist probability of (μₓ, μᵧ) being inside the region is 100 (1 — α) %. For a prediction region, we mean the probability of any one unobserved object being inside the region is 100 (1 — α) %. The inclusion region is more similar to the tolerance region. The difference between these two regions is that the former will cover all the N objects actually in the cluster while the latter will cover a specified 100 (1 — γ) % of objects that would be in the cluster as N → ∞, both with a 100 (1 — α) % probability. Among the four regions, the inclusion region is the largest and the confidence region is the smallest.

3.2 Notation

Suppose there are a total of N objects in the cluster, and that the locations of these objects are independently and identically distributed with a radially symmetric normal density with variance ρ² about the center (μₓ, μᵧ).

As was the case in the global sampling, we will use a line transect survey to collect data, in which we only observe the position of any detected object along the transect line, while the perpendicular distance is unobservable. We will again assume a Gaussian detection function \( g(r) = e^{-\frac{r^2}{2\sigma^2}} \) with a known \( \sigma \), where \( r \) is the perpendicular distance from the object to the transect. In order to determine the inclusion region, we first need to estimate the cluster center (μₓ, μᵧ). Since parallel line transects in one direction only support reliable estimation of one coordinate of the center, we will adopt a sampling plan using line transects in two perpendicular directions, say “vertical” for North-South and “horizontal” for East-West.

Let \( X_{0i}, i = 1, 2, \cdots, mₓ \), be the fixed horizontal positions of vertical line transects and \( Y_{0j}, j = 1, 2, \cdots, mᵧ \), the fixed vertical positions of horizontal line transects. Denote
by $m_x$ and $m_y$ the number of vertical and horizontal transects, respectively, used in a local survey of one cluster. Let $y_{ik}, k = 1, 2, \cdots, K_i, i = 1, 2, \cdots, m_x$, be the $k^{th}$ $y$-coordinate detected along the $i^{th}$ vertical transect; and let $x_{jl}, l = 1, 2, \cdots, L_j, j = 1, 2, \cdots, m_y$, be the $l^{th}$ $x$-coordinate detected along the $j^{th}$ horizontal transect. Let $n_v = \sum_{i=1}^{m_x} K_i$ and $n_h = \sum_{j=1}^{m_y} L_j$ be the number of objects detected along vertical and horizontal transects, respectively. Here, we assume there are no objects detected twice on different parallel transects, i.e., that the distance between neighboring parallel transects is large relative to the detection function parameter $\sigma$. Denote by $n = n_v + n_h$, the total number of detected objects from both directional transects. We allow for the possibility that some objects may be detected twice along intersecting transects, i.e., the $n$ detected objects may not all be distinct.

Let $\bar{x} = \frac{1}{n_h} \sum \sum x_{jl}$ be the average $x$-coordinate of objects detected along the horizontal transects and $\bar{y} = \frac{1}{n_v} \sum \sum y_{ik}$, the average $y$-coordinate corresponding to objects detected along the vertical transects. $(\bar{x}, \bar{y})$ is the natural estimator of $(\mu_x, \mu_y)$. Define $S_x^2 = \frac{1}{n_h-1} \sum \sum (x_{jl} - \bar{x})^2$, the corrected sum of squares for $x_{jl}$'s; $S_y^2 = \frac{1}{n_v-1} \sum \sum (y_{ik} - \bar{y})^2$ for $y_{ik}$'s, and $S^2 = \frac{1}{n-2} (\sum \sum (x_{jl} - \bar{x})^2 + \sum \sum (y_{ik} - \bar{y})^2)$, the pooled sum of squares for all observed $x$'s and $y$'s. Notice that $S^2$ is a natural estimator of the cluster dispersion parameter $\rho^2$.

Let $P_{x_i}$ be the probability that any unspecified object from this cluster is detected along the vertical line transect at $X_{0i}$; Similarly as $P_{c,x_0}$ in Eq. 2.18 of Chapter 2, we know $P_{x_i} = \sigma (\sigma^2 + \rho^2)^{-1/2} \exp \left( -\frac{(X_{0i} - \mu_x)^2}{2(\sigma^2 + \rho^2)} \right)$. Define $P_v = \sum_{i=1}^{m_x} P_{x_i}$, then $P_v$ is the approximate probability that any object is detected along any of the vertical transects if we assume the values of the $X_{0i}$'s are well-separated relative to the magnitude of $\sigma$. Similarly, we have $P_{y_j} = \sigma (\sigma^2 + \rho^2)^{-1/2} \exp \left( -\frac{(Y_{0j} - \mu_y)^2}{2(\sigma^2 + \rho^2)} \right)$ and $P_h = \sum_{j=1}^{m_y} P_{y_j}$, the approximate probability that an object is detected along any of the horizontal transects.

We condition our argument on each of $n_h$ and $n_v$ being greater than 1; then $n_h$ is distributed as a truncated Bin $(N, P_h)$ from below 2 and $n_v$ is a truncated Bin $(N, P_v)$.
from below 2. In reality, if we observe none or just one object along either set of parallel transects (cases with \( n_h \leq 1 \) or \( n_v \leq 1 \)), we might add additional transects in one or both directions to increase the number of detected objects. Let \( N_u \) be the number of undetected objects in this cluster and denote by \( n_{rec} \) the number of "recaptured" objects, i.e., those detected twice, once along a vertical transect and once along a horizontal transect, then we have the relationship \( N_u = N - n + n_{rec} \). Since the horizontal and vertical components of an object's location are independent under our model, \( P_u = 1 - P_h - P_v + P_hP_v \) is the probability that any unspecified object is not detected at all; then \( N_u \sim Bin(N, P_u) \).

Let \( \hat{P}_{x_i} = \sigma (\sigma^2 + S^2)^{-1/2} \exp \left( -\frac{(x_{0i} - \bar{x})^2}{2(\sigma^2 + S^2)} \right) \) be an estimator of \( P_{x_i} \), and \( \hat{P}_y = \sum_{i=1}^{m_y} \hat{P}_{x_i} \), an estimator of \( P_y \); similarly, we have \( \hat{P}_{y_j} = \sigma (\sigma^2 + S^2)^{-1/2} \exp \left( -\frac{(y_{0j} - \bar{y})^2}{2(\sigma^2 + S^2)} \right) \), \( \hat{P}_h = \sum_{j=1}^{m_h} \hat{P}_{y_j} \), and \( \hat{P}_u = 1 - \hat{P}_h - \hat{P}_v + \hat{P}_h\hat{P}_v. \) Since both \( \frac{n_h}{\hat{P}_h} \) and \( \frac{n_v}{\hat{P}_v} \) are estimators of \( N \), we average them to get \( \hat{N} = \frac{1}{2} \left( \frac{n_h}{\hat{P}_h} + \frac{n_v}{\hat{P}_v} \right) \).

Since \( N_u = N - n + n_{rec} \) and we already know \( n \) and an estimator of \( N \), then an exact or estimated value of \( n_{rec} \) would allow us to estimate \( N_u \). In our case, because we don’t have the information of the exact position of the detected objects, we don’t know with certainty which distinct objects have been detected more than once, so we need to estimate \( n_{rec} \). There are many possible estimators of \( n_{rec} \); we briefly describe three of them. First, consider all the \( x \)-coordinates observed along the horizontal line transects, \( x_{ji}, l = 1, 2, \ldots, L, j = 1, 2, \ldots, m_y; \) if an \( x_{jl} \) falls inside any of the intervals centered at a vertical line transect position \( X_{0i}, i = 1, 2, \ldots, m_x \), with the effective strip half-width \( \omega \) defined in Eq. 2.43 as the radius, then we label it as a recapture. The total number of all \( x_{jl} \) which are labeled as recaptures is an estimator of \( n_{rec} \). The logic of this estimator is that an object detected near the intersection of horizontal and vertical transects is close to both, and so is likely to be detected twice. Similarly, the total number of all \( y_{ik} \) which are labeled as recaptures, i.e., those detected near intersecting horizontal transects, is also an estimator of \( n_{rec} \). Averaging these two estimators provides a new one, which we
call \( \hat{n}_{\text{rec},1} \). More precisely,

\[
\hat{n}_{\text{rec},1} = \frac{1}{2} \left( \sum_{j,l,i} I_{|x_{jl} - x_{0i}| \leq \omega} + \sum_{i,k,j} I_{|y_{ik} - y_{0j}| \leq \omega} \right),
\]

where \( I \) is the indication function. Secondly, due to the independence of the two components of each object location, each of \( n_h \) objects detected along a horizontal line transect will also be detected along one of the vertical line transects with probability \( P_v \). That is to say, for \( n_h \) horizontally detected objects, \( n_h \hat{P}_v \) is an estimate of the number of recaptured objects. Similarly, for \( n_v \) vertically detected objects, \( n_v \hat{P}_h \) is also an estimate of \( n_{\text{rec}} \). Thus, the average of these two estimators, \( \frac{1}{2} (n_h \hat{P}_v + n_v \hat{P}_h) \) is a pooled estimator of \( n_{\text{rec}} \); call it \( \hat{n}_{\text{rec},2} \). Thirdly, for each \( x_{jl}, l = 1, 2, \cdots, L_j, j = 1, 2, \cdots, m_y \), detected along a horizontal transect, the vertical transect along which it is most likely redetected is the closest one. Let \( P_{jl} \) be the detection probability associated with the single vertical transect closest to \( x_{jl} \), then \( \sum_{jl} \hat{P}_{jl} \) is an estimator of the number of all \( n_h \) horizontally detected objects also detected vertically, i.e., \( n_{\text{rec}} \). Similarly, we can denote by \( P_{ik} \) the detection probability of the closest horizontal transect to each vertically detected object \( y_{ik} \), and have \( \sum_{i,k} \hat{P}_{ik} \) as an estimator of \( n_{\text{rec}} \). Therefore, \( \frac{1}{2} (\sum_{jl} \hat{P}_{jl} + \sum_{i,k} \hat{P}_{ik}) \) is also an estimator of \( n_{\text{rec}} \); call it \( \hat{n}_{\text{rec},3} \).

We have compared these three estimators of \( n_{\text{rec}} \) through simulation studies, and found that \( \hat{n}_{\text{rec},1} \) has a large negative bias, although it has a smallest variation among these three estimators. The variances of \( \hat{n}_{\text{rec},2} \) and \( \hat{n}_{\text{rec},3} \) are very similar, and \( \hat{n}_{\text{rec},3} \) has slightly smaller bias than \( \hat{n}_{\text{rec},2} \). This finding is not difficult to explain, because \( \hat{n}_{\text{rec},3} \) makes use of the one-coordinate position of the observed objects while \( \hat{n}_{\text{rec},2} \) only uses the information of total number of observed objects. But overall, the performance of \( \hat{n}_{\text{rec},2} \) and \( \hat{n}_{\text{rec},3} \) is very similar. Since \( \hat{n}_{\text{rec},2} \) has a more convenient mathematical form, we adopt it as our estimator of \( n_{\text{rec}} \). By using \( \hat{N}_{\text{rec}} = \hat{n}_{\text{rec},2} \), now we can define

\[
\hat{N}_u = \bar{N} - n + \hat{N}_{\text{rec}}.
\]
3.3 Methodology for inclusion region

Let \((x, y)\) be the location of any unobserved object, then the probability density function of \((x, y)\) is

\[
 f(x, y) = \frac{1}{P_a} \left( 1 - e^{-\frac{(x-x_0)^2}{2\sigma^2}} \right) \left( 1 - e^{-\frac{(y-y_0)^2}{2\sigma^2}} \right) \frac{1}{2\pi\rho^2} e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{2\rho^2}}.
\]  

(3.2)

Note that the detection process changes the \(x\)’s marginal distribution and \(x\) is not distributed exactly as \(N(\mu_x, \rho^2)\) any more. Now the distribution of \(x\) has the general shape of \(N(\mu_x, \rho^2)\) but with reduced probability density inside the strips centered at the positions of vertical transect lines. If \(P_a\) is small and \(\sigma\) is small relative to \(\rho\), then \(x\) is still approximately distributed as \(N(\mu_x, \rho^2)\). For a given \(n_h\), the independence between \(\bar{x}\) based on detected objects and an undetected \(x\), and the fact that \(\bar{x} \sim N(\mu_x, \rho^2/n_h)\) and \(x \sim N(\mu_x, \rho^2)\) approximately, imply that \(\frac{\bar{x} - x}{\sqrt{1 + 1/n_h \rho}}\) has an approximate standard normal distribution. Thus, \(\left( \frac{\bar{x} - x}{\sqrt{1 + 1/n_h \rho}} \right)^2 \sim \chi^2_1\) for given \(n_h\). The same argument leads to \(\left( \frac{\bar{y} - y}{\sqrt{1 + 1/n_u \rho}} \right)^2 \sim \chi^2_1\) for given \(n_u\). Now we want to estimate \(\rho\) with the pooled sample standard deviation \(S = \sqrt{S^2}\), where \(S^2\) is defined in Section 3.2. Since \(\frac{(n_h + n_u - 2)S^2}{\rho^2} \sim \chi^2_{n_h + n_u - 2}\), and all three of these \(\chi^2\) variates are independent, we have

\[
 \frac{1}{2} \left[ \left( \frac{\bar{x} - x}{\sqrt{1 + 1/n_h \rho}} \right)^2 + \left( \frac{\bar{y} - y}{\sqrt{1 + 1/n_u \rho}} \right)^2 \right] = \frac{1}{n_h + n_u - 2} \left( \frac{(n_h + n_u - 2)S^2}{\rho^2} \right) \sim F_{2, n_h + n_u - 2}.
\]

(3.3)

for given \(n_h\) and \(n_u\); simplifying Eq. (3.3), we have

\[
 \left( \frac{\bar{x} - x}{\sqrt{1 + 1/n_h S}} \right)^2 + \left( \frac{\bar{y} - y}{\sqrt{1 + 1/n_u S}} \right)^2 \sim 2F_{2, n_h + n_u - 2}.
\]

(3.4)

Denote by \(C((x_0, y_0), r_x, r_y)\) the ellipse centered at \((x_0, y_0)\), with one of its axes
parallel to the X-axis with radius $r_x$ and the other axis parallel to the Y-axis with radius $r_y$, i.e.,

$$C((x_0, y_0), r_x, r_y) = \left\{ (x, y) \mid \left(\frac{x - x_0}{r_x}\right)^2 + \left(\frac{y - y_0}{r_y}\right)^2 \leq 1 \right\}. \quad (3.5)$$

For any $0 < \beta < 1$, let $F_{2,n_h+n_v-2}(\beta)$ be the $100\beta\%$ quantile of $F_{2,n_h+n_v-2}$, then

$$\Pr(\text{any unobserved } (x, y) \text{ is inside the ellipse } C((\bar{x}, \bar{y}), r_x, r_y) | n_h, n_v) = \beta, \quad (3.6)$$

where

$$r_{x,\beta} = \sqrt{1 + 1/n_h \sqrt{2F_{2,n_h+n_v-2}(\beta)}} \quad (3.7)$$

and

$$r_{y,\beta} = \sqrt{1 + 1/n_v \sqrt{2F_{2,n_h+n_v-2}(\beta)}}. \quad (3.8)$$

Now consider two unobserved objects $(x_1, y_1), (x_2, y_2)$ and two events,

$$\begin{cases} E_1: (x_1, y_1) \in C((\bar{x}, \bar{y}), r_{x,\beta}, r_{y,\beta}) \\ E_2: (x_2, y_2) \in C((\bar{x}, \bar{y}), r_{x,\beta}, r_{y,\beta}) \end{cases} \quad (3.9)$$

Apparently, both $\Pr(E_1|n_h, n_v) = \beta$ and $\Pr(E_2|n_h, n_v) = \beta$ hold. Given $n_h$ and $n_v$, $E_1$ and $E_2$ are not independent because both of them are defined in part by common random variables $\bar{x}, \bar{y}$ and $S$, although $(x_1, y_1)$ and $(x_2, y_2)$ themselves are independent. But when $n_h$ and $n_v$ are moderately large, most of the randomness in $E_1$ and $E_2$ is associated with $(x_1, y_1)$ and $(x_2, y_2)$. So $E_1$ and $E_2$ are approximately independent for large enough $n_h$ and $n_v$, hence the probability of $(x_1, y_1)$ and $(x_2, y_2)$ being inside the ellipse $C((\bar{x}, \bar{y}), r_{x,\beta}, r_{y,\beta})$ simultaneously in this case is approximately $\beta^2$. We can generalize the above reasoning to $m$ unobserved objects, therefore the probability of all of $m$ unobserved objects being inside the above defined ellipse simultaneously is approximately $\beta^m$ for given large $n_h$ and $n_v$. 
The probability that all of \( N_u \) unobserved objects are inside \( C((x, y), r_x, r_y) \) is approximately \( \beta^N \). For \( 0 < \alpha < 1 \), if \( N_u \) were known, we could choose \( \beta = (1 - \alpha)^{1/N_u} \), and have

\[
\Pr(\text{all } N_u \text{ unobserved objects are inside } C((x, y), r_x, r_y) \mid n_h, n_v) = 1 - \alpha. \tag{3.10}
\]

Since this conditional probability statement is true for any \( n_h \) and \( n_v \), then the unconditional probability is the same, i.e.,

\[
\Pr(\text{all } N_u \text{ unobserved objects are inside } C((x, y), r_x, r_y)) = 1 - \alpha. \tag{3.11}
\]

Because \( \beta = (1 - \alpha)^{1/N_u} \) involves \( N_u \), which is unknown to us, we need to replace it with its estimator. Now we define a \( 100(1 - \alpha)\% \) inclusion region \( IR = C((x, y), r_x, r_y) \), where

\[
r_x = \sqrt{1 + 1/n_h} \sqrt{\frac{2F_{2,n_h+n_v-2}((1 - \alpha)^{1/N_u})}{S}} \tag{3.12}
\]

and

\[
r_y = \sqrt{1 + 1/n_v} \sqrt{\frac{2F_{2,n_h+n_v-2}((1 - \alpha)^{1/N_u})}{S}}, \tag{3.13}
\]

namely,

\[
IR = \left\{ (x, y) \left| \frac{(x - \bar{x})^2}{1 + \frac{1}{n_h}} + \frac{(y - \bar{y})^2}{1 + \frac{1}{n_v}} \leq 2S^2F_{2,n_h+n_v-2}((1 - \alpha)^{1/N_u}) \right. \right\}. \tag{3.14}
\]

From the above reasoning, we know that \( IR \) will contain all \( N_u \) unobserved objects in the cluster with probability approximately \( 100(1 - \alpha)\% \) depending on the approximation used and the quality of estimates used in place of unknown quantities. In the next section, we will formally state and prove the asymptotic coverage property of \( IR \) when the cluster size \( N \) goes to \( \infty \).
3.4 Asymptotic coverage property of the inclusion region

Define the region $D_1$ as a ball $C \left( (\mu_x, \mu_y), \sqrt{\frac{2}{\chi_2^2((1-\alpha)^{1/N_h})}} \rho, \sqrt{\frac{2}{\chi_2^2((1-\alpha)^{1/N_u})}} \rho \right)$, namely,

$$D_1 = \left\{ (x, y) | (x - \mu_x)^2 + (y - \mu_y)^2 \leq \chi_2^2 \left( (1-\alpha)^{1/N_h} \right) \rho^2 \right\}. \tag{3.15}$$

Note that $D_1$ is almost a fixed region except that $N_u$ is a random quantity. First, we will show that our proposed inclusion region $IR$ in the above section is asymptotically identical to $D_1$ in the sense of convergence in probability as the cluster size $N \to \infty$. Then, a lower bound of the probability that $D_1$ covers all $N_u$ unobserved objects will be given. The discussion of coverage probability of the inclusion region will be followed.

**Lemma 1.** There exists $M_1$, such that when $N > M_1$, we have

$$\Pr \left( |n_h - NP_h| > \frac{1}{2} P_h N^{\frac{1}{3} + \alpha} \right) < \left( \frac{16 P_h}{P_h} + 1 \right) \frac{1}{N^{2\alpha}} \tag{3.16}$$

for any $\alpha$, where $Q_h = 1 - P_h$.

**Pf:** Since $n_h \sim \text{truncated } Bin(N, P_h)$ below 2, then

$$E(n_h) = \sum_{k=2}^{N} k \frac{\binom{N}{k} P_h^k Q_h^{N-k}}{1 - Q_h^N - NP_h Q_h^{N-1}}$$

$$= \frac{NP_h (1 - Q_h^{N-1})}{1 - Q_h^N - NP_h Q_h^{N-1}} \tag{3.17}$$

and

$$\text{Var}(n_h) = \frac{NP_h Q_h \left( 1 - Q_h^{N-2} - Q_h^{N} + Q_h^{2N} \right) + N^2 P_h^2 Q_h^{N-1} \left( 2 - NP_h - Q_h - Q_h^{N} \right)}{(1 - Q_h^N - NP_h Q_h^{N-1})^2}. \tag{3.18}$$

Note that the order of $E(n_h)$ and $\text{Var}(n_h)$ is $NP_h$ and $NP_h Q_h$, respectively; the same as those of the untruncated binomial distribution.

Since $\lim_{N \to \infty} \frac{E(n_h) - NP_h}{\frac{1}{2} P_h N^{1/2 + \alpha}} = \lim_{N \to \infty} \frac{NP_h Q_h^{N-1} \left( 1 + Q_h + NP_h \right)}{\frac{1}{2} P_h N^{1/2 + \alpha} \left( 1 - Q_h^N - NP_h Q_h^{N-1} \right)} = 0$ for any $\alpha$, then there
exists $M_{11}$ such that when $N > M_{11}$,

\[
\left| \frac{E(n_h) - NP_h}{\frac{1}{4} P_h N^{1/2+\alpha}} \right| < 1, \quad (3.19)
\]

i.e.,

\[
|E(n_h) - NP_h| < \frac{1}{4} P_h N^{1/2+\alpha}. \quad (3.20)
\]

Since

\[
\lim_{N \to \infty} \frac{\text{Var}(n_h)}{\left( \frac{1}{4} P_h N^{1/2+\alpha} \right)^2} = \lim_{N \to \infty} \frac{16 NP_h Q_h N^{2\alpha}}{P_h N^{4+2\alpha}} = \frac{16 Q_h}{P_h}, \quad \text{here we used the fact that}
\]

\[
\lim_{N \to \infty} \frac{\text{Var}(n_h)}{NP_h Q_h} = 1, \quad \text{there exists } M_{12} \text{ such that when } N > M_{12},
\]

\[
\left| \frac{\text{Var}(n_h)}{\left( \frac{1}{4} P_h N^{1/2+\alpha} \right)^2} \right| < \frac{16 Q_h}{P_h} + 1, \quad (3.21)
\]

i.e.,

\[
\frac{\text{Var}(n_h)}{\left( \frac{1}{4} P_h N^{1/2+\alpha} \right)^2} < \left( \frac{16 Q_h}{P_h} + 1 \right) \frac{1}{N^{2\alpha}}. \quad (3.22)
\]

Now let $M_1 = \max(M_{11}, M_{12})$, then when $N > M_1$, we have

\[
\Pr \left( |n_h - NP_h| > \frac{1}{2} P_h N^{\frac{1}{2}+\alpha} \right) = \Pr \left( |n_h - E(n_h) + E(n_h) - NP_h| > \frac{1}{2} P_h N^{\frac{1}{2}+\alpha} \right)
\]

\[
\leq \Pr \left( |n_h - E(n_h)| + |E(n_h) - NP_h| > \frac{1}{2} P_h N^{\frac{1}{2}+\alpha} \right)
\]

(by Eq. (3.20)) \leq \Pr \left( |n_h - E(n_h)| + \frac{1}{4} P_h N^{\frac{1}{2}+\alpha} > \frac{1}{2} P_h N^{\frac{1}{2}+\alpha} \right)

= \Pr \left( |n_h - E(n_h)| > \frac{1}{4} P_h N^{\frac{1}{2}+\alpha} \right)

(by Tchebysheff's Theorem) \leq \frac{\text{Var}(n_h)}{\left( \frac{1}{4} P_h N^{1/2+\alpha} \right)^2}

(by Eq. (3.22)) < \left( \frac{16 Q_h}{P_h} + 1 \right) \frac{1}{N^{2\alpha}}.

This completes our proof.
Theorem 1. $\bar{x}$ is a consistent estimator of $\mu_x$, i.e., $\forall \varepsilon > 0$,

$$\lim_{N \to \infty} \Pr (|\bar{x} - \mu_x| > \varepsilon) = 0.$$  \hspace{1cm} (3.23)

In fact, we'll show that there exists $M_1$, such that when $N > M_1$, for any $0 < \alpha < \frac{1}{2}$,

$$\Pr \left( |\bar{x} - \mu_x| > \frac{1}{N^{\frac{1}{2} - \alpha}} \right) \leq \frac{\left( \frac{2 + 8Q_h}{P_h} + \frac{1}{2} \right) \rho^2}{N^{2\alpha}}.$$  \hspace{1cm} (3.24)

Note that this latter statement is stronger than the consistency.

**Pf:** Since $E(\bar{x}) = E(E(\bar{x}|n_h)) = E(\mu_x) = \mu_x$ and $Var(\bar{x}) = E(Var(\bar{x}|n_h)) + Var(E(\bar{x}|n_h)) = \rho^2 E \left( \frac{1}{n_h} \right)$, then by Tchebysheff’s Theorem,

$$\Pr \left( |\bar{x} - \mu_x| > \frac{1}{N^{\frac{1}{2} - \alpha}} \right) \leq \frac{Var(\bar{x})}{\left( \frac{1}{N^{\frac{1}{2} - \alpha}} \right)^2} = \rho^2 N^{1 - 2\alpha} E \left( \frac{1}{n_h} \right).$$  \hspace{1cm} (3.25)

Next we want to control $E \left( \frac{1}{n_h} \right)$. Using Lemma 1, choose $\alpha = \frac{1}{2}$, then there exists $M_1$ such that when $N > M_1$,

$$\Pr \left( |n_h - NP_h| > \frac{1}{2} NP_h \right) < \frac{16Q_h + 1}{N}.$$  \hspace{1cm} (3.26)

In the following, we also make use of the decomposition:

$$E \left( \frac{1}{n_h} \right) = E \left( \frac{1}{n_h} I_{|n_h - NP_h| > \frac{1}{2} NP_h} \right) + E \left( \frac{1}{n_h} I_{|n_h - NP_h| \leq \frac{1}{2} NP_h} \right).$$  \hspace{1cm} (3.27)
By the fact \( n_h \geq 2 \) and Eq. (3.26), we have

\[
E \left( \frac{1}{n_h} I_{|n_h - NP_h| > 2 NP_h} \right) \leq \frac{1}{2} \Pr \left( |n_h - NP_h| > \frac{1}{2} NP_h \right) < \frac{8Q_h + \frac{1}{2}}{2 NP_h}.
\]  

(3.28)

When \( |n_h - NP_h| \leq \frac{1}{2} NP_h \), we have \( \frac{1}{2} NP_h \leq n_h \leq \frac{3}{2} NP_h \), then

\[
E \left( \frac{1}{n_h} I_{|n_h - NP_h| \leq \frac{1}{2} NP_h} \right) \leq \frac{2}{NP_h} E \left( I_{|n_h - NP_h| \leq \frac{1}{2} NP_h} \right) \leq \frac{2}{NP_h}.
\]  

(3.29)

Combining Eq. (3.25), Eq. (3.27), Eq. (3.28) and Eq. (3.29) leads to the completion of the proof, i.e.,

\[
\Pr \left( |\bar{x} - \mu_x| > \frac{1}{N^{\frac{1}{2} - \alpha}} \right) < \rho^2 N^{1 - 2\alpha} \left( \frac{8Q_h + \frac{1}{2}}{2 NP_h} \right) \frac{2}{NP_h} = \frac{\left( \frac{2 + 8Q_h}{P_h} + \frac{1}{2} \right) \rho^2}{N^{2\alpha}}.
\]

Following the structure of Theorem 1, an analogous statement can be made about \( \bar{y} \) as an estimator of \( \mu_y \):

**Theorem 2.** There exists \( M_2 \) such that when \( N > M_1 \), for any \( 0 < \alpha < \frac{1}{2} \),

\[
\Pr \left( |\bar{y} - \mu_y| > \frac{1}{N^{\frac{1}{2} - \alpha}} \right) < \frac{\left( \frac{2 + 8Q_y}{P_y} + \frac{1}{2} \right) \rho^2}{N^{2\alpha}}
\]  

(3.30)

and \( \bar{y} \) is a consistent estimator of \( \mu_y \).
**Lemma 2.** There exists $M_3$ such that when $N > M_3$,

$$E\left( \frac{1}{n_h + n_v - 2} \right) < \frac{2}{N(P_h + P_v) - 4} + \frac{8 \left( \frac{Q_h}{P_h} + \frac{Q_v}{P_v} \right) + 1}{N}. \quad (3.31)$$

**Pf:** Similar to the argument at Eq. (3.26), there exists $M_2$ such that when $N > M_2$,

$$\Pr \left( |n_v - NP_v| > \frac{1}{2} NP_v \right) < \frac{16 Q_v}{P_v} + \frac{1}{N}. \quad (3.32)$$

Also note that $E\left( \frac{1}{n_h + n_v - 2} \right)$ can be decomposed as:

$$E\left( \frac{1}{n_h + n_v - 2} \right) = E\left( \frac{1}{n_h + n_v - 2} I_{|n_h - NP_h| \leq \frac{1}{2} NP_h \cap |n_v - NP_v| \leq \frac{1}{2} NP_v} \right) + E\left( \frac{1}{n_h + n_v - 2} I_{|n_h - NP_h| > \frac{1}{2} NP_h \cup |n_v - NP_v| > \frac{1}{2} NP_v} \right). \quad (3.33)$$

When $|n_h - NP_h| \leq \frac{1}{2} NP_h$ and $|n_v - NP_v| \leq \frac{1}{2} NP_v$, then $n_h + n_v - 2 \geq \frac{1}{2} N(P_h + P_v) - 2$, thus

$$E\left( \frac{1}{n_h + n_v - 2} I_{|n_h - NP_h| \leq \frac{1}{2} NP_h \cap |n_v - NP_v| \leq \frac{1}{2} NP_v} \right) \leq \frac{2}{N(P_h + P_v) - 4}. \quad (3.34)$$

Choose $M_3 = \max(M_1, M_2)$, where $M_1$ is as discussed in the proof of Lemma 1 and Theorem 1. When $N > M_3$, then by the fact $n_h + n_v - 2 \geq 2$ and by Eq. (3.26) and Eq. (3.32), we have

$$E\left( \frac{1}{n_h + n_v - 2} I_{|n_h - NP_h| > \frac{1}{2} NP_h \cup |n_v - NP_v| > \frac{1}{2} NP_v} \right) \leq \frac{1}{2} \left( \Pr \left( |n_h - NP_h| > \frac{1}{2} NP_h \right) + \Pr \left( |n_v - NP_v| > \frac{1}{2} NP_v \right) \right) + \frac{8 \left( \frac{Q_h}{P_h} + \frac{Q_v}{P_v} \right) + 1}{N}. \quad (3.35)$$
The proof is completed by combining Eq. (3.33), Eq. (3.34) and Eq. (3.35).

**Theorem 3.** $S^2$ is a consistent estimator of $\rho^2$. In fact, we will show that there exists $M_3$ such that when $N > M_3$, for any $0 < \alpha < \frac{1}{2}$,

$$\text{Pr}\left( |S^2 - \rho^2| > \frac{1}{N^{1/2-\alpha}} \right) < 2\rho^4 \left( \frac{2}{(P_h + P_v)N^{2\alpha} - 4N^{2\alpha - 1}} + \frac{8(Q_h + Q_v)}{N^{2\alpha}} + 1 \right).$$

(3.36)

As in **Theorem 1 and 2** concerning $\bar{x}$ and $\bar{y}$, the latter statement is stronger than consistency.

**Pf:** Since $\frac{n_h + n_v - 2}{\rho^2}S^2 \mid n_h, n_v \sim X^2_{n_h + n_v - 2}$, then $E(S^2) = E(E(S^2 \mid n_h, n_v)) = \rho^2$, $\text{Var}(S^2 \mid n_h, n_v) = \frac{2\rho^4}{n_h + n_v - 2}$, and $\text{Var}(S^2) = 2\rho^4E\left( \frac{1}{n_h + n_v - 2} \right)$. Now by Tchebysheff's Theorem,

$$\text{Pr}\left( |S^2 - \rho^2| > \frac{1}{N^{1/2-\alpha}} \right) \leq \frac{\text{Var}(S^2)}{\left( \frac{1}{N^{1/2-\alpha}} \right)^2} = 2\rho^4N^{1-2\alpha}E\left( \frac{1}{n_h + n_v - 2} \right)$$

(3.37)

Directly applying **Lemma 2** to Eq. (3.37) results in Eq. (3.36), which completes the proof.

**Lemma 3.** (Proposition 49.3 in [35]) Suppose $X_n, n = 1, 2, \cdots$, are random vectors on $\mathbb{R}^d$, $X$ is a random vector on $\mathbb{R}^d$, and $Y, Y_n, n = 1, 2, \cdots$, are random vectors on $\mathbb{R}^m$, then $X_n \xrightarrow{p} X$ and $Y_n \xrightarrow{p} Y$ iff $(X_n) \xrightarrow{p} (Y)$.\)

**Lemma 4.** (Proposition 49.5 in [35]) Let $\{X_n\}$ be random vectors on $\mathbb{R}^d$ and assume $X_n \xrightarrow{p} X$, where $X$ is a random vector on $\mathbb{R}^d$. Let $f$ be a Borel function on $\mathbb{R}^d$ such that $P(X \in C(f)) = 1$, then $f(X_n) \xrightarrow{p} f(X)$. Here $C(f)$ is the continuous region of $f$.

Proposition 49.3 and 49.5 in [35] are often used together and can yield a variety of results. In the statistical literature the combination of these two is often called "Sultzky's
Theorem for Convergence in Probability”.

**Theorem 4.** For any \( i = 1, 2, \ldots, m_x \), \( \hat{P}_{x_i} \) is a consistent estimator of \( P_{x_i} \); for any \( j = 1, 2, \ldots, m_y \), \( \hat{P}_{y_j} \) is a consistent estimator of \( P_{y_j} \).

**Pf:** Since \( S^2 \xrightarrow{P} \sigma^2 \), \( \bar{x} \xrightarrow{P} \mu_x \) and \( \bar{y} \xrightarrow{P} \mu_y \), then direct application of Slutzky’s Theorem yields \( \hat{P}_{x_i} \xrightarrow{P} P_{x_i} \) and \( \hat{P}_{y_j} \xrightarrow{P} P_{y_j} \).

**Theorem 5.** \( \hat{P}_h \) and \( \hat{P}_v \) are consistent estimators of \( P_h \) and \( P_v \), respectively.

**Pf:** Since \( \hat{P}_h = \sum \hat{P}_{y_j} \) and \( \hat{P}_{y_j} \xrightarrow{P} P_{y_j} \), then by Slutzky’s Theorem, \( \hat{P}_h \xrightarrow{P} P_h \); by the same argument \( \hat{P}_v \xrightarrow{P} P_v \).

**Lemma 5.** Show that \( \forall \epsilon > 0, \)

\[
\lim_{N \to \infty} \Pr \left( \left| \chi^2 \left( (1 - \alpha)^{1/N_u} \right) - \chi^2 \left( (1 - \alpha)^{1/N_u} \right) \right| > \epsilon \right) = 0. \tag{3.38}
\]

**Pf:** First note that \( \chi^2 \) is an exponential distribution with mean 2, hence we can write out the explicit form of its quantile,

\[
\chi^2 \left( (1 - \alpha)^{1/N_u} \right) = -2 \log \left( 1 - (1 - \alpha)^{1/N_u} \right) \tag{3.39}
\]

and

\[
\chi^2 \left( (1 - \alpha)^{1/N_u} \right) = -2 \log \left( 1 - (1 - \alpha)^{1/N_u} \right). \tag{3.40}
\]

Thus

\[
\Pr \left( \left| \chi^2 \left( (1 - \alpha)^{1/N_u} \right) - \chi^2 \left( (1 - \alpha)^{1/N_u} \right) \right| > \epsilon \right) = \Pr \left( \left| \log \frac{1 - (1 - \alpha)^{1/N_u}}{1 - (1 - \alpha)^{1/N_u}} \right| > \frac{\epsilon}{2} \right). \tag{3.41}
\]

Here \( N_u = \hat{N} - n + \hat{n}_{rec} \), so substituting \( \hat{N} = \frac{1}{2} \left( n_h \hat{P}_h + n_v \hat{P}_v \right) \) and \( \hat{n}_{rec} = \frac{1}{2} \left( n_h \hat{P}_v + n_v \hat{P}_h \right) \)
into this expression and rearranging terms, we get

\[ \hat{N}_u = \left( \frac{1}{2P_h} + \frac{\hat{P}_v}{2} - 1 \right) n_h + \left( \frac{1}{2P_v} + \frac{\hat{P}_h}{2} - 1 \right) n_v. \quad (3.42) \]

In our case, \( P_h \) and \( P_v \) are not very large, say \( 0 < P_h < \frac{1}{2} \) and \( 0 < P_v < \frac{1}{2} \); then \( \frac{1}{2P_h} + \frac{\hat{P}_h}{2} - 1 > 0 \) and \( \frac{1}{2P_v} + \frac{\hat{P}_v}{2} - 1 > 0 \). Since \( \frac{\hat{P}_h}{P_h} \to P_h \) and \( \frac{\hat{P}_v}{P_v} \to P_v \), then \( \frac{1}{2P_h} + \frac{\hat{P}_h}{2} - 1 \to \frac{1}{2P_h} + \frac{P_h}{2} - 1 \) and \( \frac{1}{2P_v} + \frac{\hat{P}_v}{2} - 1 \to \frac{1}{2P_v} + \frac{P_v}{2} - 1 \). Now we can always choose \( \varepsilon_1 > 0 \) satisfying \( \frac{1}{2P_h} + \frac{P_h}{2} - 1 - \varepsilon_1 > 0 \) and \( \frac{1}{2P_v} + \frac{P_v}{2} - 1 - \varepsilon_1 > 0 \). For the chosen \( \varepsilon_1 \), we also have

\[ \lim_{N \to \infty} \Pr \left( \left| \left( \frac{1}{2P_h} + \frac{\hat{P}_v}{2} - 1 \right) - \left( \frac{1}{2P_h} + \frac{P_v}{2} - 1 \right) \right| > \varepsilon_1 \right) = 0 \quad (3.43) \]

and

\[ \lim_{N \to \infty} \Pr \left( \left| \left( \frac{1}{2P_v} + \frac{\hat{P}_h}{2} - 1 \right) - \left( \frac{1}{2P_v} + \frac{P_h}{2} - 1 \right) \right| > \varepsilon_1 \right) = 0. \quad (3.44) \]

By Lemma 1, choosing \( \alpha = \frac{1}{6} \), we know there exist \( M_1, M_2 \) and \( M_3 \), such that when \( N > M_1 \),

\[ \Pr \left( \left| n_h - N P_h \right| > \frac{1}{2} N^\frac{3}{2} P_h \right) < \frac{16Q_h + 1}{N^\frac{3}{2}}; \quad (3.45) \]

when \( N > M_2 \),

\[ \Pr \left( \left| n_v - N P_v \right| > \frac{1}{2} N^\frac{3}{2} P_v \right) < \frac{16Q_v + 1}{N^\frac{3}{2}}; \quad (3.46) \]

and when \( N > M_3 \),

\[ \Pr \left( \left| N_u - N P_u \right| > \frac{1}{2} N^\frac{3}{2} P_u \right) < \frac{16Q_u + 1}{N^\frac{3}{2}}. \quad (3.47) \]

Now when all of \( \left| n_h - N P_h \right| \leq \frac{1}{2} N^\frac{3}{2} P_h \), \( \left| n_v - N P_v \right| \leq \frac{1}{2} N^\frac{3}{2} P_v \), \( \left| N_u - N P_u \right| \leq \frac{1}{2} N^\frac{3}{2} P_u \), \( \left| \left( \frac{1}{2P_h} + \frac{\hat{P}_h}{2} - 1 \right) - \left( \frac{1}{2P_v} + \frac{\hat{P}_v}{2} - 1 \right) \right| \leq \varepsilon_1 \) and \( \left| \left( \frac{1}{2P_v} + \frac{\hat{P}_v}{2} - 1 \right) - \left( \frac{1}{2P_h} + \frac{\hat{P}_h}{2} - 1 \right) \right| \leq \varepsilon_1 \) are satisfied (we abbreviate these as the “five inequalities”), we have the control of \( N_u \).
based solely on the third inequality as

\[
NP_u - \frac{1}{2}N^\frac{2}{3}P_u \leq N_u \leq NP_u + \frac{1}{2}N^\frac{2}{3}P_u,
\]

(3.48)

and the control of \( \hat{N}_u \) based on the combination of Eq. (3.42), the fact that both \( \frac{1}{2P_h} + \frac{P_u}{2} - 1 - \epsilon_1 \) and \( \frac{1}{2P_v} + \frac{P_h}{2} - 1 - \epsilon_1 \) are positive, and the above "five inequalities" as follows:

\[
\hat{N}_u \leq \left( \frac{1}{2P_h} + \frac{P_v}{2} - 1 + \epsilon_1 \right) \left( NP_h + \frac{1}{2}N^\frac{2}{3}P_h \right) + \\
\left( \frac{1}{2P_v} + \frac{P_h}{2} - 1 + \epsilon_1 \right) \left( NP_v + \frac{1}{2}N^\frac{2}{3}P_v \right) \\
= (P_u + \epsilon_1 (P_h + P_v)) \left( N + \frac{1}{2}N^\frac{3}{2} \right), \tag{3.49}
\]

and

\[
\hat{N}_u \geq \left( \frac{1}{2P_h} + \frac{P_v}{2} - 1 - \epsilon_1 \right) \left( NP_h - \frac{1}{2}N^\frac{2}{3}P_h \right) + \\
\left( \frac{1}{2P_v} + \frac{P_h}{2} - 1 - \epsilon_1 \right) \left( NP_v - \frac{1}{2}N^\frac{2}{3}P_v \right) \\
= (P_u - \epsilon_1 (P_h + P_v)) \left( N - \frac{1}{2}N^\frac{3}{2} \right). \tag{3.50}
\]

Since the function \( f(x) = 1 - (1 - \alpha)^\frac{1}{x} \) \( (x > 0) \) is a decreasing function of \( x \) and for any positive \( a \) and \( b \),

\[
\lim_{N \to \infty} \frac{1 - (1 - \alpha)^\frac{1}{(N+\frac{1}{2}N^\frac{3}{2})}}{1 - (1 - \alpha)^\frac{1}{(N+\frac{1}{2}N^\frac{3}{2})}} = \frac{b}{a}, \tag{3.51}
\]

then by Eq. (3.48), Eq. (3.49) and Eq. (3.50) we have

\[
\frac{1 - (1 - \alpha)^{1/N_u}}{1 - (1 - \alpha)^{1/\hat{N}_u}} \geq \frac{1 - (1 - \alpha)^{P_u (N+\frac{1}{2}N^\frac{3}{2})}}{1 - (1 - \alpha)^{(P_u - \epsilon_1 (P_h + P_v))(N-\frac{1}{2}N^\frac{3}{2})}} \xrightarrow{N \to \infty} \frac{P_u - \epsilon_1 (P_h + P_v)}{P_u}. \tag{3.52}
\]
and

$$\frac{1 - (1 - \alpha)^{1/N_u}}{1 - (1 - \alpha)^{1/N_u}} \leq \frac{1}{1 - (1 - \alpha)\left\{\frac{P_u}{(N^0.5 - N^0.5)}\right\} N \to \infty} P_u + \varepsilon_1 (P_h + P_v). \quad (3.53)$$

From Eq. (3.52) and Eq. (3.53), for any $\varepsilon > 0$, we can always choose sufficiently large $N$ and small $\varepsilon_1$ to make $\frac{1 - (1 - \alpha)^{1/N_u}}{1 - (1 - \alpha)^{1/N_u}}$ closed enough to 1 such that $\{(\log \frac{1 - (1 - \alpha)^{1/N_u}}{1 - (1 - \alpha)^{1/N_u}}) > \frac{\varepsilon}{2}\}$ will be a null event, that is to say,

$$\lim_{N \to \infty} \Pr \left( \left| \log \frac{1 - (1 - \alpha)^{1/N_u}}{1 - (1 - \alpha)^{1/N_u}} \right| > \frac{\varepsilon}{2} \cap \text{"five inequalities"} \right) = 0. \quad (3.54)$$

Since

$$\Pr \left( \left| \log \frac{1 - (1 - \alpha)^{1/N_u}}{1 - (1 - \alpha)^{1/N_u}} \right| > \frac{\varepsilon}{2} \right) \leq \Pr \left( \left| \log \frac{1 - (1 - \alpha)^{1/N_u}}{1 - (1 - \alpha)^{1/N_u}} \right| > \frac{\varepsilon}{2} \cap \text{"five inequalities"} \right)$$

$$+ \Pr \left( \left| \left( \frac{1}{2P_h} + \frac{P_v}{2} - 1 \right) - \left( \frac{1}{2P_h} + \frac{P_v}{2} - 1 \right) \right| > \varepsilon_1 \right)$$

$$+ \Pr \left( \left| \left( \frac{1}{2P_v} + \frac{P_h}{2} - 1 \right) - \left( \frac{1}{2P_v} + \frac{P_h}{2} - 1 \right) \right| > \varepsilon_1 \right)$$

$$+ \Pr \left( |n_h - NP_h| > \frac{1}{2} N^0.5 P_h \right)$$

$$+ \Pr \left( |n_v - NP_v| > \frac{1}{2} N^0.5 P_v \right)$$

$$+ \Pr \left( |N_u - NP_u| > \frac{1}{2} N^0.5 P_u \right), \quad (3.55)$$

then combining Eq. (3.41), Eq. (3.43), Eq. (3.44), Eq. (3.45), Eq. (3.46), Eq. (3.47), Eq. (3.54) and Eq. (3.55) completes the proof.

**Lemma 6.** For any $\varepsilon > 0$,

$$\lim_{N \to \infty} \Pr \left( \left| \chi_2^2 \left( (1 - \alpha)^{1/N_u} \right) - 2F_{2,n_h + n_v} \left( (1 - \alpha)^{1/N_u} \right) \right| > \varepsilon \right) = 0 \quad (3.56)$$
First, we derive the explicit form of $2F_{2,nh+nv-2}\left((1 - \alpha)^{1/N_u}\right)$. Let $X \sim F_{2,n}$, then its density and CDF are $f(x) = \left(1 + \frac{2x}{n}\right)^{-1 - \frac{n}{2}}$ and $F(x) = 1 - \left(1 + \frac{2x}{n}\right)^{-\frac{n}{2}}$, respectively. Denote by $q_{\beta,n}$ the $\beta$-quantile of $F_{2,n}$, it’s not difficult to get $q_{\beta,n} = \frac{n}{2} \left((1 - \beta)^{-\frac{n}{2}} - 1\right)$ from the expression of CDF. Also after some algebraic work, we know $q_{\beta,n}$ is a decreasing function of $n$.

Now denote by $Q$ the expression $\chi^2\left((1 - \alpha)^{\frac{1}{N_u}}\right) - 2F_{2,nh+nv-2}\left((1 - \alpha)^{\frac{1}{N_u}}\right)$, then we have the explicit form of $Q$ as the following:

$$Q = \underbrace{-2 \log \left(1 - (1 - \alpha)^{1/N_u}\right)}_{Q_1} - \underbrace{(nh + nv - 2) \left[\left(1 - (1 - \alpha)^{1/N_u}\right)^{-\frac{2}{nh+nv-2}} - 1\right]}_{Q_2}$$

$$\overset{\text{def}}{=} Q_1 - Q_2$$

(3.57)

As in the proof of Lemma 5, we first assume that the “five inequalities” hold, then try to find the lower and upper bounds of $Q$ and the limiting probabilities of those bounds.

Conditioning on the “five inequalities”, we have the bound of $nh + nv - 2$ as

$$\left(N - \frac{1}{2}N^\frac{3}{2}\right)(P_h + P_v) - 2 \leq nh + nv - 2 \leq \left(N + \frac{1}{2}N^\frac{3}{2}\right)(P_h + P_v) - 2$$

(3.58)

and the bound of $N_u$ in Eq. (3.49) and Eq. (3.50).

Combining the form of $Q_1$ with Eq. (3.49) and Eq. (3.50), we have

$$Q_{1,lb} \leq Q_1 \leq Q_{1,ub}$$

(3.59)

where

$$Q_{1,lb} = -2 \log \left(1 - (1 - \alpha)\left(\frac{P_u - \epsilon_1(P_h + P_v)}{(N + \frac{1}{2}N^\frac{3}{2})}\right)^{\frac{1}{N^\frac{3}{2}}}(N - \frac{1}{2}N^\frac{3}{2})\right)$$

(3.60)
and

\[ Q_{1, ub} = -2 \log \left( 1 - \left( 1 - \alpha \right) \left( \frac{1}{(p_{n+1} + p_n)} \left( N + \frac{1}{4} N^2 \right) \right) \right). \]  \hspace{1cm} (3.61)

For \( Q_2 \), using the fact that the quantile \( q_{\beta, n} \) is a decreasing function of \( n \) and Eq. (3.58), we get

\[ Q_2 \geq \left( \left( N + \frac{1}{2} N^2 \right) (P_h + P_v) - 2 \right) \left( 1 - \left( 1 - \alpha \right)^{1/N_u} \right)^{-2} \left( \frac{2}{(P_h + P_v)^2} - 1 \right) \]  \hspace{1cm} (3.62)

and

\[ Q_2 \leq \left( \left( N - \frac{1}{2} N^2 \right) (P_h + P_v) - 2 \right) \left( 1 - \left( 1 - \alpha \right)^{1/N_u} \right)^{-2} \left( \frac{2}{(P_h + P_v)^2} - 1 \right). \]  \hspace{1cm} (3.63)

Next applying the fact that \( \left( 1 - (1 - \alpha)^{1/N_u} \right)^{-b} \) is an increasing function of \( N_u \), Eq. (3.49) and Eq. (3.50) to the above two inequalities, we now have the bound of \( Q_2 \) as

\[ Q_{2, lb} \leq Q_2 \leq Q_{2, ub}, \]  \hspace{1cm} (3.64)

where

\[ Q_{2, lb} = \left( \left( N + \frac{1}{2} N^2 \right) (P_h + P_v) - 2 \right) \left( 1 - \left( 1 - \alpha \right)^{1/(p_{n+1} + p_n)} \left( N - \frac{1}{4} N^2 \right) \right)^{-2} \left( \frac{2}{(P_h + P_v)^2} - 1 \right). \]  \hspace{1cm} (3.65)
and

\[ Q_{2,ub} = \left( \left( N - \frac{1}{2} N^{\frac{3}{2}} \right) (P_h + P_v) - 2 \right) \cdot \left( \frac{1}{1 - (1 - \alpha)^{\left( P_h + P_v \right) \left( N + \frac{1}{2} N^{\frac{3}{2}} \right)} - \left( N - \frac{1}{2} N^{\frac{3}{2}} \right) (P_h + P_v)^{-2}}{Q_{2,ub,r}} \right) \].

(3.66)

Therefore, we get bounds on \( Q \) conditioning on the "five inequalities" as:

\[ Q_{1,lb} - Q_{2,ub} = Q_{lb} \leq Q \leq Q_{ub} = Q_{1,ub} - Q_{2,lb} \].

(3.67)

Next we want find \( \lim_{N \to \infty} Q_{lb} \) and \( \lim_{N \to \infty} Q_{ub} \). In order to achieve this, we first show that \( \lim_{N \to \infty} Q_{2,ub} = \lim_{N \to \infty} Q_{1,ub} \) and \( \lim_{N \to \infty} Q_{2,lb} = \lim_{N \to \infty} Q_{1,lb} \). It's not difficult to show that for any \( b > 0 \),

\[ \lim_{N \to \infty} \frac{Q_{1,lb}}{N^b} = 0 \]

(3.68)

and

\[ \lim_{N \to \infty} \frac{Q_{1,ub}}{N^b} = 0. \]

(3.69)

Let \( \theta = \frac{Q_{1,ub}}{\left( N - \frac{1}{2} N^{\frac{3}{2}} \right) (P_h + P_v)^{-2}} \), then \( \theta > 0 \) and \( \lim_{N \to \infty} \theta = 0 \) by Eq. (3.69). Note that \( Q_{2,ub,r} = e^\theta \), then by Taylor expansion theory, for every \( N \) there exists \( 0 < \theta_{\xi,N} < \theta \) satisfying \( \lim_{N \to \infty} \theta_{\xi,N} = 0 \) such that

\[ Q_{2,ub,r} = 1 + \theta + \frac{\theta^2 e^{\theta_{\xi,N}}}{2!} \]

(3.70)
Thus from Eq. (3.66) and Eq. (3.70),

\[
\lim_{N \to \infty} Q_{2,ub} = \left( \left( N - \frac{1}{2} N^\frac{3}{2} \right) (P_h + P_v) - 2 \right) \left( \theta + \frac{\theta^2 e^{\theta_{\epsilon,N}}}{2} \right) \\
= \lim_{N \to \infty} Q_{1,ub} + \lim_{N \to \infty} \frac{Q_{1,ub} e^{\theta_{\epsilon,N}}}{2} \left( \left( N - \frac{1}{2} N^\frac{3}{2} \right) (P_h + P_v) - 2 \right) \\
= \lim_{N \to \infty} Q_{1,ub} .
\]  

(3.71)

The last equality holds because of Eq. (3.68) and \( \lim e^{\theta_{\epsilon,N}} = 1 \). Similar reasoning can lead to \( \lim Q_{2,lb} = \lim Q_{1,lb} \). Therefore by Eq. (3.51),

\[
\lim_{N \to \infty} Q_{lb} = \lim_{N \to \infty} (Q_{1,lb} - Q_{2,ub}) \\
= \lim_{N \to \infty} (Q_{1,lb} - Q_{1,ub}) \\
= 2 \log \left( \lim_{N \to \infty} \frac{1 - (1 - \alpha) (P_u + \epsilon_1 (P_h + P_v)) (N + \frac{1}{2} N^\frac{3}{2})}{1 - (1 - \alpha) (P_u - \epsilon_1 (P_h + P_v)) (N - \frac{1}{2} N^\frac{3}{2})} \right) \\
= 2 \log \frac{P_u - \epsilon_1 (P_h + P_v)}{P_u + \epsilon_1 (P_h + P_v)}
\]  

(3.72)

and similarly,

\[
\lim_{N \to \infty} Q_{ub} = 2 \log \frac{P_u + \epsilon_1 (P_h + P_v)}{P_u - \epsilon_1 (P_h + P_v)} .
\]  

(3.73)

So conditioning on the "five inequalities", for any \( \epsilon > 0 \), we can always choose sufficiently small \( \epsilon_1 \) and large \( N \) such that \( \{ |Q| > \epsilon \} \) will be a null event. The rest of proof is analogous to that of Lemma 5.

**Lemma 7.**

\[
|S^2 - \rho^2| \chi^2 \left( (1 - \alpha) \frac{N}{2} \right) \xrightarrow{P} 0
\]  

(3.74)
Pf: In Theorem 3, choose \( \alpha = \frac{1}{6} \); then there exists \( M_3 \) such that when \( N > M_3 \),

\[
\Pr \left( \left| S^2 - \rho^2 \right| > \frac{1}{N^{\frac{1}{3}}} \right) < 2\rho^4 \left( \frac{2}{(P_h + P_u)N^{\frac{1}{3}} - 4N^{-\frac{2}{3}}} + \frac{8\left( \frac{Q_h}{P_h} + \frac{Q_u}{P_u} \right)}{N^{\frac{1}{3}}} \right),
\]

(3.75)

where clearly \( \lim_{N \to \infty} g(N) = 0 \).

Let

\[
Q = \Pr \left( \left| S^2 - \rho^2 \right| \log \left( 1 - (1 - \alpha)^{\frac{1}{6}} \right) > \frac{\varepsilon}{2} \cap \left| S^2 - \rho^2 \right| \leq \frac{1}{N^{\frac{1}{3}}} \right),
\]

(3.76)

then

\[
Q \leq \Pr \left( \left| \log \left( 1 - (1 - \alpha)^{\frac{1}{6}} \right) \right| > \frac{\varepsilon}{N^{\frac{1}{3}}} \right).
\]

(3.77)

Since \( \lim_{N \to \infty} N^{-\frac{1}{3}} \log \left( 1 - (1 - \alpha)^{\frac{1}{6}} \right) = 0 \), then for any \( \varepsilon > 0 \) when \( N \) is sufficiently large, \( \left\{ N^{-\frac{1}{3}} \left| \log \left( 1 - (1 - \alpha)^{\frac{1}{6}} \right) \right| > \frac{\varepsilon}{2} \right\} \) will be a null event. Thus,

\[
\lim_{N \to \infty} Q = 0.
\]

(3.78)

Since for any \( \varepsilon > 0 \), when \( N > M_3 \),

\[
\Pr \left( \left| S^2 - \rho^2 \right| \chi^2_2 \left( 1 - \alpha \right) \frac{1}{N^\frac{1}{3}} > \varepsilon \right) = \Pr \left( \left| S^2 - \rho^2 \right| \log \left( 1 - (1 - \alpha)^{\frac{1}{6}} \right) > \frac{\varepsilon}{2} \right)
\]

(since \( N_u < N \))

\[
\leq Q + \Pr \left( \left| S^2 - \rho^2 \right| > \frac{1}{N^{\frac{1}{3}}} \right)
\]

(by Eq. (3.75))

\[
< Q + g(N),
\]

(3.79)

then we finish the proof based on \( \lim_{N \to \infty} Q = 0 \) and \( \lim_{N \to \infty} g(N) = 0 \).
**Lemma 8.** For any $\varepsilon > 0$,

$$
\lim_{N \to \infty} \Pr \left( \left| \rho^2 \chi^2_2 \left( (1 - \alpha)^{1/N_u} \right) - 2S^2 F_{2, n_h + n_v - 2} \left( (1 - \alpha)^{1/N_u} \right) \right| > \varepsilon \right) = 0.
$$

**(Pf.**) Let

$$Q = \rho^2 \chi^2_2 \left( (1 - \alpha)^{1/N_u} \right) - 2S^2 F_{2, n_h + n_v - 2} \left( (1 - \alpha)^{1/N_u} \right),$$

then we can decompose $Q$ into

$$Q = \frac{(\rho^2 - S^2) \chi^2_2 \left( (1 - \alpha)^{1/N_u} \right)}{Q_1} + S^2 \frac{\chi^2_2 \left( (1 - \alpha)^{1/N_u} \right) - \chi^2_2 \left( (1 - \alpha)^{1/N_u} \right)}{Q_2} + S^2 \frac{\chi^2_2 \left( (1 - \alpha)^{1/N_u} \right) - 2F_{2, n_h + n_v - 2} \left( (1 - \alpha)^{1/N_u} \right)}{Q_3}.
$$

Hence,

$$\Pr (|Q| > \varepsilon) \leq \Pr \left( |Q_1| > \frac{\varepsilon}{3} \right) + \Pr \left( |Q_2| > \frac{\varepsilon}{3} \right) + \Pr \left( |Q_3| > \frac{\varepsilon}{3} \right).$$

Now by **Lemma 7**, we have

$$\lim_{N \to \infty} \Pr \left( |Q_1| > \frac{\varepsilon}{3} \right) = 0.$$

An immediate result of **Theorem 3** is

$$\lim_{N \to \infty} \Pr (S^2 \leq \rho^2 + 1) = 1.$$
By the combination of Lemma 5 and Eq. (3.85),

$$\lim_{N \to \infty} \Pr \left( |Q_2| > \frac{\varepsilon}{3} \right) = 0,$$

(3.86)

and by the combination of Lemma 6 and Eq. (3.85),

$$\lim_{N \to \infty} \Pr \left( |Q_3| > \frac{\varepsilon}{3} \right) = 0.$$

(3.87)

Therefore this lemma is the direct result of Eq. (3.83), Eq. (3.84), Eq. (3.86) and Eq. (3.87).

**Lemma 9.** \( \frac{1}{n_h} \xrightarrow{p} 0 \) and \( \frac{1}{n_v} \xrightarrow{p} 0 \).

**Pf:** We prove the lemma for \( n_h \). From Lemma 1, there exists \( M_1 \) such that when \( N > M_1 \),

$$\Pr \left( |n_h - NP_h| > \frac{1}{2} NP_h \right) < \left( \frac{16Q_h}{P_h} + 1 \right) \frac{1}{N}. \quad (3.88)$$

For any \( \varepsilon > 0 \), when \( N > M_1 \),

$$\Pr \left( \left| \frac{1}{n_h} \right| > \varepsilon \right) = \Pr \left( \left| \frac{1}{n_h} \right| > \varepsilon \cap |n_h - NP_h| \leq \frac{1}{2} NP_h \right) + \Pr \left( \left| \frac{1}{n_h} \right| > \varepsilon \cap |n_h - NP_h| > \frac{1}{2} NP_h \right)$$

(by Eq. (3.88)) \(< \Pr \left( \left| \frac{1}{n_h} \right| > \varepsilon \cap |n_h - NP_h| \leq \frac{1}{2} NP_h \right) + \left( \frac{16Q_h}{P_h} + 1 \right) \frac{1}{N}. \quad (3.89)$$

When \( |n_h - NP_h| \leq \frac{1}{2} NP_h \) holds, then \( \frac{1}{n_h} \leq \frac{2}{NP_h} \to 0 \). Hence when \( N \) is sufficiently large, \( \left\{ \left| \frac{1}{n_h} \right| > \varepsilon \cap |n_h - NP_h| \leq \frac{1}{2} NP_h \right\} \) will be a null event, i.e., the limit of its probability is 0. Therefore, from Eq. (3.89), we have shown \( \frac{1}{n_h} \xrightarrow{p} 0 \).

**Lemma 10.** \( \min \left( 1 + \frac{1}{n_h}, 1 + \frac{1}{n_v} \right) \xrightarrow{p} 1 \) and \( \max \left( 1 + \frac{1}{n_h}, 1 + \frac{1}{n_v} \right) \xrightarrow{p} 1 \).

**Pf:** It is equivalent to show that \( \min \left( \frac{1}{n_h}, \frac{1}{n_v} \right) \xrightarrow{p} 0 \) and \( \max \left( \frac{1}{n_h}, \frac{1}{n_v} \right) \xrightarrow{p} 0 \). Since \( 0 < \min \left( \frac{1}{n_h}, \frac{1}{n_v} \right) \leq \frac{1}{n_h} \) and \( 0 < \max \left( \frac{1}{n_h}, \frac{1}{n_v} \right) \leq \frac{1}{n_h} + \frac{1}{n_v} \), then by Lemma 9 we have
\[ \min \left( \frac{1}{n_h}, \frac{1}{n_w} \right) \xrightarrow{P} 0 \quad \text{and} \quad \max \left( \frac{1}{n_h}, \frac{1}{n_w} \right) \xrightarrow{P} 0. \]

**Lemma 11.** For every \( \varepsilon > 0 \),

\[
\lim_{N \to \infty} \Pr \left( \rho^2 \chi^2_2((1 - \alpha)^{1/N_u}) - \min(1 + \frac{1}{n_h}, 1 + \frac{1}{n_w}) 2S^2 F_{2,n_h+n_w-2}((1 - \alpha)^{1/N_u}) > \varepsilon \right) = 0 \quad (3.90)
\]

and

\[
\lim_{N \to \infty} \Pr \left( \rho^2 \chi^2_2((1 - \alpha)^{1/N_u}) - \max(1 + \frac{1}{n_h}, 1 + \frac{1}{n_w}) 2S^2 F_{2,n_h+n_w-2}((1 - \alpha)^{1/N_u}) > \varepsilon \right) = 0. \quad (3.91)
\]

**Pf:** Applying Slutzky’s Theorem to **Lemma 8** and **Lemma 10** results in the above two equalities.

**Theorem 6.** Let \( D_1 \triangle IR \) be the random difference set between \( D_1 \) and \( IR \), i.e.,

\[
D_1 \triangle IR = (D_1 \setminus IR) \cup (IR \setminus D_1).
\]

Then for any \( \varepsilon > 0 \),

\[
\lim_{N \to \infty} \Pr \left( \text{area}(D_1 \triangle IR) > \varepsilon \right) = 0. \quad (3.92)
\]

**Pf:** Recall that \( D_1 \) was defined in Eq. (3.15) as

\[
D_1 = \left\{ (x, y) \mid (x - \mu_x)^2 + (y - \mu_y)^2 \leq \chi^2_2 \left( (1 - \alpha)^{1/N_u} \right) \rho^2 \right\}
\]

and \( IR \) is the region defined as

\[
IR = \left\{ (x, y) \left\| \frac{(x - \bar{x})^2}{1 + \frac{1}{n_h}} + \frac{(y - \bar{y})^2}{1 + \frac{1}{n_w}} \leq 2S^2 F_{2,n_h+n_w-2}((1 - \alpha)^{1/N_u}) \right\}.
\]
Since \( \text{area}(D_1 \Delta IR) = \text{area}(D_1 \setminus IR) + \text{area}(IR \setminus D_1) \), then it is enough to show that

\[
\lim_{N \to \infty} \Pr(\text{area}\,(D_1 \setminus IR) > \varepsilon) = 0 \quad (3.93)
\]

and

\[
\lim_{N \to \infty} \Pr(\text{area}\,(IR \setminus D_1) > \varepsilon) = 0. \quad (3.94)
\]

From Theorem 1 and Theorem 2, there exists \( M \) such that when \( N > M \),

\[
\Pr\left(|\tilde{x} - \mu_x| > \frac{1}{N^\frac{1}{4}}\right) < \frac{\left(\frac{2 + 8Q_h}{p_h} + \frac{1}{2}\right) \rho^2}{N^\frac{1}{2}} \quad (3.95)
\]

and

\[
\Pr\left(|\tilde{y} - \mu_y| > \frac{1}{N^\frac{1}{4}}\right) < \frac{\left(\frac{2 + 8Q_h}{p_h} + \frac{1}{2}\right) \rho^2}{N^\frac{1}{2}}. \quad (3.96)
\]

Define \( D_2 \) and \( D_3 \) as the following regions:

\[
D_2 = \left\{(x, y) \left| (x - \tilde{x})^2 + (y - \tilde{y})^2 \leq \min\left(1 + \frac{1}{n_h}, 1 + \frac{1}{n_v}\right) 2 S^2 F_{2n_h+n_v-2} \left(1 - \alpha\right)^{1/2} \right\} \quad (3.97)
\]

and

\[
D_3 = \left\{(x, y) \left| (x - \tilde{x})^2 + (y - \tilde{y})^2 \leq \left(\sqrt{\rho^2 \chi^2_{2}((1 - \alpha)^{1/2})} + \frac{1}{N^\frac{1}{4}}\right)^2 \right\}. \quad (3.98)
\]

Apparently, \( D_2 \subset IR \). And when both \( |\tilde{x} - \mu_x| \leq \frac{1}{N^\frac{1}{4}} \) and \( |\tilde{y} - \mu_y| \leq \frac{1}{N^\frac{1}{4}} \) hold, we have \( D_1 \subset D_3 \) by the Triangle Inequality. This implies \( (D_1 \setminus IR) \subset (D_3 \setminus D_2) \), then \( \text{area}(D_1 \setminus IR) \leq \text{area}(D_3 \setminus D_2) \). Note that \( D_2 \) and \( D_3 \) are circles centered at the same
center \((\bar{x}, \bar{y})\), so we can compute

\[
\text{area}(D_3 \setminus D_2) = \pi \left( \sqrt{\rho^2 \chi_2^2 ((1 - \alpha)^{1/N_u}) + \frac{1}{N^4}} \right)^2 - \\
\pi \left( \min \left( 1 + \frac{1}{n_h}, 1 + \frac{1}{n_v} \right) 2S^2 F_{2,n_h+n_v-2} \left( (1 - \alpha)^{1/N_u} \right) \right) \]

\[
= \pi \left( \rho^2 \chi_2^2 ((1 - \alpha)^{1/N_u}) - \min \left( 1 + \frac{1}{n_h}, 1 + \frac{1}{n_v} \right) 2S^2 F_{2,n_h+n_v-2} \left( (1 - \alpha)^{1/N_u} \right) \right) + 2\pi \sqrt{\rho^2 \chi_2^2 ((1 - \alpha)^{1/N_u})} N^4 + \frac{1}{N^\frac{3}{2}}. \tag{3.99}
\]

Since \(\lim_{N \to \infty} 2\pi \sqrt{\rho^2 \chi_2^2 ((1 - \alpha)^{1/N_u}) N^4} = 0\) and \(\lim_{N \to \infty} \frac{\pi}{N^\frac{3}{2}} = 0\), then applying Lemma 11 to (3.99) results in

\[
\lim_{N \to \infty} \Pr(\text{area}(D_3 \setminus D_2) > \varepsilon) = 0. \tag{3.100}
\]

Now when \(N > M\), for any \(\varepsilon > 0\),

\[
\Pr(\text{area}(D_1 \setminus IR) > \varepsilon) \leq \Pr \left( \text{area}(D_1 \setminus IR) > \varepsilon, |\bar{x} - \mu_x| \leq \frac{1}{N^\frac{1}{4}}, |\bar{y} - \mu_y| \leq \frac{1}{N^\frac{1}{4}} \right) + \\
\Pr \left( |\bar{x} - \mu_x| > \frac{1}{N^\frac{1}{4}} \right) + \Pr \left( |\bar{y} - \mu_y| > \frac{1}{N^\frac{1}{4}} \right) \leq \Pr(\text{area}(D_3 \setminus D_2) > \varepsilon) + \frac{2 \left( \frac{2 + 8Q_h}{P_h} + \frac{1}{2} \right) \rho^2}{N^\frac{3}{2}} \tag{3.101}
\]

Both terms of the above right side go to 0 when \(N \to \infty\), therefore we have shown that \(\lim_{N \to \infty} \Pr(\text{area}(D_1 \setminus IR) > \varepsilon) = 0\).

For the proof of \(\lim_{N \to \infty} \Pr(\text{area}(IR \setminus D_1) > \varepsilon) = 0\), we can similarly define

\[
D_4 = \left\{ (x, y) \left| (x - \bar{x})^2 + (y - \bar{y})^2 \leq \max(1 + \frac{1}{n_h}, 1 + \frac{1}{n_v}) 2S^2 F_{2,n_h+n_v-2} \left( (1 - \alpha)^{1/N_u} \right) \right. \right\} \tag{3.102}
\]
and

\[ D_5 = \left\{ (x, y) \left| (x - \bar{x})^2 + (y - \bar{y})^2 \leq \left( \sqrt{\rho^2 \chi_2^2 ((1 - \alpha)^{1/N_u}) - \frac{1}{N^4}} \right)^2 \right. \right\}, \tag{3.103} \]

then \( IR \subset D_4 \). And when both \( |\bar{x} - \mu_x| \leq \frac{1}{N^4} \) and \( |\bar{y} - \mu_y| \leq \frac{1}{N^4} \) hold, we have \( D_5 \subset D_1 \). Hence, \( \text{area} (IR \setminus D_1) \leq \text{area} (D_4 \setminus D_5) \). The rest of the proof is similar to that of \( \lim_{N \to \infty} \text{Pr} (\text{area} (D_1 \setminus IR) > \epsilon) = 0 \). Thus, we finish the proof of this theorem.

Now let's find the probability that \( D_1 \) covers all \( N_u \) unobserved objects, and denote this probability by \( \pi_d \). Let \((x_i, y_i), i = 1, 2, \cdots, N, \) be the positions of all \( N \) objects in the cluster. Define \( O_i \) as the event that \((x_i, y_i)\) is observed, \( O^c_i \) the event that it is unobserved, and \( R_i \) the event that it falls inside \( D_1, i = 1, 2, \cdots, N \). Then, \( \text{Pr}(O^c_i) = P_u \) and \( \text{Pr}(R_i) = (1 - \alpha)^{1/N_u} \) for given \( N_u \). For any object, the probability that it falls inside \( D_1 \) conditional on being unobserved is

\[
\text{Pr}(R_i \mid O^c_i) = \frac{\text{Pr}(R_i) - \text{Pr}(R_i O_i)}{\text{Pr}(O^c_i)} \\
\geq \frac{\text{Pr}(R_i) - \text{Pr}(O_i)}{\text{Pr}(O^c_i)} \\
= \frac{(1 - \alpha)^{1/N_u} - (1 - P_u)}{P_u} \\
= 1 - \frac{1 - (1 - \alpha)^{1/N_u}}{P_u}, \tag{3.104}
\]

for \( i = 1, 2, \cdots, N \). Without loss of generality, let the index \( i \) from 1 to \( N_u \) represent the \( N_u \) unobserved objects. Since the detection process is independent for every object, then the events \( R_i \mid O^c_i \) are independent with each other for every two values of \( i \),
\[ i = 1, 2, \ldots, N_u. \] Therefore, a lower bound for \( \pi_d \) can be derived as

\[
\pi_d = \Pr(R_1, R_2, \ldots, R_{N_u} \mid O_{1_u}^c, O_{2_u}^c, \ldots, O_{N_u}^c) \\
= \left( \Pr(R_1 \mid O_1^c) \right)^{N_u} \\
\geq \left( 1 - \frac{1 - (1 - \alpha)^{1/N_u}}{P_u} \right)^{N_u}.
\] (3.105)

Denote this lower bound as \( Q(N_u, P_u) \). It is an increasing function of both \( N_u \) and \( P_u \), and its limit is \( (1 - \alpha)^{1/P_u} \) when \( N_u \) goes to \( \infty \). Values of \( Q(N_u, P_u) \) for combinations of \( \alpha = 0.01, 0.05 \), \( P_u = 0.5, 0.6, 0.7, 0.8 \) (reasonable values for our application) and \( N_u \) from 5 to \( \infty \) are listed in Table 3.1. From the table, we can see that the values of \( Q(N_u, P_u) \) even when \( N_u = 5 \) are very close to the corresponding limiting values for any combination of \( P_u \) and \( 1 - \alpha \). This suggests \( Q(N_u, P_u) \) approaches its limiting value very fast as a function of \( N_u \). When \( P_u \) is large, \( Q(N_u, P_u) \) is close to the nominal coverage \( 1 - \alpha \).

<table>
<thead>
<tr>
<th>( 1 - \alpha )</th>
<th>( P_u )</th>
<th>( N_u )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>0.95</td>
<td>0.5</td>
<td>0.90202</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.91779</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.92919</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.93782</td>
</tr>
<tr>
<td>0.99</td>
<td>0.5</td>
<td>0.98008</td>
</tr>
<tr>
<td></td>
<td>0.6</td>
<td>0.98338</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.98574</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.98751</td>
</tr>
</tbody>
</table>

Table 3.1 Some values of \( Q(N_u, P_u) \)

From Theorem 6, we can see \( IR \) is asymptotically identical to \( D_1 \) in the sense of convergence in probability as the cluster size \( N \to \infty \). So asymptotically, the two probabilities that \( IR \) and \( D_1 \) covers all the \( N_u \) unobserved objects are the same. Thus, \( IR \) has an asymptotically lower coverage bound \( (1 - \alpha)^{1/P_u} \) for covering all unobserved
objects when the cluster size \( N \to \infty \). The simulation results in the next section will show that the actual coverage is greater than \( 1 - \alpha \) in most cases. That means \( (1 - \alpha)^{1/P_u} \) may not be the best lower bound. But finding the exact coverage probability of \( D_1 \) for given \( N_u \) is impossible, and the bound \( (1 - \alpha)^{1/P_u} \) is the best we can obtain.

### 3.5 Simulation

To investigate the coverage performance of our proposed IR, we conduct a simulation study. In addition, we will evaluate the behavior of \( \hat{\mu}_x, \hat{\mu}_y, \hat{N} \) and \( \hat{\rho} \), and compare the three estimates of \( n_{rec} \) defined in Section 3.2.

![Figure 3.1 Center line transect position. The dot is the location of the cluster center \((\mu_x, \mu_y)\).](image)
In our simulation, we fixed the cluster center at point (50, 50) and the cluster dispersion parameter $\rho = 1$, and considered two cluster sizes $N = 250$ and 125, and three values of the detection function parameter $\sigma$, 0.1, 0.05 and 0.025. The numbers of vertical and horizontal transects were each set to 9, and the distance between neighboring transects was fixed at 1 to correspond to $\rho$. If $(\mu_x, \mu_y)$ were known, it would clearly be best to center the transect pattern on this point. However, because we might not have an accurate estimate of each cluster center from the global sampling, we considered five cases of transect positions. In the first case, called line position 1 (center line transect position), both middle vertical and horizontal transects go through the cluster center (50, 50), i.e., $(X_{01}, X_{02}, \cdots, X_{09}) = (46, 47, \cdots, 54)$ and $(Y_{01}, Y_{02}, \cdots, Y_{09}) = (46, 47, \cdots, 54)$; this is the position we would select if $(\mu_x, \mu_y)$ were known. In line position 2, the vertical transects are shifted right by a distance of $\frac{1}{2}$ from the center position while the horizontal transects remain unchanged, i.e., $(X_{01}, X_{02}, \cdots, X_{09}) = (46.5, 47.5, \cdots, 54.5)$ and $(Y_{01}, Y_{02}, \cdots, Y_{09}) = (46, 47, \cdots, 54)$. In line position 3, the vertical transects are shifted right by a distance of $\frac{1}{2}$ and the horizontal transects are shifted up by a distance of $\frac{1}{2}$ from the center position, i.e., $(X_{01}, X_{02}, \cdots, X_{09}) = (46.5, 47.5, \cdots, 54.5)$ and $(Y_{01}, Y_{02}, \cdots, Y_{09}) = (46.5, 47.5, \cdots, 54.5)$. In line position 4, the vertical transects are shifted right by a distance of 4 from the center position while the horizontal transects remain unchanged, i.e., $(X_{01}, X_{02}, \cdots, X_{09}) = (50, 51, \cdots, 58)$ and $(Y_{01}, Y_{02}, \cdots, Y_{09}) = (46, 47, \cdots, 54)$. In line position 5, the vertical transects are shifted right by a distance of 4 and the horizontal transects are shifted up by a distance of 4 from the center position, i.e., $(X_{01}, X_{02}, \cdots, X_{09}) = (50, 51, \cdots, 58)$ and $(Y_{01}, Y_{02}, \cdots, Y_{09}) = (50, 51, \cdots, 58)$. Fig. 3.1 is the illustration of line position 1 and Fig. 3.2 displays the other four transect positions.

For each combination of $N$ and $\sigma$, positions of these $N$ objects were generated from a symmetrical bivariate normal distribution with variance $\rho^2$. According to each case of the transect position, the detection process was executed and the detected objects were
recorded. Then $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\bar{N}}$ and $\hat{\rho}$ were calculated. We considered two nominal coverage probabilities, 95% and 99% ($\alpha = .05$ and .01), of $IR$. Then for each nominal coverage, $IR$ was computed and the result of whether $IR$ includes all of the unobserved objects, either 0 or 1, called the single coverage, was recorded.

The above procedure was repeated 10,000 times; the average of all single coverages is the simulated coverage of $IR$ which we can compare to the nominal coverage.

Table 3.2 is the summary of the standard deviation of $\hat{\mu}_x$ and $\hat{\mu}_y$. In this table, columns are combinations of $N$ and $\sigma$, rows are line transect positions; each cell has two numbers, the top one is the standard deviation for $\hat{\mu}_x$ and the bottom one is for $\hat{\mu}_y$. Note that there is no combination of $(N, \sigma) = (125, 0.025)$ in the table because no
objects were observed in some iterations of the simulation for this combination. From the table, we can see all the standard deviations are very small (less than 0.32) relative to the value of \( \rho \). Also, the average bias of \( \hat{\mu}_x \) and \( \hat{\mu}_y \) is less than 0.01 in each case of line transect positions, and so is not tabulated. Therefore the center is very well estimated across all scenarios. The two standard deviations in each cell for line transect positions 1, 3 and 5 are almost identical as expected, because the line transects in these positions are located symmetrically relative to \( \mu_x \) and \( \mu_y \). The standard deviations in each cell for line transect positions 2 and 3 are similar to their counterparts in position 1 because the asymmetry in transect location is minor. This suggests that shifting one or both sets of transects by a distance of \( \frac{\rho}{2} \) does not practically degrade the quality of the estimate of the center. For line transect position 4, we do find that the standard deviation for \( \mu_y \) is larger than that of \( \mu_x \). This is not difficult to explain; shifting the vertical transects right by a distance of \( 4\rho \) leads to fewer detected objects along vertical transects than along horizontal ones. This of course will make \( \hat{\mu}_y \) less stable than \( \hat{\mu}_x \). For line transect position 5, the two standard deviations in each cell are similar as the corresponding standard deviation of \( \mu_y \) in line transect position 4 for each case of \((N, \sigma)\). The standard deviations of \( \mu_x \) and \( \mu_y \) increase when \( \sigma \) or \( N \) decreases.

<table>
<thead>
<tr>
<th>line pos.</th>
<th>((N, \sigma))</th>
<th>((250, 0.1))</th>
<th>((250, 0.05))</th>
<th>((250, 0.025))</th>
<th>((125, 0.1))</th>
<th>(125, 0.05))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.127</td>
<td>0.182</td>
<td>0.262</td>
<td>0.181</td>
<td>0.259</td>
<td>0.259</td>
</tr>
<tr>
<td>1</td>
<td>0.129</td>
<td>0.183</td>
<td>0.260</td>
<td>0.181</td>
<td>0.259</td>
<td>0.259</td>
</tr>
<tr>
<td>2</td>
<td>0.127</td>
<td>0.182</td>
<td>0.262</td>
<td>0.181</td>
<td>0.259</td>
<td>0.258</td>
</tr>
<tr>
<td>2</td>
<td>0.127</td>
<td>0.183</td>
<td>0.262</td>
<td>0.181</td>
<td>0.263</td>
<td>0.258</td>
</tr>
<tr>
<td>3</td>
<td>0.128</td>
<td>0.182</td>
<td>0.266</td>
<td>0.181</td>
<td>0.263</td>
<td>0.258</td>
</tr>
<tr>
<td>3</td>
<td>0.127</td>
<td>0.183</td>
<td>0.266</td>
<td>0.181</td>
<td>0.263</td>
<td>0.258</td>
</tr>
<tr>
<td>4</td>
<td>0.126</td>
<td>0.180</td>
<td>0.261</td>
<td>0.180</td>
<td>0.263</td>
<td>0.319</td>
</tr>
<tr>
<td>4</td>
<td>0.152</td>
<td>0.219</td>
<td>0.319</td>
<td>0.216</td>
<td>0.319</td>
<td>0.319</td>
</tr>
<tr>
<td>5</td>
<td>0.151</td>
<td>0.219</td>
<td>0.316</td>
<td>0.219</td>
<td>0.319</td>
<td>0.319</td>
</tr>
<tr>
<td>5</td>
<td>0.151</td>
<td>0.224</td>
<td>0.318</td>
<td>0.219</td>
<td>0.319</td>
<td>0.318</td>
</tr>
</tbody>
</table>

Table 3.2 Standard deviations of \( \hat{\mu}_x \) and \( \hat{\mu}_y \)
Tables 3.3 through 3.9 have similar structure to Table 3.2 and summarize other results from the simulation study. Table 3.3 is the summary of the average and standard deviation of the estimate of the cluster size $N$. In each cell, the top number is the average estimate and the bottom one is the standard deviation. Overall, the average estimates are very closed to the true $N$. The behavior of $\hat{N}$ is similar for line positions 1, 2 and 3. The standard deviations for line positions 4 and 5 are larger than those for line positions 1, 2 and 3 because there is fewer observed data. It's not surprising that the standard deviation is larger when $\sigma$ is smaller for the same cluster size since this also results in fewer observed objects.

![Table 3.3](image)

Table 3.4 is the summary of the standard deviation of $\hat{p}^2$. The average bias of $\hat{p}^2$ is
less than 0.01 (1% of the correct value) in each case, so we did not tabulate these values.

The patterns of the standard deviation of \( \hat{\rho}^2 \) is similar to those of \( \hat{N} \) in Table 3.3.

<table>
<thead>
<tr>
<th>line pos.</th>
<th>((N, \sigma))</th>
<th>((250, 0.1))</th>
<th>((250, 0.05))</th>
<th>((250, 0.025))</th>
<th>((125, 0.1))</th>
<th>((125, 0.05))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>11.49 (3.269)</td>
<td>2.61 (1.619)</td>
<td>0.63 (0.796)</td>
<td>5.43 (2.263)</td>
<td>1.26 (1.119)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.71 (1.204)</td>
<td>3.93 (0.465)</td>
<td>0.98 (0.171)</td>
<td>7.87 (0.857)</td>
<td>1.97 (0.328)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.73 (2.929)</td>
<td>3.94 (1.531)</td>
<td>0.98 (0.776)</td>
<td>7.87 (2.064)</td>
<td>1.97 (1.073)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.75 (3.829)</td>
<td>3.94 (1.972)</td>
<td>0.99 (1.000)</td>
<td>7.88 (2.727)</td>
<td>1.96 (1.402)</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>11.44 (3.289)</td>
<td>2.58 (1.629)</td>
<td>0.63 (0.793)</td>
<td>5.42 (2.259)</td>
<td>1.27 (1.126)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.70 (1.219)</td>
<td>3.92 (0.475)</td>
<td>0.98 (0.172)</td>
<td>7.86 (0.850)</td>
<td>1.97 (0.328)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.67 (2.942)</td>
<td>3.91 (1.546)</td>
<td>0.98 (0.775)</td>
<td>7.85 (2.039)</td>
<td>1.97 (1.074)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.67 (3.884)</td>
<td>3.91 (2.007)</td>
<td>0.99 (1.011)</td>
<td>7.85 (2.694)</td>
<td>1.97 (1.397)</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>11.46 (3.313)</td>
<td>2.60 (1.581)</td>
<td>0.63 (0.789)</td>
<td>5.44 (2.270)</td>
<td>1.28 (1.124)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.72 (1.230)</td>
<td>3.93 (0.465)</td>
<td>0.98 (0.171)</td>
<td>7.86 (0.866)</td>
<td>1.97 (0.332)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.70 (2.931)</td>
<td>3.92 (1.512)</td>
<td>0.98 (0.761)</td>
<td>7.87 (2.058)</td>
<td>1.98 (1.078)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>15.68 (3.862)</td>
<td>3.93 (1.954)</td>
<td>0.97 (0.976)</td>
<td>7.87 (2.714)</td>
<td>1.97 (1.392)</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>8.13 (2.788)</td>
<td>1.83 (1.350)</td>
<td>0.44 (0.667)</td>
<td>3.82 (1.905)</td>
<td>0.88 (0.934)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.98 (1.061)</td>
<td>2.74 (0.392)</td>
<td>0.68 (0.141)</td>
<td>5.49 (0.753)</td>
<td>1.37 (0.276)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10.99 (2.495)</td>
<td>2.75 (1.289)</td>
<td>0.69 (0.651)</td>
<td>5.48 (1.745)</td>
<td>1.36 (0.902)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11.00 (3.271)</td>
<td>2.73 (1.652)</td>
<td>0.69 (0.838)</td>
<td>5.49 (2.287)</td>
<td>1.36 (1.173)</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>5.76 (2.363)</td>
<td>1.29 (1.139)</td>
<td>0.31 (0.555)</td>
<td>2.68 (1.603)</td>
<td>0.61 (0.776)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.67 (0.931)</td>
<td>1.92 (0.330)</td>
<td>0.48 (0.117)</td>
<td>3.82 (0.646)</td>
<td>0.95 (0.232)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.69 (2.113)</td>
<td>1.93 (1.081)</td>
<td>0.48 (0.540)</td>
<td>3.81 (1.473)</td>
<td>0.95 (0.756)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.69 (2.742)</td>
<td>1.92 (1.394)</td>
<td>0.49 (0.697)</td>
<td>3.81 (1.918)</td>
<td>0.95 (0.972)</td>
</tr>
</tbody>
</table>

Table 3.5 Estimate and standard deviation of \( n_{\text{rec}} \)

Table 3.5 is the summary of the average of the three estimates of \( n_{\text{rec}} \) and the true value, along with their corresponding standard deviations. In each cell, there are four rows; the first row is for \( \hat{n}_{\text{rec},1} \), the second row for \( \hat{n}_{\text{rec},2} \), our defined \( \hat{n}_{\text{rec}} \), the third row for \( \hat{n}_{\text{rec},3} \), and the fourth row for the true value of \( n_{\text{rec}} \); the first entry in each line is the simulation average, and the numbers in parentheses are the corresponding standard deviations. For all the cells, average values of \( \hat{n}_{\text{rec},2} \) and \( \hat{n}_{\text{rec},3} \) are almost identical to the average value of \( n_{\text{rec}} \), while the standard deviation of \( \hat{n}_{\text{rec},3} \) is closer to that of \( n_{\text{rec}} \); \( \hat{n}_{\text{rec},1} \) has the largest bias among the three estimators. When \( \sigma \) decreases, the number of recaptures decreases dramatically.
The coverage of 95% IR is summarized in Table 3.6. For example, the number in the first cell is 0.9492; this means out of 10,000 times there are 9,492 times that our proposed IR includes all unobserved objects in the cluster. Coverages from all scenarios are very close to the nominal coverage probability 0.95. The coverage is closer to 0.95 and smaller when σ is larger, holding N and line transect position unchanged. There is no apparent difference for the coverage over the five line transect positions.

<table>
<thead>
<tr>
<th>line pos.</th>
<th>(250, 0.1)</th>
<th>(250, 0.05)</th>
<th>(250, 0.025)</th>
<th>(125, 0.1)</th>
<th>(125, 0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9492</td>
<td>0.9527</td>
<td>0.9642</td>
<td>0.9515</td>
<td>0.9608</td>
</tr>
<tr>
<td>2</td>
<td>0.9501</td>
<td>0.9563</td>
<td>0.9620</td>
<td>0.9539</td>
<td>0.9617</td>
</tr>
<tr>
<td>3</td>
<td>0.9525</td>
<td>0.9581</td>
<td>0.9631</td>
<td>0.9509</td>
<td>0.9588</td>
</tr>
<tr>
<td>4</td>
<td>0.9497</td>
<td>0.9549</td>
<td>0.9649</td>
<td>0.9498</td>
<td>0.9608</td>
</tr>
<tr>
<td>5</td>
<td>0.9450</td>
<td>0.9559</td>
<td>0.9690</td>
<td>0.9467</td>
<td>0.9624</td>
</tr>
</tbody>
</table>

Table 3.6 Coverage of 95% IR

Table 3.7 is the summary of coverage for 99% IR. All coverages are very close to 0.99. As in the case of 95% IR, the coverage decreases when σ increases, holding N and line transect position unchanged.

We also recorded the coverages of IR when \( \hat{n}_{\text{necl}} \) and \( \hat{n}_{\text{necl}} \) are used instead of \( \hat{n}_{\text{necl}} \); these averages are summarized in Table 3.8 and Table 3.9. In these two tables, the top number in each cell is for IR using \( \hat{n}_{\text{necl}} \) and the bottom one is that of \( \hat{n}_{\text{necl}} \). Comparing Tables 3.6 and 3.7 with Tables 3.8 and 3.9, we find there is little difference among the
Table 3.8 Coverage of 95% IR using $\hat{n}_{rec,1}$ and $\hat{n}_{rec,3}$

<table>
<thead>
<tr>
<th>line pos.</th>
<th>$(N, \sigma)$</th>
<th>$(250, 0.1)$</th>
<th>$(250, 0.05)$</th>
<th>$(250, 0.025)$</th>
<th>$(125, 0.1)$</th>
<th>$(125, 0.05)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9471</td>
<td>0.9524</td>
<td>0.9641</td>
<td>0.9491</td>
<td>0.9601</td>
<td>0.9605</td>
</tr>
<tr>
<td></td>
<td>0.9491</td>
<td>0.9529</td>
<td>0.9642</td>
<td>0.9509</td>
<td>0.9614</td>
<td>0.9615</td>
</tr>
<tr>
<td>2</td>
<td>0.9486</td>
<td>0.9561</td>
<td>0.9619</td>
<td>0.9531</td>
<td>0.9614</td>
<td>0.9615</td>
</tr>
<tr>
<td></td>
<td>0.9499</td>
<td>0.9562</td>
<td>0.9620</td>
<td>0.9536</td>
<td>0.9615</td>
<td>0.9615</td>
</tr>
<tr>
<td>3</td>
<td>0.9511</td>
<td>0.9578</td>
<td>0.9631</td>
<td>0.9495</td>
<td>0.9585</td>
<td>0.9587</td>
</tr>
<tr>
<td></td>
<td>0.9522</td>
<td>0.9581</td>
<td>0.9631</td>
<td>0.9508</td>
<td>0.9587</td>
<td>0.9587</td>
</tr>
<tr>
<td>4</td>
<td>0.9491</td>
<td>0.9545</td>
<td>0.9648</td>
<td>0.9488</td>
<td>0.9605</td>
<td>0.9605</td>
</tr>
<tr>
<td></td>
<td>0.9498</td>
<td>0.9549</td>
<td>0.9648</td>
<td>0.9498</td>
<td>0.9605</td>
<td>0.9605</td>
</tr>
<tr>
<td>5</td>
<td>0.9446</td>
<td>0.9557</td>
<td>0.9690</td>
<td>0.9462</td>
<td>0.9623</td>
<td>0.9627</td>
</tr>
<tr>
<td></td>
<td>0.9452</td>
<td>0.9559</td>
<td>0.9690</td>
<td>0.9468</td>
<td>0.9627</td>
<td>0.9627</td>
</tr>
</tbody>
</table>

Table 3.9 Coverage of 99% IR using $\hat{n}_{rec,1}$ and $\hat{n}_{rec,3}$

<table>
<thead>
<tr>
<th>line pos.</th>
<th>$(N, \sigma)$</th>
<th>$(250, 0.1)$</th>
<th>$(250, 0.05)$</th>
<th>$(250, 0.025)$</th>
<th>$(125, 0.1)$</th>
<th>$(125, 0.05)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9883</td>
<td>0.9900</td>
<td>0.9916</td>
<td>0.9892</td>
<td>0.9920</td>
<td>0.9920</td>
</tr>
<tr>
<td></td>
<td>0.9886</td>
<td>0.9901</td>
<td>0.9916</td>
<td>0.9897</td>
<td>0.9920</td>
<td>0.9920</td>
</tr>
<tr>
<td>2</td>
<td>0.9891</td>
<td>0.9904</td>
<td>0.9924</td>
<td>0.9917</td>
<td>0.9923</td>
<td>0.9923</td>
</tr>
<tr>
<td></td>
<td>0.9893</td>
<td>0.9905</td>
<td>0.9924</td>
<td>0.9921</td>
<td>0.9923</td>
<td>0.9923</td>
</tr>
<tr>
<td>3</td>
<td>0.9900</td>
<td>0.9906</td>
<td>0.9899</td>
<td>0.9891</td>
<td>0.9915</td>
<td>0.9915</td>
</tr>
<tr>
<td></td>
<td>0.9904</td>
<td>0.9908</td>
<td>0.9899</td>
<td>0.9894</td>
<td>0.9915</td>
<td>0.9915</td>
</tr>
<tr>
<td>4</td>
<td>0.9886</td>
<td>0.9893</td>
<td>0.9933</td>
<td>0.9891</td>
<td>0.9919</td>
<td>0.9919</td>
</tr>
<tr>
<td></td>
<td>0.9889</td>
<td>0.9894</td>
<td>0.9933</td>
<td>0.9896</td>
<td>0.9919</td>
<td>0.9919</td>
</tr>
<tr>
<td>5</td>
<td>0.9884</td>
<td>0.9918</td>
<td>0.9934</td>
<td>0.9870</td>
<td>0.9917</td>
<td>0.9917</td>
</tr>
<tr>
<td></td>
<td>0.9888</td>
<td>0.9918</td>
<td>0.9934</td>
<td>0.9871</td>
<td>0.9919</td>
<td>0.9919</td>
</tr>
</tbody>
</table>

Before implementing the line transect survey for the local sampling, we have to decide where to position those line transects. In our case, we have 9 parallel transects in each of two directions. One way to address this problem in practice is to first position preliminary transects relative to the vague estimate of the center from the global sampling, collect preliminary data, and then shift transects to positions along which more objects are likely to be detected. However, the simulated coverages of IR for our 5 line
transect positions are not very different, even when some transects are far away from the true cluster center. This may suggest that the precise positioning of transects relative to the cluster center is not critical to effective sampling. The simulation results further suggest IR coverage may be close to the nominal level if the line transects go through at least part of the cluster.

Defining IR requires estimating the number of "recaptured" objects. We proposed three estimators for it. We computed coverages of IR by using these three estimators in the simulation and showed no practical difference in performance. \( \hat{n}_{\text{rec}} \) appears in the formula of IR through \( \frac{1}{N_u} \) with \( \tilde{N}_u = \tilde{N} - n + \hat{n}_{\text{rec}} \). Because in our simulation, the detection probability for each object is very small and the cluster size is medium-to-large, the value of \( \tilde{N}_u \) is dominated by \( \tilde{N} \). So \( \frac{1}{N_u} \) varies little regardless of the estimate used for \( n_{\text{rec}} \). This suggests the coverage IR is insensitive to the choice of estimator of the number of "recaptured" objects when the cluster size is medium-to-large and the detection probability is small.

In conclusion, the cluster center, size and dispersion are all estimated well within the range of conditions included in the simulation study. This provides a solid foundation for accurate coverage of IR. The coverage of IR tends to be close to its nominal level. We also found that the estimation of "recapture" objects does not have a major impact on the coverage of IR. If the line transects are not positioned too far away from the cluster center, estimates of the cluster center, size and dispersion are reliable, hence the coverage of IR is near the nominal level.
CHAPTER 4  CONCLUSION

4.1 Summary of results

Military bases that have been used for weapons testing and training usually are con­
taminated with UXO. These bases can be returned to public use only after cleaning up
those UXO. The cleaning-up procedure are usually very expensive and time-consuming.
This demands statistical methods to provide more effective sampling survey, to charac­
terize the UXO distribution, and to give a probability confidence statement about UXO.
Based on the physical characteristics of UXO deposition, we use a line transect survey to
collect data on individual object locations and adopt a simplified Neyman-Scott process
to model the distribution of UXO over a contaminated site.

In Chapter 2, we applied a one-directional sparse parallel line transects to screen
the whole site. We extended Brown and Cowling’s [30] approach to estimate the cluster
intensity, mean cluster size and cluster dispersion from their case where data are available
on the complete position of any detected object to our case in which only the coordinate
corresponding to the direction of the transect can be observed. A likelihood estimator for
the parameter $\theta$ of the Neyman-Scott line transect distribution was given and was shown
to be a more efficient estimator than the method of moments estimator provided in [30].
For estimation of a large true $\theta$ value, we discussed the difficulty of numerical evaluation
and provided an ad hoc modified method to deal with this issue. The theoretical variance
estimators of all the cluster parameters were also given. Simulation studies showed
that all the cluster parameter estimators perform well, and their theoretical variance
estimators match their corresponding sample variances very well.

In Chapter 3, a two perpendicular directional line transect survey was described for augmenting information on each cluster detected in the global sampling. First, the estimators of the cluster center, size, and dispersion, and the number of “recapture” objects, were provided. We showed that all parameter estimators are consistent as the cluster size goes to infinity. Then based on these parameter estimators, we proposed a \((1 - \alpha)100\%\) inclusion region which will cover all undetected objects in this cluster with a target probability of \(1 - \alpha\). By showing that the inclusion region is asymptotically equivalent to an almost fixed region in the sense of convergence in probability, we gave the lower bound of coverage. Simulation studies were conducted and showed that the coverage of the proposed inclusion region is very close to the nominal level. We provided three estimators of number of “recapture” objects and it turned out that the choice of these estimators had minimal effect on the coverage of the inclusion region. In the simulation, we also considered five choices of the line transect positions and the results indicated that the coverages were similar, all pointed to the nominal level. This suggested the coverage is robust to the position of line transects provided that the transects go through at least part of the cluster.

4.2 Limitations and further research

In our model of a simplified Neyman-Scott process, we assume that “parents” are from a homogeneous Poisson process. If some sub-areas in the study region were used more than others, this assumption would not be appropriate. If this were the case, then we should model parents as an inhomogeneous Poisson process. So our model would be a general Neyman-Scott process. In addition, if we knew the use record of the study region, this could provide valuable prior information about the distribution of the parents. Then a general Neyman-Scott process should make use of this prior
information.

In the line transect survey we used, objects at a given perpendicular distance from a line transect are assumed to be detected with an equal probability. This assumption is not necessarily true in reality because larger objects are usually easier to detect than smaller ones and some kinds of metal are easier to detect than others. One way to deal with this issue is to model the detection function with some covariates of detected objects besides the distance information. There are lots of literature like Barry et al. [18] and Mack et al. [28] talking about modeling the detection function with covariates.

In the local sampling, we assume that clusters are identifiable. If all the clusters are well-separated, this assumption is generally acceptable. If the overlapped clusters consist of different types of UXO, the measurement technology may still tell the difference between them. But if two overlapped clusters consist of the same type of UXO, the identifiable assumption will no longer be appropriate. Brown and Cowling [30] proposed a criterion based on the likelihood ratio test to decide whether an initial cluster should be separated into two clusters or two neighboring clusters should be combined into one cluster.

For the distribution of the offspring positions around the parents, we assumed a radially symmetric bivariate normal distribution, i.e., the variance structure of the offspring position \((x, y)\) is \(\rho^2 I\), where \(I\) is a two-dimensional identity matrix. We can generalize the variance matrix from \(\rho^2 I\) to \(\Sigma\) with \(\Sigma = \begin{bmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{bmatrix}\), where \(\sigma_x^2\) and \(\sigma_y^2\) are marginal variances for \(x\) and \(y\), respectively; and \(\rho\) is the correlation between \(x\) and \(y\). Now there are three unknown parameters in \(\Sigma\). Since we don't have the distance information in the UXO line transect survey, one directional line transect data can only estimate one unknown parameter in \(\Sigma\) (or, one function of the three parameters, depending on the orientation of the transects). So we need to execute a three-directional (say vertical, horizontal, and diagonal) line transect survey to observe the data needed
to estimate all three unknown parameters in $\Sigma$. One way to estimate those parameters is to solve a system of three equations, the data observed from one set of transects lead to an equation by setting the marginal variance of that direction equal to the corresponding sample variance. We did a preliminary simulation study on the estimation of this general $\Sigma$, the results indicated that the estimators were very unstable. If we know the correlation $r$, then there are only two unknown parameters, the marginal variances of $x$ and $y$. In this situation, the estimation is much better. So when the variance structure for the object location is a general $\Sigma$, more work is needed to understand how effective transect surveys should be designed and how the resulting data should best be used for estimation.

Finally, one goal of the first stage sampling is to detect as many clusters as possible. If there is some prior information about the parent distribution, an important practical issue is how to take advantage of this prior information to design the sample paths under the restriction of a fixed length to minimize the number of missed clusters. While this aspect of the UXO sampling problem has not been explored in this dissertation, some research has been discussed in [10]. Further work might be undertaken to incorporate their ideas into our context.


