

## INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

# UMI

A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA  
313/761-4700 800/521-0600



MULTINOMIAL SMOOTHING AND INVESTIGATING  
ELEVATED DISEASE INCIDENCE BY PENALIZED  
LIKELIHOOD

By  
Robert Glen Downer

SUBMITTED IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY  
AT  
DALHOUSIE UNIVERSITY  
HALIFAX, NOVA SCOTIA  
APRIL 1997

© Copyright by Robert Glen Downer, 1997



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file Votre référence*

*Our file Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-24736-8

Canada

**DALHOUSIE UNIVERSITY**

**FACULTY OF GRADUATE STUDIES**

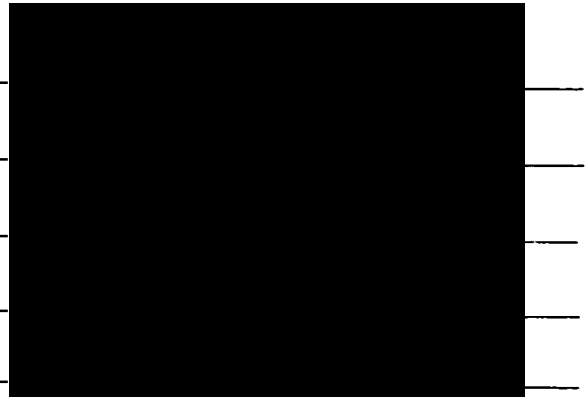
The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled “Multinomial Smoothing and Investigating Elevated Disease Incidence by Penalized Likelihood”

by Robert Glen Downer

in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dated: April 4, 1997

External Examiner  
Research Supervisor  
Examining Committee

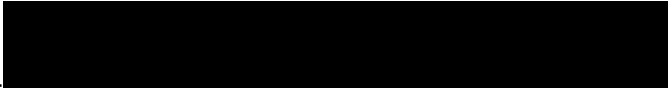


DALHOUSIE UNIVERSITY

Date: April 1997

Author: Robert Glen Downer  
Title: Multinomial Smoothing and Investigating Elevated  
Disease Incidence by Penalized Likelihood  
Department: Mathematics Statistics and Computing Science  
Degree: Ph.D. Convocation: May Year: 1997

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.



Signature of Author

THE AUTHOR RESERVES OTHER PUBLICATION RIGHTS, AND NEITHER THE THESIS NOR EXTENSIVE EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT THE AUTHOR'S WRITTEN PERMISSION.

THE AUTHOR ATTESTS THAT PERMISSION HAS BEEN OBTAINED FOR THE USE OF ANY COPYRIGHTED MATERIAL APPEARING IN THIS THESIS (OTHER THAN BRIEF EXCERPTS REQUIRING ONLY PROPER ACKNOWLEDGEMENT IN SCHOLARLY WRITING) AND THAT ALL SUCH USE IS CLEARLY ACKNOWLEDGED.

*To my father. Raymond Winsor Downer*

# Contents

<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>xi</b>
<b>Abstract</b>	<b>xii</b>
<b>Acknowledgements</b>	<b>xiii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	1
1.2 Tests of Clustering . . . . .	2
1.3 Spatial Autocorrelation methods . . . . .	4
1.4 Smoothing . . . . .	6
1.4.1 Background . . . . .	7
1.4.2 Smoothing Incidence Rates . . . . .	8
1.5 Outline of the Thesis . . . . .	10
<b>2 Smoothed Estimates of Incidence by Penalized Likelihood</b>	<b>11</b>
2.1 Poisson and Penalized Likelihood . . . . .	12
2.2 Multinomial Estimation . . . . .	14
2.3 Examples of Smoothed Estimates . . . . .	18
2.4 An Approximation for the Smoothed Estimate . . . . .	26
2.5 Moments . . . . .	30



2.5.1	Moments of the Unsmoothed Estimate . . . . .	30
2.5.2	Approximate Moments of $\hat{\gamma}_{sm}$ . . . . .	31
2.6	Bias and MSE . . . . .	34
2.7	Simulation Examples - Bias and MSE . . . . .	38
2.8	Further Investigations . . . . .	45
2.8.1	Reduced Estimators . . . . .	45
2.8.2	Normality of $\hat{\gamma}_{sm}$ . . . . .	51
<b>3</b>	<b>Testing for Elevated Rates</b>	<b>55</b>
3.1	Testing Strategy and Cluster Definition . . . . .	55
3.2	Testing with no smoothing . . . . .	56
3.3	Testing with smoothing . . . . .	59
3.3.1	Motivation . . . . .	59
3.3.2	Refinement of the test statistic for smoothed estimates . . . . .	60
3.3.3	Determining the critical region for $Z_{\lambda,red,i}$ . . . . .	64
3.4	Description of the Simulations . . . . .	70
3.5	Results - Size of Tests . . . . .	73
3.6	Results - Detection of Elevated Rates . . . . .	77
<b>4</b>	<b>Choice of Smoothing Parameter</b>	<b>94</b>
4.1	Introduction . . . . .	94
4.2	An MSE choice for $\lambda$ . . . . .	95
4.3	Case Deletion Criteria . . . . .	97
4.3.1	Predictive Loss . . . . .	98
4.3.2	A testing-based choice for $\lambda$ . . . . .	100
4.4	A combined strategy for choosing $\lambda$ . . . . .	102
4.5	Results - Combined Strategy . . . . .	106
<b>5</b>	<b>Application</b>	<b>113</b>
5.1	Estimation and Testing with fixed $\lambda$ . . . . .	113
5.2	Analysis using data-based selection of $\lambda$ . . . . .	119

<b>6 Discussion</b>	<b>123</b>
6.1 Summary . . . . .	123
6.2 Further Work . . . . .	124
<b>A Chapter 2 Appendices</b>	<b>127</b>
<b>B Chapter 3 Appendices</b>	<b>151</b>
<b>C Chapter 4 Appendices</b>	<b>180</b>
<b>D Chapter 5 Appendices</b>	<b>191</b>
<b>Bibliography</b>	<b>194</b>

# List of Tables

2.1	An example of smoothed estimates. Given are the estimates for 10 sites in a circle configuration with 100 cases and equal populations of 1000. . . . .	22
2.2	An example of smoothed estimates. Given are the estimates for 10 sites in a circle configuration with 100 cases and site 10 has a smaller population at $N_{10} = 500$ . . . . .	23
2.3	An example of smoothed estimates. Given are the estimates for 10 sites in a circle + 1 configuration with 100 cases. . . . .	24
2.4	An example of smoothed estimates. Given are the estimates for 10 sites in a circle + 1 configuration with 100 cases and site 10 has a smaller population at $N_{10} = 500$ . . . . .	25
2.5	Illustration of numerically smoothed estimate $\hat{\gamma}_{sm}$ and the approximate estimate $\hat{\gamma}_\lambda$ for a circle configuration of $k = 10$ sites with $m = 100$ cases. . . . .	29
2.6a	Observed Variance and Observed Biases calculated from a simulation of 5000 repetitions involving 10 sites, 100 cases in the circle + 1 configuration. Site 10 has elevated incidence. . . . .	36
2.6b	Observed Total Variance and Total Bias from the same simulation as described in Table 2.6a. . . . .	37
2.7a	Observed and Approximate Bias for Two Sites from a simulation of 5000 repetitions under complete homogeneity. There are 10 sites, 100 cases and site 10 is distant in the circle + 1 configuration. . . . .	40

2.7b	Total Observed and Approximate Mean Squared Error for the same simulation as described in Table 2.7a. . . . .	40
2.8a	Observed and Approximate Bias for Two Sites from a simulation of 5000 repetitions under a small departure from homogeneity. There are 10 sites. 100 cases and the distant site 10 has elevated incidence in the circle + 1 configuration. . . . .	42
2.8b	Total Observed and Approximate Mean Squared Error for the same simulation as described in Table 2.8a. . . . .	42
2.9a	Observed and Approximate Bias for Two Sites from a simulation of 5000 repetitions under a moderate departure from homogeneity. There are 10 sites. 100 cases and the distant site 10 has elevated incidence in the circle + 1 configuration. . . . .	43
2.9b	Total Observed and Approximate Mean Squared Error for the same simulation as described in Table 2.9a . . . . .	43
2.10	Total MSE of Reduced Estimators under a small departure from homogeneity. Results are taken from the same simulation as Tables 2.8a and 2.8b . . . . .	47
2.11	Total MSE of Reduced Estimators under a moderate departure from homogeneity. Results are taken from the same simulation as Tables 2.9a and 2.9b . . . . .	47
3.1	Homogeneity Simulations: Circle configuration. 20 sites. Presented are the fraction of 5000 repetitions with at least one test rejected and the maximum fraction of tests rejected at the site level . . . . .	75
3.2	Homogeneity Simulations: Circle + 1 configuration. 20 sites. Presented are the fraction of 5000 repetitions with at least one test rejected and the maximum fraction of tests rejected at the site level . . . . .	76

3.3	Ability to Detect Single Elevated Rate: Circle Configuration. 20 sites and 200 cases. Presented are fraction of 5000 repetitions with at least one test rejected, the fraction which include the elevated rate among those declared elevated and the fraction in which only the elevated rate is declared elevated . . . . .	79
3.4	Ability to Detect Single Elevated Rate: Circle + 1 Configuration. 20 sites and 200 cases. . . . .	80
3.5	Ability to Detect Two Contiguous Elevated Rates : Circle configuration. 20 sites, 200 cases. Equal Populations. Presented are fraction of 5000 repetitions with at least one test rejected, the fraction which include the first elevated rate among those declared elevated, the fraction which include the second elevated rate among those declared elevated and the fraction in which only the elevated rates are declared elevated	85
3.6	Ability to Detect Two Contiguous Elevated Rates : Circle + 1 configuration. 20 sites, 200 cases. Equal Populations. . . . .	86
3.7	Ability to Detect Single Elevated Rate : Random populations and locations Configuration. 20 sites for 200 and 400 cases . . . . .	90
3.8	Ability to Detect Many Elevated Rates : Bump configuration of seven contiguous elevated rates. 20 sites and 200 cases. Presented are the fraction of repetitions with at least one test rejected and the fraction with more than half of the elevated rates declared elevated. . . . .	93
4.1	Homogeneity Simulations with Data-based Choice of $\lambda$ for 20 sites and 200 cases . . . . .	107
4.2	Ability to Detect Single Elevated Rate with Data-based choice of $\lambda$ for 20 sites and 200 cases in two main configurations . . . . .	109
4.3	Ability to Detect Single Elevated Rate with Data-based choice of $\lambda$ for 20 sites and 200 cases in random populations/locations configuration	110
4.4	Ability to Detect Two Contiguous Elevated Rates with Data-based choice of $\lambda$ for 20 sites 200 cases in four configurations . . . . .	112

5.1	Case counts, populations and estimates for Nova Scotia gastric cancer incidence for five year period ending December 31, 1981. . . . .	115
5.2	Approximate Estimates and Test Statistics for Sites Declared Elevated for Nova Scotia gastric cancer data . . . . .	117

# List of Figures

2.1	Spatial configurations and populations for Tables 2.1 through 2.4 . . .	20
2.2	Observed and approximate total MSE from Tables 2.7b,2.8b,2.9b . . .	44
2.3	Observed MSE for smoothed estimators for simulation with one elevated Rate. small departure from homogeneity (results in Table 2.10)	48
2.4	Observed MSE for smoothed estimators for simulation with one elevated Rate. moderate departure from homogeneity ( results in Table 2.11 ) . . . . .	49
2.5a	Q-Q plots - Complete homogeneity: Site on circle in circle + 1 configuration. 20 sites and 200 cases . . . . .	53
2.5b	Q-Q plots - Complete homogeneity: Distant site in circle + 1 configuration. 20 sites and 200 cases . . . . .	54
3.1	Fraction of at least 1 Rejection. Single Elevated Rate for 20 sites. 200 and 400 cases. . . . .	82
3.2	Fraction of at least 1 Rejection. Two Contiguous Elevated Rates (from Table 3.5) . . . . .	88
4.1	Examples of CVPW function . . . . .	104
4.2	Examples of CVPW function . . . . .	105
5.1	Nova Scotia Map and municipalities: Sites with incidence declared elevated indicated by an x . . . . .	118
5.2	Change in estimates (with smoothing) versus maximum likelihood estimates and populations . . . . .	120

# Abstract

A compromise is needed between the presentation of highly variable estimates of incidence for subregions and an overall estimate for the region as a whole. Smoothed estimates are obtained by adding an inter-site distance penalty to a constrained multinomial likelihood. An approximation is developed for the smoothed estimate and approximate moments are obtained.

A test statistic is developed based on the smoothed estimates for detecting elevated incidence at a site relative to the others. An appropriate critical region for the test statistic is established. Clusters are defined to be contiguous elevated rates and the technique is evaluated through simulation. Significant increases in detection ability are observed in many situations and the size of the test is controlled. Techniques are also discussed for determining the smoothing parameter from the data.

The overall method is applied to gastric cancer incidence in Nova Scotia, Canada.



# Acknowledgements

I must express an eternal gratitude and appreciation to my parents, Raymond and Gloria. They have motivated my education and inspired my will to persevere. I would like to thank my family and friends for their constant encouragement and support. I would also like to thank my supervisor, David Hamilton, for his patience and advice.

# Chapter 1

## Introduction

In this chapter, the motivation and background literature for the thesis problem are described. Brief overviews of tests of clustering, spatial autocorrelation, and smoothing are given. The remaining chapters of the thesis are also outlined.

### 1.1 Motivation

This thesis was motivated by an analysis of gastric cancer incidence data for Nova Scotia, Canada. Interest was expressed in the variation of incidence rates across subregions and a method to detect elevated rates was required. Recent attention in gastric cancer research has focussed on the micro-organism *Helicobacter pylori* (*H.pylori*). Infection with *H.pylori* has been linked to chronic gastritis and ulcers which are established precursors for intestinal gastric carcinoma.

Although motivated by health outcome data, the methodology developed in this thesis is relevant to other aggregated spatial counts for which a risk population can be established for a given subregion. These applications include traffic accidents, forest infestation and disease prevalence in animal populations.

## 1.2 Tests of Clustering

Spatial disease data can be categorized according to two basic forms. The most common consists of case counts which have been aggregated over a known geographic unit. The other possibility, which is less common for human health outcomes, is to have a specific  $(x, y)$  location associated with each case. Data in the latter form could be considered in the alternative framework of spatial point processes (see Cliff and Ord, 1981; Diggle, 1983). With the data in one of these two forms, tests for clustering can be classified according to the overall purpose of the method. Most techniques indicate if there is clustering somewhere within the region of interest, while others give an indication of whether clustering has occurred near a specific hazard. Some methods focus on temporal as well as spatial clustering. There is a large body of literature in this area. The methods given in this section are only representative examples of the research. For detailed reviews, see Marshall (1991) or Hills and Alexander (1989).

Besag (1991) defines tests for detecting clusters as tests that look for 'individual hot spots of disease which merit further investigation and more detailed study'. This definition should be kept in mind in that such tests should generate subsequent investigation. A test of clustering is meant to be an initial step. Many apparent clusters can be easily explained by other factors or existing facts about the population.

One of the most commonly used statistics was one of the earliest developed. Mantel (1967) expressed the need to develop a sensitive procedure for detecting any non-random component of disease. He suggested the form  $\sum_{i=1}^k \sum_{j=1}^k f(x_{ij})g(y_{ij})$  for the detection of clustering in both space and time. In this expression,  $f$  is a measure of spatial closeness,  $g$  is a measure of temporal closeness,  $x_{ij}$  is Euclidean distance between the  $i$ 'th and  $j$ 'th cases and  $y_{ij}$  is the difference in the  $i$ 'th and  $j$ 'th case

onset times. The sum is over all possible pairs of cases. A null hypothesis of no clustering is equivalent to random spatial locations matched with random case onset times. Testing is done against a permutation distribution under the null hypothesis. The cell occupancy method of Ederer, Mantel, and Myers (1964) assumes the spatial data is in the first form mentioned and there are associated fixed time periods for the data aggregation. The test statistic reveals clustering if there is a tendency for some space-time cells to contain an excess beyond expectation under a hypothesis of no clustering. While temporal clustering alone is a related subject that will only be mentioned briefly here, the scan statistic (Naus, 1965 ; and Wallenstein, 1980) is commonly used for temporal disease data. Its concept is similar to the cell occupancy method in that a cell or interval of time is fixed. The test statistic is the maximum number of cases observed for intervals of this length over the entire time period.

While Mantel's statistic has been popular due to its simplicity, it has been criticized because it does not incorporate the underlying population structure (Roberson, 1990). Mantel's statistic motivated the development of many other methods. While using a multinomial model, Whittemore *et.al* (1987) developed a general test statistic in the Mantel form which utilizes subregion populations. The index of Tango (1984) is a special case of the Whittemore statistic.

Besag and Newell (1991) developed a method applicable when the location of each case is known. To test the null hypothesis that the distribution of cases is random, the authors count the number of administrative zones occurring within circles centered at each case. The risk population of small administrative zones is required. For data which is case location specific, Cuzick and Edwards (1990) suggested drawing a control set from the associated risk population and their test statistic counts the number of cases among  $k$  nearest neighbour individuals.

The geographical analysis machine (GAM), first introduced by Openshaw *et.al*

(1988) has recently become very popular as an exploratory visual technique. Although recent changes have been made, circles were originally displayed if there was an excess relative to the expected Poisson count. Criticism of this approach has involved the difficult estimation of the risk population for circular regions and complications associated with overlapping circles.

Raubertas (1985) developed a method based on a generalized linear model for data formatted into a table of time periods (rows) and geographical regions (columns). Departures from expectation are modelled into region, time and region-time effects. An overall test of clustering is given but the noteworthy difference with this technique is the attempt to identify the significance of contributions made by subregions. Unfortunately subjective definitions of regional neighbourhoods must be made to achieve this purpose.

In focussed tests of clustering, a smaller subregion is chosen for investigation due to some previously hypothesized hazard. Hills and Alexander (1989) propose a test statistic based on expected and observed mean distances from the source. Bithell and Stone (1989) suggest testing for monotonic decay of risk with increasing distance. The underlying logic and justification of these tests has been questioned (e.g. Besag and Newell, 1991). These discussions have centered on possible preselection bias of the subregion.

### **1.3 Spatial Autocorrelation methods**

In general, tests of clustering such as the ones given in the previous subsection assume independence of the observations. The spatial configuration is included after making

this assumption. Attempting to model spatial dependence is another alternative. Cliff and Ord (1981), Ripley (1981) and Cressie (1991) give extensive treatments of the subject in their texts on spatial data.

Cliff and Ord (1981) define spatial autocorrelation as systematic spatial variation. They describe various indices of spatial autocorrelation which attempt to determine whether an observed spatial pattern is significant and worth interpreting. The index of Moran (1948) is probably the most commonly used and is given by :

$$I = \frac{\bar{w} z' W z}{z' z} .$$

where  $W = ( w_{ij} )$  is a matrix of weights describing the closeness of region  $i$  to region  $j$  and  $\bar{w}$  is the average of the  $w_{ij}$ 's. The elements of the vector  $z$  are  $z_i = x_i - \bar{x}$  and  $x_i$  is the incidence rate of region  $i$ .  $I$  has been shown to be normally distributed for only moderately large number of subregions ( $n \geq 20$ ) but calculation of the first two moments requires either an assumption that the  $n$  subregional incidences are a random sample from a normal distribution or the randomization assumption in which one assumes each of the permutations of  $z$  are equally likely. This second assumption has been questioned (e.g. Besag, 1991). The incidence rates,  $x_i$ , of more rural counties will be more extreme and variable than those for more urban counties. In a series of simulations, Walter (1992) demonstrated the effect of different population sizes on the distribution of  $I$ . As a result, he adjusts the critical value of  $I$  prior to evaluating the spatial autocorrelation of the incidence of lip, stomach, and thyroid cancer in the counties of Ontario.

Cressie and Read (1989) suggested detailed modelling, which consists of several steps, in their analysis of Sudden Infant Death Syndrome (SIDS) data. The analysis consists of analyzing case counts within a table of rows and columns corresponding to divisions of latitude and longitude. The authors suggest removing possible row and column effects from their fitted model by using a median polish (Tukey, 1977) and

focus on a set of stationary residuals. Residual analysis makes use of the geostatistical ideas of the variogram and covariogram (Cressie, 1991). The variogram indicates the lag distance on the map for which independence can be assumed.

Modelling spatial dependence may also be considered within a regression context. Covariate data may be available at each location. The usual assumption of uncorrelated errors is replaced by the assumption that  $\text{Var}(\epsilon) = A\sigma^2$  (Cook and Pocock, 1983). The covariance or off-diagonal elements of  $A$  are usually a decreasing function of distance. The SIDS data was further analyzed by Cressie and Chan (1989) using a conditional autoregressive model (Besag, 1974). In this model, the modelling of the mean count of a particular subregion is conditional on the count of its neighbours. Space-time autoregressive moving average models have also been developed (Martin and Oeppen, 1975 ; Pfeifer and Deutch, 1980)

## 1.4 Smoothing

The idea of smoothing has a broad scope within the field of statistics. While the work within this thesis pertains to the topics of penalized likelihood, multinomial smoothing and disease mapping, other areas include density estimation, nonparametric and ridge regression and image processing. A comprehensive treatment of the subject can be found in Titterton (1985). The common link within the methodology is the belief that the true parameter(s) or function of interest has a smooth form. Some further background on smoothing and its context within this thesis are presented in this section.

### 1.4.1 Background

Raw unsmoothed estimates such as the observed incidence rates of case count divided by risk population constitute ‘ultra-rough’ estimates. Reporting an overall constant rate ignores the observed data and is ‘ultra-smooth’ (Titterington, 1985). The form or prescription for smoothing is defined by the smoothing technique chosen (Stone, 1974). The degree of smoothing is governed by a smoothing parameter.

The classical trade-off in smoothing is between the possible gains resulting from decreased variance in the estimate and the possible losses incurred from introducing bias. A common criterion for assessing the sampling properties of the estimates and choice of smoothing parameter is to minimize the total mean squared error given by the sum of the variances and the squared biases.

For a vector  $\mathbf{y}$  of length  $k$  distributed as multinomial with  $m$  cases and cell probability vector  $\boldsymbol{\theta}$ . Good (1965) was the first to suggest smoothing the maximum likelihood estimates for the cell probabilities. While the maximum likelihood estimate  $y_j/m$  is ‘ultra-rough’, the ‘ultra-smooth’ estimate for  $\hat{\theta}_j$  is the constant  $1/k$ . For  $0 < \alpha < 1$ , his simple prescription of convex smoothing takes the form

$$\hat{\theta}_{j,\alpha} = \alpha \left( \frac{1}{k} \right) + (1 - \alpha) \hat{\theta}_j \quad . \quad (1.1)$$

This smoothing prescription has some intuitive appeal due to its simplicity and will be discussed again later in the thesis. In a Bayesian framework, Fienberg and Holland (1973) as well as Leonard (1977) obtain smoothed multinomial estimates through the use of this type of smoothing prescription.

In techniques which use penalized distance or penalized roughness, estimation



is performed under some constraint. In multinomial smoothing, some examples of penalties are  $\sum_{i=1}^k \sum_{j=1}^k (\theta_i - \theta_j)^2$  and  $\sum_{i=1}^{k-1} (\theta_i - \theta_{i+1})^2$ . The penalty chosen depends on the context. In these two penalties, the former concentrates on general roughness while the latter assumes a meaningful ordering among the cells and penalizes locally through first differences. In a spatial context, the approach of smoothing via penalized distance will be an integral part of the smoothing technique presented in Chapter 2.

Bayesian estimation is a large fraction of the smoothing literature and will be discussed in the next subsection in the context of smoothing incidence rates. The issue of the choice of smoothing parameter will be addressed in Chapter 4 and will involve some of the concepts outlined above.

### 1.4.2 Smoothing Incidence Rates

Producing geographic estimates of incidence has become quite common. Atlases of incidence and/or mortality are now becoming more widely available (see Walter and Birnie, 1991). Case counts are usually aggregated to a regional level such as municipality or county and census data is referenced to establish the risk population.

A compromise is needed between the presentation of raw incidence rates for the subregions and an overall mean rate for the entire map. Highly variable standardized rates from a subregion of low population may hide the true underlying pattern. Believing that the true unknown incidence for the map takes on a smooth form and that these subregion estimates are not independent are reasonable assumptions. Borrowing information from geographically close areas is a logical step towards producing a map of more stable estimates.

Empirical Bayes (EB) methods have been popular for the purpose of stabilizing the observed rates. A comprehensive review can be found in Clayton and Bernardinelli (1992). Assume we have a case count of  $y_j$  for subregion/site with population  $N_j$ . For the  $k$  sites, denote the observed raw incidence vector by  $\hat{\mathbf{p}}$ . The EB approach uses a multivariate prior for the underlying incidence vector  $\mathbf{p}$ . Maximization of the resulting posterior gives the EB estimates for the map. Various techniques may be used to estimate the parameters of the prior for  $\mathbf{p}$ . These techniques may be computationally intensive but they have generally been preferred to the numerical methods required to maximize the posterior distribution of a fully Bayesian approach. Application of the Gibbs sampler (Geman and Geman, 1984) has generated some renewed interest in the fully Bayesian approach.

Approaches within the EB framework can generally be classified by the type of prior model used. Global approaches shrink the estimates towards a global expectation. If there is no spatial dependence included in the prior or sampling distributions, these estimates remain unchanged by a rearrangement of spatial locations (see Devine *et al.*, 1994). Local approaches shrink the estimate towards a local mean and have generally used spatial auto-correlation models to model spatial dependence among the  $p_i$ . A weighting or adjacency matrix relates incidences  $p_i$  and  $p_j$  as in the conditional autoregressive model (CAR) described by Besag (1974). In a fully Bayesian set-up, Besag (1989) models  $p_i$  as the sum of a spatially defined global variable (the realization of a spatial process) and another localized variable which has no spatial definition. Marshall (1991) proposed a local moment based EB estimator. This method relies on the reasonable assumption that local mean incidence is constant over a certain predefined neighbourhood.

## 1.5 Outline of the Thesis

In general, tests of clustering do not highlight specific areas contributing to clustering and do not attempt to test for elevated rates. Furthermore, these tests typically don't give useful subregional estimates of incidence. The idea of smoothing to produce an estimate with better sampling properties was considered appropriate in the context of the motivating problem. However, smoothing by itself does not go far enough to address the question of elevated rates or clusters. A testing strategy based on smoothed estimates is developed.

In Chapter 2, smoothed estimation of incidence rates by penalized likelihood is proposed. Conditional on the total number of cases, smoothed estimates are obtained by adding an inter-site distance penalty to a constrained multinomial likelihood. An approximation for the smoothed estimate is developed and approximate moments of the smoothed estimate are discussed. Chapter 3 outlines a testing strategy to identify elevated rates and to define clusters. Testing without smoothing and testing based on smoothed estimates is described. Simulations evaluate the methodology. The problem of choosing the smoothing parameter is addressed in Chapter 4 and a strategy for this choice is suggested. The overall method is applied to Nova Scotia gastric cancer data in Chapter 5. In Chapter 6, the thesis is briefly reviewed and future work is outlined.

## Chapter 2

# Smoothed Estimates of Incidence by Penalized Likelihood

In this chapter, smoothed estimation of incidence rates by penalized likelihood is presented. In Section 2.1, the form of the data and the Poisson likelihood approach to estimation are reviewed. Penalized likelihood estimation is introduced. Conditional on the total number of cases, a reparametrization and smoothed estimation from a penalized multinomial likelihood are presented in Section 2.2. Examples of the smoothed estimate are given in Section 2.3 and an approximation for the smoothed estimate is developed in Section 2.4. Moments of the unsmoothed estimate and approximate moments of the smoothed estimate are presented in Section 2.5. The sampling properties of the smoothed estimate are addressed through bias and MSE in Section 2.6. Simulations compare observed bias and MSE to their approximations in Section 2.7. Other investigations which discuss simplified (reduced) estimators and the normality of the smoothed estimate are presented in Section 2.8.

## 2.1 Poisson and Penalized Likelihood

For the time period of interest, assume we have data of the form  $(r_i, s_i, y_i, N_i)$ ,  $i = 1, \dots, k$ , where  $r_i$  and  $s_i$  are the coordinates of site  $i$ ,  $y_i$  is the case count, and  $N_i$  is the associated risk population. For example, the Nova Scotia gastric cancer data consists of case counts at the municipality or county level over specific time periods. Census information was referenced to ascertain the relevant risk populations. The assigned coordinates were the latitude and longitude of the municipality/county seat. Another example of rare event data at known spatial locations is traffic accidents. Estimates of incidence rates can be found by maximizing a Poisson likelihood. In this section the case counts,  $y_i$ , are considered to be independent Poisson random variables and the concept of smoothing by penalized likelihood is introduced. However, most of the work of this thesis considers the distribution of the case vector  $\mathbf{y} = (y_1, \dots, y_k)$  conditional on the case total ( $m = \sum_{i=1}^k y_i$ ) for the region. The issue of overdispersion is not addressed in this thesis.

Assuming that the  $y_i$  are distributed as independent Poisson random variables with mean  $N_i p_i$ , then the log likelihood up to an additive constant is

$$l(\mathbf{p}) = - \sum_{i=1}^k N_i p_i + \sum_{i=1}^k y_i \log(N_i p_i) \quad (2.1)$$

and the maximum likelihood estimate for  $p_j$  is  $y_j/N_j$ .

Assume now that the unknown incidence rates  $\mathbf{p} = (p_1 \dots p_k)'$  have a smooth form as discussed in Chapter 1. The spatial aspect of the problem can be introduced through a roughness penalty which incorporates the inter-site distances and the differences among the rates. For any pair of sites, the contribution to the penalty should decrease with increasing inter-site distance  $d_{ij}$  and increase with the absolute

difference in the rates  $p_i$  and  $p_j$ . Therefore, the suggested penalty is

$$pen(\mathbf{p}) = \sum_i \sum_{j \neq i} \frac{1}{d_{ij}} (p_i - p_j)^2 \quad . \quad (2.2)$$

The penalty  $pen(\mathbf{p})$  can be written as a quadratic form as

$$pen(\mathbf{p}) = 2\mathbf{p}'(\mathbf{B} - \mathbf{E})\mathbf{p} \quad . \quad (2.3)$$

where the matrix  $\mathbf{E}$  has  $\epsilon_{ij}$  as its  $(i, j)$ th element and diagonal element  $\epsilon_{ii}$  defined to be 0 and  $\mathbf{B}$  is a diagonal matrix with  $(i, i)$ th element  $b_i = \sum_{j \neq i} \epsilon_{ij}$  where  $\epsilon_{ij} = 1/d_{ij}$ .

Up to an additive constant, the penalized log likelihood can be written as

$$l_{pen}(\mathbf{p}) = l(\mathbf{p}) - h \sum_i \sum_{j \neq i} \frac{1}{d_{ij}} (p_i - p_j)^2 \quad . \quad (2.4)$$

where  $h$  is the smoothing parameter. Assuming the same relative positions for the sites, distance measured in different units gives the same penalized estimates as before since a constant will factor out of the sum of squares.

The estimates  $\hat{\mathbf{p}}_{sm}$ , obtained by maximizing this function, are a compromise between the maximum likelihood estimates,  $\hat{p}_{ml,j} = y_j/N_j$ , and a constant estimate which gives a zero penalty. Denote the total population  $\sum_{i=1}^k N_i$  as  $N_+$ . Although any constant  $p_o$  makes the penalty zero, as  $h$  tends to infinity, the estimate  $\hat{p}_{f,j} = m/N_+$  maximizes  $l_{pen}(\mathbf{p})$  evaluated at  $\mathbf{p}_o = (p_o, \dots, p_o)'$ . The ultra-smooth or flat estimate  $\hat{p}_{f,j}$  puts no faith in the data. When  $h = 0$ , the maximum likelihood estimates are unbiased while the ultra-smooth estimates have expectation

$$\begin{aligned} \mathbf{E} \left[ \frac{m}{.N_+} \right] &= \frac{\sum_{i=1}^k .N_i p_i}{.N_+} \\ &= \bar{p} \quad . \end{aligned} \tag{2.5}$$

which is an average rate, weighted by population.

Downer (1996), using the penalized likelihood  $L_{pen}(\mathbf{p})$ , studied the normality and sampling properties of the resulting smoothed estimate vector  $\hat{\mathbf{p}}_{sm}$ . Penalized likelihood can also be viewed in a Bayesian or Empirical Bayes context. Within the Empirical Bayes framework, Clayton (1990), Bernardinelli and Montomoli (1992) as well as Clayton and Bernardinelli (1992) have discussed maximizing the posterior distribution as maximization of a penalized likelihood.  $L_{pen}(\mathbf{p}) = L(\mathbf{p}|y)\pi(\mathbf{p})$  is the penalized likelihood and the penalty corresponds to the logarithm of the multivariate prior distribution  $\pi(\mathbf{p})$  for  $\mathbf{p}$ . The penalty given in (2.3) can be interpreted as a multivariate normal prior for  $\mathbf{p}$  centered at  $\mathbf{p} = \mathbf{0}$ . This prior is degenerate because  $(\mathbf{B} - \mathbf{E}) \mathbf{1} = \mathbf{0}$  where  $\mathbf{1}$  is the vector of ones.

## 2.2 Multinomial Estimation

The eventual goal of this thesis is testing for elevated rates. The magnitudes of the individual rates carry no information about the differences among them, so it is generally accepted practice to condition on the total number of cases, which is an ancillary statistic (Lehmann, 1986 ; McCullagh and Nelder, 1989). The distribution of the case vector  $\mathbf{y}$  conditional on the total number of cases is multinomial with index

$m = \sum_{i=1}^k y_i$  and cell probabilities  $\theta_i = .N_i p_i / \sum_j .N_j p_j$ ,  $i = 1, \dots, k$ . The multinomial probabilities  $\boldsymbol{\theta} = (\theta_1, \dots, \theta_k)'$  depend on the rates and the risk populations. Although there are  $k$  distinct rates, the multinomial probabilities sum to one and this implies a constraint on the rates. To simplify the dependence of the cell probabilities on the rates, define the rate ratio

$$\gamma_i = \frac{p_i}{\bar{p}} \quad . \quad (2.6)$$

where  $\bar{p}$  is as in (2.5), so

$$\theta_i = \frac{.N_i \gamma_i}{.N_+} \quad . \quad (2.7)$$

A rate ratio  $\gamma_i$  which is greater than one indicates a rate which is elevated relative to the weighted average.

The multinomial log likelihood for  $\boldsymbol{\gamma}$  is, to an additive constant,

$$l(\boldsymbol{\gamma}) = \sum_{i=1}^k y_i \log \left( \frac{.N_i \gamma_i}{.N_+} \right) \quad . \quad (2.8)$$

subject to  $\sum_{i=1}^k .N_i \gamma_i / .N_+ = 1$ . To an additive constant, this becomes

$$l(\boldsymbol{\gamma}) = \sum_{i=1}^k y_i \log(\gamma_i) \quad .$$

subject to the constraint  $\sum_{i=1}^k .N_i \gamma_i / .N_+ = 1$ . In Appendix 2.1, it is shown that the maximum likelihood estimate for  $\gamma_i$  is  $\hat{\gamma}_{ml,i} = (y_i / .N_i) / (m / .N_+) = \hat{p}_{ml,i} / \hat{p}_f$ .

The reparametrization to  $\boldsymbol{\gamma}$  can be utilized to form an equivalent penalty based on inter-site distance and rate ratio difference. From (2.2),



$$pen(\mathbf{p}) = \bar{p}^2 \sum_i \sum_{j \neq i} \frac{1}{d_{ij}} \left( \frac{p_i}{\bar{p}} - \frac{p_j}{\bar{p}} \right)^2 = \bar{p}^2 \sum_i \sum_{j \neq i} \frac{1}{d_{ij}} (\gamma_i - \gamma_j)^2 \quad . \quad (2.9)$$

Therefore a reasonable choice for the penalty as a function of  $\gamma$  is

$$pen(\gamma) = \sum_i \sum_{j \neq i} \frac{1}{d_{ij}} (\gamma_i - \gamma_j)^2 \quad . \quad (2.10)$$

or in matrix form.

$$pen(\gamma) = 2\gamma'(\mathbf{B} - \mathbf{E})\gamma \quad . \quad (2.11)$$

where  $\mathbf{B}$  and  $\mathbf{E}$  are as in (2.3). The penalty is zero if all the rate ratios  $\gamma_i$  equal one (i.e. complete homogeneity). Assuming the same relative positions for the sites. distance measured in different units gives the same penalized estimates as before since a constant will factor out of the sum of squares.

An alternative definition of the penalty involves multiplication by the average inter-site distance  $\bar{d}$ . Multiplication by  $\bar{d}$  may be advantageous as both  $l(\gamma)$  and  $pen(\gamma)$  are then dimensionless and the choice of smoothing parameter does not depend on the units of measurement of the inter-site distances.

The penalty is another constraint on the estimation and hence subtracting the penalty (2.10), weighted by a smoothing parameter  $\lambda$ , forms the penalized log likelihood. Subject to the multinomial constraint  $\sum_{i=1}^k (N_i \gamma_i / N_+) = 1$ , the penalized log likelihood is

$$\sum_{i=1}^k y_i \log(\gamma_i) - \lambda \sum_i \sum_{j \neq i} \frac{1}{d_{ij}} (\gamma_i - \gamma_j)^2 \quad .$$

The constraint equation  $\sum_{i=1}^k \left( \frac{N_i \gamma_i}{N_+} \right) - 1 = 0$  can be included as a Lagrangian giving

$$l_{pen}(\boldsymbol{\gamma}) = \sum_{i=1}^k y_i \log(\gamma_i) - o \left( \sum_{i=1}^k \left( \frac{N_i \gamma_i}{N_+} \right) - 1 \right) - \lambda \sum_i \sum_{j \neq i} \frac{1}{d_{ij}} (\gamma_i - \gamma_j)^2. \quad (2.12)$$

$\lambda = 0$  gives the unpenalized likelihood and no smoothing is performed. For  $\lambda > 0$ , the smoothed estimates,  $\hat{\gamma}_{sm,i}$ , which optimize  $l_{pen}$  are a compromise between the maximum likelihood estimates and the completely smoothed or flat estimate  $\hat{\gamma}_{f,i} = 1$ . The smoothed estimates are the solutions to the penalized likelihood equations obtained by differentiating  $l_{pen}$  with respect to  $\gamma_i$  (for  $i = 1, \dots, k$ ) and  $o$ , and equating to 0. Differentiating with respect to  $o$  gives

$$\frac{\partial l_{pen}}{\partial o} = - \sum_{i=1}^k \left( \frac{N_i \gamma_i}{N_+} \right) + 1. \quad (2.13)$$

and solving gives  $\sum_{i=1}^k N_i \gamma_i = N_+$ . Differentiating with respect to  $\gamma_i$  gives

$$\frac{\partial l_{pen}}{\partial \gamma_i} = \frac{y_i}{\gamma_i} - o \left( \frac{N_i}{N_+} \right) - \lambda \frac{\partial pen}{\partial \gamma_i}. \quad (2.14)$$

Multiplying by  $\gamma_i$  and summing over  $i = 1, \dots, k$  gives

$$m - o \sum_{i=1}^k \left( \frac{N_i \gamma_i}{N_+} \right) - \lambda \sum_{j=1}^k \gamma_j \frac{\partial pen}{\partial \gamma_j} = 0.$$

and equating to zero gives

$$o = m - \lambda \sum_{j=1}^k \gamma_j \frac{\partial pen}{\partial \gamma_j}. \quad (2.15)$$

Substituting (2.15) into (2.14) and equating to zero gives

$$\frac{\partial l_{pen}}{\partial \gamma_i} = \frac{y_i}{\gamma_i} - \frac{m N_i}{N_+} - \lambda \left( \frac{\partial pen}{\partial \gamma_i} - \frac{N_i}{N_+} \sum_{j=1}^k \gamma_j \frac{\partial pen}{\partial \gamma_j} \right). \quad (2.16)$$

where the derivative of the penalty with respect to  $\gamma_i$  is

$$\frac{\partial pen}{\partial \gamma_i} = -b_i \gamma_i - \sum_{j \neq i} \epsilon_{ij} \gamma_j \quad (2.17)$$

and  $b_i$  and  $\epsilon_{ij}$  are as described in Section 2.1.

The  $k$  equations  $\partial l_{pen} / \partial \gamma_i = 0$  for  $i = 1, \dots, k$  are nonlinear in the  $\gamma_i$ 's and an explicit solution isn't possible. A solution can be found numerically. A Fortran quasi-Newton optimization routine (Gill and Murray, 1976) was used to obtain the solution, which is denoted as  $\hat{\gamma}_{sm}$ . Examples of these numerically obtained smoothed estimates are given in the next section.

A smoothed estimate for site  $i$  can also be obtained via the convex smoothing prescription given by (1.1). The approximate relationship between this convex estimator and the one proposed in this section is discussed in Appendix 2.2.

## 2.3 Examples of Smoothed Estimates

The effect of smoothing on the estimates depends on several factors: the spatial configuration, the populations  $N_j$ , the number of cases  $m$ , the number of sites  $k$  and the parameter vector  $\gamma$ .

The examples given in this section display the effect of smoothing for two different spatial configurations and two different population vectors. The number of cases and sites remain constant. Only one incidence rate vector  $\mathbf{p}$  was used and by definition the rate ratio vector  $\gamma$  changes with any different population.

Many possibilities come to mind for levels of the spatial configuration factor. The two levels chosen minimize and maximize the coefficient of variation of the distances  $d_{ij}$  for  $k \geq 8$ . The minimum corresponds to  $k-1$  sites which are equidistant on a circle with a  $k$ 'th site at the center. The maximum corresponds to  $k-1$  sites equidistant on a circle while the  $k$ 'th site is infinitely far from the others. The position of this distant site was arbitrary but the two spatial configurations are very dissimilar using the coefficient of variation criterion.

A logical choice for the population factor was equal populations. The other choice changes just one of these populations to half the others. Estimates at the site with the smaller population have a larger variance and hence the effect of smoothing is more pronounced.

Tables 2.1 through 2.4 present examples of the effect of smoothing on a single case vector for  $k = 10$  sites with  $m = 100$  cases. The true rate at site 10 is actually slightly elevated at  $p_k = .0075$  while  $p_j = .005$  for sites 1 through 9. In Tables 2.1 and 2.3, the populations are all 1000. Reparametrization leads to rate ratios  $\gamma_j = 0.952$  for  $j = 1, \dots, 9$  and  $\gamma_{10} = 1.43$ . In Tables 2.2 and 2.4, in which site 10 has population 500, reparametrization gives rate ratios  $\gamma_j = 0.974$  for  $j = 1, \dots, 9$  and  $\gamma_{10} = 1.46$ . The spatial configuration/population combinations are shown in Figure 2.1 and given below

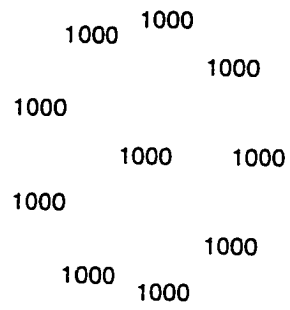
Table 2.1: Circle of 9 sites with site 10 in the center at  $N_j = 1000$  for  $j = 1, \dots, 10$ :

Table 2.2: Circle of 9 sites with site 10 in the center.  $N_{10} = 500$  and  $N_j = 1000$  for  $j = 1, \dots, 9$ :

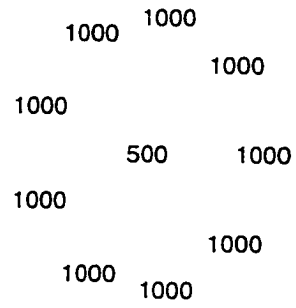
Table 2.3: Circle of 9 sites centered at (1.1) with site 10 located at (1.5).  $N_j = 1000$  for  $j = 1, \dots, 10$ :

Table 2.4: Circle of 9 sites centered at (1.1) with site 10 located at (1.5).  $N_{10} = 500$  and  $N_j = 1000$  for  $j = 1, \dots, 9$ .

Configuration for Table 2.1

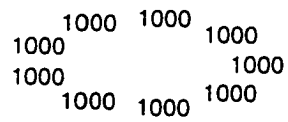


Configuration for Table 2.2



Configuration for Table 2.3

1000



Configuration for Table 2.4

500

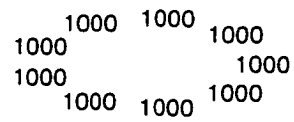


Figure 2.1: Spatial Configurations and Populations for Tables 2.1 through 2.4

The smoothing parameters used in these examples are somewhat arbitrary. The values of  $\lambda$  were chosen simply to illustrate various amounts of smoothing. In all four tables, the smoothed estimates move toward 1.0 as  $\lambda$  increases. If  $\hat{\gamma}_{ml,i} > 1$ , then  $\hat{\gamma}_{sm,i}$  is less than  $\hat{\gamma}_{ml,i}$ . The effect of any smoothing is largest for those sites in which  $|\hat{\gamma}_{ml} - 1|$  is large. In Table 2.1, the largest  $\hat{\gamma}_{ml,i} = 1.5$  at site 10 (the elevated site) and  $\lambda = .025$  gives a smoothed estimate  $\hat{\gamma}_{sm,10} = 1.446$ . At site 8, only six cases were observed and  $\hat{\gamma}_{ml,8} = 0.60$  is smoothed to  $\hat{\gamma}_{sm,8} = 0.622$ . Most of the simulations presented in this thesis represent no more than a small departure from homogeneity and involve a set of similar populations. In these situations,  $\hat{\gamma}_{ml,i}$  greater than 1.0 are generally smoothed down and  $\hat{\gamma}_{ml,i}$  less than 1.0 are smoothed up, but situations have been observed where these relations do not hold (see Chapter 5).

The spatial configuration is the same in Tables 2.1 and 2.2. The center site 10 has elevated incidence in both tables but has population 500 in Table 2.2. In Table 2.1,  $\hat{\gamma}_{ml,10} = 1.50$  and  $\hat{\gamma}_{sm,10} = 1.446$  at  $\lambda = .025$ . However, the estimate drops from  $\hat{\gamma}_{ml,10} = 2.09$  to  $\hat{\gamma}_{ml,10} = 1.94$  at  $\lambda = .01$  in Table 2.2. The effect of smoothing is more pronounced when site 10 has the smaller population. Sites with population 1000 are smoothed by approximately the same amount for the same value of  $\lambda$ . One can also observe that the site with the smallest population exhibits the greatest amount of smoothing in Tables 2.3 and 2.4.

Comparing the effect of smoothing on the large estimate at site 10 in Tables 2.1 and 2.3 illustrates the effect of spatial configuration. In Table 2.3, site 10 is removed from the other nine on the circle and is at a distance of five units from the center. A larger  $\lambda$  must be applied to see the same effect on this estimate in Table 2.3, as is seen in Table 2.1. Examples with these same two spatial configurations and populations are given for  $m = 200$  cases in Appendix 2.3.

Table 2.1

## Example of Smoothed Estimates

Circle Configuration: 10 sites, 100 cases, Equal Populations

Site	Coords	$\gamma$	$N$	$y$	$\lambda$			
					0	.025	.050	.256
					$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$
1	(1.76.1.64)	0.95	1000	9	0.900	0.913	0.924	0.967
2	(1.17.1.98)	0.95	1000	11	1.100	1.099	1.097	1.070
3	(0.5.1.86)	0.95	1000	12	1.200	1.189	1.179	1.116
4	(0.06.1.34)	0.95	1000	9	0.900	0.912	0.922	0.962
5	(0.06.0.66)	0.95	1000	9	0.900	0.909	0.917	0.951
6	(0.50.0.13)	0.95	1000	8	0.800	0.814	0.827	0.892
7	(1.17.0.02)	0.95	1000	9	0.900	0.907	0.913	0.941
8	(1.76.0.36)	0.95	1000	6	0.600	0.622	0.642	0.774
9	(2.00.1.00)	0.95	1000	12	1.200	1.185	1.172	1.101
10	(1.00.1.00)	1.43	1000	15	1.500	1.446	1.402	1.222

Table 2.2

## Example of Smoothed Estimates

Circle Configuration: 10 sites, 100 cases, One Smaller Population

Site	Coords	$\gamma$	N	y	$\lambda$			
					0	.01	.02	.168
					$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$
1	(1.76.1.64)	0.97	1000	10	0.950	0.959	0.965	0.987
2	(1.17.1.98)	0.97	1000	7	0.665	0.679	0.692	0.790
3	(0.5.1.86)	0.97	1000	12	1.140	1.139	1.136	1.101
4	(0.06.1.34)	0.97	1000	8	0.760	0.774	0.785	0.865
5	(0.06.0.66)	0.97	1000	10	0.950	0.960	0.967	0.996
6	(0.50.0.13)	0.97	1000	10	0.950	0.962	0.970	0.997
7	(1.17.0.02)	0.97	1000	16	1.520	1.507	1.484	1.336
8	(1.76.0.36)	0.97	1000	6	0.570	0.585	0.599	0.721
9	(2.00.1.00)	0.97	1000	10	0.950	0.959	0.966	0.987
10	(1.00.1.00)	1.46	500	11	2.090	1.943	1.849	1.412



**Table 2.3**

Example of Smoothed Estimates

Circle + 1 Configuration: 10 sites, 100 cases, Equal Populations

Site	Coords	$\gamma$	N	y	$\lambda$			
					0	.08	.16	.064
					$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$
1	(1.76.1.64)	0.95	1000	7	0.700	0.740	0.771	0.846
2	(1.17.1.98)	0.95	1000	7	0.700	0.746	0.781	0.857
3	(0.5.1.86)	0.95	1000	12	1.200	1.140	1.116	1.061
4	(0.06.1.34)	0.95	1000	9	0.900	0.922	0.935	0.957
5	(0.06.0.66)	0.95	1000	9	0.900	0.923	0.937	0.959
6	(0.50.0.13)	0.95	1000	11	1.100	1.082	1.071	1.041
7	(1.17.0.02)	0.95	1000	13	1.300	1.239	1.187	1.108
8	(1.76.0.36)	0.95	1000	9	0.900	0.916	0.923	0.949
9	(2.00.1.00)	0.95	1000	7	0.700	0.741	0.775	0.850
10	(1.00.5.00)	1.43	1000	16	1.600	1.542	1.486	1.324

Table 2.4

Example of Smoothed Estimates

Circle + 1 Configuration: 10 sites, 100 cases, One Smaller Population

Site	Coords	$\gamma$	N	y	$\lambda$			
					0	.05	.10	.495
					$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$
1	(1.76.1.64)	0.97	1000	11	1.045	1.037	1.031	1.009
2	(1.17.1.98)	0.97	1000	9	0.855	0.874	0.889	0.940
3	(0.5.1.86)	0.97	1000	12	1.140	1.126	1.140	1.068
4	(0.06.1.34)	0.97	1000	12	1.140	1.127	1.117	1.072
5	(0.06.0.66)	0.97	1000	11	1.045	1.041	1.040	1.029
6	(0.50.0.13)	0.97	1000	11	1.045	1.041	1.037	1.021
7	(1.17.0.02)	0.97	1000	9	0.855	0.8708	0.881	0.930
8	(1.76.0.36)	0.97	1000	10	0.950	0.9518	0.954	0.964
9	(2.00.1.00)	0.97	1000	7	0.665	0.694	0.726	0.831
10	(1.00.5.00)	1.46	500	8	1.52	1.426	1.388	1.269

## 2.4 An Approximation for the Smoothed Estimate

One can't directly see the dependence of the smoothed estimate  $\hat{\gamma}_{sm}$  on the various factors mentioned in the previous section. An approximation expressed as a function of the smoothing parameter  $\lambda$ , the populations, the inter-site distances and the case counts is most useful for showing the effect of these factors on smoothing. A quadratic expansion for  $\hat{\gamma}_{sm,i}$  about  $\lambda = 0$  is of the form

$$\hat{\gamma}_{\lambda,i} = \hat{c}_i + \lambda \hat{l}_i + \lambda^2 \hat{q}_i \quad . \quad (2.18)$$

where  $\hat{c}_i, \hat{l}_i, \hat{q}_i$  are constant, linear, and quadratic coefficients. Expressions for  $\hat{c}_i, \hat{l}_i$  and  $\hat{q}_i$  are obtained by substituting  $\hat{\gamma}_{\lambda,i}$  into (2.16) and equating to zero. Multiplying though by  $\hat{\gamma}_{\lambda,i}$  gives

$$\begin{aligned} y_i &= \left( \frac{m}{N_+} \right) N_i (\hat{c}_i + \lambda \hat{l}_i + \lambda^2 \hat{q}_i) \\ &- \lambda \left[ (\hat{c}_i + \lambda \hat{l}_i + \lambda^2 \hat{q}_i) 4b_i \hat{\gamma}_{\lambda,i} - \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_{\lambda,j} + \frac{N_i \hat{\gamma}_i}{N_+} \sum_{j=1}^k (\hat{c}_j + \lambda \hat{l}_j + \lambda^2 \hat{q}_j) \frac{\partial p \epsilon n}{\partial \gamma_{ij}} \right] = 0. \end{aligned}$$

Collecting terms that are constant in  $\lambda$ , these terms must equal 0 so

$$y_i - \frac{m}{N_+} N_i \hat{c}_i = 0 \quad .$$

and

$$\hat{c}_i = \frac{y_i N_i}{m N_+} \quad .$$

This constant term is  $\hat{\gamma}_{ml,i}$ , the maximum likelihood estimate for  $\gamma_i$ .

Similarly, if  $\hat{\gamma}_{ml,i}$  is denoted simply as  $\hat{\gamma}_i$ , collecting terms that are linear and quadratic in  $\lambda$  and solving gives

$$\hat{l}_i = \frac{N_+}{m} \left\{ -\frac{1}{N_i} \left[ b_i \hat{\gamma}_i^2 - \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] + \frac{1}{N_+} \left[ \hat{\gamma}_i \sum_r b_r \hat{\gamma}_r^2 - \hat{\gamma}_i \sum_r \sum_{s \neq r} \epsilon_{rs} \hat{\gamma}_r \hat{\gamma}_s \right] \right\} \quad (2.19)$$

and

$$\hat{q}_i = \frac{N_+}{m} \left\{ -\frac{1}{N_i} \left[ 2b_i \hat{\gamma}_i \hat{l}_i - \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \hat{l}_j - \hat{l}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] + \frac{1}{N_+} \left[ 2\hat{\gamma}_i \sum_j b_j \hat{\gamma}_j \hat{l}_j + \hat{\gamma}_i \sum_j b_j \hat{\gamma}_j^2 \right] - \frac{1}{N_+} \left[ \hat{\gamma}_i \sum_r \sum_{s \neq r} \epsilon_{rs} \hat{\gamma}_r \hat{l}_s \hat{\gamma}_i + \sum_r \sum_{s \neq r} \epsilon_{rs} \hat{\gamma}_i \hat{l}_r + \hat{l}_i \sum_r \sum_{s \neq r} \epsilon_{rs} \hat{\gamma}_r \hat{\gamma}_s \right] \right\} \quad (2.20)$$

The approximate smoothed estimate for site  $i$ ,  $\hat{\gamma}_{\lambda,i}$ , is the maximum likelihood estimate plus terms which constitute the smoothing. These smoothing terms involve  $\lambda$ , the inter-site distances, the populations and the maximum likelihood estimates. Again denoting  $\hat{\gamma}_{ml,i}$  by  $\hat{\gamma}_i$ , the linear term of the approximation is

$$\hat{l}_i = \frac{-N_+}{m} \hat{\gamma}_i \left[ \frac{1}{N_i} \frac{\partial \text{pen}(\hat{\gamma})}{\partial \gamma_i} - \sum_r \frac{N_r \hat{\gamma}_r}{N_+} \frac{\partial \text{pen}(\hat{\gamma})}{N_r \partial \gamma_r} \right]. \quad (2.21)$$

An examination of this expression reveals that  $\hat{l}_i$  is proportional to  $\hat{\gamma}_i$  as noticed in the examples. However this estimate is multiplied by a correction of two terms involving

the other factors. The second term in the square brackets is the same for all  $i$  and is a weighted average of the first terms for all sites. Empirical evidence has shown that the second term is much smaller than the first. Some of the contributions to this sum are positive while others are negative (they all have the form of the first term) and so this weighted average is typically be close to zero. The derivative of the penalty has the form

$$\frac{\partial \text{pen}(\hat{\gamma})}{\partial \hat{\gamma}_i} = 4 \sum_{j=1}^k \epsilon_{ij} (\hat{\gamma}_i - \hat{\gamma}_j) \quad . \quad (2.22)$$

If  $\hat{\gamma}_i$  is large ( $\hat{\gamma}_i \gg 1.0$ ), the differences  $(\hat{\gamma}_i - \hat{\gamma}_j)$  are also large and positive so an overall negative correction for smoothing results. Similarly, if  $\hat{\gamma}_i$  is small relative to the others, most of the differences  $\hat{\gamma}_i - \hat{\gamma}_j$  are negative and a positive correction for smoothing results.

The smoothed estimates from Table 2.1 at  $\lambda = .025$  are presented again in Table 2.5. Also presented are the corresponding approximate estimate vector  $\hat{\gamma}_\lambda = (\hat{\gamma}_{\lambda,1}, \dots, \hat{\gamma}_{\lambda,k})'$ . For this example, the difference between  $\hat{\gamma}_{sm,i}$  and  $\hat{\gamma}_{\lambda,i}$  is less than .001 for sites one through nine and is less than .002 for site ten. Similarly, these quantities are shown for  $m = 200$  in Appendix 2.3.

**Table 2.5**

Illustration of  $\hat{\gamma}_{sm}$  and  $\hat{\gamma}_\lambda$   
 Circle Configuration, 10 sites and 100 cases

	$\lambda = 0$	$\lambda = .025$	
Site	$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_\lambda$
1	0.9000	0.9136	0.9134
2	1.1000	1.0993	1.0991
3	1.2000	1.1895	1.1893
4	0.9000	0.9124	0.9122
5	0.9000	0.9099	0.9097
6	0.8000	0.8147	0.8145
7	0.9000	0.9077	0.9074
8	0.6000	0.6222	0.6222
9	1.2000	1.1854	1.1852
10*	1.5000	1.4448	1.4464

\* = site of elevated incidence

## 2.5 Moments

In this section, approximate moments of the estimates are derived using the expansion for  $\hat{\gamma}_{sm,i}$ . This approximation is a function of the maximum likelihood estimate, so their moments are established initially.

### 2.5.1 Moments of the Unsmoothed Estimate

The case vector  $\mathbf{y}$  is distributed as multinomial with  $m$  cases and cell probability vector  $\boldsymbol{\theta}$ , so  $y_i$  is marginally distributed as binomial with index  $m$  and success probability  $\theta_i$ . As a result,  $\mathbf{E}(y_i) = m\theta_i$  and  $\mathbf{Var}(y_i) = m\theta_i(1 - \theta_i)$ .

Denoting  $\hat{\gamma}_{ml,i}$  more simply as  $\hat{\gamma}_i$ , then  $\hat{\gamma}_i = (y_i/N_i)/(m/N_+)$  and the mean and variance for  $\hat{\gamma}_i$  are

$$\begin{aligned}\mathbf{E}(\hat{\gamma}_i) &= \frac{N_+m}{mN_i} \left( \frac{N_i\gamma_i}{N_+} \right) \\ &= \gamma_i\end{aligned}\tag{2.23}$$

and

$$\begin{aligned}\mathbf{Var}(\hat{\gamma}_i) &= \left( \frac{N_+^2m}{m^2N_i^2} \right) \left( \frac{N_i}{N_+} \gamma_i \right) \left( 1 - \frac{N_i}{N_+} \gamma_i \right) \\ &= \left( \frac{N_+}{mN_i} \right) \gamma_i \left( 1 - \frac{N_i}{N_+} \gamma_i \right).\end{aligned}\tag{2.24}$$

Hence the maximum likelihood estimate  $\hat{\gamma}_i$  is unbiased. This is not unexpected as it is the unbiased multinomial cell estimate  $\hat{\theta}_i$  multiplied by fixed constants. The

variance of  $\hat{\gamma}_i$  decreases as  $m$  increases and increases with  $N_+/N_i$ . As expected, estimates at sites with smaller population have larger variance. Higher moments about 0 of the form,  $\mathbf{E}(\hat{\gamma}_i^r)$  are used in the next section and can be found in Appendix 2.4 for  $r = 1, \dots, 6$ .

### 2.5.2 Approximate Moments of $\hat{\gamma}_{ism}$

The mean and variance of the approximate estimate  $\hat{\gamma}_{\Lambda,i}$  can be used to better understand the distribution of the numerically smoothed estimate  $\hat{\gamma}_{ism,i}$ .

Taking the expectation of the approximate estimate given by (2.17),  $\mathbf{E}[\hat{\gamma}_{\Lambda}] = \gamma_i + \lambda \mathbf{E}[\hat{l}_i] + \lambda^2 \mathbf{E}[\hat{q}_i]$ . For moments, it is convenient to rewrite  $\hat{l}_i$  as

$$\begin{aligned} \hat{l}_i = \frac{N_+}{m} & \left\{ -\frac{4}{N_i} \left[ b_i \hat{\gamma}_i^2 - \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] \right. \\ & + \frac{4}{N_+} \left[ b_i \hat{\gamma}_i^3 - \hat{\gamma}_i^2 \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] \\ & \left. + \frac{4}{N_+} \left[ \hat{\gamma}_i \sum_{r \neq i} b_r \hat{\gamma}_r^2 - \hat{\gamma}_i \sum_{r \neq i} \sum_{s \neq r} \epsilon_{rs} \hat{\gamma}_r \hat{\gamma}_s \right] \right\} \end{aligned} \quad (2.25)$$

Hence,

$$\mathbf{E}(\hat{l}_i) = \frac{N_+}{m} \left\{ -\frac{4}{N_i} \left[ b_i \mathbf{E}(\hat{\gamma}_i^2) - \sum_{j \neq i} \epsilon_{ij} \mathbf{E}(\hat{\gamma}_i \hat{\gamma}_j) \right] \right.$$



$$\begin{aligned}
& + \frac{4}{N_+} \left[ b_i \mathbf{E}(\hat{\gamma}_i^3) - \sum_{j \neq i} \epsilon_{ij} \mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j) \right] \\
& + \frac{4}{N_+} \left[ \sum_{r \neq i} b_r \mathbf{E}(\hat{\gamma}_i \hat{\gamma}_r^2) - \sum_{r \neq i, s \neq r} \epsilon_{rs} \mathbf{E}(\hat{\gamma}_i \hat{\gamma}_r \hat{\gamma}_s) \right] \Big\}
\end{aligned}$$

Expectations of mixed products of the form  $y_i^{r_1} y_j^{r_2}$  are required. Such mixed moments for the multinomial are given by Johnson and Kotz (1969).  $\mathbf{E}(y_i^{r_1} y_j^{r_2}) = (m)^{[r_1+r_2]} \theta_i^{r_1} \theta_j^{r_2}$ , where  $(m)^{[r]}$  denotes the product  $m(m-1)\dots(m-r+1)$ . Derivations of some mixed moments of the form  $\mathbf{E}(\hat{\gamma}_i^{r_1} \hat{\gamma}_j^{r_2} \hat{\gamma}_l^{r_3})$  are given in Appendix 2.5.

The expression for  $\mathbf{E}(\hat{l}_i)$  resembles the one for  $\hat{l}_i$  but is a function of the parameter vector  $\gamma$ . Suppose one considers the first line of  $\hat{l}_i$  as given by (2.25). The expectation of this line is

$$\frac{-N_+}{m} \left( \frac{4}{N_i} \mathbf{E} \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} (\hat{\gamma}_i - \hat{\gamma}_j) \right).$$

Substitution for  $\mathbf{E}(\hat{\gamma}_i^2)$  and  $\mathbf{E}(\hat{\gamma}_i \hat{\gamma}_j)$  (found in Appendices 2.4 and 2.5 respectively) gives

$$\frac{-N_+}{m} \left( \frac{4\hat{\gamma}_i}{N_i} \sum_{j \neq i} \epsilon_{ij} \left( \frac{m-1}{m} (\gamma_i - \gamma_j) + \frac{N_+}{m N_i} \right) \right).$$

This expression can be interpreted in similar manner as  $\hat{l}_i$  (discussed in the previous section). It is the true parameter  $\gamma_i$  weighted by its population  $N_i$  and multiplied by a correction consisting of two terms. The first term  $((m-1)/m) \sum_{j \neq i} \epsilon_{ij} (\gamma_i - \gamma_j)$  is negative if  $\gamma_i$  is large relative to the others and positive if it is small relative to the others. The second term is independent of  $\gamma$  and increases with the number of sites  $k$ .

Prior to taking expectation, it is advantageous to rewrite  $\hat{q}_i$  as given by (2.20) so that  $\hat{\gamma}_i$  and  $\hat{l}_i$  are expressed separately. Some rearrangement gives

$$\begin{aligned} \hat{q}_i = \frac{N_+}{m} & \left\{ -\frac{1}{N_i} \left[ 2b_i \hat{\gamma}_i \hat{l}_i - \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \hat{l}_j - \hat{l}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] \right. \\ & + \frac{1}{N_+} \left[ 2b_i \hat{\gamma}_i^2 \hat{l}_i + b_i \hat{\gamma}_i^2 \hat{l}_i - 2\hat{\gamma}_i \hat{l}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j - \hat{\gamma}_i^2 \sum_{j \neq i} \epsilon_{ij} \hat{l}_j \right] \\ & + \frac{1}{N_+} \left[ 2\hat{\gamma}_i \sum_{r \neq i} b_r \hat{\gamma}_r \hat{l}_r + \hat{l}_i \sum_{r \neq i} b_r \hat{\gamma}_r^2 \right] \\ & \left. - \frac{1}{N_+} \left[ \hat{\gamma}_i \sum_{r \neq i} \sum_{s \neq r} \epsilon_{rs} \hat{\gamma}_r \hat{l}_s + \hat{\gamma}_i \sum_{r \neq i} \sum_{s \neq r} \epsilon_{rs} \hat{l}_r \hat{\gamma}_s + \hat{l}_i \sum_{r \neq i} \sum_{s \neq r} \epsilon_{rs} \hat{\gamma}_r \hat{\gamma}_s \right] \right\} \end{aligned}$$

Hence,

$$\begin{aligned} \mathbf{E}(q_i) = \frac{N_+}{m} & \left\{ -\frac{1}{N_i} \left[ 2b_i \mathbf{E}(\hat{\gamma}_i \hat{l}_i) - \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \mathbf{E}(\hat{\gamma}_i \hat{l}_j) - \sum_{j \neq i} \epsilon_{ij} \mathbf{E}(\hat{\gamma}_j \hat{l}_i) \right] \right. \\ & + \frac{1}{N_+} \left[ 2b_i \mathbf{E}(\hat{\gamma}_i^2 \hat{l}_i) + b_i \mathbf{E}(\hat{\gamma}_i^2 \hat{l}_i) - 2 \sum_{j \neq i} \epsilon_{ij} \sum_{j \neq i} \epsilon_{ij} \mathbf{E}(\hat{\gamma}_i \hat{\gamma}_j \hat{l}_i) \right] \\ & + \frac{1}{N_+} \left[ 2 \sum_{r \neq i} b_r \mathbf{E}(\hat{\gamma}_i \hat{\gamma}_r \hat{l}_r) + \hat{l}_i \sum_{r \neq i} b_r \mathbf{E}(\hat{\gamma}_r^2 \hat{l}_i) - \sum_{j \neq i} \epsilon_{ij} \mathbf{E}(\hat{\gamma}_i^2 \hat{l}_j) \right] \\ & - \frac{1}{N_+} \left[ \sum_{r \neq i} \sum_{s \neq r} \epsilon_{rs} \mathbf{E}(\hat{\gamma}_i \hat{\gamma}_r \hat{l}_s) + \sum_{r \neq i} \sum_{s \neq r} \epsilon_{rs} \mathbf{E}(\hat{\gamma}_i \hat{\gamma}_s \hat{l}_r) \right] \\ & \left. - \frac{1}{N_+} \left[ \sum_{r \neq i} \sum_{s \neq r} \epsilon_{rs} \mathbf{E}(\hat{\gamma}_r \hat{\gamma}_s \hat{l}_i) \right] \right\}. \end{aligned}$$

The term  $\hat{q}_i$  involves powers of  $\hat{\gamma}$  up to order five, the expression for  $\mathbf{E}(\hat{q}_i)$  involves mixed moments of powers of  $\hat{\gamma}$  to exponent five. Expressions such as  $\mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j \hat{\gamma}_r \hat{\gamma}_s)$  are quite involved and some examples can be found in Appendix 2.5. Fortran subroutines

were programmed for each of the required mixed moments.  $\mathbf{E}(\hat{q}_i)$  involves over sixty of these expressions and hence was not simplified.

The estimate  $\hat{\gamma}_{\lambda,i}$  is a quadratic approximation for  $\hat{\gamma}_{sm,i}$  and its variance involves powers of  $\lambda$  to order four

$$\begin{aligned} \text{Var}(\hat{\gamma}_{\lambda,i}) = & \text{Var}(\hat{\gamma}_{ml,i}) + 2\lambda\text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_i) + \lambda^2\text{Var}(\hat{l}_i) \\ & 2\lambda^2\text{Cov}(\hat{\gamma}_{ml,i}, \hat{q}_i) + 2\lambda^3\text{Cov}(\hat{l}_i, \hat{q}_i) + \lambda^4\text{Var}(\hat{q}_i) \quad . \end{aligned} \quad (2.26)$$

If  $\hat{l}_i$  is written in the full form given by (2.25), then  $\text{Var}(\hat{l}_i)$  and  $\text{Cov}(\hat{\gamma}_{ml,i}, \hat{q}_i)$  both involve powers of  $\gamma$  up to order six and are much too complicated for practical use. The expression for  $\text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_i)$  involves powers of  $\gamma$  to order four and is more reasonable computationally. As a result, a linear approximation to the variance of  $\hat{\gamma}_{\lambda,i}$  was taken to be

$$\text{Var}(\hat{\gamma}_{\lambda,i}) = \text{Var}(\hat{\gamma}_{ml,i}) + 2\lambda\text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_i) \quad . \quad (2.27)$$

## 2.6 Bias and MSE

As first mentioned in Chapter 1, the classical trade-off in smoothing is between the gains associated with reduced variance of the estimate(s) and the losses incurred from increased bias. A common indicator of the sampling properties of the estimator is the total mean squared error (MSE). For the vector estimate  $\hat{\gamma}_{sm}$ ,

$$\text{MSE} = \sum_{i=1}^k \text{Var}(\hat{\gamma}_{sm,i}) + \sum_{i=1}^k [\text{E}(\hat{\gamma}_{sm,i}) - \gamma_i]^2 \quad . \quad (2.28)$$

In general, one hopes the reduction in the total variance outweighs the increase in the total squared bias. In Tables 2.6a and 2.6b, one can see that this is indeed the case for  $\hat{\gamma}_{sm}$ . The summary estimates presented here are from a simulation of 5000 repetitions for the 'circle + 1' configuration. There are  $k = 10$  sites,  $m = 100$ ,  $\gamma_1 = \gamma_2 = \dots = \gamma_9 = 0.9524$  and  $\gamma_{10} = 1.4286$ , and equal risk populations of 1000. In Table 2.6a, the observed variance and bias are sample variance and bias for the simulation of 5000 estimates. Results for three of the five  $\lambda$  values are presented for each site in Table 2.6a. Table 2.6b presents the observed total variance, total squared bias and MSE for all five  $\lambda$  values used.

In Table 2.6a, one can see the decreasing variance with  $\lambda$  for all ten sites. The absolute magnitude of the bias is increasing with  $\lambda$ . The average smoothed estimate at the elevated site (site 10) is below the true  $\gamma_{10}$  for the positive  $\lambda$ 's shown and hence the observed bias for this site is negative. In Table 2.6b, the total observed variance and total squared bias are presented for all five of the  $\lambda$ 's used. As expected, the MSE decreases with  $\lambda$ . For this configuration and small departure from homogeneity, the observed MSE is dominated by the total variance for these small  $\lambda$ .

**Table 2.6a**

## Observed Variances and Biases

Simulation of 5000 repetitions

Circle + 1 Configuration, 10 sites and 100 cases

$$\gamma_1 = \gamma_2 = \dots = \gamma_9 = 0.9524, \gamma_{10} = 1.4286$$

Site	$\lambda = 0$		$\lambda = .0167$		$\lambda = .0333$	
	Ob. Var	Ob. Bias	Ob. Var	Ob. Bias	Ob. Var	Ob. Bias
1	.08648	.00153	.07925	.00091	.07011	.00132
2	.08661	-.00062	.08010	.00078	.07153	.00119
3	.08294	-.00134	.07821	.00051	.07317	.00169
4	.08861	-.00046	.07639	.00067	.07515	.00162
5	.08432	.00006	.07895	.00056	.07256	.00086
6	.08563	.00088	.07931	-.00015	.07627	.00205
7	.08636	.00029	.07870	.00085	.07269	-.00013
8	.08659	-.00112	.07827	.00139	.07256	.00253
9	.08863	.00145	.07036	-.00065	.07320	.00146
10*	.12177	-.00069	.11733	-.00048	.11203	-.01260

\* = site of elevated incidence

**Table 2.6b**  
 Observed Total Variances and Total Bias  
 Simulation of 5000 repetitions  
 Circle + 1 Configuration. 10 sites and 100 cases

$\lambda$	Ob.Tot.Var	Ob.Tot.Sq.Bias	MSE
0	.89795	$9.4932 \times 10^{-6}$	.89796
.00833	.85304	$1.5291 \times 10^{-5}$	.85306
.01667	.82356	$2.9324 \times 10^{-5}$	.82359
.02500	.79001	$9.0438 \times 10^{-5}$	.79018
.03333	.76412	$1.8708 \times 10^{-4}$	.76431

For a single site  $i$ , the bias of the smoothed estimate  $\hat{\gamma}_{sm,i}$  can be approximated by using the quadratic approximation  $\hat{\gamma}_{\lambda,i}$

$$\begin{aligned}
 \text{bias}(\hat{\gamma}_{\lambda,i}) &= E[\hat{\gamma}_{sm,i} - \gamma_i] \\
 &\approx E[\hat{\gamma}_{\lambda,i} - \gamma_i] \\
 &= \lambda E(\hat{l}_i) + \lambda^2 E(\hat{q}_i) \quad .
 \end{aligned}
 \tag{2.29}$$

and the total mean squared error for the smoothed estimate vector (over all sites) can be approximated as

$$\text{MSE}(\hat{\gamma}_{\Lambda,i}) = \sum_{i=1}^k \text{Var}(\hat{\gamma}_{\Lambda,i}) + \sum_{i=1}^k [\text{E}(\hat{\gamma}_{\Lambda,i}) - \gamma_i]^2. \quad (2.30)$$

The contribution of the  $i$ 'th site to the total squared bias in the MSE (approximated by the square of (2.29)) contains powers of  $\lambda$  to order four. However, the approximation for  $\text{Var}(\hat{\gamma}_{sm,i})$  is only linear in  $\lambda$  as given by (2.27). There are no terms which are linear in  $\lambda$  in the total squared bias so the approximate MSE is approximated by total variance. One therefore expects that the approximation is decreasing in  $\lambda$ .

## 2.7 Simulation Examples - Bias and MSE

In this section, the approximations for bias and MSE given in the previous section are compared to their observed counterparts from simulations. In each of 5000 repetitions, the smoothed estimate vector  $\hat{\gamma}_{sm}$  was obtained numerically for each case vector. The overall sample average and sample variance of these estimates was recorded for each site. The total sample variance added to the total squared sample bias gives the observed MSE for the simulation. The  $\lambda$ 's chosen are at equal intervals but are smaller than in the examples of Section 2.3. The expansion of  $\hat{\gamma}_{sm,i}$  is about  $\lambda = 0$  and this is reflected in the limits of the accuracy of the approximation. In each of the three simulations presented in this section, the 'circle + 1' configuration is used with  $k = 10$  sites, equal populations of  $N_j = 1000$  and  $m = 100$  cases.

First consider the case of complete homogeneity (i.e.  $\gamma_i = 1$  for all  $i$ ). In this situation, the variance of the observed estimates is decreased by the smoothing as desired. Presented in Table 2.7a are the observed and approximate biases for sites 1 and 10 for the 'circle + 1' configuration. Both the observed and approximate biases are small. The standard error of the observed bias in the simulation is much larger than the difference between the two. In Table 2.7b, the observed and approximate MSE are presented. The approximate MSE to linear terms in  $\lambda$  is also the approximate total variance as described in the previous section. As desired, the approximate and observed MSE are close. The observed MSE is essentially the observed total variance as there is very little bias. Hence the closeness of these values for each  $\lambda$  is a verification of the accuracy of the linear variance approximation.



**Table 2.7a**

Observed and Approximate Biases for Two Sites

Circle + 1 configuration. 100 cases

Complete homogeneity. site 10 is distant

$\lambda$	Site 1		Site 10	
	Ob. Bias	Approx. Bias	Ob. Bias	Approx. Bias
0	-.00112	0.0	.00780	0.0
.00746	-.00091	-.00259	.00065	.00014
.01492	-.00036	-.00965	.00322	-.00202
.02238	-.00763	-.02360	.00195	-.00653
.02984	-.00668	-.04210	-.00646	-.01322

**Table 2.7b**

Observed and approximate total mean squared error

(same simulation as Table 2.7a)

$\lambda$	Obs. MSE	Approx. MSE
0	0.9088	0.900
.00746	0.8627	0.8667
.01492	0.8329	0.8334
.02238	0.8084	0.8002
.02984	0.7742	0.7669

Presented in Table 2.8a are the observed biases for sites 1 and 10 for the same simulation which produced the results presented in Table 2.6a. The incidence rate at site 10 is slightly elevated,  $\gamma_{10} = 1.46$ , while the other nine rate ratios are 0.95. The departure from homogeneity is very small. Once again, the observed and approximate biases are small and the difference between the two is smaller than the standard error of the observed bias. In Table 2.8b, one can note that the influence of bias on the total MSE is again only slight and is decreasing through the four positive  $\lambda$ 's used. The linear approximation to the MSE again does reasonably well.

The results given in Tables 2.9a and 2.9b are for a simulation in which the incidence rate at site 10 is four times the other nine. This results in a  $\gamma_{10}$  of 3.07 while the other rate ratios are 0.76. Once again the 'circle + 1' configuration was used for  $m = 100$  cases. For a larger departure from homogeneity, one expects an increased bias with smoothing. The observed and approximate biases for sites 1 and 10 are larger and in close agreement, particularly for larger  $\lambda$ . In Table 2.9b an initial decrease in the observed MSE is evident but the MSE at the final  $\lambda$  of .00872 is larger than the observed MSE for  $\lambda = .01492$ . The linear approximation to the MSE is less effective at this larger  $\lambda$ .

Presented in Figure 2.1 are the observed MSE and approximate MSE which were given in Tables 2.7b, 2.8b and 2.9b respectively. Note that the vertical scales of the three plots are the same while the horizontal scales are not. The departures from homogeneity for the three simulations could be classified as none, low and moderately high respectively. The overall drop in MSE is not as large in the data of Table 2.9b and the increase in observed MSE at large  $\lambda$  has already been noted. Except for at large  $\lambda$  in the high departure simulation, the approximate MSE agrees with the observed MSE. The same pattern was observed in other simulations. Appendix 2.6 gives the observed and approximate MSE for the same  $m$ ,  $k$  and spatial configuration but the population at the distant site has been changed to 500 (while the others remain the same at 1000).

**Table 2.8a**

Observed and Approximate Biases for Two Sites

Circle + 1 configuration. 100 cases

Small departure from homogeneity. site 10 is distant

$\lambda$	Site 1		Site 10	
	Ob. Bias	Approx. Bias	Ob. Bias	Approx. Bias
0	.00153	0.0	-.00069	0.0
.008333	-.00124	-.00234	.00084	-.0048
.01667	.00091	-.01054	-.00048	-.01383
.02500	-.00067	-.02457	-.00755	-.02688
.03333	-.01320	-.04446	-.01260	-.04402

**Table 2.8b**

Total observed and approximate mean squared error

(same simulation as Table 2.8a)

$\lambda$	Obs. MSE	Approx. MSE
0	0.89796	0.89795
.008333	0.85306	0.86281
.01667	0.82359	0.82767
.02500	0.79018	0.79252
.03333	0.76431	0.75737

**Table 2.9a**

Observed and Approximate Biases for Two Sites

Circle + 1 Configuration, 100 cases

Moderate departure from homogeneity, site 10 is distant

$\lambda$	Site 1		Site 10	
	Ob. Bias	Approx. Bias	Ob. Bias	Approx. Bias
0	-.00047	0.0	-.00022	0.0
.00218	-.00030	.00096	-.00734	-.00906
.00436	.00382	.00167	-.01822	-.01849
.00654	.00447	.00216	-.02521	-.02828
.00872	.00493	.00430	-.03693	-.03845

**Table 2.9b**

Observed and approximate total mean squared error

(same simulation as Table 2.9a)

$\lambda$	Obs. MSE	Approx. MSE
0	0.85940	0.85207
.00218	0.84558	0.84579
.00436	0.83316	0.83953
.00654	0.82906	0.83326
.00872	0.83440	0.82698

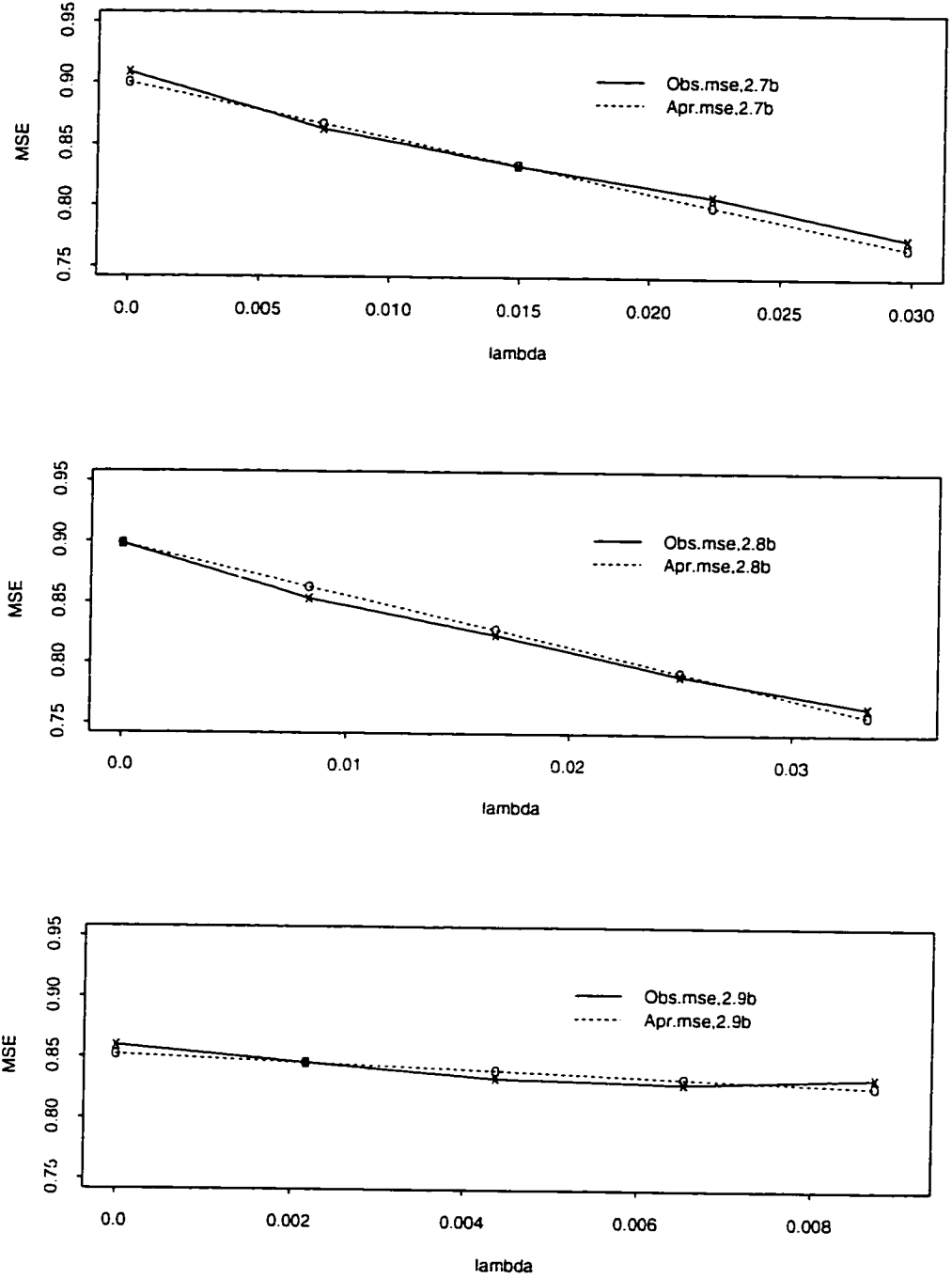


Figure 2.2: Observed and Approximate Total MSE  
from Tables 2.7b,2.8b,2.9b

## 2.8 Further Investigations

This section investigates simplified or 'reduced' estimators and the normality of the estimate  $\hat{\gamma}_{sm}$ . Application of these topics will appear in subsequent chapters.

### 2.8.1 Reduced Estimators

The approximate estimate  $\hat{\gamma}_{\lambda,i} = \hat{\gamma}_{ml,i} + \lambda \hat{l}_i + \lambda^2 \hat{q}_i$  involves powers of  $\hat{\gamma}_{ml,i}$  up to and including order five. It is logical to consider the linear approximation  $\hat{\gamma}_{ml,i} + \lambda \hat{l}_i$  as a simpler approximation. The term

$$\hat{l}_i = \frac{N_+}{m} \left\{ -\frac{4}{N_i} \left[ b_i \hat{\gamma}_i^2 - \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] + \frac{4}{N_+} \left[ b_i \hat{\gamma}_i^3 - \hat{\gamma}_i^2 \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] + \frac{4}{N_+} \left[ \hat{\gamma}_i \sum_{r \neq i} b_r \hat{\gamma}_r^2 - \hat{\gamma}_i \sum_{r \neq i} \sum_{s \neq r} \epsilon_{rs} \hat{\gamma}_r \hat{\gamma}_s \right] \right\}$$

(where the subscript "ml" has been omitted) is still rather complicated and involves powers of  $\hat{\gamma}_{ml,i}$  up to and including order three. Its variance involves terms with powers of  $\hat{\gamma}_{ml,i}$  up to and including order six. The first line of  $\hat{l}_i$ ,

$$\hat{l}_{red,i} = \left( \frac{N_+}{m} \right) \frac{-4}{N_i} \left[ b_i \hat{\gamma}_i^2 - \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] \quad (2.31)$$

was used to form a simpler or reduced linear estimator. Using these terms which are linear in  $\lambda$ , the estimators  $\hat{\gamma}_{\lambda,lin,i} = \hat{\gamma}_{ml,i} + \lambda \hat{l}_i$  and  $\hat{\gamma}_{\lambda,red,i} = \hat{\gamma}_{ml,i} + \lambda \hat{l}_{red,i}$

are investigated. The ‘reduced estimator’  $\hat{\gamma}_{\Lambda,red}$  is appealing due to its simple interpretation (given in Section 2.4) and because its moments are more tractable. The reduced estimates for the same examples presented in Tables 2.1 through 2.4 are given in Appendix 2.7.

The observed MSE for  $\hat{\gamma}_{sm}$ ,  $\hat{\gamma}_{\Lambda}$ ,  $\hat{\gamma}_{\Lambda,lin}$  and  $\hat{\gamma}_{\Lambda,red}$  were investigated empirically. In Table 2.10 and Figure 2.3, the observed MSE is presented for the same simulation described in Table 2.8 (one elevated rate,  $\gamma_{10} = 1.46$ ). The observed MSEs are very close for all the estimators. The lowest MSE is actually for the simplest estimator  $\hat{\gamma}_{\Lambda,red}$ . The observed MSE among the estimators is also similar for the configuration of higher departure from homogeneity ( $\gamma_{10} = 3.07$ , same simulation as Table 2.9). These results are given in Table 2.11 and Figure 2.4. Other simulations were performed and the observed MSEs of the four estimators are very close in each of them. The observed MSEs of the three approximate estimators is almost the same for  $k = 20$  with only a small departure from homogeneity.

**Table 2.10**

Total Mean Squared Error of Reduced Estimators  
 Small Departure From Homogeneity  
 (same simulation as Tables 2.8a and 2.8b)

$\lambda$	MSE			
	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{\Lambda}$	$\hat{\gamma}_{\Lambda,lin}$	$\hat{\gamma}_{\Lambda,red}$
0	0.89796	†	†	†
.008333	0.85306	0.85310	0.85208	0.84742
.01667	0.82359	0.82391	0.81985	0.78231
.02500	0.79018	0.79115	0.78231	0.76922
.03333	0.76431	0.76644	0.75114	0.73442

† indicates all elements of row are same because  $\lambda = 0$

**Table 2.11**

Total Mean Squared Error for Reduced Estimators  
 (same simulation as Tables 2.9a and 2.9b)

$\lambda$	MSE			
	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{\Lambda}$	$\hat{\gamma}_{\Lambda,lin}$	$\hat{\gamma}_{\Lambda,red}$
0	0.85940	†	†	†
.00218	0.84558	0.84562	0.84559	0.84227
.00436	0.83316	0.83321	0.83312	0.82682
.00654	0.82906	0.82909	0.82888	0.81970
.00872	0.83440	0.83444	0.83410	0.82229

† indicates all elements of row are same because  $\lambda = 0$

The estimators can also be compared through simple pairwise correlations. The similarity of the values of the estimators was quite evident as the correlation between



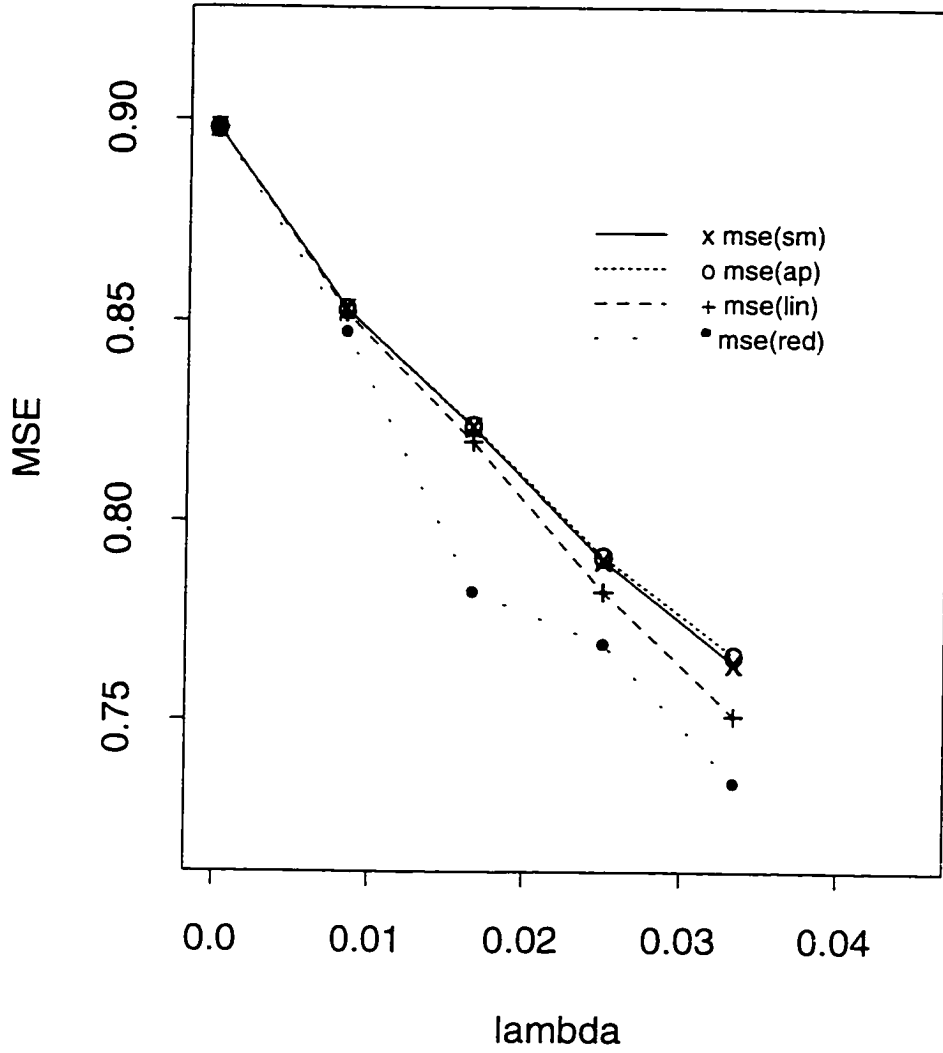


Figure 2.3: Observed MSE for smoothed estimators

1 Elevated Rate, Small Departure

(from Table 2.10)

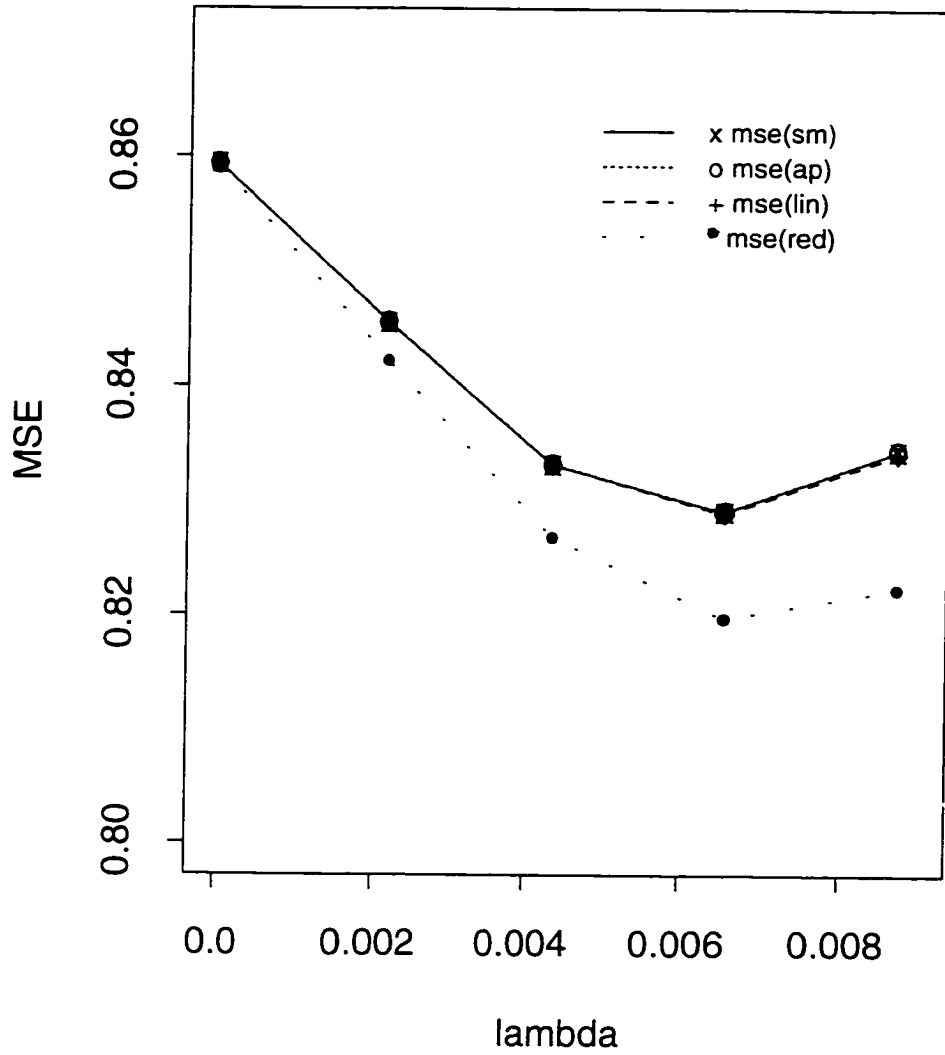


Figure 2.4: Observed MSE for smoothed estimators

1 Elevated Rate, Moderate Departure

(from Table 2.11)

any pair was usually 1.0 and at least 0.99 for any simulation performed (regardless of the departure from homogeneity).

Pairs of estimators were also compared with respect to Pitman Nearness (PN) (Pitman, 1937). If  $\hat{\theta}_1$  and  $\hat{\theta}_2$  are two univariate estimators for the parameter  $\theta$ , then the Pitman Nearness, PN is defined as

$$PN = P(|\hat{\theta}_1 - \theta| > |\hat{\theta}_2 - \theta|)$$

where the absolute distance  $|\hat{\theta} - \theta|$  is the chosen measure of distance between  $\hat{\theta}$  and  $\theta$ . If the vector  $\hat{\theta}$  is an estimator for the parameter vector  $\theta$ , then the distance between the estimator and the parameter vector can be taken as Euclidean and given by  $\|\hat{\theta} - \theta\| = \sqrt{(\hat{\theta} - \theta)'(\hat{\theta} - \theta)}$ . For a given simulation for which  $\lambda > 0$ , the fraction of case vectors in which one estimator was closer to  $\gamma$  than another was taken to be an estimator of PN.

Prior to pairwise comparisons among each of the smoothed estimators, each of  $\hat{\gamma}_{sm}$ ,  $\hat{\gamma}_\lambda$ ,  $\hat{\gamma}_{\lambda,lin}$ ,  $\hat{\gamma}_{\lambda,red}$  were compared to the unsmoothed estimate  $\hat{\gamma}_{ml}$ . For small to moderate departures from homogeneity and  $k \geq 20$ , each of these smoothed estimators was always closer to  $\gamma$  (PN = 1.0). Once the departure gets very large, the PN decreases (e.g. PN was approximately 0.75 for each smoothed estimator in the simulations of Tables 2.9 and 2.11).

The full approximation  $\hat{\gamma}_\lambda$  is slightly more distant from  $\gamma$  than the numerically smoothed estimate  $\hat{\gamma}_{sm}$ . As a result the fraction PN was low (close to 0.0) for the comparison between  $\hat{\gamma}_\lambda$  and  $\hat{\gamma}_{sm}$ . However  $\hat{\gamma}_{\lambda,lin}$  was generally closer than  $\hat{\gamma}_\lambda$  (PN close to 1.0) and  $\hat{\gamma}_{\lambda,red}$  was closer than  $\hat{\gamma}_{\lambda,lin}$ .

There is little or no loss in estimating precision by using the simple estimator  $\hat{\gamma}_{\lambda,red}$ . In Chapter 3,  $\hat{\gamma}_{\lambda,red}$  and its approximate variance are used to develop a test

statistic. This reduced estimate is later involved in criteria for the choice of  $\lambda$ , to be addressed in Chapter 4.

## 2.8.2 Normality of $\hat{\gamma}_{ism}$

The maximum likelihood estimate for the probability vector  $\theta$  is known to be asymptotically normal as  $m$  goes to infinity (Kendall and Stuart (1957), Bishop, Holland and Fienberg (1975)). The populations  $N_i$  are fixed and because  $\hat{\gamma}_{ml,i} = N_+ \hat{\theta}_i / N_i$ , we expect the maximum likelihood estimate vector  $\hat{\gamma}_{ml}$  to be asymptotically normal as  $m$  goes to infinity. The normal approximation for  $\hat{\gamma}_{ml,i}$  is examined in more detail in the context of its quantiles in Chapter 3.

Brief investigations were performed to verify the approximate normality of  $\hat{\gamma}_{sm}$ . For small  $\lambda$ , normality of  $\hat{\gamma}_{sm,i}$  was expected because the approximation for  $\hat{\gamma}_{sm,i}$  is the maximum likelihood estimate plus small correction terms as a function of  $\lambda$ . Under complete homogeneity,  $\gamma_i = 1$  for  $i = 1, \dots, k$ . The expected number of cases at each site is  $mN_i/N_+$ . If the expected number of cases is large enough (five or ten is suggested by different authors), normality of  $\hat{\gamma}_{sm,i}$  should result. Figures 2.4a and 2.4b present Q-Q plots for  $\hat{\gamma}_{sm,i}$  for the 'circle + 1' configuration under complete homogeneity. 5000 repetitions were performed for  $k = 20$  and  $m = 200$  for equal populations.

The discreteness of the estimates is evident from the figures. Presented in the top left plot are the unsmoothed estimates  $\hat{\gamma}_{ml,i}$ . In this plot, each horizontal band represents values of  $\hat{\gamma}_{sm,i}$  corresponding to a particular value of  $y_i$ . As  $\lambda$  increases, the range of possible  $\hat{\gamma}_{sm,i}$  changes, and hence the shape and slope of each 'step' in a plot

also changes. The estimates involved in Figure 2.5a are for a site on the circle which is not contiguous to the distant site. The estimates in the tail of the distribution tend to follow those of a normal distribution as more smoothing is applied. Figure 2.5b presents estimates for the same  $\lambda$  for the same simulation at the geographically distant site. Under complete homogeneity, there is little difference between the Q-Q plots of the estimates presented in Figures 2.5a and 2.5b.

Q-Q plots for additional simulations are given in Appendix 2.8. In general, smoothing tends to bring the estimates closer to normality. The estimates for the site with elevated incidence correspond to a much higher expected count and the Q-Q plots indicate approximate normality for all the  $\lambda$ 's used including 0. For simulations involving larger  $m$ , the Q-Q plots indicate approximate normality for each  $\lambda$  used.

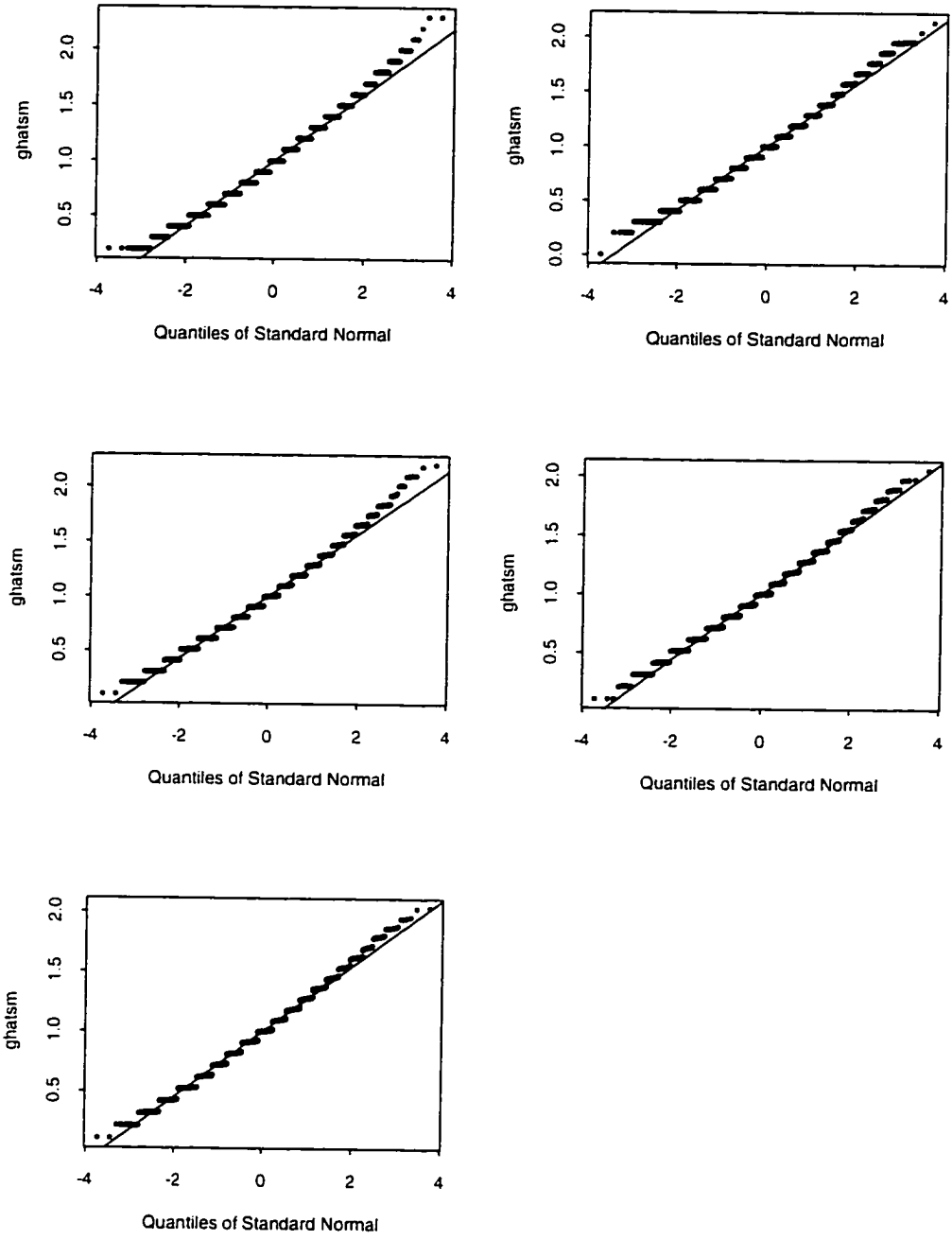


Figure 2.5a: Q-Q plots - Complete homogeneity ( $k=20, m=200$ )

Site on circle, circle + 1 configuration

Across Rows, lambdas : 0, .0029, .0058, .0087, .016

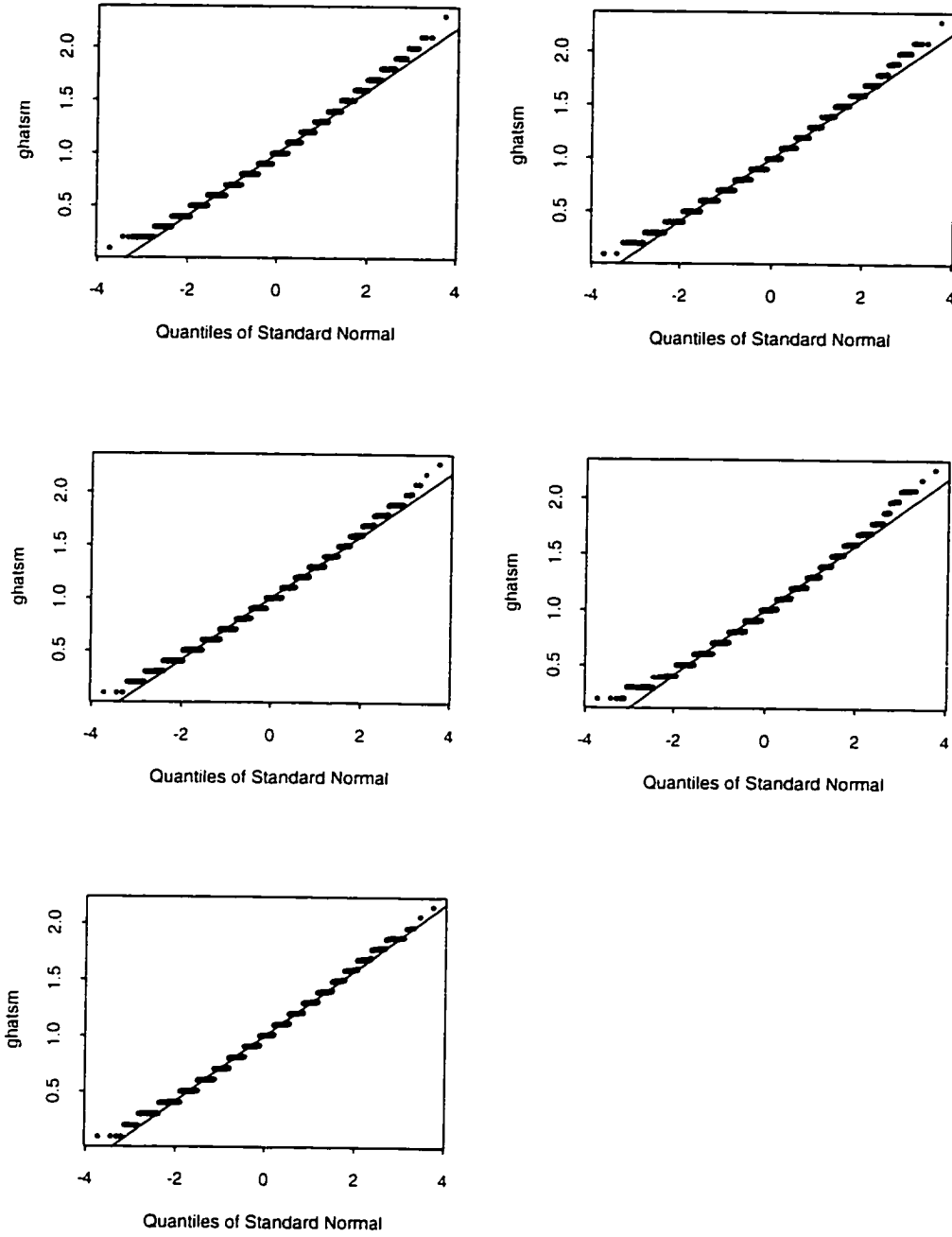


Figure 2.5b: Q-Q plots - Complete homogeneity ( $k=20, m=200$ )

Distant site, circle + 1 configuration

Across Rows, lambdas : 0, .0029, .0058, .0087, .016

## Chapter 3

# Testing for Elevated Rates

In this chapter, a testing strategy is outlined to identify elevated rates and define clusters. In Section 3.1, the testing strategy is outlined and clusters are defined. Section 3.2 describes testing without smoothing and testing based on smoothed estimates is presented in Section 3.3. The remaining sections of the chapter involve simulations which evaluate the testing methodology. The simulations are described in Section 3.4. Results on the size of the tests are given in Section 3.5 while results on the method's ability to detect elevated rates are given in Section 3.6.

### 3.1 Testing Strategy and Cluster Definition

The presentation of estimates does not specify whether the incidence at a particular site is elevated relative to the others. Testing for elevated incidence at an individual site  $i$  corresponds to a null hypothesis of  $H_{oi} : \gamma_i = 1$  where  $\gamma_i = p_i/\bar{p}$  is the rate ratio defined in Chapter 2. Complete homogeneity corresponds to all  $H_{oi}$  true for



$i = 1, \dots, k$ . Except under  $H_o$ , at least one  $p_j$  must be less than  $\bar{p}$ , so at least one rate ratio  $\gamma_j$  must be less than 1.0. A priori, the multinomial constraint implies that one estimate will be less than 1.0. Under complete homogeneity, a mistake of incorrectly rejecting  $H_o$  can only be made at  $k - 1$  sites. As a result, only  $k - 1$  tests must be performed for a given case vector  $\mathbf{y}$ . To maintain an overall error rate of  $\alpha$ , a correction for multiple testing is required. For most of this thesis, a simple Bonferroni correction is applied and each test is performed at site level  $\alpha/(k - 1)$ . Bonferroni corrections are known to be conservative and a modified procedure (Hochberg, 1988) is discussed for the application in Chapter 5.

To define elevated rates, the following procedure is proposed. After ranking the  $k$  estimates  $\hat{\gamma}_j, j = 1, \dots, k$ , drop the lowest and perform  $k - 1$  tests using these estimates. Sites with the null hypothesis rejected are identified as being elevated rates.

Sites or regions are defined to be contiguous if they share a common border. From the spatial contiguities, spatial groups or clusters of elevated rates are formed. The number of sites in a cluster defines its size. A safeguard is proposed which drops the lowest ranking group if the total sites declared elevated exceeds  $k/2$ . This situation was never observed in any simulations subsequently performed but can occur.

## 3.2 Testing with no smoothing

The strategy outlined for identifying elevated rates and clusters requires a test statistic which is to be defined for each individual site. Without smoothing, the approximate

normality of  $\hat{\gamma}_{ml,i}$  (investigated in Section 2.8) suggests the test statistic

$$Z_{ml,i} = \frac{\hat{\gamma}_{ml,i} - 1}{v_{o,i}^{1/2}}$$

for testing the null  $H_{o,i} : \gamma_i = 1$  versus the alternative  $H_{a,i} : \gamma_i > 1$ . The variance of  $\hat{\gamma}_{ml,i}$  under  $H_{o,i}$  is given by

$$v_{o,i} = \text{Var}(\hat{\gamma}_{ml,i}) = \frac{1}{m} \frac{N_+}{N_i} \left( 1 - \frac{N_i}{N_+} \right) .$$

Geography is not involved with the test statistic at  $\lambda = 0$  but the spatial contiguity of the sites is required for cluster definition.

For testing at site level  $\alpha^* = \alpha/(k-1)$ , we need the critical value  $C_{\alpha^*,i}$  such that  $P(Z_{ml,i} \geq C_{\alpha^*,i}) \leq \alpha^*$ . The test statistic  $Z_{ml,i}$  is discrete and takes on values from  $(-1/v_{o,i}^{1/2})$  to  $(N_+/N_i - 1)/v_{o,i}^{1/2}$ , as  $y_i$  ranges from 0 to  $m$ , with the corresponding binomial probabilities. Given in Example 3.1 are the first 22 possible values for  $Z_{ml,i}$  for a site of population  $N_i = 500$  in a region of twenty sites with total population  $N_+ = 19500$  and  $m = 200$  cases. Also shown are the case counts,  $y_i$ , and the value of the cumulative binomial probability.

### Example 3.1

$y_i$	$Z_{ml}$	cum.prob.
0	-2.2941573	0.005543677
1	-1.84679666	0.03472092
2	-1.39943598	0.11111924
3	-0.95207530	0.24381106
4	-0.50471461	0.41578664
5	0.05735393	0.59319302
6	0.39000675	0.74492217
7	0.83736743	0.85558177
8	1.28472811	0.92583606
9	1.73208879	0.96527706
10	2.17944947	0.98510136
11	2.62681015	0.99411240
12	3.07417083	0.99784724
13	3.52153151	0.99926860
14	3.96889220	0.99976821
15	4.41625288	0.99993124
16	4.86361356	0.99998085
17	5.31097424	0.99999497
18	5.75833492	0.99999876
19	6.20569560	0.99999971
20	6.65305628	0.99999994
21	7.10041696	0.99999999
22	7.54777764	1.00000000

To achieve the desired overall test level of  $\alpha = .05$ , the probability of exceeding the required  $C_{\alpha^*,i}$  must be less than  $.05/19 = .00263$ . In this example,  $C_{\alpha^*,i} = 3.52$  and the achievable site level is .0021. It is apparent that the normal approximation for  $Z_{ml,i}$  is inadequate. The standard normal quantile  $Z_{\alpha^*}$  is 2.79 and because  $P(Z_{ml,i} \geq 2.79) = P(Z_{ml,i} \geq 3.07) = .0059$ , the site and overall levels are not controlled. A

continuity correction is not advantageous in this context either. Details are discussed in Appendix 3.1.

Given  $m$ ,  $N_i$ , and  $N_+$ , appropriate statistical software can be used to generate the exact critical value for  $\lambda = 0$ . For the simulations of this chapter, these exact critical values were obtained using the Splus function `pbinom`. The normal approximation with a continuity correction will be most accurate when  $m\theta_i > 5$  and the cell probability is near 0.5 (Molenaar, 1970). Accurate transformations have been presented for a small success probability (e.g. Peizer and Pratt (1968)), but these transformations were not investigated due to the availability of the exact critical values.

### 3.3 Testing with smoothing

In this section testing based on smoothed estimates is presented. The motivation for testing with smoothing is given in Section 3.3.1. The test statistic with smoothed estimates is developed and refined in 3.3.2. The critical region for the test statistic is determined in Section 3.3.3.

#### 3.3.1 Motivation

The reduced variance and MSE of the smoothed estimate vector  $\hat{\gamma}_{sm}$  were demonstrated in Chapter 2. It is logical to investigate whether testing based on the individual smoothed estimates is better able to detect elevated rates or clusters than testing without smoothing. A test statistic based on  $\hat{\gamma}_{sm,i}$  incorporates the spatial configuration of the sites through the estimate itself. The  $i$ 'th element of the numerically

obtained smoothed estimate vector gives the test statistic

$$Z_{sm,i} = \frac{\hat{\gamma}_{sm,i} - 1}{\sqrt{\text{Var}(\hat{\gamma}_{sm,i})}} \quad (3.1)$$

for the test of  $H_{oi} : \gamma_i = 1$  versus  $H_{a,i} : \gamma_i > 1$ .

### 3.3.2 Refinement of the test statistic for smoothed estimates

For a single estimate vector,  $\hat{\gamma}_{sm}$ ,  $\text{Var}(\hat{\gamma}_{sm,i})$  in (3.1) is unknown. Rather than using numerical optimization to obtain  $\hat{\gamma}_{sm}$ , it is also more practical to use the approximate vector  $\hat{\gamma}_\lambda$  which has  $i$ 'th element given by the approximation (2.18). This argument suggests the test statistic

$$Z_{\lambda,i} = \frac{\hat{\gamma}_{\lambda,i} - 1}{\sqrt{\hat{\text{Var}}(\hat{\gamma}_{\lambda,i})}} \quad (3.2)$$

where  $\hat{\text{Var}}(\hat{\gamma}_{\lambda,i})$  is the linear approximation to  $\text{Var}(\hat{\gamma}_{sm,i})$  evaluated at current estimates under  $H_o$ .

The method for determining the critical region for  $Z_{\lambda,i}$  is outlined in the next subsection. Further simplification of the test statistic  $Z_{\lambda,i}$  is required to accomplish this goal and an approximate test statistic which is linear in  $\lambda$  is derived. Consider  $Z_{\lambda,i}$  above with the following components:

the linear in  $\lambda$  estimate of the  $i$ 'th rate ratio

$$\hat{\gamma}_{\lambda, lin, i} = \hat{\gamma}_{ml, i} + \lambda \hat{l}_i \quad . \quad (3.3)$$

with smoothing component

$$\begin{aligned} \hat{l}_i = \left( \frac{1}{m \cdot N_+} \right) & \left\{ -\frac{1}{N_i} \left[ b_i \hat{\gamma}_i^2 - \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] \right. \\ & + \frac{1}{N_+} \left[ b_i \hat{\gamma}_i^3 - \hat{\gamma}_i^2 \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right] \\ & \left. + \frac{1}{N_+} \left[ \hat{\gamma}_i \sum_{r \neq i} b_r \hat{\gamma}_r^2 - \hat{\gamma}_i \sum_{r \neq i} \sum_{s \neq r} \epsilon_{rs} \hat{\gamma}_r \hat{\gamma}_s \right] \right\} \end{aligned} \quad (3.4)$$

and linear in  $\lambda$  variance estimate

$$\hat{\text{Var}}(\hat{\gamma}_{\lambda, i}) = v_{o, i} + \hat{v}_{1, i} \lambda \quad . \quad (3.5)$$

where

$$\begin{aligned} v_{o, i} &= \text{Var}(\hat{\gamma}_{ml, i} | \gamma_i = 1) \\ &= \frac{1}{m} \frac{N_+}{N_i} \left( 1 - \frac{N_i}{N_+} \right) \end{aligned} \quad (3.6)$$

and

$$\hat{v}_{1, i} = \hat{\text{Cov}}(\hat{\gamma}_{ml, i}, \hat{l}_i | \gamma_i = 1) \quad . \quad (3.7)$$

Maximum likelihood estimates are obtained under  $H_o : \gamma_i = 1$  and are given by

$$\frac{y_j}{N_j} / \frac{m - y_i}{N_+ - N_i}$$

for  $j \neq i$  (details given in Appendix 3.2). These estimates and  $\hat{\gamma}_{ml,i} = 1$  are used to evaluate  $\text{Var}(\hat{\gamma}_{\lambda,i})$  under  $H_o$ . Furthermore,  $\hat{\text{Var}}(\hat{\gamma}_{\lambda,i})$  can be re-written as  $v_{o,i}^{\frac{1}{2}}(1 + (\hat{v}_{1,i}\lambda)/v_{o,i})^{\frac{1}{2}}$  giving the test statistic

$$Z_{\lambda,i} = \frac{\hat{\gamma}_{ml,i} + \hat{l}_i\lambda - 1}{v_{o,i}^{\frac{1}{2}}(1 + \frac{\hat{v}_{1,i}\lambda}{v_{o,i}})^{\frac{1}{2}}} \quad . \quad (3.8)$$

and the statistic which is linear in  $\lambda$  is

$$Z_{\lambda,i} = \left(1 + \frac{\hat{v}_{1,i}\lambda}{v_{o,i}}\right)^{-\frac{1}{2}} Z_{ml,i} + \frac{\left(1 + \frac{\hat{v}_{1,i}\lambda}{v_{o,i}}\right)^{-\frac{1}{2}} \hat{l}_i\lambda}{v_{o,i}^{\frac{1}{2}}} \quad . \quad (3.9)$$

where  $Z_{ml,i}$  is as given in Section 3.2.

A one-term Taylor approximation (about  $\lambda = 0$ ) for  $(1 + \hat{v}_{1,i}\lambda)^{\frac{1}{2}}$  is  $1 - 0.5(\hat{v}_{1,i}\lambda)/v_{o,i}$ . and retaining only terms which are linear in  $\lambda$ , the second term of (3.9) is  $\hat{l}_i\lambda$ , giving the approximate test statistic

$$\begin{aligned} Z_{\lambda,lin,i} &= \left(1 - 0.5\frac{\hat{v}_{1,i}\lambda}{v_{o,i}}\right) Z_{ml,i} + \frac{\hat{l}_i\lambda}{v_{o,i}^{\frac{1}{2}}} \\ &= \hat{s}_{\lambda,i} Z_{ml,i} + \frac{\hat{l}_i\lambda}{v_{o,i}^{\frac{1}{2}}} \quad . \end{aligned} \quad (3.10)$$

Within  $\hat{l}_i$ , the  $\hat{\gamma}_{ij,ml}$  for  $j \neq i$  are re-estimated under  $H_o$ . Extensive empirical investigation revealed that  $\hat{s}_{\lambda,i}$  is relatively constant for a given site and population/geography configuration. Henceforth,  $\hat{s}_{\lambda,i}$  is considered non-random. Although linear in  $\lambda$ ,  $\hat{l}_i$  as given by (2.25) and (3.4) involves powers of  $\hat{\gamma}_{ml,j}$  up to order 3. This fact and the

form of (3.4) complicates the investigation of the distribution of  $Z_{\lambda,lin,i}$ . As a simplification, the linear term of  $\hat{\gamma}_{\lambda,i}$  was reduced to powers of  $\hat{\gamma}_{ml}$  to order 2 (as previously described in Section 2.8). The linear coefficient of  $\lambda$  in the approximate variance is  $\hat{v}_{1,i} = \text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_i)$ . It is adjusted in a similar fashion to include terms to power 3 in  $\hat{\gamma}_{ml}$  giving

$$\hat{l}_{i,red} = t_{i,\lambda} (-b_i \hat{\gamma}_i^2 + \hat{\gamma}_i w_i) \quad (3.11)$$

where

$$t_{i,\lambda} = \frac{4\lambda}{N_i(m/N_+)}$$

and

$$w_i = \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \quad .$$

Unless otherwise indicated, testing with smoothed estimates now involves the reduced test statistic given by

$$Z_{\lambda,red,i} = \hat{s}_{\lambda,red,i} Z_{ml,i} + \frac{\hat{l}_{red,i} \lambda}{v_o^{\frac{1}{2}}} \quad . \quad (3.12)$$

where

$$\hat{s}_{\lambda,red,i} = 1 - 0.5 \hat{v}_{1,red,i} / v_o$$

and

$$\hat{v}_{1,red,i} = \text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_{red,i})$$



### 3.3.3 Determining the critical region for $Z_{\lambda,red,i}$

At  $\lambda = 0$ , exact critical values for  $Z_{ml,i}$  are available as described in Section 3.2.1. While tests at  $\lambda = 0$  may not achieve size  $\alpha^* = \alpha/(k-1)$  due to discreteness of the test statistic, approximate tests at any level  $\alpha^*$  for  $\lambda > 0$  are possible. Although  $Z_{\lambda,red,i}$  is also discrete, a large number of values near  $Z_{ml,i}$  are generated for each  $y_i$  by the possible distributions of the remaining  $m - y_i$  cases. An approximate critical value for  $Z_{\lambda,red,i}$  for level  $\alpha^*$  needs to be determined. Denoting the  $1 - \alpha^*$  quantile for  $Z_{\lambda,red,i}$  by  $C_{\alpha^*,i}$  then  $C_{\alpha^*,i}$  must satisfy

$$P(Z_{\lambda,red,i} \geq C_{\alpha^*,i}) \leq 1 - \alpha^* . \quad (3.13)$$

A conditional approach is used to find  $C_{\alpha^*,i}$ . A possible alternative approach is the use of a Cornish-Fisher expansion. A description of this technique is given in Appendix 3.3

The left hand side of (3.13) is equivalent to

$$\sum_{\forall x_i} P(Z_{\lambda,red,i} \geq C_{\alpha^*,i} | \hat{\gamma}_{ml,i} = x_i) P(\hat{\gamma}_{ml,i} = x_i) . \quad (3.14)$$

The exact values for  $P(\hat{\gamma}_{ml,i} = x_i)$  are available from the marginal binomial distribution for  $\hat{\gamma}_{ml,i}$  and  $\hat{\gamma}_{ml,i} = x_i$  ranges from 0 to  $N_+/N_i$  as  $y_i$  ranges from 0 to  $m$ . The first term of the summand in (3.14) is

$$P\left(\hat{s}_{\lambda,red,i}Z_{ml,i} + \frac{\hat{l}_{red,i}\lambda}{v_{\sigma,i}^{\frac{1}{2}}} \geq C_{\alpha^*,i} | \hat{\gamma}_{ml,i} = x_i\right) \quad (3.15)$$

which can be written as

$$P\left(A_i \hat{\gamma}_{ml,i}^2 + D_i w_i \hat{\gamma}_{ml,i} \geq H_i | \hat{\gamma}_{ml,i} = x_i\right) \quad (3.16)$$

to emphasize the dependence on  $w_i$  (the only term considered random given the conditioning), where

$$A_i = \frac{-t_{i,\lambda} b_i}{v_{\sigma,i}^{\frac{1}{2}}} \quad .$$

$$D_i = \frac{(\hat{s}_{\lambda,red,i} + t_{i,\lambda})}{v_{\sigma,i}^{\frac{1}{2}}} \quad .$$

and

$$H_i = C_{\alpha^*,i} - \frac{\hat{s}_{\lambda,red}}{v_{\sigma,i}^{\frac{1}{2}}} \quad .$$

Substituting  $x_i$  for  $\hat{\gamma}_{ml,i}$ ,  $P(Z_{\lambda,red,i} > C_{\alpha^*,i})$  equals

$$\sum_{\forall x_i} P\left(w_i \geq \frac{H_i - A_i x_i^2}{D_i x_i} | \hat{\gamma}_i = x_i\right) P(\hat{\gamma}_{ml,i} = x_i) \quad . \quad (3.17)$$

We need to find  $C_{\alpha^*,i}$  such that this sum is less than or equal to  $\alpha^*$ . For a particular  $C_{\alpha^*,i}$ , this can be computed if we can appropriately specify the distribution of  $w_i$  given  $\hat{\gamma}_{ml,i}$ .

Denote the  $k-1$  dimensional vector of maximum likelihood estimates by  $\hat{\gamma}_{(i)}$ , the  $k-1$  dimensional vector  $\epsilon_{ij}$  (for  $j \neq i$ ) by  $\mathbf{e}'_{(i)}$ , and the variance-covariance matrix of  $\hat{\gamma}_{(i)}$  by  $\Sigma_{\hat{\gamma}_{(i)}}$ . The quantity  $w_i = \mathbf{e}'_{(i)}\hat{\gamma}_{(i)}$  is approximately normal with mean  $\mathbf{e}'_{(i)}\gamma_{(i)}$  and variance  $\mathbf{e}'_{(i)}\Sigma_{\hat{\gamma}_{(i)}}\mathbf{e}_{(i)}$ . The distribution of  $w_i|\hat{\gamma}_{ml,i}$  is approximately normal with mean and variance

$$\mu_{w_i|\hat{\gamma}_{ml,i}} = \mu_{w_i} + \frac{\text{Cov}(w_i, \hat{\gamma}_{ml,i}) (\hat{\gamma}_{ml,i} - \mu_{\hat{\gamma}_{ml,i}})}{\sigma_{\hat{\gamma}_{ml,i}}^2} \quad (3.18)$$

and

$$\sigma_{w_i|\hat{\gamma}_{ml,i}}^2 = \sigma_{w_i}^2 - \frac{[\text{Cov}(w_i, \hat{\gamma}_{ml,i})]^2}{\sigma_{\hat{\gamma}_{ml,i}}^2} \quad (3.19)$$

The unconditional mean and variance of  $w_i$  are

$$\begin{aligned} \mu_{w_i} &= E\left(\sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_{ml,j}\right) \\ &= \sum_{j \neq i} \epsilon_{ij} \gamma_j \end{aligned} \quad (3.20)$$

and

$$\sigma_{w_i}^2 = \mathbf{e}'_{(i)}\Sigma_{\hat{\gamma}_{(i)}}\mathbf{e}_{(i)}$$

which simplifies to

$$\sigma_{w_i}^2 = \sum_{j \neq i} \epsilon_{ij}^2 \frac{1}{m} \frac{N_+}{N_j} \gamma_j \left(1 - \frac{N_j}{N_+} \gamma_j\right) - \frac{1}{m} \sum_{j \neq i} \sum_{q \neq j,i} \epsilon_{ij} \epsilon_{iq} \gamma_j \gamma_q \quad .$$

Also,

$$\begin{aligned} \text{Cov}(w_i, \hat{\gamma}_{ml,i}) &= \sum_{j \neq i} \epsilon_{ij} \text{Cov}(\hat{\gamma}_{ml,j}, \hat{\gamma}_{ml,i}) \\ &= -\frac{1}{m} \sum_{j \neq i} \epsilon_{ij} \gamma_i \gamma_j \end{aligned}$$

which becomes  $-(1/m) \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j$ . under  $H_0 : \gamma_i = 1$ . Finally,  $\sigma_{\hat{\gamma}_{ml,i}}^2 = (.N_+/m.N_i)(1 - .N_+/N_i) = v_{o,i}$  under the null hypothesis.

The rate ratios  $\gamma_j$ . for  $j \neq i$  are unknown and not specified under  $H_0$  and so they must be estimated under the constraint  $\gamma_i = 1$ . As mentioned in the previous section and described in Appendix 3.2. these estimates without smoothing are equal to  $\hat{\gamma}_{j(i)} = (y_j/N_j)/((m - y_i)/(N_+ - N_i))$  for  $j \neq i$ . Substituting these estimates into the expressions for  $\mu_{w_i|\hat{\gamma}_{ml,i}}$  and  $\sigma_{w_i|\hat{\gamma}_{ml,i}}^2$  gives  $\hat{\mu}_{w_i|\hat{\gamma}_{ml,i}}$  and  $\hat{\sigma}_{w_i|\hat{\gamma}_{ml,i}}^2$  gives the following estimates after some rearrangement shown in Appendix 3.4

$$\hat{\mu}_{w_i|\hat{\gamma}_{ml,i}} = \frac{.N_+ - .N_i}{m - y_i} \sum_{j \neq i} \epsilon_{ij} \left( \frac{y_j}{.N_j} \right) \left[ 1 + \left( \frac{.N_i}{.N_+ - .N_i} \right) (\hat{\gamma}_{ml,i} - 1) \right] \quad (3.21)$$

and

$$\begin{aligned} \hat{\sigma}_{w_i|\hat{\gamma}_{ml,i}}^2 &= \frac{.N_+ (.N_+ - .N_i)}{m(m - y_i)} \sum_{j \neq i} \epsilon_{ij}^2 \left( \frac{y_j}{.N_j} \right) \left[ 1 - \left( \frac{.N_+ - .N_i}{m - y_i} \right) \frac{y_j}{.N_+} \right] \\ &\quad - \frac{1}{m} \left( \frac{.N_+ - .N_i}{m - y_i} \right)^2 \left[ \sum_{j \neq i} \sum_{q \neq j,i} \epsilon_{ij} \epsilon_{iq} \frac{y_j y_q}{.N_j .N_q} \right] \left[ 1 - \frac{.N_i}{.N_+ - .N_i} \right] \\ &\quad - \frac{.N_i .N_+^2}{m^2 (.N_+ - .N_i)} \left[ \sum_{j \neq i} \epsilon_{ij}^2 \frac{y_j^2}{.N_j^2} \right] \end{aligned} \quad (3.22)$$

Both  $\hat{\mu}_{w_i|y_i}$  and  $\hat{\sigma}_{w_i|y_i}^2$  depend on the the fixed value of  $y_i$  as well as  $N_i$ . Finally then, the cumulative probability (3.14) can be evaluated as

$$\sum_{\forall y_i} \left[ 1 - \Phi \left( \frac{C_{\alpha^*} - \hat{\mu}_{u_i|y_i}}{\hat{\sigma}_{u_i|y_i}} \right) \right] P(y_i) \quad . \quad (3.23)$$

where  $P(y_i)$  is the binomial probability.

The problem of finding the required critical value for  $Z_{\lambda,red,i}$  has been reduced to a one-dimensional numerical problem in  $C_{\alpha^*,i}$ . A solution  $C_{\alpha^*,i}$  to the equation

$$\alpha^* - \sum_{y_i} \left[ 1 - \Phi \left( \frac{C_{\alpha^*,i} - \hat{\mu}_{u_i|y_i}}{\hat{\sigma}_{u_i|y_i}} \right) \right] P(y_i) = 0 \quad (3.24)$$

is required.

Extensive preliminary simulations under  $H_0$  revealed that the quantiles of  $Z_{\lambda,red,i}$  were smaller than the quantiles of  $Z_{m,i}$  for the same  $\alpha^*$ . Hence  $q_{\alpha^*}$ , the exact critical value for  $\lambda = 0$  is a logical starting point for a step-down algorithm. Denoting the left hand side of (3.24) by  $g(C_{\alpha^*,i})$ ,  $C_{\alpha^*,i}$  is decreased sequentially until  $g$  becomes negative. Once this occurs, the previous  $C_{\alpha^*,i}$  is used as a new starting point for the algorithm and the decrement is reduced. There is a range of possible  $Z_{\lambda,red,i}$  values for each  $y_i$  so  $g$  is a step function. The change in  $g$  will very small for some steps. The amount of probability accumulated at a particular  $C_{\alpha^*,i}$  depends on the value of  $y_i$ . The behavior of  $g$  is illustrated in Example 3.2. There are  $m = 200$  cases,  $k = 20$  sites and a total population of  $N_+ = 20000$ . The given site has population 1000 and the critical value under  $H_{0i}$  for  $\alpha = .05$  is 3.56. This site is one of nineteen with  $\gamma_i = 0.95$ . It is contiguous to a site with elevated rate ratio  $\gamma_j = 1.43$ . In the case vector for this example,  $y_i = 8$  is observed while  $y_j = 18$  at the neighbouring site. Breaks in the table indicate that the  $g$  is not changing to the decimal place shown. The algorithm was subsequently restarted at 3.02 with a decrement of .001 and a final  $C_{\alpha^*,i}$  of 3.015 was obtained.

**Example 3.2**

$C_{\alpha \cdot i}$	$g$
3.56	.0021506
3.54	.0021505
3.52	.0021505
-	-
3.46	.0021505
3.44	.0021503
3.42	.0021487
3.40	.0020764
-	-
3.32	.0020762
3.30	.0014910
3.28	.0014741
3.26	.0014718
3.24	.0014717
-	-
3.14	.0014701
3.12	.0014537
3.10	.0013522
3.08	.0010651
3.06	.0005933
3.04	.0001940
3.02	.0000146
3.00	-.000274

The final solution is a function of the data and involves a normal approximation so the solution of the algorithm is actually an estimate of quantile and can be denoted as  $\hat{C}_{\alpha \cdot i}$ . In some simple situations, a large final step in the algorithm to find  $\hat{C}_{\alpha \cdot i}$  gives a more conservative choice for the critical value.

### 3.4 Description of the Simulations

The approximate estimate  $\hat{\gamma}_{\lambda,red,i}$  and the test statistic  $Z_{\lambda,red,i}$  depend on the number of sites  $k$ , the number of cases  $m$ , the populations  $\mathbf{N}$ , the geographical configuration  $\{d_{ij}\}$ , the rate ratios  $\gamma$  and the smoothing parameter  $\lambda$ . An investigation of these factors in the context of testing with and without smoothing was carried out using simulation. The test statistic's ability to detect and identify existing elevated rates and to control the number of false rejections under the null hypothesis was investigated for each of these factors. Of particular interest is the performance of the test statistic with smoothing ( $\lambda > 0$ ) versus its performance without smoothing ( $Z_{\lambda,i} = Z_{ml,i}$  for  $\lambda = 0$ ).

Many levels of the factors could be examined but the following levels were chosen to simplify the interpretation while still maximizing the information that could be obtained. The set of simulations investigated each of the factors (i) to (iv) given below at two levels. The levels of  $\gamma$  (factor (v)) are unique to the type of cluster/elevation being investigated. While not part of the complete simulations set, some additional simulations were performed and will be described in Section 3.5

- (i)  $k$ : Two levels were chosen for the number of sites: 20 and 40. Although arbitrary, these choices may correspond to data aggregation at the county and municipal level respectively.
- (ii)  $m$ : This factor could actually be combined with (i) and written as  $m/k$ .  $m$  was chosen to correspond to an expected case count of 10 and 20 at each level of  $k$ . (i.e.  $m$  was chosen to be 200 and 400 for  $k = 20$ , to be 400 and 800 for  $k = 40$ ). The lower level corresponding to the smaller case count may be more interesting as it is more representative of rare diseases.

(iii) **N**: Many levels were possible for the population vector. The 'equal' level was a population of 1000 at each site. The '1 different population' level had  $k - 1$  sites with population 1000 while the other had population 500.

(iv)  $\{d_{ij}\}$ : Once again, many levels for geographical configuration are possible. Minimizing and maximizing the coefficient of variation (CV) of the inter-site distances was deemed to be appropriate.

The  $k$  sites arranged with  $k - 1$  on a circle and one in the center minimize  $CV(d_{ij})$  for  $k \geq 8$ . However, for a given  $\hat{\gamma}_{ml,i}$ , the value of  $w_i = \mathbf{e}'_{(i)} \hat{\gamma}_{(i)}$  given  $y_i$  in (3.17) is not random if  $i$  is the center site because the  $\epsilon_{ij}$  are equal. This results in a fixed value for  $Z_{\lambda,red,i}$  and  $\hat{C}_{\sigma^*,i}$  at this site. To ensure a range of values for  $Z_{\lambda,red,i}$  and  $\hat{C}_{\sigma^*,i}$  at all sites, the  $k$ 'th site was moved onto the circle with the other  $k - 1$ .  $CV(d_{ij})$  for the circle configuration is very close to the 'circle + center' configuration for  $k = 20$  and  $k = 40$ . The  $k - 1$  sites arranged as a circle with the other site outside the circle at a distance maximizes  $CV(d_{ij})$  so this configuration and the circle were taken as the two spatial configurations. A random set of locations was also used in a smaller number of simulations to verify the methodology for finding  $\hat{C}_{\sigma^*,i}$ .

(v)  $\gamma$ : The elevated rate(s) or cluster(s) depend on the true vector  $\gamma$  (or rates  $\mathbf{p}$ ). Many levels of this factor were investigated: homogeneity (overall null), a single elevated rate, two contiguous elevated rates as a cluster and two geographically separate elevated rates.

(vi)  $\lambda$ : For the simulations presented in this chapter, a fixed set of  $\lambda$ 's was used to govern the amount of smoothing applied. In other words, the choices for  $\lambda$  were made beforehand, and the values were not data dependent. The range for  $\lambda$  was determined by establishing an approximate upper bound. For each site,  $\hat{\text{Var}}(\hat{\gamma}_{i,\lambda})$  tends to 0 as  $\lambda$  increases. Using the known true  $\gamma$  vector as input, the value for which  $\text{Var}(\hat{\gamma}_{i,\lambda})$  is approximately zero can be used as an upper bound for  $\lambda$  at each site. The one term approximation is  $\hat{\text{Var}}(\hat{\gamma}_{\lambda,i}) = \text{Var}(\hat{\gamma}_{ml,i}) + 2\lambda \text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_{red,i})$ , and equating to



zero gives  $\lambda_{ub,i} = -\text{Var}(\hat{\gamma}_{ml,i})/2\text{Cov}(\hat{\gamma}_{ml,i}, \hat{I}_{red,i})$ . The upper bound for  $\lambda$  must be applicable to all sites, the minimum of  $\lambda_{ub,i}$  from  $i = 1, \dots, k$  was taken as  $\lambda_{UB}$ . For observed values of  $\hat{\gamma}_{ml,i}$  far from  $\gamma_i$ ,  $\lambda$  close to  $\lambda_{UB}$  may still give a negative  $\text{Var}(\hat{\gamma}_{i,\lambda})$  for some  $i$ . As a result, only the first four fifths of the range from 0 to  $\lambda_{UB}$  was used. A very small  $\lambda$  of  $\lambda_{UB}/10$  was chosen to see the initial impact of smoothing and the chosen  $\lambda$  set was: 0,  $\lambda_{UB}/10$ ,  $\lambda_{UB}/5$ ,  $2\lambda_{UB}/5$ ,  $3\lambda_{UB}/5$ ,  $4\lambda_{UB}/5$ . The set of  $\lambda$ 's chosen is unique to a given simulation because  $\lambda_{ub,i}$  depends on  $N_i$ .  $\lambda$ 's chosen for the larger value of  $m$  were usually smaller. The overall  $\lambda_{UB}$  must be an upper bound for  $\lambda$  for the lower population site. Approximate variance estimates at this site go to zero much more quickly. Results on the detection of elevated rates in Section 3.6 should be considered with this in mind. Figures 3.1 and 3.6 of that section display the ranges of  $\lambda$  for several simulations.

A new simulation set of case vectors was generated for each  $\lambda$ . The observed response in all simulations is the fraction of tests rejected. For each generated case vector, whether or not the test of  $H_0: \gamma_i = 1$  was rejected using  $Z_{\lambda, red,i}$  was tabulated for  $i = 1, \dots, k$ . The total number of tests rejected for site  $i$  divided by the number of repetitions of the simulation gives the error rate for that site. The overall error rate is the fraction of repetitions in which at least one site was declared elevated (i.e. at least one test was rejected for  $i = 1, \dots, k$ ). The site error rate was examined closely when  $H_0$  was true for all  $i$  (homogeneity) as it might be possible for a site error rate to be beyond the acceptable range while the overall size of the test is still controlled. Controlling the size of the test corresponds to control of the number of false positives. Simulations of size 10000 were used for these simulations of homogeneity while simulations of size 5000 were used for 'alternative' simulations in which  $H_{0i}$  was not true for at least one  $i$ .

### 3.5 Results - Size of Tests

Homogeneity or complete null simulations were performed ( $H_0$  true for all  $i$ ) to investigate the ability of the method to properly control the number of false rejections. For a given case vector, the estimate  $\hat{\gamma}_{\lambda,j}$  and test statistic  $Z_{\lambda,red,j}$  are obtained. The algorithm described in Section 3.3.3 is used to obtain  $\hat{C}_{\lambda,\alpha^*,j}$ . If  $Z_{\lambda,red,j} \geq \hat{C}_{\lambda,\alpha^*,j}$ , then  $H_0$  is rejected and the incidence at site  $j$  is declared 'elevated'. For  $\lambda = 0$ , the exact critical values for  $Z_{ml,i}$  are described in Section 3.2. When more than one site is declared elevated, the spatial contiguity, as defined by a common border, determines whether the sites form a cluster. Complete null or homogeneity simulations were performed for each of the levels of  $k$ ,  $d_{i,j}$  and  $\mathbf{N}$  described in the previous section. In addition, a random set of populations was assigned to a set of random locations from a 10 by 10 grid for each level of  $k$  to further verify the methodology under homogeneity. The method produced rejection rates that controlled the overall and site error rates under all homogeneity simulations.

Tables 3.1 and 3.2 display the percentage of tests rejected for 'complete null' or 'homogeneity' simulations for the two spatial configurations for  $k = 20$ . Given are the fraction of tests rejected at the site and overall level based on 10000 repetitions.

At  $\lambda = 0$  the critical values obtained are exact. The achieved level which is denoted by  $\alpha_{ach}$  is described in Section 3.2. The upper probability limit for the percentage of tests rejected (at the site level) is approximately  $\alpha_{ach} + 2\sqrt{\alpha_{ach}(1 - \alpha_{ach})/10000}$ . Footnoted between the tables are acceptable upper bounds for the percent rejected for  $\lambda > 0$ . Lower bounds are not listed since Bonferroni test levels are known to be conservative and hence we are only concerned with not exceeding the upper bound. There are two possible upper bounds for the ' $N$  different configuration', the larger

was used. The standard error is based on a normal approximation for the proportion of tests rejected and this bound was not exceeded in any simulation. The observed percent rejected is useful for comparison in simulations with at least some departure from homogeneity. For  $\lambda = 0$ , the upper probability limits for Tables 3.1 and 3.2 (footnoted as (1) through (8)) are : .0019, .0258, .0024, .0363, .0031, .0438, .0025 and .0359. For  $\lambda > 0$ , the upper probability limits at the site and overall levels for Tables 3.1 and 3.2 are .0037 and .0544 respectively.

The critical values at  $\lambda = 0$  are exact: it is the percentage of tests rejected for  $\lambda > 0$  that is most relevant for establishing the utility of the methodology which gives  $\hat{C}_{\alpha, \lambda, i}$ . With  $\alpha^* = \alpha / (k - 1)$ , and once again using a standard error based on a normal approximation, the upper probability limit for the acceptable fraction of tests rejected at the site level is approximately  $\alpha^* + 2\sqrt{\alpha^*(1 - \alpha^*)/10000}$ . In Tables 3.1 and 3.2 as well as all other null simulations, the upper bound is not exceeded for any  $\lambda$ , and the fraction of repetitions with at least one test rejected is quite conservative in many instances. One might note an increasing trend with  $\lambda$  for some simulations for the overall level but it remains less than the upper bound. Appendix 3.5 presents the circle configuration null simulations for  $k = 40$ .

Table 3.1

## Homogeneity Simulations

Circle Configuration. 20 sites

	$m = 200$				$m = 400$			
	N's equal		1 $N_j$ different		N's equal		1 $N_j$ different	
	max site frej.	overall frej.	max site frej.	overall frej.	max site frej.	overall frej.	max site frej.	overall frej.
$\lambda$								
0	.0017 (1)	.0248 (2)	.0022 (3)	.0327 (4)	.0028 (5)	.0403 (6)	.0023 (7)	.0326 (8)
$\lambda_1$	.0034	.0518	.0025	.0381	.0030	.0388	.0030	.0432
$\lambda_2$	.0032	.0462	.0024	.0346	.0033	.0413	.0030	.0456
$\lambda_3$	.0033	.0506	.0027	.0404	.0034	.0426	.0031	.0477
$\lambda_4$	.0032	.0484	.0033	.0503	.0030	.0436	.0030	.0498
$\lambda_5$	.0034	.0511	.0034	.0508	.0027	.0420	.0031	.0511

max site frej. = maximum fraction of tests rejected at any site  
 overall frej. = fraction of repetitions with at least one test rejected  
 Upper probability limits (1) through (8) are given in text

Table 3.2

## Homogeneity Simulations

Circle + 1 Configuration, 20 sites

	$m = 200$				$m = 400$			
	N's equal		1 $N_j$ different		N's equal		1 $N_j$ different	
$\lambda$	max site frej.	overall frej.	max site frej.	overall frej.	max site frej.	overall frej.	max site frej.	overall frej.
0	.0015 (1)	.0213 (2)	.0020 (3)	.0349 (4)	.0027 (5)	.0410 (6)	.0016 (7)	.0287 (8)
$\lambda_1$	.0033	.0510	.0028	.0387	.0028	.0396	.0031	.0478
$\lambda_2$	.0031	.0508	.0028	.0397	.0030	.0407	.0032	.0488
$\lambda_3$	.0029	.0481	.0030	.0502	.0030	.0410	.0034	.0528
$\lambda_4$	.0034	.0518	.0026	.0378	.0028	.0402	.0032	.0498
$\lambda_5$	.0030	.0500	.0026	.0345	.0029	.0412	.0032	.0488

max site frej. = maximum fraction of tests rejected at any site

overall frej. = fraction of repetitions with at least one test rejected

Upper probability limits (1) through (8) are given in text

### 3.6 Results - Detection of Elevated Rates

The simulation results under complete homogeneity show that the site and overall error rates are controlled under  $H_o$ . As a result, simulations under  $H_a$  can be considered. In these simulations,  $H_o : \gamma_i = 1$  is not true for at least one site. The true  $\gamma$  vector is known in simulation so one can express  $\gamma$  in terms of a departure from overall homogeneity. Under  $H_a$ , the traditional Pearson statistic has a noncentral chisquare distribution with noncentrality parameter  $\delta = m \sum_{j=1}^k (\theta_j - \theta_{oj})^2 / \theta_{oj}$ . In the present context, this corresponds to  $\delta = m \sum_{j=1}^k N_j (\gamma_j - 1)^2 / N_+$ . A wide range of alternatives was possible for the simulations but low to moderate departures were considered. As an illustration, consider 20 equal populations  $N_j = 1000$  with case total  $m = 200$ . The incidence at site  $k$  is elevated with rate  $p_k = .0075$  and  $p_j = .005$  for  $j \neq k$ , giving rate ratios  $\gamma_k = 1.46$  and  $\gamma_j = 0.977$  for  $j \neq k$  and a departure from homogeneity of  $\delta = 2.79$ . Two elevated rates at sites  $k$  and  $k - 1$  with incidence  $p_{k-1} = p_k = .0075$  and  $p_j = .005$  for  $j \neq k, k - 1$  give rate ratios  $\gamma_k = \gamma_{k-1} = 1.43$  and  $\gamma_j = .953$  for  $j \neq k, k - 1$  and a departure from homogeneity of  $\delta = 4.03$ . All simulations under an alternative were of size 5000.

Presented in Table 3.3, are the results of four simulations involving a single elevated site on a circle. They involve simulations for  $k = 20$  sites for each of the two choices for  $N$ . In those with rate ratios  $\gamma_{el} = 1.46$  or  $\gamma_{el} = 1.48$ , the incidence rate at the site of the elevated rate is one and a half times the incidence rate of the others. The differing populations in two of the simulations causes  $\gamma_{el}$  to be slightly different in those simulations although the incidence rates are the same. Within each table entry is the following: 'frej.' - the fraction of repetitions with at least 1 site declared elevated; 'ftcor.' - the fraction of repetitions in which only the elevated rate was declared elevated, and 'fincl.' - the fraction of repetitions in which the elevated site was included among those declared elevated. As an illustration, consider the top left

entry in the table. In  $354/5000 = .0778$  repetitions, at least one site was declared elevated. In  $301/5000 = .0602$  repetitions, the site of the elevated rate was the only one declared elevated and in  $304/5000 = .0608$  the site with elevated incidence was among those sites which had the null hypothesis rejected.

Several observations can be made from this table which also were made for other simulations/factor combinations which involve a single elevated site. A significant increase in detection ability is observed when  $\lambda$  is increased from  $\lambda = 0$  (testing at level  $\alpha = .05$ ). This observation is relevant for the fraction of repetitions with at least one rejection and the fraction totally correct. A significant increase in the observed fraction from  $\lambda = 0$  to the first  $\lambda$  is indicated by (\*\*) while (\*) indicates that the observed fraction is significantly greater than that observed at  $\lambda = 0$  but not from the previous  $\lambda$ . Significant increases in detection ability with smoothing were observed but no significant decreases in detection ability were found with smoothing. This important result held true throughout the entire simulation set.

The simulation results presented in Table 3.4 involve the same  $\gamma$ ,  $m$  and  $N$  as those of Table 3.3 but the 'circle + 1' configuration was used. In this configuration, the distant site is located at (1.5) while the other 19 are equidistant on a circle centered at (1.1). See Appendix 3.7 for their locations and spatial contiguity. The observed maximum achieved for the fraction with at least one rejection and fraction totally correct are similar but a significant increase from  $\lambda = 0$  occurred at the second  $\lambda$  of the set. This is intuitively sensible since it should require more smoothing within the proposed range to bring the estimate of incidence at the geographically separate site towards the others. Hence a larger  $\lambda$  is required to achieve the benefit of smoothing with respect to inference. In both tables, no significant increase in detection ability is found when the elevated incidence occurs at the lower population site.

**Table 3.3**

Ability to Detect Single Elevated Rate

Circle Configuration, 20 sites and 200 cases

$\gamma_{el}$	N's equal		1 $N_j$ different	
	1.46	1.90	1.48	1.95
	frej. ftcor. fincl.	frej. ftcor. fincl.	frej. ftcor. fincl.	frej. ftcor. fincl.
0	.0778 .0602 .0608	.3648 .3536 .3560	.0730 .0450 .0452	.2174 .1964 .1990
$\lambda_1$	.1278** .0934** .0950**	.4674** .4382** .4462**	.0720 .0432 .0446	.2248 .1996 .2040
$\lambda_2$	.1334* .0986* .0994*	.4674* .4392* .4444*	.0691 .0420 .0422	.2188 .1962 .1990
$\lambda_3$	.1298* .0952* .0964*	.4594* .4348* .4380*	.0726 .0434 .0442	.2364 .2092 .2112
$\lambda_4$	.1408* .1008* .1012*	.4640* .4370* .4444*	.0752 .0460 .0470	.2262 .1968 .1990
$\lambda_5$	.1284* .0910* .0920*	.4636* .4346* .4548*	.0710 .0444 .0458	.2254 .1948 .1956

frej. = fraction of repetitions with at least one test rejected

ftcor. = fraction of repetitions in which only elevated site is declared elevated

fincl. = fraction of repetitions in which elevated site is among those declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$ \* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$



**Table 3.4**

Ability to Detect Single Elevated Rate  
Circle + 1 Configuration, 20 sites and 200 cases

$\gamma_{el}$	N's equal		1 $N_j$ different	
	1.46	1.90	1.48	1.95
	frej. ftcor. fincl.	frej. ftcor. fincl.	frej. ftcor. fincl.	frej. ftcor. fincl.
0	.0810 .0640 .0642	.3580 .3476 .3492	.0678 .0392 .0396	.2104 .1894 .1916
$\lambda_1$	.0986 .0660 .0670	.3722 .3494 .3526	.0682 .0394 .0900	.2242 .1988 .2014
$\lambda_2$	.1244** .0874** .0886**	.4446** .4206** .4248**	.0732 .0410 .0422	.2344 .2094 .2188
$\lambda_3$	.1260* .0926* .0936*	.4426* .4204* .4254*	.0846 .0446 .0448	.2372 .2076 .2108
$\lambda_4$	.1332* .0980* .0988*	.4680* .4490* .4536*	.0776 .0432 .0436	.2370 .2052 .2078
$\lambda_5$	.1284* .1010* .1020*	.4452* .4344* .4372*	.0754 .0436 .0438	.2392 .2106 .2384

frej. = fraction of repetitions with at least one test rejected

ftcor. = fraction of repetitions in which only elevated site is declared elevated

fincl. = fraction of repetitions in which elevated site is among those declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

\* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$

More results for a single elevated rate are presented in Appendix 3.6 based on simulations for  $k = 20$  and  $m = 400$  as well as  $k = 40$  and  $m = 400$  are presented. Figure 3.1 displays results from simulations for twenty sites in the circle configuration. The top panel shows the fraction with at least 1 rejection for  $m = 200$  (from Table 3.3) while the lower plot shows the fraction with at least one rejection for  $m = 400$  (from Appendix 4.6). Points with labels A through D correspond to columns 1 through 4 for each simulation. Significant increases in detection ability were found less often with the higher  $m/k$  ratio. The plot is relatively flat and this observation holds for both  $k = 20, m = 400$  and  $k = 40, m = 800$ . It is also worth noting that simulations involving the one differing  $N_j$  (with points labelled C and D) have a somewhat smaller  $\lambda$  range.

Without smoothing, only one critical value exists for each value of  $y_i$ . As described earlier,  $Z_{\lambda, red, i}$  can assume a large number of values due to the possible distributions of the remaining case counts at the other sites. Corresponding values of  $\hat{C}_{\alpha^*, i}$  are also generated and rejection may now occur at a  $y_i$  not previously included in the rejection region. As a result, testing with smoothing is at a level closer to  $\alpha^* = \alpha/(k-1)$  than is achieved at  $\lambda = 0$  (see Section 3.2). Therefore, tests are performed at a more liberal level with smoothing, leading to the observed increases in detection ability. Furthermore, no significant decreases in detection ability have been observed with smoothing.

Presented in Tables 3.5 and 3.6 are the inference results for two elevated contiguous rates (a cluster of size two) for  $k = 20$  and  $m = 200$ . The departure from homogeneity has now increased because of the two elevated rates. The departure, populations and  $\gamma$  are as follows:

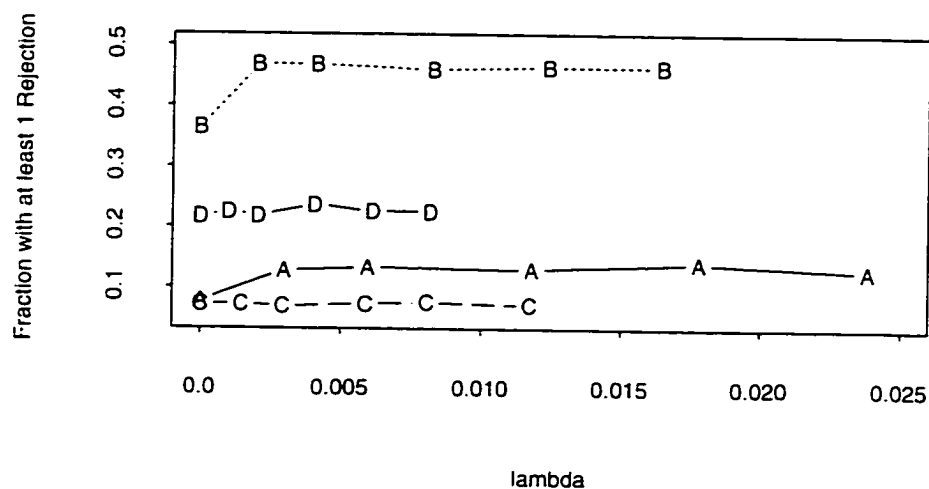
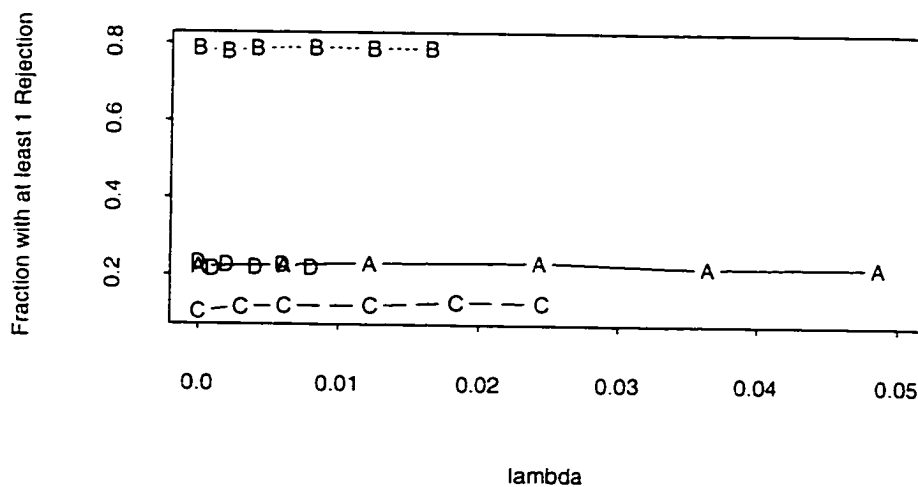
Single Elevated Rate, circle,  $m = 200$ , (from Table 3.3)Single Elevated Rate, circle,  $m = 400$  (from Appendix 3.6)

Figure 3.1 : Fraction of at least 1 Rejection, Single Elevated Rate

**equal N**

A: departure: 4.08.  $\gamma_{el1} = \gamma_{el2} = 1.43$

B: departure: 9.84.  $\gamma_{el1} = 1.86$ .  $\gamma_{el2} = 1.40$

C: departure: 9.84.  $\gamma_{el1} = 1.40$ .  $\gamma_{el2} = 1.86$

**1 N different**

D: departure: 3.29.  $\gamma_{el1} = \gamma_{el2} 1.44$

E: departure: 6.48.  $\gamma_{el1} = 1.90$ .  $\gamma_{el2} = 1.43$

F: departure: 6.48.  $\gamma_{el1} = 1.43$ .  $\gamma_{el2} = 1.90$

In the simulations described in Table 3.5, the two elevated rates are contiguous neighbours among the twenty sites which are equally spaced on a circle of radius one. In the simulations summarized in Table 3.6, one of the elevated rates is always at the distant site while the other is on the same side of the circle as the distant site. In a table's column headings, 'frej.' is the fraction of repetitions in which at least one site was declared elevated 'f(1)in.' represents the fraction of repetitions which include the first elevated site among those declared elevated (similarly for 'f(2)in.'). The abbreviation 'ftcor.' represents the fraction totally correct or the fraction of repetitions which had only the two elevated sites declared as elevated. Asterisks have the same meaning as in Tables 3.3 and 3.4 while a '\*x' indicates a significant increase in the observed fraction from the previous  $\lambda$ .

The results of both tables are very similar. As with the single elevated rate(s), the largest gains in detection for simulations A through C' can be observed in the change from  $\lambda = 0$  to  $\lambda = \lambda_1$ . The maximum detection ability achieved for simulations B and C' is similar within both tables. Comparing simulations B and C' between Tables 3.5 and 3.6, we see that a slightly higher detection ability is achieved with the two elevated contiguous rates are closer together (i.e. in Table 3.5). This observation can not be made between tables for simulations E and F. In simulations D, E and F of both tables, if a significant increase from an observed fraction at  $\lambda = 0$  is observed,

then it is achieved at a larger  $\lambda$  in the set than in simulations A through C. This observation can be interpreted as meaning that more smoothing may be required to obtain the same gains in inference if the two populations in the cluster of size two are very different. The observed fraction totally correct is very low due to the small departure from homogeneity and small probability of observing a large number of cases at each of the elevated sites. The results for a cluster of size two for the circle configuration and  $k = 40$  are presented in Appendix 3.8.

**Table 3.5**  
Ability to Detect Two Contiguous Elevated Rates

Circle,  $k = 20, m = 200$

$\lambda$	A	B	C	D	E	F
	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.
0	.1040 .0494 .0472 .0016	.3438 .3086 .0376 .0098	.3520 .0422 .3128 .0096	.1136 .0618 .0380 .0012	.3962 .3640 .0298 .0120	.2440 .0602 .1782 .0074
$\lambda_1$	.1846** .0792** .0834** .0062**	.4574** .3962** .0630** .0174**	.4506** .0630** .3848** .0154**	.1190 .0670 .0318 .0014	.4086 .3744 .0314 .0126	.2554 .0666 .1850 .0104
$\lambda_2$	.1876* .0842* .0808* .0058*	.4664* .4074* .0694* .0250*	.4580* .0648* .4030* .0234*x	.1242 .0710 .0334 .0014	.4152 .3834** .0322 .0128	.2598 .0652 .1900 .0097
$\lambda_3$	.1768* .0794* .0820* .0062*	.4636* .4054* .0670* .0226*	.4616* .0712* .3998* .0238*	.1282** .0768** .0338 .0016	.4176** .3850* .0319 .0128	.2532 .0648 .1848 .0094
$\lambda_4$	.1844* .0850* .0748* .0052*	.4456* .3918* .0620* .0206*	.4596* .0688* .3986* .0244*	.1314* .0796* .0364 .0020	.4342* .3998* .0324 .0130	.2514 .0638 .1802 .0106
$\lambda_5$	.1904* .0790* .0882* .0046*	.4626* .3946* .0706* .0198*	.4646* .0700* .4010* .0022*	.1324* .0826* .0320 .0222	.4334* .4086* .0316 .0126	.2568 .0690 .1826 .0016

frej. = fraction of repetitions with at least one test rejected

f(1)in. = fraction of repetitions in which first elevated site is among those declared elevated

f(2)in. = fraction of repetitions in which second elevated site is among those declared elevated

ftcor. = fraction of repetitions in which only elevated sites are declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

\* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$

\*x indicates significant difference from observed fraction at  $\lambda = 0$  and previous  $\lambda$

**Table 3.6**  
Ability to Detect Two Contiguous Elevated Rates

Circle + 1.  $k = 20$ .  $m = 200$

$\lambda$	A	B	C	D	E	F
	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.
0	.1108 .0540 .0480 .0016	.3480 .3050 .0422 .0098	.3550 .0414 .3160 .0100	.1282 .0762 .0352 .0016	.3950 .3638 .0314 .0096	.2502 .0652 .1768 .0084
$\lambda_1$	.1674** .0800* .0634* .0030*	.4448** .4001* .0452* .0142*	.4236** .0708* .3636* .0244*	.1290 .0708 .0420 .0022	.4020 .3660 .0356 .0086	.2590 .0646 .1884 .0094
$\lambda_2$	.1490* .0760* .0518* .0026*	.4408* .3980* .0412* .0140*	.4046* .0642* .3322* .0156*	.1194 .0676 .0388 .0018	.4166** .3798** .0376 .0088	.2472 .0632 .1755 .0082
$\lambda_3$	.1618* .0780* .0612* .0044	.4406* .3986* .0378* .0114*	.3982* .0722* .3086* .0172*	.1310 .0780 .0384 .0024	.4074 .3768 .0382 .0132	.2498 .0648 .1784 .0088
$\lambda_4$	.1870*x .0828* .0810* .0044*	.4470* .4002* .0552*x .0120*	.4010* .0666* .3360*x .0180*	.1442** .0850 .0396 .0024	.4298*x .3926*x .0396 .0103	.2696** .0794** .1862 .0130**
$\lambda_5$	.1788* .0810* .0736* .0032*	.4442* .3962* .0454* .0126*	.4062* .0672* .3546* .0220*	.1436 .0860 .0389 .0024	.4286* .3942* .0328 .0094	.2716* .0798* .1868 .0134*

frej. = fraction of repetitions with at least one test rejected

f(1)in. = fraction of repetitions in which first elevated site is among those declared elevated

f(2)in. = fraction of repetitions in which second elevated site is among those declared elevated

ftcor. = fraction of repetitions in which only elevated sites are declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

\* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$

\*x indicates significant difference from observed fraction at  $\lambda = 0$  and previous  $\lambda$

Presented in Figure 3.2 are the entries corresponding to the fraction of repetitions with at least one rejection in Table 3.5. The large initial increase in detection ability in the simulations with equal populations is the most striking feature of the upper plot. The chosen range of  $\lambda$  in the simulations D, E, F of the lower plot is much smaller. As described in Section 3.4, the approximate variance for the low population site tends towards zero much more quickly.

The results for two noncontiguous elevated rates for  $k = 20, m = 200$  in the 'circle + 1' configuration are presented in Appendix 3.8. The magnitude of the true  $\gamma_j$ 's and the departure from homogeneity for simulations A - F are the same as described above for Tables 3.5 and 3.6. The results for this configuration are similar to the contiguous rates of Table 3.6. However, the initial impact of smoothing on inference is not as large in these simulations as when the two elevated rates are equal (simulations A and D in Table 3.6). Perhaps this implies that less smoothing is required when the two elevated rates are very close together and the same. Results for two elevated noncontiguous rates for  $k = 40, m = 400$  are also presented in Appendix 3.8. A small drop in detection ability can be observed with the increase in  $k$  but not in all situations.

As a practical check of the methodology, some simulations were run with a set of randomly chosen populations between 500 and 1000, assigned to 20 randomly chosen locations on a 10 by 10 grid. Regions were formed around each  $(x, y)$  location. The populations, locations and sets of contiguous sites are given in Appendix 3.9. Complete null or homogeneity simulations for this configuration controlled the  $\alpha$  level at the site and overall level for all  $\lambda$ 's considered.

Presented in Table 3.7 are results for a single elevated site for this new configuration at  $m = 200$  and  $m = 400$ . Although an arbitrary choice, site 13 in Table 3.7 was



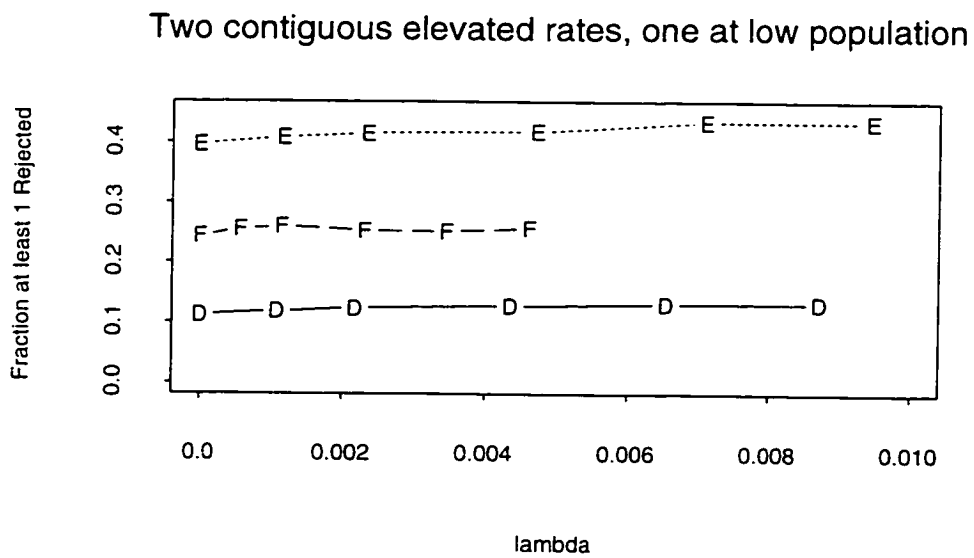
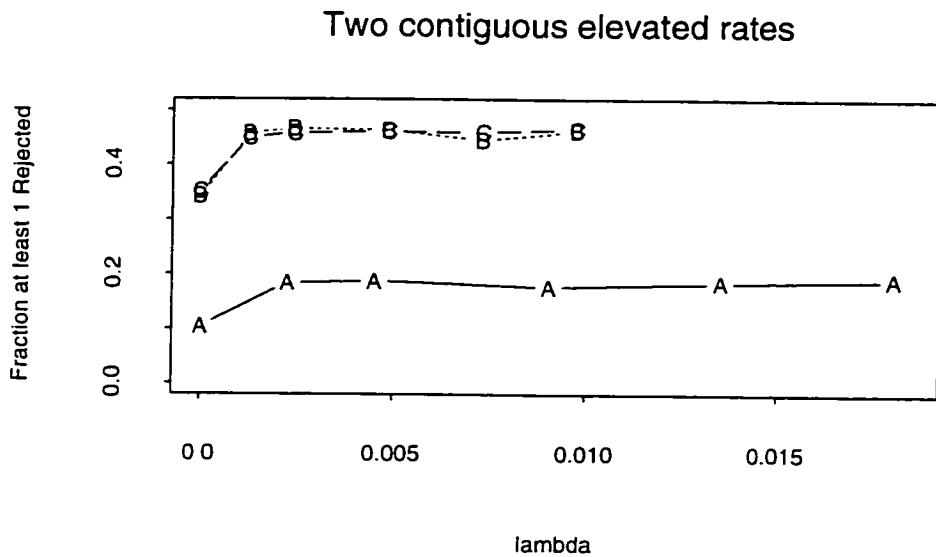


Figure 3.2 : Two contiguous Elevated Rates (from Table 3.5), Fraction with at least 1 rejection

chosen as the elevated site. It is geographically near at least two others. The elevated rates for these four simulations are: (i)  $\gamma_{el} = 1.47$ . (ii)  $\gamma_{el} = 1.93$ . (iii)  $\gamma_{el} = 1.47$ . (iv)  $\gamma_{el} = 1.93$ .

As in the equal population simulations of Tables 3.3 and 3.4, an increase in detection ability is observed from  $\lambda = 0$  to  $\lambda = \lambda_1$  and this ability is maintained throughout the set. However, there are noteworthy differences. There are significant increases for  $m = 400$ . At  $m = 400$ , the simulations presented in Tables 3.3 and 3.4 in which the elevated incidence rate was twice the others showed no significant improvement in detection ability with smoothing. However  $\gamma_{el}$  is virtually the same in the simulations of Table 3.8. In the simulations of Table 3.3 and Table 3.4,  $\gamma_{el}$  was 1.90 while  $\gamma_{el}$  was 1.93 in these simulations. Perhaps this implies some limitations to the effect of smoothing on inference with the chosen spatial configurations of the simulation set. The maximum observed fraction is an increase of approximately 15 percent from  $\lambda = 0$ . Significant increases in observed fractions occur between positive  $\lambda$ 's in the chosen set (indicated by \*x). Results for two elevated rates and  $m = 200$  cases for this new configuration are presented in Appendix 3.8.

**Table 3.7**  
 Ability to Detect Single Elevated Rate  
 Random populations/locations Configuration

	$k = 20, m = 200$		$k = 20, m = 400$	
	i	ii	iii	iv
	frej. fincl. ftcor.	frej. fincl. ftcor.	frej. fincl. ftcor.	frej. fincl. ftcor.
0	.0680 .0378 .0372	.2402 .2260 .2234	.1122 .0856 .0842	.5298 .5208 .5138
$\lambda_1$	.0931** .0638** .0626**	.3368** .3162** .3116**	.1580** .1292** .1270**	.6244** .6148** .6020**
$\lambda_2$	.0988* .0690* .0674*	.3368* .3190* .3126*	.1814*x .1504*x .1468*x	.6356* .6236* .6088*
$\lambda_3$	.1026* .0692* .0680*	.3390* .3208* .3148*	.1954*x .1634*x .1598*x	.6792*x .6682*x .6532*x
$\lambda_4$	.0940* .0702* .0686*	.3380* .3170* .3120*	.2004* .1710* .1666*	.6816* .6720* .6626*
$\lambda_5$	.0938* .0644* .0630*	.3442* .3218* .3160*	.1912* .1664* .1632*	.6808* .6714* .6624*

frej. = fraction of repetitions with at least one test rejected

fincl. = fraction of repetitions in which elevated site is among those declared elevated

ftcor. = fraction of repetitions in which only elevated site is declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

\* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$

\*x indicates significant difference from observed fraction at  $\lambda = 0$  and previous  $\lambda$

Finally, as another practical check of the methodology, a large 'bump' cluster of contiguous elevated rates was considered. It was suspected that smoothing would increase the inclusion rate of several elevated rates into the cluster. Or in other words, smoothing should be advantageous in identifying elevated rates within one existing large cluster. Presented in Table 3.8 are four simulations with  $k = 20$ ,  $m = 200$  in which seven out of twenty sites are indeed elevated in the following form. The incidence rate at the highest point on the bump is five times the common thirteen while pairs of neighbouring sites in the cluster have incidence which are four times, three times, and twice as elevated as the common thirteen. The locations and contiguity can be found in Appendices 3.7 and 3.10. The populations and  $\gamma$  values are as follows

(I) Random Population/Location Set (see Appendix 3.10 for populations) :

$$\gamma_{el1} = 2.81, \gamma_{el2} = \gamma_{el3} = 2.25, \gamma_{el4} = \gamma_{el5} = 1.69, \gamma_{el6} = \gamma_{el7} = 1.12.$$

(II) Circle, equal populations, all  $N_j = 1000$ :

$$\gamma_{el1} = 2.78, \gamma_{el2} = \gamma_{el3} = 2.22, \gamma_{el4} = \gamma_{el5} = 1.67, \gamma_{el6} = \gamma_{el7} = 1.11.$$

(III) Circle,  $N_{el1} = 500$ , others 1000:

$$\gamma_{el1} = 2.91, \gamma_{el2} = \gamma_{el3} = 2.33, \gamma_{el4} = \gamma_{el5} = 1.74, \gamma_{el6} = \gamma_{el7} = 1.16.$$

(IV) Circle + 1, equal populations of 1000,  $\gamma_{el1}$  at distant site:

$$\gamma_{el1} = 2.78, \gamma_{el2} = \gamma_{el3} = 2.22, \gamma_{el4} = \gamma_{el5} = 1.67, \gamma_{el6} = \gamma_{el7} = 1.11.$$

Entries in the table are the fraction of tests in which there was at least one rejection (frej.) and the percent of tests in which at least four of the seven sites are included in the declared cluster. The abbreviation '> hf.in.' represents the fraction of case vectors in which greater than half of the elevated rates are declared elevated. Since the fraction of tests with at least one rejection is approximately 1.0, we are most interested in the inclusion of more than half of the elevated sites in the cluster. As denoted previously, (\*\*) indicates the initial significant increase from  $\lambda = 0$ . (\*)

indicates that the observed fraction is still significantly greater than the observed fraction at  $\lambda = 0$  while ( $\times$ ) indicates that the observed fraction is significantly greater than the observed fraction at  $\lambda = 0$  and the observed fraction at the previous  $\lambda$ .

Simulations (I) and (IV) show gains in the inclusion rate with each successive  $\lambda$ . For simulation (I), these gains are significant up to the third  $\lambda$  in the set. Simulation (I) is the simulation using the random population/location set and the observed inclusion fraction levels off at approximately twice the initial inclusion rate. Simulations (II) and (III) with circle configurations show similar patterns to those discovered earlier in the main simulation set. In the equal population simulation (II), there is a large significant initial gain with smoothing and this inclusion rate is maintained. Although the initial gain is smaller, the same pattern exists in (III). With the peak of the 'bump' separated from the circle in simulation (IV), the observed inclusion rate at the final  $\lambda$  is almost four times that which was observed at  $\lambda = 0$ , but is still at only two percent.

**Table 3.8**  
 Ability to Detect Many Elevated Rates  
 Bump Situation. Four Configurations

$k = 20, m = 200$				
$\bar{r}$ elevated contiguous rates				
	I	II	III	IV
	frej. > hf.in.	frej. > hf.in.	frej. > hf.in.	frej. > hf.in.
0	.9978 .0814	.9982 .1022	.9984 .1648	.9992 .0052
$\lambda_1$	.9998 .1234**	1.0 .2082**	.9984 .1876**	.9992 .0158**
$\lambda_2$	.9992 .1424*x	.9998 .2062*	.9986 .2030*	.9978 .0170*
$\lambda_3$	.9994 .1592*x	1.0 .1998*	.9988 .2082*	.9998 .0178*
$\lambda_4$	.9992 .1684*	.9996 .2004*	.9986 .2044*	.9998 .0190*
$\lambda_5$	.9998 .1666*	.9996 .2090*	.9988 .2070*	.9998 .0200*

Configurations (I) to (IV) described in text  
 frej. = fraction of repetitions with at least one test rejected  
 > hf in. = fraction of repetitions in which more than half of elevated rates are included

## Chapter 4

# Choice of Smoothing Parameter

In this chapter, the problem of choosing the smoothing parameter is addressed. The relevant issues associated with this choice are introduced in Section 4.1. In Section 4.2, an MSE based criterion is given. Two case deletion criteria are presented in Section 4.3. Section 4.4 outlines a combined strategy for the choice of  $\lambda$  and Section 4.5 presents results for this suggested strategy.

### 4.1 Introduction

In the previous chapter, the smoothing parameter  $\lambda$  was considered fixed. For each of the simulations performed, an upper bound for  $\lambda$  was established using the known  $\gamma$  vector. In practice, the true  $\gamma$  vector is unknown and the observed data consists of only have a single case vector  $\mathbf{y}$ . It is unlikely that one has any prior knowledge about the appropriate degree of smoothing. An estimate  $\hat{\lambda}$  must be made from the data according to certain criteria. The smoothed estimate vector  $\hat{\gamma}_\lambda$  has better sampling

properties than its unsmoothed counterpart  $\hat{\gamma}_{ml}$  but any estimation-based criterion must be considered in conjunction with criteria which consider the ability to detect elevated rates or clusters.

In general, various strategies are used to choose a smoothing parameter. The total mean squared error is an estimation-based criterion as the reduction in total variance is balanced with increase in bias. The minimum MSE can be viewed as minimization of a quadratic loss function. A predictive loss function criterion emphasizes the ability of the smoothing prescription to predict a future observation. For predictive loss functions such as the one presented in Section 4.3, the actual choice of smoothing parameter is usually determined by cross-validation. In the Bayesian framework, the smoothing parameter is contained within the specified prior density. If appropriate hyperparameters can't be established, Empirical Bayes methods are usually employed. Typically, these empirical estimates are found by maximizing a posterior for the smoothing parameter.

## 4.2 An MSE choice for $\lambda$

As outlined in Section 2.6, the total MSE for the numerically obtained smoothed estimate vector  $\hat{\gamma}_{sm}$  is

$$\text{TOTMSE}(\hat{\gamma}_{sm}) = \sum_{i=1}^k \text{Var}(\hat{\gamma}_{sm,i}) + \sum_{i=1}^k [E(\hat{\gamma}_{sm,i}) - \gamma_i]^2. \quad (4.1)$$

This function can be minimized numerically over a specified range for  $\lambda$ . However, the moments of  $\hat{\gamma}_{sm}$  are not known exactly but only approximately through expansion as described in Section 2.5.



Alternatively, an approximation to  $\hat{\gamma}_{sm}$  can be used and its mean squared error given exactly. For example, if the approximate estimate given by (2.18) is linear in  $\lambda$ , the overall expression for the total mean squared error is quadratic in  $\lambda$ .

The first or variance term of (4.1) is

$$\sum_{i=1}^k \text{Var}(\hat{\gamma}_{ml,i}) + 2\lambda \sum_{i=1}^k \text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_i) + \lambda^2 \sum_{i=1}^k \text{Var}(\hat{l}_i). \quad (4.2)$$

Details concerning  $\text{Var}(\hat{l}_i)$  are given in Appendix 4.1.

The contribution of site  $i$  to the second or bias term of (4.1) is

$$E(\hat{\gamma}_{\lambda,i}) - \gamma_i = \gamma_i + \lambda E(\hat{l}_i) - \gamma_i.$$

The sum of the squared biases is simply  $\lambda^2 \sum_{i=1}^k [E(\hat{l}_i)]^2$ . Hence the total mean squared error for  $\hat{\gamma}_{\lambda}$  is

$$\begin{aligned} \text{TOTMSE}(\hat{\gamma}_{\lambda}) &= \sum_{i=1}^k \text{Var}(\hat{\gamma}_{ml,i}) + 2\lambda \sum_{i=1}^k \text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_i) \\ &+ \lambda^2 \sum_{i=1}^k \text{Var}(\hat{l}_i) + \lambda^2 \sum_{i=1}^k [E(\hat{l}_i)]^2. \end{aligned} \quad (4.3)$$

The minimum of this quadratic occurs at

$$\lambda_{\min(MSE)} = \frac{-\sum_{i=1}^k \text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_i)}{\sum_{i=1}^k \text{Var}(\hat{l}_i) + \sum_{i=1}^k [E(\hat{l}_i)]^2}. \quad (4.4)$$

Differentiating  $\text{TOTMSE}(\hat{\gamma}_\lambda)$  twice and evaluating at  $\lambda_{\min(MSE)}$  confirms that the critical point is indeed a minimum. In general, sites with a large  $\hat{\gamma}_{ml,i}$  are smoothed down due to a negative  $\hat{l}_i$  while low  $\hat{\gamma}_{ml,i}$  are smoothed up with a positive  $\hat{l}_i$ . This results in an overall negative  $\sum_{i=1}^k \text{Cov}(\hat{\gamma}_{ml,i}, \hat{l}_i)$  and a positive solution for  $\hat{\lambda}_{\min(MSE)}$ .

The expression for TOTMSE as given by (4.3) and the location of the minimum given by (4.4) depend on the unknown  $\gamma$  vector. Hence they must be evaluated at the current estimates. Evaluation of the solution by using the current estimates gives a possible data-based choice for  $\lambda$ . Recommending this choice implies a complete emphasis on an estimation-based criterion. The MSE choice for  $\lambda$  is incorporated into a combined strategy to be outlined in Section 4.4.

### 4.3 Case Deletion Criteria

Deleting a case or subset of the data and using the reduced data set for prediction has wide application within the area of smoothing. The subset of the data is used within some criterion so that the choice of smoothing parameter is determined or validated by the data. In ridge regression, for example, one has the overall solution  $\hat{\beta} = (\mathbf{X}'\mathbf{X} + h\mathbf{I})^{-1} \mathbf{X}'\mathbf{y}$  where  $h$  is the smoothing parameter. Deleting the  $i$ 'th row of the design matrix  $\mathbf{X}$  and the  $i$ 'th element of the observation vector  $\mathbf{y}$ , one obtains the deleted estimate vector  $\hat{\beta}_{(i)}(h)$  and the case deleted prediction  $\hat{y}_{(i)}$  of the removed observation. Denoting the predicted observation from the full data set by  $\hat{y}_i(h)$ , the sum of squares  $\sum_i (\hat{y}_i - \hat{y}_{(i)}(h))^2$  can be minimized with respect to  $h$ . The solution is the cross-validatory choice for the smoothing parameter. Two procedures are considered for the choice of  $\lambda$  in this section.

### 4.3.1 Predictive Loss

The convex smoothing prescription for the multinomial cell probability estimate is given by  $\hat{\theta}_{i,i} = a(1/k) + (1 - a)\hat{\theta}_i$  and was first introduced in Section 1.4.1. The approximate equivalence of this smoothing prescription and the one proposed in this thesis is presented in Appendix 2.2.

Stone (1974) suggested a cross validatory choice for the smoothing parameter based on convex smoothing and quadratic loss. The data of  $m$  cases among  $k$  cells can be viewed as an  $m$  by  $k$  table in which the  $i$ 'th row has only one non zero element in the cell in which case  $i$  occurred. The rows indexed by  $i$  can be used sequentially generate case deletions from the vector  $\mathbf{y}$ . In a deleted vector  $\mathbf{y}_{(i)}$ , there are  $y_i - 1$  cases at cell  $i$  while the counts at the other sites remain the same. The deleted vector  $\mathbf{y}_{(i)}$  follows a multinomial distribution with  $m - 1$  cases and  $k - 1$  dimensional cell probability vector  $\boldsymbol{\theta}$ . If  $\boldsymbol{\delta}_{(i)}$  is the vector which is the  $i$ 'th row of this table then  $\delta_{ij}$  is the  $j$ 'th element of this vector. Following Stone, the cross validatory predictive loss function

$$\text{CVPL}(\lambda) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k (\delta_{ij} - \hat{\theta}_{sm,j,(i)})^2. \quad (4.5)$$

where  $\hat{\theta}_{sm,j,(i)} = \frac{N_{ij}}{N_{+}} \hat{\gamma}_{sm,j}$  is the numerically smoothed estimate with smoothing parameter  $\lambda$  for the  $i$ 'th case deletion.

There are  $y_i$  repetitions of each row in the table and re-estimation for the  $i$ 'th case deletion occurs  $y_i$  times. For table entries with  $j = i$ , a smoothed estimate  $\hat{\theta}_{sm,j,(i)}$  which is close to one gives a small contribution to the loss function. Averaged over all  $m$  cases, the function is an indicator of the smoothing method's ability to predict the deleted observation. Minimization over  $\lambda$  gives the cross-validatory choice for  $\lambda$  based on predictive loss.

The CVPL criterion was considered using the numerically-obtained smoothed cell estimate  $\hat{\theta}_{sm,j}$ . Empirical investigation involving complete homogeneity to moderate departure revealed that  $CVPL(\lambda)$  is not always a function which initially decreases and then increases. Frequently it is monotonically decreasing function over the entire range for  $\lambda$ . Minimization chooses  $\lambda$  at infinity. Using  $m - 1$  cases,  $k$  re-estimations are required to obtain the sum given by (4.5). For simulations in which only one elevated rate existed, the contribution to (4.5) from the  $k - 1$  non elevated sites decreased with  $\lambda$ . This decrease was not offset by an increase at the site with the elevated rate. Such results led to further investigation of the function  $CVPL(\lambda)$  which is given in Appendix 4.2.

In Appendix 4.3, the CVPL function is discussed in terms of the convex multinomial smoothing prescription. It is shown that some smoothing is always beneficial with respect to predictive loss. Under equal populations and homogeneity,  $E[CVPL]$  is shown to be uniformly decreasing with the smoothing parameter. The CVPL function has shown strictly decreasing behavior in fairly general situations (such as a homogeneity and a low departure from homogeneity). In these situations, adapting the CVPL criterion suggests taking the smoothing parameter at infinity. This CVPL criterion emphasizes the ability to predict a deleted or future observation. However testing for elevated rates is of primary interest in this thesis. The testing results of Chapter 3 indicate that although some smoothing is often advantageous, an infinitely large amount of smoothing is not necessary.

### 4.3.2 A testing-based choice for $\lambda$

The results of the  $\lambda$ -fixed simulations of Chapter 3 indicate that as  $\lambda$  increased, significant increases were observed in the ability to detect elevated rates. It was logical to establish a criterion for selecting the smoothing parameter based on the testing procedure developed.

Power is the probability of rejecting the null hypothesis when it is false. The ability to detect elevated rates or clusters is a related quantity based on correct decision(s) following the rejection of  $H_0$ . Power and detection ability are a function of the unknown parameter vector  $\gamma$ . If smoothing is applied, expressions for these functions can't be found because the critical region at site  $i$  involves the complicated determination of  $\hat{C}_{\alpha^*,i}$ , as described in Section 3.3.3.

The cross validation algorithm of Stone (1974) outlined in Section 4.3.1 can be adapted to the testing procedure. The existence of elevated rates is unknown and a correct or incorrect decision in this sense can't be recorded. However, the number of tests rejected out of  $k$  can be recorded for each case deletion. The distribution of the deleted case vector  $\mathbf{y}_{(i)}$  was discussed in Section 4.3.1. For each  $\lambda$ , a case from each of the  $i$  sites is deleted sequentially. For the deletion of a case at site  $i$ , the maximum likelihood vector  $\hat{\gamma}_{(i)}$  is estimated and the approximate smoothed estimates at each site are obtained. For this case deletion at site  $i$ , the test statistics  $Z_{\lambda,red,j}$ , and critical values  $\hat{C}_{\alpha^*,j}$  are determined for  $j = 1, \dots, k$ . The fraction rejected out of  $k$ ,  $fr_{\lambda,(i)}$ , is then calculated. The following criterion is calculated over the  $k$  required case deletions

$$CVPW(\lambda) = \frac{1}{m} \sum_{i=1}^k y_i fr_{\lambda,(i)} \quad . \quad (4.6)$$

Averaged over the entire data set, this quantity indicates the power to reject with an observation removed. The function  $CVPW(\lambda)$  is maximized numerically over a set of  $\lambda$  to give the choice of smoothing parameter.

A maximum of  $k$  case deletions are required at each  $\lambda$ . However, empirical investigation showed that  $k$  deletions were not necessary for computing  $CVPW(\lambda)$ . Only a large  $\hat{\gamma}_{\lambda,i}$  can lead to a rejection of  $H_{o,i}$ , and hence only case deletions corresponding to sites having a  $\hat{\gamma}_{ml,i}$  greater than 1 made any contribution to the overall sum in (4.6). This observation is helpful for reducing the amount of computing if a large number of  $\lambda$ 's are to be used in the numerical maximization.

The same method for establishing the upper bound for  $\lambda$  in the simulations of Chapter 3 can be used to set an estimated upper bound for the grid of  $\lambda$  values to be used in the maximization of  $CVPW(\lambda)$ . In Section 3.4, the upper bound was described as being that value of  $\lambda$  at which the approximate variance for  $\hat{\gamma}_{\lambda}$  equals 0. This solution depends on the unknown  $\gamma$  vector. The maximum likelihood estimate vector is the logical choice for substitution into this solution to obtain an estimated upper bound for  $\lambda$ .

$CVPW(\lambda)$  is step function with a range of  $[0,1]$  and increments as small as  $1/mk$ . A unique  $\lambda$  to maximize  $CVPW(\lambda)$  may not be found. A maximum level or plateau is more likely. The width of the interval of maximum level could depend on the number of  $\lambda$  in the grid search. Some possible examples of the form of the step function are given in the next section.

The  $CVPW(\lambda)$  criterion is appropriate in the inference setting which has been presented. Due to its step function form, this criterion can be combined with the minimum MSE choice for  $\lambda$  to establish a more comprehensive strategy for selection

of the smoothing parameter. This strategy is outlined in the next section.

## 4.4 A combined strategy for choosing $\lambda$

The power criterion CVPW emphasizes power in testing as a criterion for choosing the smoothing parameter. This emphasis can be maintained while still recognizing the sampling properties of the estimate vector  $\hat{\gamma}_\lambda$ . Even with a large number of  $\lambda$  in the grid search, the CVPW function is likely to give an interval choice for  $\lambda$ . The minimum MSE choice as outlined in Section 4.2 can be considered after this interval has been found. The following strategy is proposed.

- (1) Using the established  $\lambda_{LB}$ , select  $n + 1$  equally spaced values for  $\lambda$  between 0 and  $\lambda_{LB}$ .
- (2) For each value of  $\lambda$ , calculate  $CVPW(\lambda_j)$  for  $j = 1, \dots, n + 1$ , where  $\lambda_1 = 0$ .
- (3) Find  $\max\{CVPW(\lambda_j)\}$ .
- (4) Establish where  $\max\{CVPW(\lambda_j)\}$  occurs. In other words, search to verify if  $\max\{CVPW(\lambda_j)\}$  occurs at a single  $\lambda$ . If it does not, find the  $\lambda$  limits  $[\lambda_a, \lambda_b]$  of this maximum plateau.
- 5(a) If the maximum occurs at a single point  $\lambda_{max}$ , perform a new search from the previous  $\lambda$  to  $\lambda_{max}$ . If a plateau is now found, check to see if  $\hat{\lambda}_{min(MSE)}$  is contained within the  $\lambda$  range. If it is not, take the average  $\lambda$  of the plateau as the  $\lambda$  choice. If no new plateau is found, take  $\lambda_{max}$  as the choice for  $\lambda$ .

**5(b)** If the original search gives an interval  $[\lambda_a, \lambda_b]$  for the  $\max[\text{CVPW}(\lambda_j)]$  then perform a more refined search over the range. If no new plateau is found, check to see if the  $\min(\text{MSE})$  choice for  $\lambda$  is within the interval  $[\lambda_a, \lambda_b]$ . If it is, take the  $\min(\text{MSE})$  choice for  $\lambda$ . If it is not, take the midpoint  $(\lambda_a + \lambda_b)/2$  as the choice for  $\lambda$ .

**5(c)** If more than one plateau occurs, consider the  $\lambda$  range of the highest plateau only and perform the operations described in (b).

**5(d)** If the original search shows that the CVPW function is flat over the entire range for  $\lambda$  then take the  $\min(\text{MSE})$  choice for  $\lambda$ . This option is quite likely for situations which are a small departure or no departure from homogeneity. The recommendation of the  $\min(\text{MSE})$  choice for smoothing parameter suggests taking a  $\lambda$  that gives an estimate with better sampling properties, although the ability to detect elevated rates may not be improved. The simulations of fixed  $\lambda$  in Chapter 3 imply that there is no significant loss in the ability to detect elevated rates if some smoothing is performed.

The four examples in Figures 4.1 and 4.2 illustrate some of the possible outcomes for the CVPW function. The estimated upper bound,  $\hat{\lambda}_{UB}$  is approximately .02 for each. In plot 1 of Figure 4.1, the  $\lambda$  interval  $[\lambda_a, \lambda_b]$  is checked for  $\hat{\lambda}_{\min(\text{MSE})}$ . If it is not within this range, the midpoint .015 would be chosen for  $\lambda$ . In the second plot of Figure 4.1, the highest plateau is approximately on the interval  $[\lambda_a, \lambda_b]$ . If no new plateau is found on a more refined search of this interval, and if  $\hat{\lambda}_{\min(\text{MSE})}$  is not found within this interval, .014 is chosen for  $\lambda$ . In the first plot of Figure 4.2, the highest plateau, approximately on the interval  $[\lambda_a, \lambda_b]$ , is considered. If no new plateau is found and if  $\hat{\lambda}_{\min(\text{MSE})}$  is not found within this interval, .016 is chosen for  $\lambda$ . The CVPW function is flat over the entire  $\lambda$  range in the second plot of Figure 4.2. The proposed strategy suggests taking  $\hat{\lambda}_{\min(\text{MSE})}$  in this situation.



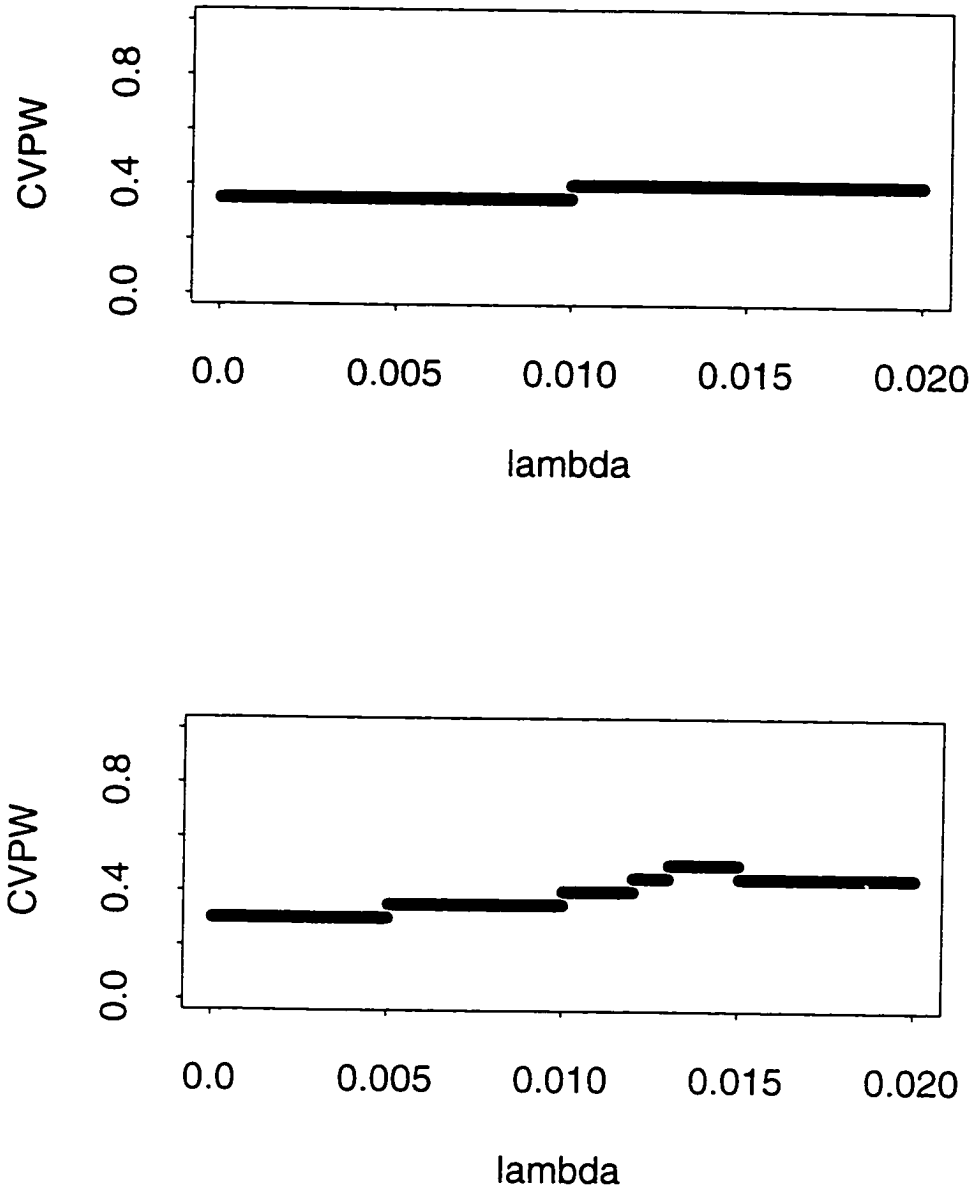


Figure 4.1 : Examples of CVPW function

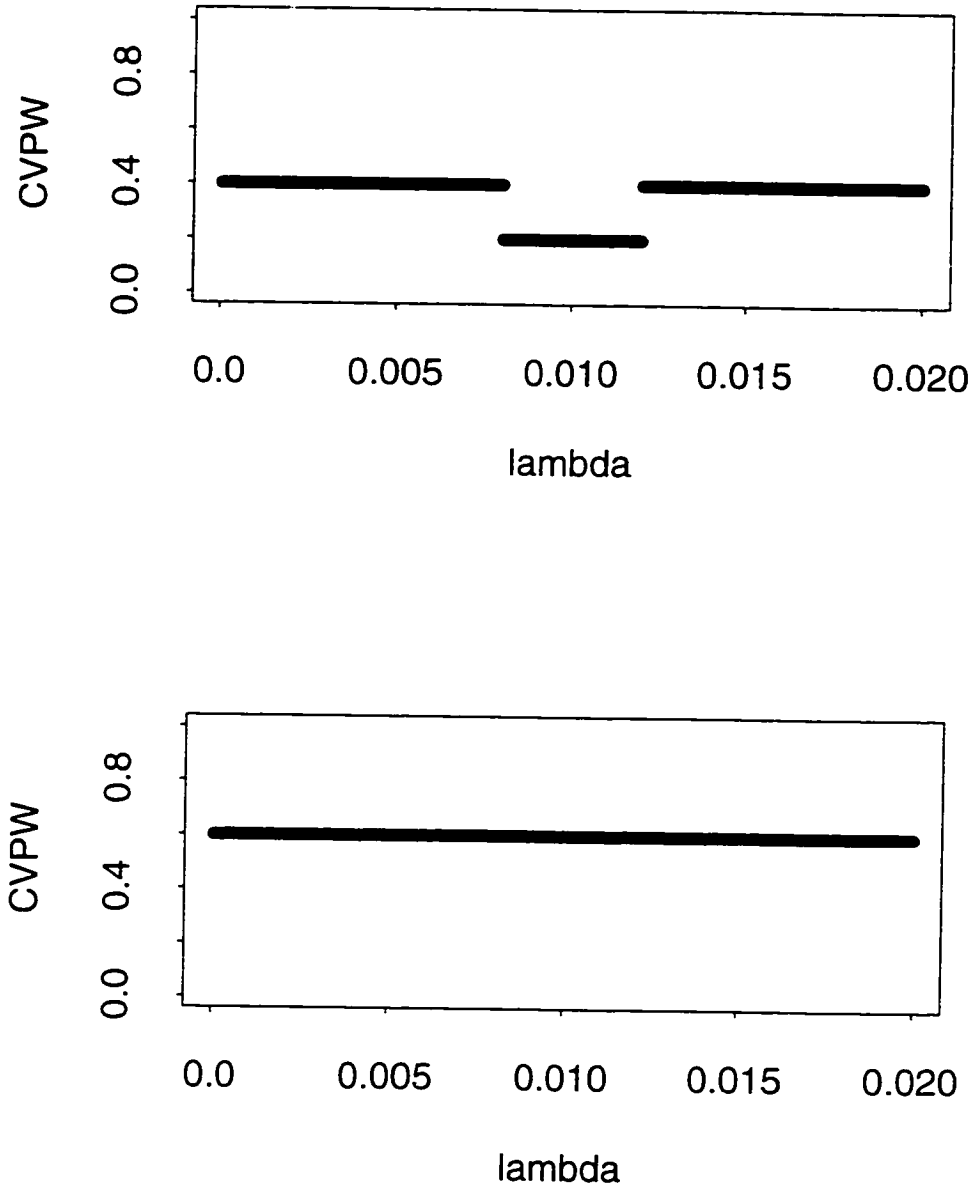


Figure 4.2 : Examples of CVPW function

## 4.5 Results - Combined Strategy

The strategy proposed in Section 4.4 is evaluated using simulation in this section. In practice one would have only one case vector at the known locations and no prior knowledge of the amount of smoothing required. In the simulations discussed in this section, five thousand case vectors were generated. For each case vector, the smoothing parameter  $\lambda$  was chosen according to the strategy proposed in the previous section. Approximate smoothed estimates for  $\gamma$  were then obtained using the chosen  $\lambda$ . A critical value was obtained for the test statistic  $Z_{\lambda, red, t}$  and testing was performed in the same manner as described in Chapter 3. For comparison, testing was also done with  $\lambda = 0$ . Preliminary simulations were performed to assess the computing time of these extensive simulations. Based on these simulations, it was decided that an original grid search of twenty-five  $\lambda$  values was sufficient for each repetition. In practice, with only one data vector, many more  $\lambda$  values could be used.

Firstly, homogeneity or complete null simulations were carried out with  $\gamma_1 = \dots = \gamma_k = 1.0$ . The upper probability limit for the fraction of tests rejected for  $k = 20, m = 200$ , equal populations of 1000 and one different population situation were given in Section 3.5. These limits at the site and overall level remain the same for the complete null simulations summarized in Table 4.1 given below. The limits for this Table are (a) .0037, (b) .0019, (c) .0544, (d) .0258, (e) .0024, (f) .0363. As mentioned in Chapter 3, for  $\lambda = 0$ , there are two possible upper bounds for the 'N different configuration', the higher was used. As can be seen in the table, the upper limit was not exceeded and the achieved levels are quite conservative.

**Table 4.1**Homogeneity Simulations with Data-based Choice of  $\lambda$ 

20 Sites, 200 Cases

	Circle + 1		Circle	
	Smoothed	Unsmoothed	Smoothed	Unsmoothed
	Site	Site	Site	Site
	Overall	Overall	Overall	Overall
Equal Popns	.0024 (a)	.0016 (b)	.0020 (a)	.0018 (b)
	.0432 (c)	.0202 (d)	.0476 (c)	.0234 (d)
1 Popn Different	.0028 (a)	.0020 (e)	.0030 (a)	.0020 (e)
	.0472 (c)	.0334 (f)	.0402 (c)	.0316 (f)

Site = max. fraction of tests rejected at any site  
 Overall = fraction of repetitions with at least one rejection  
 Probability limits (a) through (f) given in text

A complete null simulation was also carried out for the 'Random/location Population' configuration. The site and overall levels were again conservative for the tests with and without smoothing.

Simulations with elevated incidence at one site were performed for these three configurations and are presented in Tables 4.2 and 4.3. The fraction of tests rejected and the fraction totally correct are given. The fraction totally correct is the fraction of tests in which the only site declared elevated is the elevated site. In Table 4.2, the elevated incidence rate in each simulation is one and a half and two times the

true incidence rate of the other nineteen. The elevated site in the other simulation of Table 4.2 is once again the distant site in the 'circle + 1' configuration, and it has population 500 in the two simulations in which the populations are not equal. The elevated rate ratio  $\gamma_{el}$  is indicated in the left column. The simulations with a single elevated site for the 'Random Population/Location' configuration are given in Table 4.3. The row corresponding to  $\gamma_{el} = 2.79$  are results for a simulation in which the elevated site has incidence rate three times that of the common nineteen.

**Table 4.2**Ability to Detect Single Elevated Rate (Data-based choice of  $\lambda$ )

20 sites and 200 cases

	Circle + 1		Circle	
	Smoothed	Unsmoothed	Smoothed	Unsmoothed
Populations and $\gamma_{el}$	f.rej. f.corr.	f.rej. f.corr.	f.rej. f.corr.	f.rej. f.corr.
Equal Popns $\gamma_{el} = 1.46$	.1222* .0888*	.0700 .0534	.1320* .0978*	.0642 .0510
Equal Popns $\gamma_{el} = 1.90$	.4500* .4428*	.3608 .3522	.4640* .4382*	.3520 .3408
1 Different Popn $\gamma_{el} = 1.48$	.0782 .0474	.0742 .0454	.0730 .0436	.0726 .0430
1 Different Popn $\gamma_{el} = 1.95$	.2218 .2024	.2182 .1968	.2252 .1948	.2162 .1918

f rej. = fraction of repetitions with at least one test rejected

f corr. = fraction of repetitions in which only elevated site is declared elevated

\* indicates significant difference from observed fraction without smoothing

**Table 4.3**Ability to Detect Single Elevated Rate (Data-based choice of  $\lambda$ )

Random Populations/Locations. 20 sites and 200 cases

Single Elevated Rate		
k = 20. m = 200		
	Random Populations/Locations	
	Smoothed	Unsmoothed
$\lambda_{el}$	f.rej.	f.rej.
	f.corr.	f.corr.
1.47	.0980*	.0496
	.0826*	.0414
1.93	.3216*	.2386
	.2958*	.2200
2.79	.8770*	.8066
	.8602*	.7934

f rej. = fraction of repetitions with at least one test rejected

f corr. = fraction of repetitions in which only elevated site is declared elevated

\* indicates significant difference from observed fraction without smoothing

The results presented in Table 4.2 are similar to those presented in Chapter 3 for a single elevated rate. In Table 4.2, the observed detection ability (the fraction with at least one rejection or fraction correctly declared elevated) with smoothing performed is similar to the the maximum observed fraction with fixed positive  $\lambda$  in the simulations of Table 3.3 and 3.4. An asterisk indicates that the observed fraction with smoothing is a significant increase from the observed fraction with  $\lambda = 0$ . The

fraction with at least one rejection or totally correct from the tests with smoothing is significantly greater than the observed fraction with no smoothing in the simulations with equal populations. There is no significant difference in power when the elevated site has the lower population (as in the simulations of fixed  $\lambda$  in Chapter 3). The elevated site for the 'Random Populations/Locations' configuration of Table 4.3 is the same as for the simulation presented in Table 3.7. In the first two simulations ( $\gamma_{el} = 1.47$ ,  $\gamma_{el} = 1.93$ ), the fraction of tests rejected and fraction totally correct with smoothing applied are similar to the highest achieved previously with fixed  $\lambda$ . Although the fraction totally correct without smoothing is near eighty percent in the third simulation, the observed fraction with smoothing (0.8770) is also a significant increase.

Simulations with the 'circle' and 'circle + 1' configurations which use the data based choice for  $\lambda$  reveal results which are very similar to those obtained in Chapter 3. As a result, only the 'circle + 1' configuration and the 'Random Population/Location' configurations are used in simulations with two elevated contiguous rates. As in Chapter 3, true  $\gamma$  vectors which represent a small departure from homogeneity were chosen. The results for two contiguous elevated rates with  $k = 20$ ,  $m = 200$  for the 'circle + 1' configuration for simulations with fixed  $\lambda$  appeared in Table 3.7. For simulations in which the elevated (distant) rate had the differing population (previously labelled simulations D through F), no significant difference exists in the observed detection ability between tests with and without smoothing (both fraction with at least one rejection or totally correct). The results for equal populations are given in Table 4.4 along with simulations for the 'Random Population/Location' configuration. The elevated rate ratios and departure measure  $\delta$  for these simulations are as follows

Circle + 1 (Equal Populations)

$$A: \delta = 4.08, \gamma_{el1} = \gamma_{el2} = 1.43$$



B:  $\delta = 9.84$ ,  $\gamma_{el1} = 1.86$ ,  $\gamma_{el2} = 1.40$

C:  $\delta = 9.84$ ,  $\gamma_{el1} = 1.40$ ,  $\gamma_{el2} = 1.86$

#### Random Populations/Locations

A:  $\delta = 3.20$ ,  $\gamma_{el1} = \gamma_{el2} = 1.45$

B:  $\delta = 7.64$ ,  $\gamma_{el1} = 1.89$ ,  $\gamma_{el2} = 1.42$

C:  $\delta = 7.90$ ,  $\gamma_{el1} = 1.42$ ,  $\gamma_{el2} = 1.89$ .

In each of the simulations representing a cluster of size two summarized in Table 4.4, the observed fraction of tests with at least one rejection with smoothing applied is significantly different from the unsmoothed fraction. The corresponding observed fractions totally correct were not always significantly different.

**Table 4.4**

Ability to Detect Two Contiguous Elevated Rates (Data-based choice of  $\lambda$ )

20 sites and 200 cases

	Circle + 1		Random Popns/Locus	
	Smoothed	Unsmoothed	Smoothed	Unsmoothed
	f.rej.	f.rej.	f.rej.	f.rej.
	f.corr.	f.corr.	f.corr.	f.corr.
A	.1582 *	.1064	.1288 *	.1012
	.0030	.0014	.0028	.0010
B	.4330 *	.3502	.3330 *	.2608
	.0132	.0096	.0086 *	.0064
C	.4400 *	.3508	.3628 *	.3292
	.0256 *	.0102	.0114 *	.0056

f.rej. = fraction of repetitions with at least one test rejected

f.corr. = fraction of repetitions in which only elevated sites are declared elevated

\* indicates significant difference from observed fraction without smoothing

# Chapter 5

## Application

For an ongoing analysis of gastric cancer rates in Nova Scotia, assessing the significance of elevated rates or clusters was required. This goal is achieved by applying the method developed in this thesis. The resulting smoothed estimates of incidence possess better sampling properties than estimates obtained if smoothing is not considered. Sixty-six municipalities are used by the Nova Scotia Cancer Registry. Eight municipalities were combined because of the closeness of their co-ordinates. The latitudes and longitudes as well as the contiguity of the  $k = 58$  sites or regions, are given in Appendix 5.1. The diagnosed cases of gastric cancer for the population aged 65 and over were aggregated for the five year period ending December 31, 1981. The risk population was determined from 1981 Statistics Canada census data. The total risk population is  $N_+ = 92385$  and the total number of cases is  $m = 214$ .

### 5.1 Estimation and Testing with fixed $\lambda$

Case counts, risk populations and estimates for the 58 sites are presented in Table 5.1. At  $\lambda = 0$ , the maximum likelihood estimates  $\hat{\gamma}_{ml,i}$  of the rate ratio  $\gamma_i$  range from 0.0 (occurring at 13 sites) to  $\hat{\gamma}_{ml,20} = 15.69$  (Guysborough Municipality). Nearly half of the maximum likelihood estimates are greater than 1.0, and over half of these sites

have a risk population less than 1000. Metropolitan Halifax (site 55) has close to 20 percent of the risk population and a maximum likelihood estimate of  $1.3373 = (55/17755)/(214/92385)$ .

The smoothed estimate vector  $\hat{\gamma}_{sm}$  is presented for four smoothing parameters. The range for these  $\lambda$ 's was chosen in the same manner as in the fixed  $\lambda$  simulations of Chapter 3. The true  $\gamma$  is not known so an upper bound  $\lambda_{U.B.i}$  for each site is found by using the value at which the approximate variance (evaluated under  $H_0$ ) is approximately zero. The presence of relatively small  $N_i$ 's led to an overall  $\lambda_{U.B}$  which is small for other sites with larger population. The range of  $\lambda$  used is smaller than the range of the simulations presented in this thesis because the average  $d_{ij}$  is now much less (defined by latitude and longitude rather than on a 10 by 10 grid).

As expected, the effect of smoothing is most pronounced on the extreme estimate at site 20 (Guysborough municipality). The estimate for site 20 is smoothed down from 15.5983 to 11.7909 for the set of  $\lambda$ 's used. Smoothed estimates corresponding to sites with zero cases are actually less than that listed (e.g. an estimate of  $5 \times 10^{-9}$  is given as  $10^{-8}$ ). It is interesting to note that some smoothed estimates which are greater than 1.0 are larger than their unsmoothed counterparts. For example, site 5 (Antigonish) is contiguous and very close to site 4 (rural Antigonish municipality). The maximum likelihood estimate for rural Antigonish (site 4) is 3.6482 and the estimate without smoothing for the town of Antigonish (site 5) is 1.1589. The smoothed estimate at site 5 increases for each successive  $\lambda$  used. Such a result was not seen in the simulations presented in earlier chapters. The increase with smoothing of an estimate greater than 1.0 can be attributed to the higher estimates observed in this data set and the smaller inter-site distances. Despite the large estimate at  $\lambda = 0$ , the estimates at site 4 do not quickly decrease with smoothing. They decrease from 3.6482 to 3.6388 (from  $\lambda = 0$  to  $\lambda = 5.305 \times 10^{-5}$ ) while the estimate for site 11 decreases from 3.1978 to 3.1230 with the same amount of smoothing applied. The estimate at site 4 may not be smoothed down because of its proximity to the very high estimate at site 20.

**Table 5.1**  
 Nova Scotia Gastric Cancer Data  
 Cases, Populations and Estimates using fixed  $\lambda$

Site	$y$	$N$	$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$			
$\lambda \times 10^{-5}$			0.0	.884	1.768	3.537	5.305
1	1	1375	0.3139	0.3146	0.3152	0.3163	0.3172
2	1	2625	0.1644	0.1648	0.1652	0.1657	0.1662
3	1	175	2.4669	2.4608	2.4539	2.4391	2.4233
4	9	1065	3.6482	3.6478	3.6441	3.6414	3.6388
5	2	745	1.1589	1.1617	1.1641	1.1681	1.1716
6	2	1215	0.7106	0.7119	0.7131	0.7149	0.7165
7	1	755	0.5717	0.5728	0.5738	0.5753	0.5765
8	1	395	1.0929	1.0953	1.0973	1.1007	1.1035
9	3	245	5.2862	5.2291	5.1733	5.0672	4.9686
10	3	850	1.5237	1.5254	1.5268	1.5286	1.5297
11	1	135	3.1978	3.1878	3.1763	3.1504	3.1230
12	12	3795	1.3651	1.3674	1.3694	1.3727	1.3754
13	1	1315	0.3283	0.3290	0.3297	0.3308	0.3318
14	2	1380	0.6256	0.6269	0.6279	0.6296	0.6311
15	0	135	0.0	$10^{-8}$	$10^{-8}$	$10^{-7}$	$10^{-7}$
16	2	3110	0.2776	0.2782	0.2786	0.2794	0.2800
17	5	2525	0.8548	0.8563	0.8577	0.8598	0.8616
18	2	1270	0.6799	0.6811	0.6823	0.6842	0.6858
19	0	465	0.0	$10^{-9}$	$10^{-8}$	$10^{-8}$	$10^{-8}$
20	4	110	15.6983	14.6697	13.8683	12.6690	11.7909
21	12	5690	0.9104	0.9121	0.9134	0.9157	0.9175
22	2	1340	0.6443	0.6457	0.6468	0.6487	0.6504
23	0	1285	0.0	$10^{-9}$	$10^{-9}$	$10^{-9}$	$10^{-9}$
24	0	160	0.0	$10^{-8}$	$10^{-7}$	$10^{-7}$	$10^{-7}$
25	7	2370	1.2751	1.277	1.2789	1.2816	1.2837
26	4	785	2.1998	2.1943	2.1902	2.1884	2.1853
27	5	3785	0.5703	0.5714	0.5723	0.5739	0.5751
28	0	515	0.0	$10^{-9}$	$10^{-8}$	$10^{-8}$	$10^{-8}$
29	0	185	0.0	$10^{-8}$	$10^{-8}$	$10^{-8}$	$10^{-8}$
30	1	170	2.5394	2.5375	2.5347	2.5275	2.5190

Smoothed estimates for sites with zero cases are listed to next decimal place

**Table 5.1**  
 Nova Scotia Gastric Cancer Data  
 Cases, Populations and Estimates using fixed  $\lambda$

Site	$y$	$N$	$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$			
$\lambda \times 10^{-5}$			0.0	.884	1.768	3.537	5.305
31	2	3225	0.2677	0.2683	0.2687	0.2695	0.2702
32	1	670	0.6443	0.6458	0.6472	0.6495	0.6515
33	1	260	1.6604	1.6593	1.6576	1.6534	1.6486
34	3	335	3.8660	3.8481	3.8295	3.7915	3.7538
35	0	1060	0.0	$10^{-9}$	$10^{-9}$	$10^{-9}$	$10^{-9}$
36	0	290	0.0	$10^{-8}$	$10^{-8}$	$10^{-8}$	$10^{-8}$
37	2	385	2.2426	2.2409	2.2384	2.2325	2.2256
38	6	2430	1.0659	1.0678	1.0694	1.0721	1.0742
39	1	645	0.6693	0.6709	0.6722	0.6745	0.6765
40	0	205	0.0	$10^{-8}$	$10^{-8}$	$10^{-7}$	$10^{-7}$
41	3	1410	0.9185	0.9201	0.9215	0.9237	0.9256
42	5	1490	1.4487	1.4513	1.4536	1.4573	1.4603
43	0	655	0.0	$10^{-9}$	$10^{-8}$	$10^{-8}$	$10^{-8}$
44	0	285	0.0	$10^{-8}$	$10^{-8}$	$10^{-7}$	$10^{-7}$
45	5	830	2.6006	2.5933	2.5908	2.5888	2.5838
46	1	160	2.6982	2.6850	2.6711	2.6431	2.6155
47	0	490	0.0	$10^{-9}$	$10^{-8}$	$10^{-8}$	$10^{-8}$
48	14	4080	1.4813	1.4839	1.4859	1.4895	1.4923
49	7	1995	1.5148	1.5172	1.5193	1.5227	1.5253
50	2	1195	0.7225	0.7240	0.7253	0.7275	0.7279
51	3	755	1.7154	1.7166	1.7172	1.7177	1.7170
52	0	550	0.0	$10^{-9}$	$10^{-8}$	$10^{-8}$	$10^{-8}$
53	3	1310	0.9886	0.9904	0.9918	0.9943	0.9963
54	4	1185	1.4572	1.4594	1.4611	1.4638	1.4630
55	55	17755	1.3373	1.3395	1.3415	1.3447	1.3473
56	6	3900	0.6642	0.6653	0.6663	0.6681	0.6695
57	3	3085	0.4198	0.4206	0.4213	0.4225	0.4235
58	3	1780	0.7276	0.7290	0.7303	0.7323	0.7341

Smoothed estimates for sites with zero cases are listed to next decimal place

Test statistics at  $\lambda = 0$  ( $Z_{ml,i}$ ) and  $\lambda > 0$  ( $Z_{\lambda,red,i}$ ) were calculated for each of the 58 sites. Exact critical values for  $\lambda = 0$  were obtained as described in Section 3.1. Approximate critical values  $\hat{C}_{\alpha^*,i}$  (at level  $\alpha^* = .05/57$ ) were obtained through the method described in Section 3.3. Site 20 (Guysborough municipality) was declared elevated for all five  $\lambda$ 's (including  $\lambda = 0$ ). Site 4 (rural Antigonish municipality) was declared elevated for the last 3  $\lambda$ 's in the set. These two sites share a common border. Hence smoothing results in the declaration of a cluster of size two while only site 20 is declared elevated without smoothing. The reduced estimates  $\hat{\gamma}_{red,i}$  and test statistics  $Z_{\lambda,red,i}$  for these two sites are given in Table 5.2. A map of Nova Scotia which displays the 58 municipalities is given in Figure 5.1. The municipalities of Guysborough (site 20) and Antigonish (site 4) are indicated by an "X" and "x" respectively.

The analysis was reperformed without metropolitan Halifax ( $y_{55} = 55$ , and  $N_{55} = 17755$  were removed). A new range of  $\lambda$ 's was used and a new set of estimates, test statistics and critical values were obtained. Sites 4 and 20 are now both declared elevated at  $\lambda \geq 0$  ( $\hat{\gamma}_{ml,4} = 3.9665$  with the smaller  $N_+$ ).

**Table 5.2**

Approximate Estimates and Test Statistics for Sites Declared Elevated

	Site 4		Site 20	
$\lambda \times 10^{-5}$	$\hat{\gamma}_{red,4}$	$Z_{\lambda,red,4}$	$\hat{\gamma}_{red,20}$	$Z_{\lambda,red,20}$
0.0	3.6482	4.1836	15.6983	7.4239 †
0.884	3.6447	4.1778	14.4843	6.9963 †
1.768	3.6412	4.1742 †	13.2703	6.5687 †
3.537	3.6341	4.1647 †	10.8423	5.7135 †
5.305	3.6272	4.1552 †	8.4143	4.8584 †

† = declared elevated

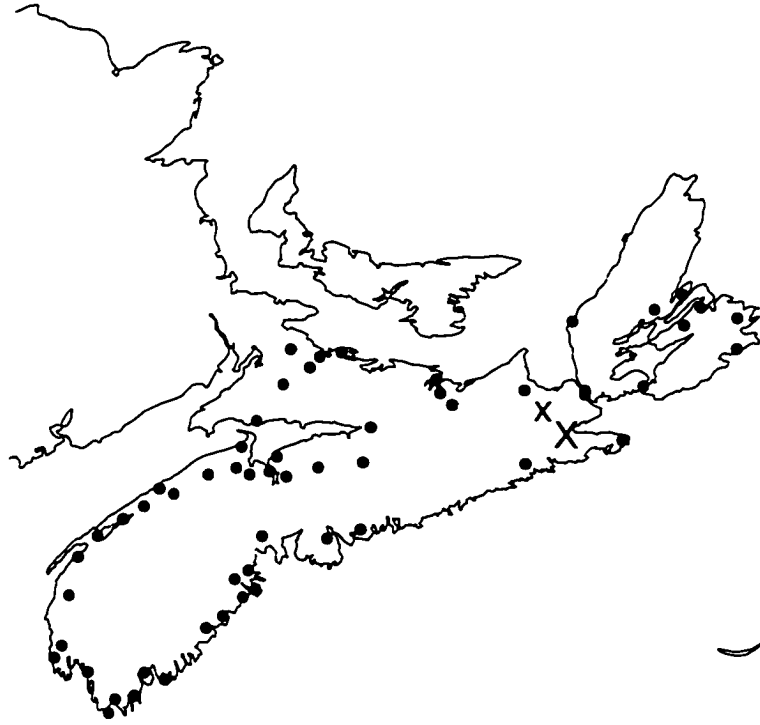


Figure 5.1: Nova Scotia Map and Municipalities  
( Elevated incidence indicated by an x )

## 5.2 Analysis using data-based selection of $\lambda$

Estimation and testing were carried out using the data-based choice of  $\lambda$  (described in Section 4.4). The upper bound for  $\lambda$  (required for the strategy's grid search) was the same as in the previous section (i.e. the minimum  $\lambda$  over all sites at which  $\hat{\text{Var}}_{H_0}(\hat{\gamma}_{red,i})$  is approximately zero). The  $\hat{\lambda}$  chosen was  $2.2193 \times 10^{-5}$ . The smoothed estimates, approximate estimates and test statistics at  $\hat{\lambda}$  are between those for  $\lambda = 1.768 \times 10^{-5}$  and  $\lambda = 3.537 \times 10^{-5}$  given in the previous section.

The top panel of Figure 5.2 presents the difference in the smoothed and maximum likelihood estimates versus the maximum likelihood estimate. The extremely large estimate and large negative difference for Guysborough municipality is not shown. Only 12/58 estimates are smoothed down (i.e.  $\hat{\gamma}_{sm,i} - \hat{\gamma}_{ml,i} < 0$ ) but as expected these sites have large maximum likelihood estimates. The increase in the other estimates with smoothing is much smaller and includes the 13 sites for which zero cases were observed. Twelve sites which have maximum likelihood estimates between 1.0 and 1.7 have their estimates increased by the smoothing. The second panel of Figure 5.2 presents the difference in estimation versus the risk population  $N_i$ . Metropolitan Halifax is not shown because its population is so large relative to the others. As was seen in previous examples, the effect of smoothing is larger for the sites with smaller populations. The large decrease with smoothing in the largest estimate (Guysborough municipality), which has a small population is evident.

Site 4 (Antigonish municipality) and site 20 (Guysborough municipality) are declared elevated at  $\hat{\lambda}$  ( $\hat{\lambda} = 2.2193 \times 10^{-5}$  is greater than first fixed  $\lambda$  at which site 4 is declared elevated). Hence the utility of smoothing and the testing method are demonstrated because only site 20 is declared elevated based on the maximum likelihood estimates. Estimation and testing were repeated without metropolitan Halifax. Sites 4 and 20 were declared elevated with and without smoothing applied.



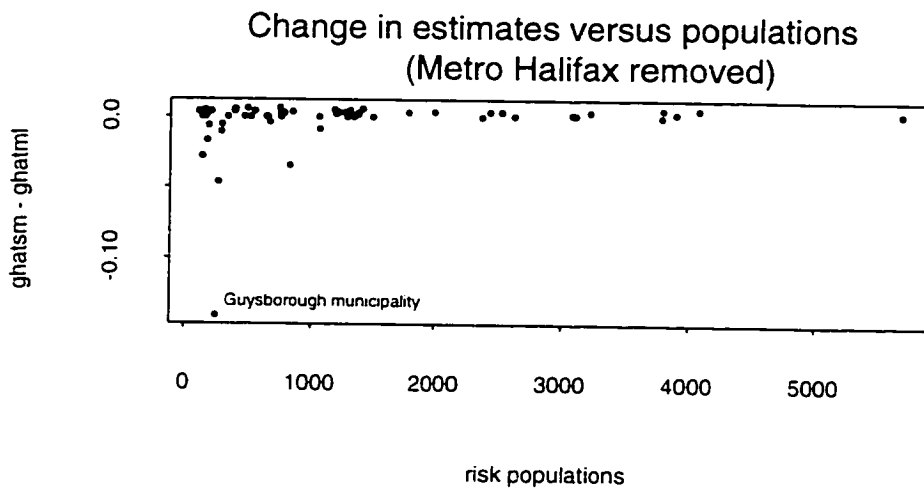
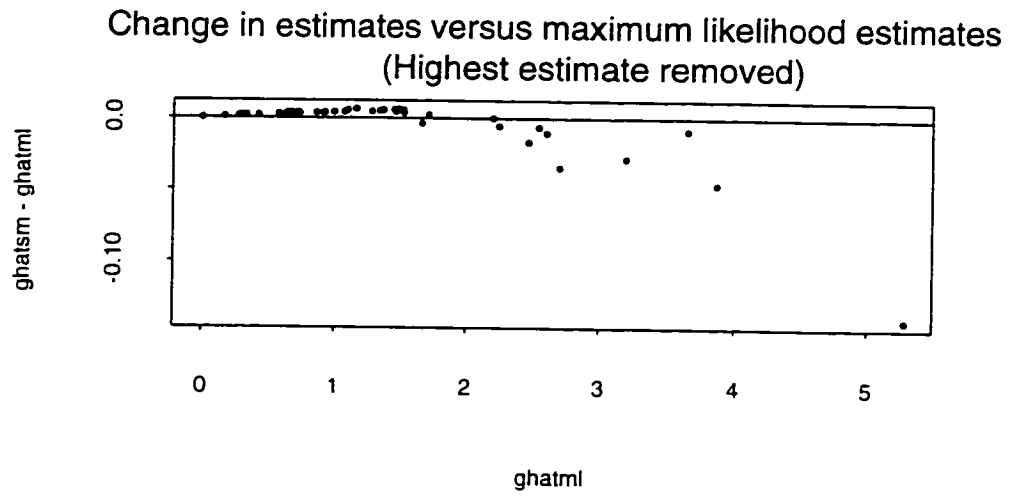


Figure 5.2 : Change in estimates (with smoothing) versus maximum likelihood estimates and populations

Classical Bonferroni corrections are known to be conservative and so the sharper correction of Hochberg (1988) was applied. Rather than use  $\alpha_{*,i} = .05/(k-1)$  for all  $(k-1)$  tests as above and as in the simulations of Chapter 3, the observed p-values are ordered. The  $Z_{m,i}$  giving the smallest p-value is tested at level  $\alpha/(k-1)$ , the  $Z_{m,i}$  giving second smallest p-value is tested at level  $\alpha/(k-2)$  and so on. The exact critical value for this new level at  $\lambda = 0$  was used as the starting point in the algorithm to find  $\hat{C}_{\alpha*,i}$ , the critical value for the test with smoothing. Even with the more liberal testing level (at sites other than 20), site 20 was still the only one declared elevated without smoothing and sites 4 and 20 were once again declared elevated at the chosen  $\hat{\lambda}$ .

## **NOTE TO USERS**

**Page(s) not included in the original manuscript are unavailable from the author or university. The manuscript was microfilmed as received.**

**UMI**

# Chapter 6

## Discussion

In this chapter, the methodology and main results are summarized, followed by a discussion of future work.

### 6.1 Summary

Smoothed estimation of incidence rates by penalized likelihood has been proposed. The parameter  $\gamma_i$ , the ratio of the incidence rate  $p_i$  to the weighted average  $\bar{p} = \sum_i N_i p_i / N_+$  was introduced because relative incidence is of interest. Conditioning on the total number of cases led to smoothed estimates based on a penalized multinomial likelihood. The resulting estimates  $\hat{\gamma}_{sm}$  are a compromise between the overall rate ratio for the region and the maximum likelihood estimates for individual sites. An approximation was developed for the smoothed estimate and approximate moments of the smoothed estimate were obtained. An approximation to the total mean squared error of  $\hat{\gamma}_{sm}$  performed well under simulation.

Detecting sites with elevated incidence is of primary interest and a test statistic based on smoothed estimates was developed to meet this objective. A method for determining the critical region of the test statistic was determined. Further refinement

of the test statistic was required to accomplish this goal. The testing procedure was evaluated through simulation. Under complete homogeneity ( $\gamma_i = 1$  for all  $i$ , no elevated rates), the level of the tests was controlled at both the site and overall levels. The ability to detect elevated rates increased significantly with smoothing in many situations while no significant decreases were found. More smoothing is required to observe benefits with respect to inference if the site of an elevated rate is distant from the others. Significant gains in detection ability were observed with smoothing for two contiguous elevated rates, particularly if these rates were very similar.

The choice of smoothing parameter  $\lambda$  was addressed by considering a minimum MSE choice and case deletion ideas. A criterion based on the testing procedure developed was combined with the MSE choice to establish a strategy for choosing  $\lambda$ . This strategy was evaluated by simulation and the observed detection ability was similar to the maximum achieved under fixed  $\lambda$  for the same configuration. Smoothing using the chosen  $\hat{\lambda}$  significantly increased the detection ability in some situations while significant decreases were not observed.

The overall method was applied to gastric cancer incidence data in Nova Scotia. For the time period considered (1977-81), two municipalities were declared elevated with smoothing applied while only one of these sites was declared elevated without smoothing.

## 6.2 Further Work

The methods proposed in this thesis have accomplished several objectives. The penalized likelihood incorporates the spatial aspect of the data and the resulting smoothed

estimates have better sampling properties than their unsmoothed counterparts. The test statistic developed demonstrated an improved ability to detect elevated rates. A data-based method for choosing the smoothing parameter has given satisfactory results.

Further investigation is possible within each of these topics. The convex smoothing prescription  $\hat{\gamma}_{con,i} = \hat{\gamma}_{ml,i}(1 - a) + a(1)$  will be pursued further. The approximate equivalence of  $\hat{\gamma}_{con,i}$  and  $\hat{\gamma}_{sm,i}$  was shown in Appendix 2.2 but the moments of  $\hat{\gamma}_{con,i}$  and a possible test statistic based on this estimate have not been considered. A Cornish-Fisher expansion produced critical values for  $Z_{\lambda,red,i}$  which were very conservative but this method may be revisited if other test statistics are developed.

More work on the choice of smoothing parameter is planned. If prediction of a future observation becomes more relevant, the cross-validators predictive loss criterion may be pursued. This thesis has not focussed on Bayesian ideas but as mentioned in Section 2.1,  $\exp\{-\lambda p_{en}(\gamma)\}$  can be interpreted as a prior for  $\gamma$  and  $\lambda$  is a hyperparameter. Techniques for the choice of  $\lambda$  in this framework have not been explored. These methods include improper hyperpriors for  $\lambda$  and estimates obtained through Empirical Bayes methods.

Allowing for case counts to occur over several time periods at each location is a natural extension of the proposed method. For  $t = 1, \dots, T$  time periods and  $i = 1, \dots, k$  sites, the case count  $y_{ti}$  is distributed as Poisson but with mean  $N_{ti}p_{ti}$ . Conditioning on the total number of cases,  $m_t$ , in each time period (conditioning on the grand total is another approach),  $\mathbf{y}_t|m_t$  is distributed as multinomial with index  $m_t$  and cell probability vector  $\theta_t$  where  $\theta_{ti} = N_{ti}p_{ti}/\sum_j N_{tj}p_{tj}$ . A possible penalty for the penalized likelihood is  $\bar{d}t \sum_t \sum_s \sum_i \sum_j 1/(d_{ij}t_{rs})(\gamma_{ir} - \gamma_{js})^2$  (multiplication by  $\bar{d}$  eliminates problems with differing units). The contribution to the penalty is large for large absolute differences in rate ratios which are temporally and spatially close together. Another possibility is having a separate penalty for time and space but this

also requires two smoothing parameters.

Administrative investigation may reveal that high estimates of incidence are attributable to external factors. If data on potential risk factors such as socio-economic status or level of medical care are available, these covariates could be included (a-priori) into the methodology within a generalized linear model framework. For  $y_i$  distributed as Poisson with mean  $N_i p_i$ , then we can model  $\log(N_i p_i)$  as the linear function  $\sum_j x_{ji} \beta_{ji}$  where  $x_{ji}$  represents the  $j$ th covariate at site  $i$  and  $\beta_{ji}$  is a parameter to be estimated for this effect. Conditioning on the total cases as before, the cell probability  $\theta_i$  or the rate ratio  $\gamma_i$  can be expressed as function of  $\sum_j x_{ji} \beta_{ji}$ . The issue of overdispersion will be considered in this new framework.

There are a vast variety of spatial configurations and levels of the factors  $N$ ,  $m$ ,  $k$  and  $\gamma$  that have not yet been investigated. The method developed can be applied to data consisting of counts other than human health outcomes. Assuming traffic volume data is collected to establish an associated risk population, smoothed estimates of accident incidence can be presented. Other applications are possible within fields such as forestry and biology if appropriate risk populations can be determined.

# Appendix A

## Chapter 2 Appendices

2.1	Multinomial Likelihood Estimation . . . . .	128
2.2	The approximate relationship between estimators obtained through convex smoothing and the proposed method . . . . .	130
2.3	Additional Examples of Smoothed Estimates . . . . .	132
2.4	Higher moments for the unsmoothed estimate . . . . .	136
2.5	Mixed Moments . . . . .	138
2.6	Additional Results: Observed and Approximate MSE . . . . .	141
2.7	Additional Results: Estimators $\hat{\gamma}_{sm}$ , $\hat{\gamma}_{\Lambda}$ , $\hat{\gamma}_{\Lambda,lin}$ , and $\hat{\gamma}_{\Lambda,red}$ . . . . .	143
2.8	Normality of $\hat{\gamma}_{sm}$ . . . . .	147



## Appendix 2.1 Multinomial Likelihood Estimation

In this appendix, maximum likelihood estimates for the multinomial likelihood  $l(\gamma)$  are obtained.

To an additive constant, the multinomial log likelihood was given by (2.8) as

$$l(\gamma) = \sum_{i=1}^k y_i \log(\gamma_i).$$

The constraint  $\sum_{i=1}^k N_i \gamma_i / N_+ - 1 = 0$  is included as a Lagrangian and the constrained multinomial likelihood becomes

$$l(\gamma) = \sum_{i=1}^k y_i \log(\gamma_i) - o \left( \sum_{i=1}^k \left( \frac{N_i \gamma_i}{N_+} \right) - 1 \right).$$

Differentiating with respect to  $o$  and equating to 0 gives

$$\frac{\partial l(\gamma)}{\partial o} = - \sum_{i=1}^k \left( \frac{N_i \gamma_i}{N_+} \right) + 1 = 0.$$

Solving gives  $\sum_{i=1}^k N_i \gamma_i = N_+$ .

Differentiating with respect to  $\gamma_i$  and equating to 0 gives

$$\frac{\partial l(\gamma)}{\partial \gamma_i} = \frac{y_i}{\gamma_i} - o \left( \frac{N_i \gamma_i}{N_+} \right) = 0.$$

Multiplying by  $\gamma_i$  and then summing over  $i = 1, \dots, k$  gives

$$m - o \sum_{i=1}^k \frac{.N_i \hat{\gamma}_i}{.N_+} = 0.$$

which implies  $\hat{o} = m$ .

Substituting into  $\partial l(\gamma)/\gamma_i$  gives

$$\frac{\partial l(\gamma)}{\partial \gamma_i} = \frac{y_i}{\gamma_i} - \frac{m.N_i}{.N_+}.$$

and solving gives the maximum likelihood estimate

$$\hat{\gamma}_{ml} = \left( \frac{y_j}{.N_j} \right) / \left( \frac{m}{.N_+} \right).$$

## Appendix 2.2

The approximate relationship between estimators obtained through convex smoothing and the proposed method

The penalized log likelihood  $l_{pen}(\boldsymbol{\gamma})$  was given by (2.10) as

$$l_{pen}(\boldsymbol{\gamma}) = \sum_{i=1}^k y_i \log(\gamma_i) - o\left(\sum_{i=1}^k \left(\frac{N_i \gamma_i}{N_+}\right) - 1\right) - \lambda \sum_{q \neq i} \frac{1}{d_{iq}} (\gamma_i - \gamma_q)^2 \quad .$$

Maximizing this function numerically gives the smoothed estimate  $\hat{\boldsymbol{\gamma}}_{sm}$ . Alternatively, one can restrict the penalized estimate  $\hat{\boldsymbol{\gamma}}_{sm,i}$  to be a convex combination of the maximum likelihood estimate  $\hat{\boldsymbol{\gamma}}_{ml,i}$  and the constant 1.  $\hat{\boldsymbol{\gamma}}_{con,i} = \hat{\boldsymbol{\gamma}}_{ml,i}(1-a) + a(1)$  (see Sections 1.4.1 and 4.3). Substituting this estimate into  $l_{pen}(\boldsymbol{\gamma})$  gives

$$\begin{aligned} l_{pen}(a) &= \sum_{i=1}^k y_i \log[(1-a)\hat{\boldsymbol{\gamma}}_{ml,i} + a] - o\left(\sum_{i=1}^k \frac{N_i [(1-a)\hat{\boldsymbol{\gamma}}_{ml,i} + a]}{N_+} - 1\right) \\ &- \lambda \sum_i \sum_{q \neq i} \frac{1}{d_{iq}} [(1-a)\hat{\boldsymbol{\gamma}}_{ml,i} - (1-a)\hat{\boldsymbol{\gamma}}_{ml,q}]^2 \quad . \quad (2.2.1) \end{aligned}$$

If one writes  $pen(\hat{\boldsymbol{\gamma}}_{ml}) = \sum_i \sum_{q \neq i} \frac{1}{d_{iq}} [\hat{\boldsymbol{\gamma}}_{ml,i} - \hat{\boldsymbol{\gamma}}_{ml,q}]^2$  then (2.2.1) reduces to

$$l_{pen}(a) = \sum_{i=1}^k y_i \log[(1-a)\hat{\boldsymbol{\gamma}}_{ml,i} + a] - \lambda(1-a)^2 pen(\hat{\boldsymbol{\gamma}}_{ml}) \quad .$$

Differentiating with respect to  $a$  gives

$$\frac{\partial l_{pen}(a)}{\partial a} = \sum_{i=1}^k y_i \frac{1 - \hat{\boldsymbol{\gamma}}_{ml,i}}{(1-a)\hat{\boldsymbol{\gamma}}_{ml,i} + a} + 2\lambda pen(\hat{\boldsymbol{\gamma}}_{ml}) \quad .$$

Equating a one-term Taylor series expansion about  $a = 0$  for the right hand side of this equation to zero gives

$$\hat{a} = \frac{\lambda \rho \epsilon n(\hat{\gamma}_{ml})}{\lambda \rho \epsilon n(\hat{\gamma}_{ml}) + \sum_{i=1}^k y_i r_i^2}$$

as an approximate solution for  $a$  in terms of  $\lambda$ , where  $r_i = (1 - \hat{\gamma}_{ml,i})/\hat{\gamma}_{ml,i}$ . Hence the choice for  $a$  in  $\hat{\gamma}_{con,i}$  is approximately equal to the estimate proposed in Section 2.2. For  $\lambda = 0$ ,  $\hat{a} = 0$  and  $\hat{\gamma}_{con,i} = \hat{\gamma}_{ml,i}$ . For  $\lambda$  approaching infinity,  $\hat{a}$  approaches 1 and  $\hat{\gamma}_{con,i}$  tends to 1.

### Appendix 2.3 Additional Examples of Smoothed Estimates

In this appendix, additional examples of the numerically smoothed estimate  $\hat{\gamma}_{sm}$  are given. There are  $k = 10$  sites and the spatial configurations and population vectors are the same as the simulations of Tables 2.1 through 2.4. The number of cases is now  $m = 200$  so the ratio  $m/k$  has increased from 10 to 20.

The following table supplements Table 2.1 in that the same  $\gamma$  vector, spatial configurations and populations are used. The upper bound for  $\lambda$  is approximately twice as high. The effect of smoothing is similar.

Site	Coords	$\gamma$	$N$	$y$	$\lambda$			
					0	.05	.10	.550
					$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$
1	(1.76.1.64)	0.95	1000	19	0.950	0.958	0.965	0.987
2	(1.17.1.98)	0.95	1000	21	1.050	1.047	1.044	1.024
3	(0.5.1.86)	0.95	1000	13	0.650	0.670	0.659	0.810
4	(0.06.1.34)	0.95	1000	19	0.950	0.955	0.958	0.974
5	(0.06.0.66)	0.95	1000	21	1.050	1.048	1.046	1.029
6	(0.50.0.13)	0.95	1000	21	1.050	1.048	1.046	1.028
7	(1.17.0.02)	0.95	1000	14	0.700	0.720	0.738	0.848
8	(1.76.0.36)	0.95	1000	22	1.100	1.095	1.091	1.057
9	(2.00.1.00)	0.95	1000	24	1.200	1.186	1.175	1.107
10	(1.00.1.00)	1.43	1000	26	1.300	1.266	1.246	1.134

The following table supplements Table 2.2 in that the same rate ratios  $\gamma$ , spatial

configurations and populations are used. The upper bound for  $\lambda$  is higher than in that simulation. The effect of smoothing is similar to that of Table 2.2.

					$\lambda$			
					0	.025	.05	.40
Site	Coords	$\gamma$	$N$	$y$	$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$
1	(1.76.1.64)	0.95	1000	20	0.950	0.953	0.955	0.971
2	(1.17.1.98)	0.95	1000	18	0.855	0.861	0.866	0.915
3	(0.5.1.86)	0.95	1000	17	0.8075	0.814	0.821	0.886
4	(0.06.1.34)	0.95	1000	18	0.855	0.861	0.867	0.919
5	(0.06.0.66)	0.95	1000	20	0.950	0.955	0.959	0.986
6	(0.50.0.13)	0.95	1000	27	1.2825	1.273	1.267	1.166
7	(1.17.0.02)	0.95	1000	22	1.045	1.043	1.041	1.031
8	(1.76.0.36)	0.95	1000	27	1.2825	1.269	1.262	1.161
9	(2.00.1.00)	0.95	1000	16	0.760	0.769	0.780	0.870
10	(1.00.1.00)	1.43	500	15	1.425	1.382	1.324	1.146

The following table supplements Table 2.3 in that the same rate ratios  $\gamma$ , spatial configurations and populations are used. The upper bound for  $\lambda$  is higher and the effect of smoothing is similar to that of Table 2.3.

Site	Coords	$\gamma$	N	y	$\lambda$			
					0	.1	.2	1.05
					$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$
1	(1.76.1.64)	0.95	1000	21	1.050	1.038	1.030	1.013
2	(1.17.1.98)	0.95	1000	22	1.100	1.074	1.056	1.027
3	(0.5.1.86)	0.95	1000	16	0.800	0.832	0.855	0.906
4	(0.06.1.34)	0.95	1000	14	0.700	0.744	0.778	0.858
5	(0.06.0.66)	0.95	1000	18	0.900	0.912	0.920	0.942
6	(0.50.0.13)	0.95	1000	19	0.950	0.957	0.962	0.970
7	(1.17.0.02)	0.95	1000	21	1.050	1.040	1.031	1.014
8	(1.76.0.36)	0.95	1000	23	1.150	1.114	1.093	1.050
9	(2.00.1.00)	0.95	1000	17	0.850	0.882	0.903	0.942
10	(1.00.5.00)	1.43	1000	29	1.450	1.406	1.369	1.312

The following table supplements Table 2.4 in that the same rate ratios  $\gamma$ , spatial configurations and populations are used. The upper bound for  $\lambda$  is higher and the effect of smoothing is similar to that of Table 2.4.

Site	Coords	$\gamma$	N	y	$\lambda$			
					0	.1	.2	.75
					$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sm}$
1	(1.76.1.64)	0.95	1000	25	1.1875	1.163	1.143	1.088
2	(1.17.1.98)	0.95	1000	19	0.9025	0.9180	0.931	0.964
3	(0.5.1.86)	0.95	1000	25	1.1875	1.163	1.142	1.086
4	(0.06.1.34)	0.95	1000	19	0.9025	0.915	0.925	0.955
5	(0.06.0.66)	0.95	1000	19	0.9025	0.913	0.922	0.949
6	(0.50.0.13)	0.95	1000	21	0.9975	0.998	0.999	0.999
7	(1.17.0.02)	0.95	1000	19	0.9025	0.912	0.918	0.946
8	(1.76.0.36)	0.95	1000	18	0.855	0.871	0.882	0.925
9	(2.00.1.00)	0.95	1000	21	0.9975	0.999	0.999	1.000
10	(1.00.5.00)	1.43	500	14	1.333	1.299	1.276	1.119



## Appendix 2.4 Higher Moments for the Unsmoothed Estimate

Higher moments of  $\hat{\gamma}_{ml,i}$  are required to obtain  $\mathbf{E}(\hat{l}_i)$  and  $\mathbf{E}(\hat{q}_i)$  in  $\mathbf{E}(\hat{\gamma}_{\Lambda,i})$ . For  $y_i$  distributed as binomial with index  $m$  and success probability  $\theta_i$ , the  $r$ 'th moment about 0 can be found through the use of the moment generating function  $M(t) = ((1 - \theta_i) + \theta_i e^t)^m$ . The  $r$ 'th derivative of  $M(t)$  evaluated at 0 gives the  $r$ 'th moment about 0. Using the relationships  $\hat{\gamma}_{ml,i} = (y_i/N_i)/(m/N_+)$  and  $\theta_i = N_i \gamma_i / N_+$ ,  $M^{(r)}(0)$  for the binomial distribution was used to obtain the higher moments for  $\hat{\gamma}_{ml,i}$  (denoted as  $\hat{\gamma}_i$  below). In the following expressions,  $m^{[r]}$  is the factorial product  $m(m-1)\dots(m-r+1)$

$$\mathbf{E}(\hat{\gamma}_i^2) = \frac{1}{m} \left[ \frac{N_+}{N_i} \gamma_i + (m-1) \gamma_i^2 \right]$$

$$\mathbf{E}(\hat{\gamma}_i^3) = \frac{1}{m^2} \left[ \left( \frac{N_+}{N_i} \right)^2 \gamma_i + 3(m-1) \frac{N_+}{N_i} \gamma_i^2 + \frac{m^{[3]}}{m} \gamma_i^3 \right]$$

$$\begin{aligned} \mathbf{E}(\hat{\gamma}_i^4) &= \frac{1}{m^3} \left[ \left( \frac{N_+}{N_i} \right)^3 \gamma_i + 7(m-1) \left( \frac{N_+}{N_i} \right)^2 \gamma_i^2 \right] \\ &+ \frac{1}{m^3} \left[ 6 \frac{m^{[3]}}{m} \left( \frac{N_+}{N_i} \right) \gamma_i^3 + \frac{m^{[4]}}{m} \gamma_i^4 \right] \end{aligned}$$

$$\begin{aligned} \mathbf{E}(\hat{\gamma}_i^5) &= \frac{1}{m^4} \left[ \left( \frac{N_+}{N_i} \right)^4 \gamma_i + 15(m-1) \left( \frac{N_+}{N_i} \right)^3 \gamma_i^2 \right] \\ &+ \frac{1}{m^4} \left[ 25 \frac{m^{[3]}}{m} \left( \frac{N_+}{N_i} \right)^2 \gamma_i^3 + 10 \frac{m^{[4]}}{m} \frac{N_+}{N_i} \gamma_i^4 \right] \\ &+ \frac{1}{m^4} \left[ \frac{m^{[5]}}{m} \gamma_i^5 \right] \end{aligned}$$

$$\begin{aligned}
\mathbf{E}(\hat{\gamma}_i^6) &= \frac{1}{m^5} \left[ \left( \frac{N_+}{N_i} \right)^5 \gamma_i + 31(m-1) \left( \frac{N_+}{N_i} \right)^4 \gamma_i^2 \right] \\
&\quad + \frac{1}{m^5} \left[ 90 \frac{m^{[3]}}{m} \left( \frac{N_+}{N_i} \right)^3 \gamma_i^3 \right] \\
&\quad + \frac{1}{m^5} \left[ 65 \frac{m^{[4]}}{m} \left( \frac{N_+}{N_i} \right)^2 \gamma_i^4 \right] \\
&\quad + \frac{1}{m^5} \left[ 15 \frac{m^{[5]}}{m} \frac{N_+}{N_i} \gamma_i^5 \right] \\
&\quad + \frac{1}{m^5} \left[ \frac{m^{[6]}}{m} \gamma_i^6 \right]
\end{aligned}$$

## Appendix 2.5 Mixed Moments

The moments of the approximate estimate  $\hat{\gamma}_{\lambda,i} = \hat{\gamma}_{ml,i} + \lambda \hat{l}_i + \lambda^2 \hat{q}_i$  require expectations of the form  $\mathbf{E}(\hat{\gamma}_i^{r_1} \hat{\gamma}_j^{r_2} \hat{\gamma}_l^{r_3})$  where  $\hat{\gamma}_i$  is the maximum likelihood estimate for site  $i$ .

For  $\mathbf{y}$  distributed as multinomial with  $m$  cases and probability vector  $\boldsymbol{\theta}$ .  $\text{Cov}(y_i, y_j) = -m\theta_i\theta_j$  and  $\mathbf{E}(y_i y_j) = m(m-1)\theta_i\theta_j$ . Writing  $\hat{\gamma}_{ml,i}$  as  $\hat{\gamma}_i = (y_i/N_i)/(m/N_+)$  and  $\hat{\gamma}_{ml,j} = (y_j/N_j)/(m/N_+)$ , it can be shown that  $\text{Cov}(\hat{\gamma}_{ml,i}, \hat{\gamma}_{ml,j}) = -(1/m)\gamma_i\gamma_j$  and  $\mathbf{E}(\hat{\gamma}_{ml,i} \cdot \hat{\gamma}_j) = ((m-1)/m)\gamma_i\gamma_j$ .

For expectations of products involving powers of  $\hat{\gamma}_{ml,i}$  and  $\hat{\gamma}_{ml,j}$  which are greater than one, a more general expression for the mixed moment is required. Johnson and Kotz (1967) give the expression

$$\mathbf{E}(y_1^{[r_1]} y_2^{[r_2]} \dots y_s^{[r_s]}) = m \binom{s}{\sum_{j=1}^s r_j} \prod_j \theta_j^{r_j}$$

where  $m^{[r]}$  is the factorial product  $m(m-1)\dots(m-r+1)$ . Similarly,  $y^{[r]}$  is  $y(y-1)\dots(y-r+1)$ . The required mixed moment can be found by rearrangement of this expression for the factorial moment. For example,

$$\mathbf{E}[y_i^{[2]} y_j] = E[y_i(y_i-1)y_j]$$

$$m^{[3]}\theta_i^2\theta_j = E[y_i^2 y_j] + E[y_i y_j]$$

$$m^{[3]}\theta_i^2\theta_j - m(m-1)\theta_i\theta_j = E[y_i^2 y_j] \quad .$$

Substituting  $\hat{\gamma}_{ml,i} = (y_i/N_i)/(m/N_+)$  and  $\hat{\gamma}_{ml,j} = (y_j/N_j)/(m/N_+)$ , one obtains

$$\mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j) = \frac{(m-1)(m-2)}{m^2} \gamma_i^2 \gamma_j - \frac{m-1}{m} \left( \frac{N_+}{N_i} \right) \gamma_i \gamma_j \quad .$$

Denote the quotient  $(m-1)\dots(m-r+1)/m^{r-1}$  as  $m(r)$ . Some other examples of required mixed moments which were obtained in a similar manner are given below.

$$\begin{aligned} \mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j^2) &= m(4) \gamma_i^2 \gamma_j^2 + \frac{N_+}{m N_i} \mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j) \\ &+ \frac{N_+}{m N_j} \mathbf{E}(\hat{\gamma}_j^2 \hat{\gamma}_i) + \frac{m(2)}{m^2} \frac{N_+^2}{N_i N_j} \gamma_i \gamma_j \end{aligned}$$

$$\mathbf{E}(\hat{\gamma}_i^3 \hat{\gamma}_i) = m(4) \gamma_i^3 \gamma_j + 3 \frac{N_+}{m N_i} \mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j) - 2 \frac{m(2)}{m^2} \left( \frac{N_+}{N_i} \right)^2 \gamma_i \gamma_j$$

$$\begin{aligned} \mathbf{E}(\hat{\gamma}_i^3 \hat{\gamma}_j^2) &= m(5) \gamma_i^3 \gamma_j^2 + 3 \frac{N_+}{m N_i} \mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j^2) - \left( \frac{N_+}{m N_i} \right)^2 \mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j) \\ &+ \frac{N_+}{m N_i} \mathbf{E}(\hat{\gamma}_i^3 \hat{\gamma}_j) - 3 \frac{N_+^2}{m^2 N_i N_j} \mathbf{E}(\hat{\gamma}_j^2 \hat{\gamma}_i) + 2 \frac{m(2)}{m^3} \frac{N_+^3}{N_i^2 N_j} \gamma_i \gamma_j \end{aligned}$$

$$\mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j \hat{\gamma}_l) = m(4) \gamma_i^2 \gamma_j \gamma_l + m(3) \frac{N_+}{m N_i} \gamma_i \gamma_j \gamma_l$$

$$\begin{aligned} \mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j^2 \hat{\gamma}_l) &= m(5) \gamma_i^2 \gamma_j^2 \gamma_l + \frac{N_+}{m N_i} \mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j \hat{\gamma}_l) \\ &+ \frac{N_+}{m N_j} \mathbf{E}(\hat{\gamma}_j^2 \hat{\gamma}_i \hat{\gamma}_l) + \frac{m(3)}{m^2} \frac{N_+^2}{N_i N_j} \gamma_i \gamma_j \gamma_l \end{aligned}$$

$$\mathbf{E}(\hat{\gamma}_i^2 \hat{\gamma}_j \hat{\gamma}_l \hat{\gamma}_s) = m(5) \gamma_i^2 \gamma_j \gamma_l \gamma_s + m(4) \frac{N_+}{m N_i} \gamma_i \gamma_j \gamma_l \gamma_s$$

### Appendix 2.6 Additional Results: Observed and Approximate MSE

The following tables compare the observed and approximate MSE for simulations involving the same  $m$ ,  $k$  and spatial configuration as for the results of Tables 2.7 through 2.9. The 'circle + 1' configuration was used for  $k = 10$ , and  $m = 100$ . In the simulations presented in Section 2.7, the populations were equal ( $N_j = 1000$ ). The population at the distant site is 500 in the simulations summarized below.

$\gamma_1 = \dots = \gamma_{10} = 1.0$		
$\lambda$	Obs. MSE	Approx. MSE
0	0.95322	0.94500
.00784	0.90029	0.90986
.01569	0.86414	0.87471
.02353	0.85369	0.83957
.03138	0.82355	0.80442

$\gamma_1 = \dots = \gamma_9 = 0.97, \gamma_{10} = 1.46$		
$\lambda$	Obs. MSE	Approx. MSE
0	1.00623	1.00396
.00459	0.98254	0.97998
.00918	0.94663	0.95601
.01377	0.94396	0.93203
.01837	0.93040	0.90805

$\gamma_1 = \dots = \gamma_9 = 0.86, \gamma_{10} = 3.45$		
$\lambda$	Obs. MSE	Approx. MSE
0	1.18902	1.20831
.00076	1.18776	1.20031
.00152	1.18265	1.19232
.00227	1.16761	1.18432
.00303	1.17933	1.17633

**Appendix 2.7 Additional Examples: Estimators  $\hat{\gamma}_{sm}$ ,  $\hat{\gamma}_{\lambda}$ ,  $\hat{\gamma}_{\lambda,lin}$ , and  $\hat{\gamma}_{\lambda,red}$**

The following tables presents several forms of the approximate estimates for the same examples presented in Table 2.1. through 2.4. In the table below the following are given at  $\lambda = .025$ : the numerically smoothed estimate  $\hat{\gamma}_{sm}$ ; the full approximation  $\hat{\gamma}_{\lambda} = \hat{\gamma}_{ml} + \lambda \hat{l}_i + \lambda^2 \hat{q}_i$ ; the linear approximation  $\hat{\gamma}_{ml,i} + \lambda \hat{l}_i$ ; and the reduced linear approximation  $\hat{\gamma}_{ml,i} - 4/N_i(m/N_+) (b_i \hat{\gamma}_i^2 - \hat{\gamma}_i \sum_{j \neq i} e_{ij} \hat{\gamma}_j)$ . The reduced estimators  $\hat{\gamma}_{\lambda,lin}$  and  $\hat{\gamma}_{\lambda,red}$  were introduced in Section 2.9.1

	$\lambda = 0$	$\lambda = .025$			
Site	$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{\lambda}$	$\hat{\gamma}_{\lambda,lin}$	$\hat{\gamma}_{\lambda,red}$
1	0.9000	0.9136	0.9134	0.9155	0.9109
2	1.1000	1.0993	1.0991	1.1005	1.0949
3	1.2000	1.1895	1.1893	1.1896	1.1835
4	0.9000	0.9124	0.9122	0.9141	0.9095
5	0.9000	0.9099	0.9097	0.9112	0.9067
6	0.8000	0.8147	0.8145	0.8160	0.8120
7	0.9000	0.9077	0.9074	0.9088	0.9043
8	0.6000	0.6222	0.6222	0.6231	0.6201
9	1.2000	1.1854	1.1852	1.1847	1.1787
10	1.5000	1.4448	1.4464	1.4361	1.4285



The following table presents several forms of the approximate estimates for the same example presented in Table 2.2. For  $\lambda = .01$ , the following are given : the numerically smoothed estimate  $\hat{\gamma}_{sm}$ ; the full approximation  $\hat{\gamma}_{\lambda} = \hat{\gamma}_{ml,i} + \lambda \hat{l}_i + \lambda^2 \hat{q}_i$ ; the linear approximation  $\hat{\gamma}_{ml,i} + \lambda \hat{l}_i$ ; and the reduced linear approximation  $\hat{\gamma}_{ml,i} - 4/N_i(m/N_+) (b_i \hat{\gamma}_i^2 - \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j)$ . The reduced estimators  $\hat{\gamma}_{\lambda,lin}$  and  $\hat{\gamma}_{\lambda,red}$  were introduced in Section 2.9.1

	$\lambda = 0$	$\lambda = .01$			
Site	$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{\lambda}$	$\hat{\gamma}_{\lambda,lin}$	$\hat{\gamma}_{\lambda,red}$
1	0.9500	0.9591	0.9586	0.9611	0.9533
2	0.6650	0.6795	0.6792	0.6808	0.6753
3	1.1400	1.1425	1.1418	1.1447	1.1353
4	0.7600	0.7740	0.7736	0.7756	0.7693
5	0.9500	0.9602	0.9597	0.9622	0.9545
6	0.9500	0.9619	0.9614	0.9640	0.9562
7	1.5200	1.5041	1.5032	1.5059	1.4934
8	0.5700	0.5855	0.5852	0.5865	0.5819
9	0.9500	0.9592	0.9587	0.9612	0.9534
10	2.0900	1.9475	1.9565	1.9151	1.8980

The following table presents several forms of the approximate estimates for the same example presented in Table 2.3. For  $\lambda = .08$ , the following are given : the numerically smoothed estimate  $\hat{\gamma}_{sm}$ ; the full approximation  $\hat{\gamma}_\lambda = \hat{\gamma}_{ml,i} + \lambda \hat{l}_i + \lambda^2 \hat{q}_i$ ; the linear approximation  $\hat{\gamma}_{ml,i} + \lambda \hat{l}_i$  ; and the reduced linear approximation  $\hat{\gamma}_{ml,i} - 4/N_i(m/N_+) (b_i \hat{\gamma}_i^2 - \hat{\gamma}_i \sum_{i \neq j} \epsilon_{ij} \hat{\gamma}_j)$ . The reduced estimators  $\hat{\gamma}_{\lambda,lin}$  and  $\hat{\gamma}_{\lambda,red}$  were introduced in Section 2.9.i

	$\lambda = 0$	$\lambda = .08$			
Site	$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_\lambda$	$\hat{\gamma}_{\lambda,lin}$	$\hat{\gamma}_{\lambda,red}$
1	0.7000	0.7402	0.7394	0.7458	0.7366
2	0.7000	0.7465	0.7453	0.7540	0.7448
3	1.2000	1.1481	1.1539	1.1311	1.1154
4	0.9000	0.9226	0.9199	0.9309	0.9191
5	0.9000	0.9237	0.9211	0.9320	0.9202
6	1.1000	1.0848	1.0828	1.0855	1.0710
7	1.3000	1.2303	1.2391	1.2075	1.1904
8	0.9000	0.9188	0.9160	0.9261	0.9143
9	0.7000	0.7425	0.7416	0.7488	0.7396
10	1.6000	1.5418	1.5404	1.5378	1.5169

The following table presents several forms of the approximate estimates for the same example presented in Table 2.4. For  $\lambda = .05$ , the following are given : the numerically smoothed estimate  $\hat{\gamma}_{sm}$ ; the full approximation  $\hat{\gamma}_\lambda = \hat{\gamma}_{ml,i} + \lambda \hat{l}_i + \lambda^2 \hat{q}_i$ ; the linear approximation  $\hat{\gamma}_{ml,i} + \lambda \hat{l}_i$ ; and the reduced linear approximation  $\hat{\gamma}_{ml,i} - 4/N_i(m/N_+) (b_i \hat{\gamma}_i^2 - \hat{\gamma}_i \sum_{j \neq i} e_{ij} \hat{\gamma}_j)$ . The reduced estimators  $\hat{\gamma}_{\lambda,lin}$  and  $\hat{\gamma}_{\lambda,red}$  were introduced in Section 2.9.1

	$\lambda = 0$	$\lambda = .05$			
Site	$\hat{\gamma}_{ml}$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_\lambda$	$\hat{\gamma}_{\lambda,lin}$	$\hat{\gamma}_{\lambda,red}$
1	1.0450	1.0365	1.0366	1.0353	1.0313
2	0.8550	0.8740	0.8739	0.8769	0.8736
3	1.1400	1.1257	1.1258	1.1241	1.1198
4	1.1400	1.1276	1.1276	1.1266	1.1222
5	1.0450	1.0447	1.0444	1.0456	1.0416
6	1.0450	1.0410	1.0409	1.0409	1.0369
7	0.8550	0.8707	0.8705	0.8727	0.8694
8	0.9500	0.9518	0.9518	0.9521	0.9484
9	0.6650	0.6937	0.6937	0.6954	0.6928
10	1.5200	1.4673	1.4685	1.4598	1.4539

### Appendix 2.8 Further Investigation: Normality of $\hat{\gamma}_{sm}$

This appendix gives additional Q-Q plots investigating the normality of  $\hat{\gamma}_{sm}$ . In the simulation for Figure A28.1, the expected number of cases out of 200 at the distant site is larger because of its elevated incidence. More smoothing is required to bring the estimates at other sites closer to normality. In the simulations for Figures A28.2 and A28.3,  $m = 400$  cases are used and the estimates for all  $\lambda$  (including 0) are approximately normal.

**Figure A28.1** presents Q-Q plots for the estimates at the distant site in the 'circle + 1' configuration for  $k = 20$  sites and  $m = 200$  cases. The incidence rate at the distant site is twice that of the common nineteen, resulting in rate ratios  $\gamma_1 = \gamma_2 = \dots \gamma_{19} = 0.95$  and  $\gamma_{20} = 1.90$ .

**Figure A28.2** presents Q-Q plots for the estimates at the distant site under complete homogeneity, (as in Figure 2.5b) but twice the number of cases ( $m = 400$ ) There are  $k = 20$  sites in the 'circle + 1' configuration.

**Figure A28.3** presents Q-Q plots for the estimates at the elevated/distant site as in Figure A28.1 but twice the number of cases ( $m = 400$ ) have now been used. At this site,  $\gamma_i = 1.90$ . There are  $k = 20$  sites in the 'circle + 1' configuration.

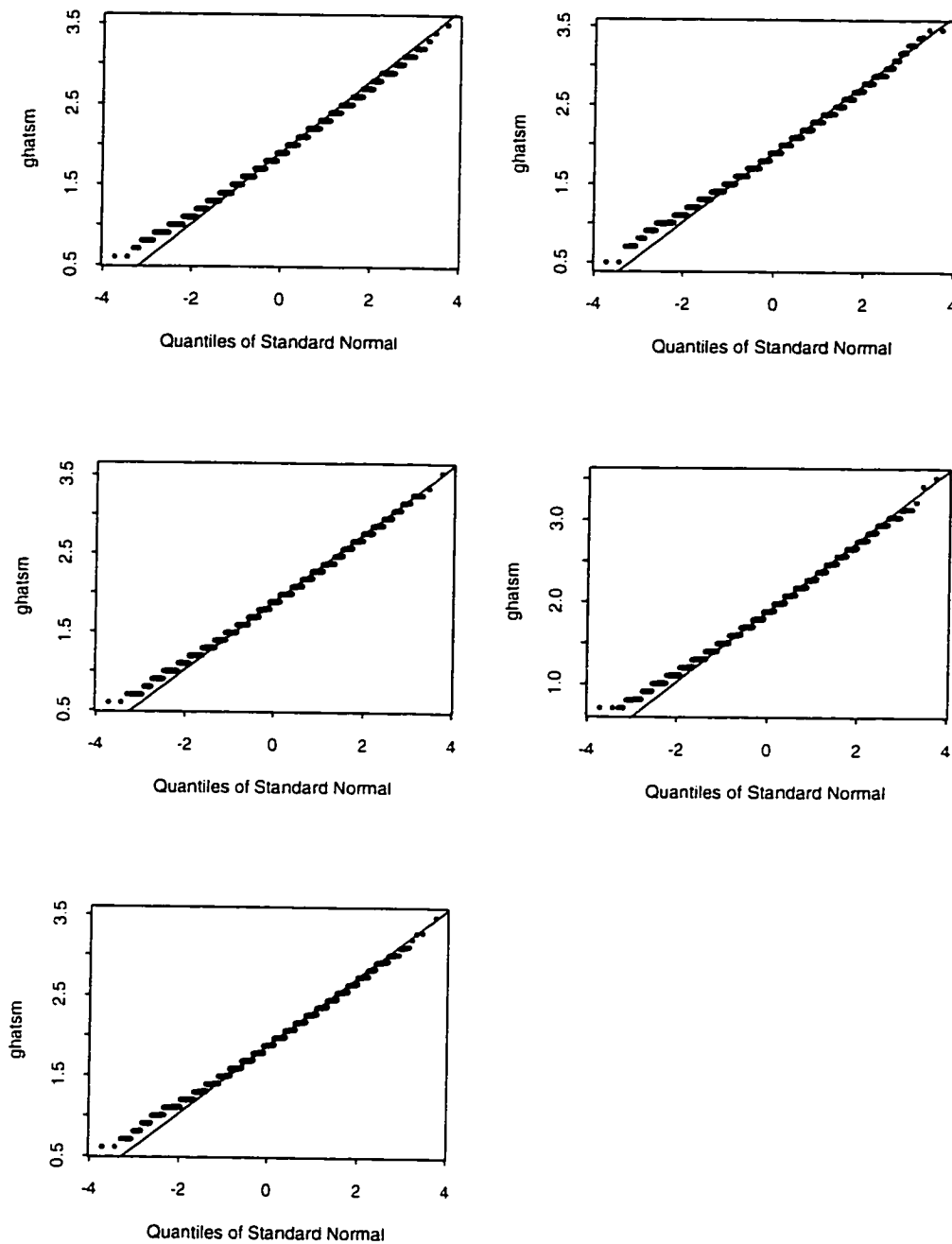


Figure A28.1: Q-Q plots - 1 Elevated Rate ( $k=20, m=200$ )

Elevated rate at distant site

Across Rows, lambdas : 0, .0023, .0046, .0073, .0094

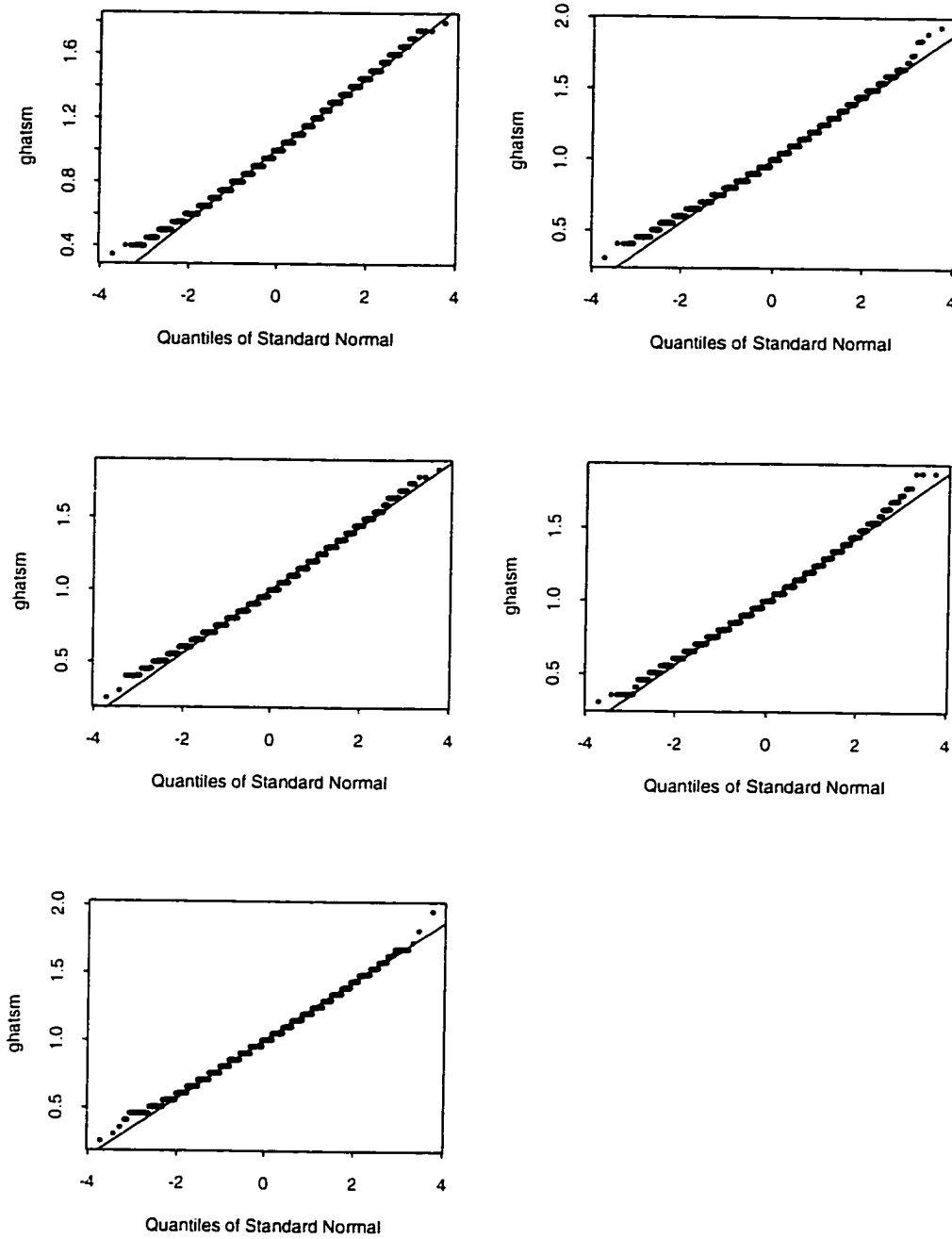


Figure A28.2: Q-Q plots - Complete Homogeneity ( $k=20, m=400$ )

Distant site, circle + 1 configuration

Across Rows, lambdas : 0,.006,.012,.018,.024

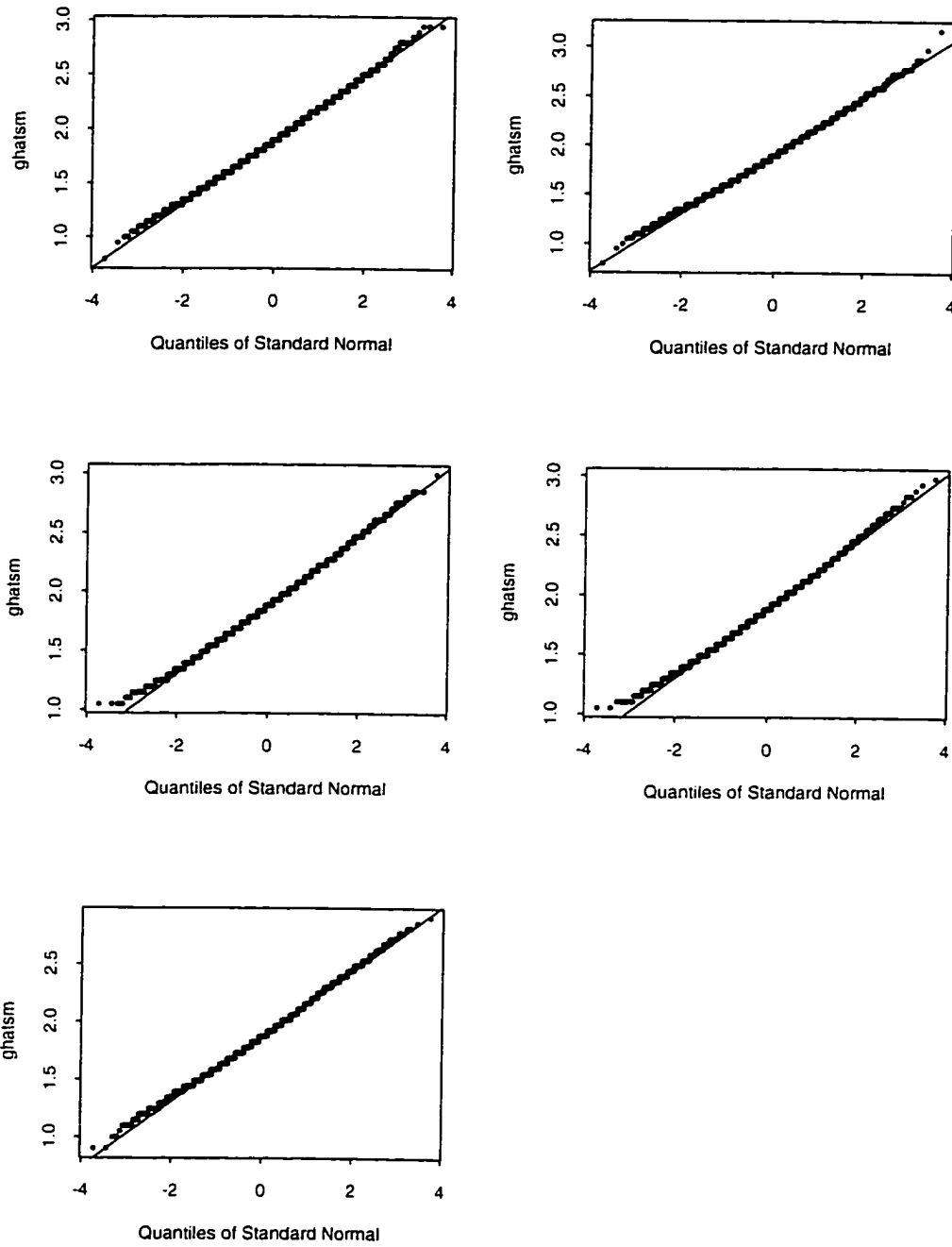


Figure A28.3: Q-Q plots - 1 Elevated Rate (k=20,m=400)

Distant site, circle + 1 configuration

Across Rows, lambdas : 0,.0048,.0096,.0144,.0188

# Appendix B

## Chapter 3 Appendices

3.1	Use of a Continuity Correction. . . . .	152
3.2	Estimation under the constraint $\gamma_i = 1$ . . . . .	154
3.3	Use of a Cornish-Fisher expansion. . . . .	159
3.4	Conditional moments of $w_i$ given $\hat{\gamma}_{ml,i}$ . . . . .	162
3.5	Additional Results: Complete Homogeneity, $k = 40$ sites. . . . .	166
3.6	Additional results: Single elevated rate. . . . .	167
3.7	Locations and Contiguities: Main Spatial Configurations . . . . .	171
3.8	Additional Results: Two elevated rates. . . . .	172
3.9	Locations and Contiguities: 'Random' Configuration . . . . .	179



### Appendix 3.1

#### Use of a continuity correction

With no smoothing, the test statistic

$$Z_{ml,i} = \frac{\hat{\gamma}_{ml,i} - 1}{v_{\sigma,i}^{1/2}}$$

has been proposed for testing  $H_{\sigma i} : \gamma_i = 1$  versus  $H_{\alpha i} : \gamma_i > 1$ , where  $v_{\sigma,i} = \text{Var}(\hat{\gamma}_{ml,i}) = (1/m)(N_+/N_i)(1 - (N_i/N_+))$  under  $H_{\sigma,i}$ .

For testing at level  $\alpha^* = \alpha/(k - 1)$ , the critical value  $C_{\alpha^*,i}$  such that  $P(Z_{ml,i} \geq C_{\alpha^*,i}) \leq \alpha^*$  is required.

One can attempt to use the standard normal quantile  $Z_{\alpha^*}$  as  $C_{\alpha^*,i}$  because  $\hat{\gamma}_{ml,i}$  is approximately  $N(1, v_{\sigma,i})$  under  $H_{\sigma}$ . In Section 3.2, the standard normal approximation was shown to be inadequate in this context. Too many rejections under  $H_{\sigma}$  occur at the site and overall levels.

In an attempt to improve the normal approximation, continuity corrections are often used. If  $X$  is a discrete random variable, the general idea is to replace the point probability  $P(X = x)$  by the interval probability  $P(x - 1/2 \leq X \leq x + 1/2)$  from the approximating continuous distribution ( see Johnson and Kotz(1970)). The estimator  $\hat{\gamma}_{ml,i} = (y_i/N_i)/(m/N_+)$  has increments of  $1/(m.N_i/N_+)$  as  $y_i$  ranges from 0 to  $m$ . The test statistic  $Z_{ml,i}$  has a discrete distribution with probabilities corresponding to the binomial  $y_i$  as described in Section 3.1. The distance between possible values of  $Z_{ml,i}$  is  $1/(v_{\sigma,i}^{1/2} m.N_i/N_+)$ . Taking half this distance as a continuity correction (cc), a continuity-corrected normal approximation for the required probability is

$$P(Z_{ml,i} \geq C_{\alpha^*,i}) = 1 - \Phi(C_{\alpha^*,i} - cc) \quad .$$

Solving for  $C_{\alpha^*,i}$  in

$$P(Z_{ml,i} \geq C_{\alpha^*,i}) = \alpha^*$$

or

$$\Phi(C_{\alpha^*,i} - cc) = 1 - \alpha^*$$

gives

$$C_{\alpha^*,i} - cc = Z_{\alpha^*}$$

or

$$C_{\alpha^*,i} = Z_{\alpha^*} + cc \quad .$$

If  $C_{\alpha^*,i}$  must be a possible value for  $Z_{ml,i}$ , then to ensure  $P(Z_{ml,i} \geq C_{\alpha^*,i}) \leq \alpha^*$ ,  $C_{\alpha^*,i}$  should be the next value of  $Z_{ml,i}$  which exceeds  $Z_{\alpha^*} + cc$

Example 3.1 illustrated that the continuity corrected standard normal quantile is also inadequate. For  $\alpha^* = .05/19 = .00263$ , the standard normal quantile  $Z_{\alpha^*,i}$  is 2.79. The distance between consecutive  $Z_{ml,i}$  values is approximately 0.45 and taking 0.225 as the continuity correction, one finds that the next possible  $Z_{ml,i}$  value beyond  $Z_{\alpha^*} + 0.225$  is 3.07. If 3.07 is taken to be the critical value for  $Z_{ml,i}$ , then  $P(Z_{ml,i} \geq 3.07) = 1 - 0.9941 = .0059$ . This probability is more than twice as large as  $\alpha^*$  and hence the site level is not achieved. As described in Section 3.1, an exact critical value of 3.52 conservatively achieves the required site and overall error rates.

### Appendix 3.2

Estimation under the constraint  $\gamma_i = 1$

#### (1) Maximum Likelihood Estimation (Unpenalized)

The multinomial likelihood  $\text{Mult}(m, \theta)$  is described in Section 2.2. The  $j$ th cell probability  $\theta_j$  is  $N_j \gamma_j / N_+$ .

With  $\gamma_i = 1$  under  $H_{oi}$ , the likelihood becomes

$$L_{(i)} = \prod_{j \neq i} \left( \frac{N_j \gamma_j}{N_+} \right)^{y_j} \left( \frac{N_i}{N_+} \right)^{y_i} .$$

and to an additive constant the log likelihood is simply

$$l_{(i)} = \sum_{j \neq i} y_j \log(\gamma_j) .$$

The multinomial constraint  $\sum_{j=1}^k \theta_j = 1$  under  $H_o$  can be written as  $\sum_{j \neq i} (N_j \gamma_j) / N_+ + N_i / N_+ - 1 = 0$  and can be included in the likelihood using a Lagrange multiplier

$$l_{(i)} = \sum_{j \neq i} y_j \log(\gamma_j) - o \left( \sum_{j \neq i} \left( \frac{N_j \gamma_j}{N_+} \right) - \frac{N_+ - N_i}{N_+} \right) .$$

Differentiating, we find that  $\partial l_{(i)} / \partial o = 0$  returns the constraint equation and

$$\frac{\partial l_{(i)}}{\partial \gamma_j} = \frac{y_j}{\gamma_j} - o \left( \frac{N_j}{N_+} \right) .$$

Multiplying by  $\gamma_j$  and setting to 0 gives

$$y_j - o \left( \frac{.N_j}{.N_+} \right) = 0 \quad (3.2.1)$$

Summing over  $j \neq i$ ,

$$m - y_{(i)} - o \left( \frac{.N_+ - .N_i}{.N_+} \right) = 0$$

and solving gives

$$\hat{o} = \frac{(m - y_{(i)}) \cdot N_+}{.N_+ - .N_i} .$$

Substituting into (3.2.1) gives

$$y_j - \frac{(m - y_{(i)}) \cdot N_+ \cdot N_j \hat{\gamma}_j}{(.N_+ - .N_i) \cdot N_+} = 0 .$$

and solving gives the constrained maximum likelihood estimate

$$\hat{\gamma}_{j,(i)} = \frac{y_j / .N_j}{(m - y_{(i)}) / (.N_+ - .N_i)} .$$

## (2) Penalized Likelihood

As described in Section 2.2 the multinomial constrained penalized likelihood has the form

$$l_{pen}(\gamma) = \sum_{j=1}^k y_j \log(\gamma_j) - o \left[ \sum_{j=1}^k \left( \frac{.N_j \gamma_j}{.N_+} \right) - 1 \right] - \lambda \sum_{\gamma \neq j} \frac{1}{d_{j\gamma}} (\gamma_i - \gamma_\gamma)^2 .$$

The rate ratio  $\gamma_i$  is now fixed at 1.0 and the equation above becomes

$$\begin{aligned} l_{pen}(\gamma)_{(i)} &= \sum_{j \neq i}^k y_j \log(\gamma_j) - o \left( \sum_{j \neq i}^k \frac{.N_j \gamma_j}{.N_+} + \frac{.N_i}{.N_+} - 1 \right) \\ &- \lambda \left[ \sum_{\gamma \neq i} \frac{1}{d_{i\gamma}} (\gamma_\gamma - 1.0)^2 + \sum_{\gamma \neq i, j} \frac{1}{d_{j\gamma}} (\gamma_j - \gamma_\gamma)^2 \right] \end{aligned}$$

Using the same notation as in Section 2.2, rearrangements show that for  $j \neq i$ ,  $\partial l_{pen} / \partial \gamma_j = 4(b_j \gamma_j - \epsilon_{i,j} - \sum_{\gamma \neq j, i} \epsilon_{j\gamma} \gamma_\gamma)$  and, as in equation (2.16) of that section,

$$\frac{\partial l_{pen}}{\partial \gamma_j} = \frac{y_j}{\gamma_j} - o \left( \frac{.N_j}{.N_+} \right) - \lambda \frac{\partial pen}{\partial \gamma_j} . \quad (3.2.2)$$

Multiplying by  $\gamma_j$  and setting to 0 gives

$$y_j - o \left( \frac{.N_j}{.N_+} \right) - \lambda \gamma_j \frac{\partial pen}{\partial \gamma_j} = 0 .$$

Summing over  $l \neq i$ , yields

$$m - y_i - o \sum_{l \neq i} \left( \frac{N_l \gamma_{il}}{N_+} \right) - \lambda \sum_{l \neq i} \gamma_{il} \frac{\partial p \epsilon n}{\partial \gamma_{il}}$$

or

$$m - y_i - o \left( \frac{N_+ - N_i}{N_+} \right) - \lambda \sum_{l \neq i} \gamma_{il} \frac{\partial p \epsilon n}{\partial \gamma_{il}} = 0 \quad .$$

Solving, gives the Lagrangian

$$o = \left( m - y_i - \lambda \sum_{l \neq i} \gamma_{il} \frac{\partial p \epsilon n}{\partial \gamma_{il}} \right) \left( \frac{N_+}{N_+ - N_i} \right) \quad .$$

Substituting this result into equation (3.2.2).

$$\frac{y_j}{\gamma_{ij}} - \frac{(m - y_i) N_j}{N_+ - N_i} - \lambda \left[ \frac{\partial p \epsilon n}{\partial \gamma_{ij}} - \left( \frac{N_+}{N_+ - N_i} \right) \sum_{l \neq i} \gamma_{il} \frac{\partial p \epsilon n}{\partial \gamma_{il}} \right] = 0.$$

and multiplying by  $\gamma_{ij}$  gives

$$y_j - \frac{(m - y_i) N_j}{N_+ - N_i} \gamma_{ij} - \lambda \gamma_{ij} \left[ \frac{\partial p \epsilon n}{\partial \gamma_{ij}} - \left( \frac{N_+}{N_+ - N_i} \right) \sum_{l \neq i} \gamma_{il} \frac{\partial p \epsilon n}{\partial \gamma_{il}} \right] = 0 \quad (3.2.3)$$

As first outlined in Section 2.2, the  $k - 1$  equations for  $j \neq i$  are once again nonlinear in the  $\gamma_j$ . As before, assume that for small  $\lambda$  the approximate solution takes a form which is quadratic in  $\lambda$ . In other words,

$$\gamma_{j,adj(i)} \simeq c_{j(i)} + \lambda l_{j(i)} + \lambda^2 q_{j(i)} \quad .$$

Substituting the approximation into (3.2.3) gives

$$y_j - \left( \frac{m - y_i}{N_+ - N_i} \right) \cdot N_j (c_j + l_j \lambda + q_j) \\ - \lambda (c_j + l_j \lambda + q_j) \left[ \frac{\partial \text{pen}}{\partial \gamma_j} - \left( \frac{N_+}{N_+ - N_i} \right) \sum_{r \neq i} (c_r + l_r \lambda + q_r) \left( \frac{\partial \text{pen}}{\partial \gamma_r} \right) \right] = 0 \quad .$$

Collecting terms constant in  $\lambda$  (i.e. no  $\lambda$  involved), then these terms must sum to zero and hence

$$y_j - \left( \frac{N_+}{N_+ - N_i} \right) \cdot N_j c_j = 0 \quad .$$

Solving gives

$$c_j = \frac{y_j / N_j}{(m - y_i) / (N_+ - N_i)} \quad .$$

Hence the first term of the approximation under the constraint  $\gamma_i = 1$  is the maximum likelihood estimator under this constraint. This result coincides with the first term of the approximate smoothed estimate without the imposed constraint (as described in Section 2.4).

Further collection of terms gives the same form for the linear term as previously obtained. It now involves the adjusted maximum likelihood estimate  $c_j$  where  $m - y_i$  replaces  $m$  and  $N_+ - N_i$  replaces  $N_+$ . One finds

$$l_j = \left( \frac{N_+ - N_i}{m - y_i} \right) \left[ -\frac{1}{N_j} \left( c_j \frac{\partial \text{pen}}{\partial \gamma_j} \right) \right. \\ \left. + \frac{1}{N_+ - N_i} \left( c_j \sum_{r \neq j, i} c_r \frac{\partial \text{pen}}{\partial \gamma_r} \right) \right] \quad .$$

Similarly, collecting terms gives the quadratic term, which is similar in form in the previous approximation without the constraint.

### Appendix 3.3

#### Use of a Cornish-Fisher expansion

Perhaps the most common technique for finding the quantile of an unknown distribution function was introduced by Cornish and Fisher (1937). A Cornish-Fisher expansion expresses the quantile of an unknown distribution function  $g$  in terms of the cumulants and corresponding quantile for a known distribution function  $f$ .

Assume  $f(x)$  and  $g(x)$  are probability density functions and that  $g(x)$  has cumulants  $\kappa_1, \kappa_2, \dots, \kappa_r$  and  $f(x)$  has cumulants  $\kappa_1 + \epsilon_1, \kappa_2 + \epsilon_2, \dots, \kappa_r + \epsilon_r$ . As given in Johnson and Kotz (1970), one can write:

$$g(x) = \exp \left[ \sum_{j=1}^{\infty} \epsilon_j \{(-D)^j / j!\} \right] f(x) \quad . \quad (3.3.1)$$

where  $D^j f(x) = \frac{d^j}{dx^j} f(x)$  and the exponent must be expanded as an operator and applied to  $f(x)$ . In other words, the exponent must first be expanded as

$$\sum_{i=0}^{\infty} \left[ \sum_{j=1}^{\infty} \epsilon_j \{(-D)^j / j!\} \right]^i / i!$$

and then applied to  $f(x)$ . As is typically the case, assume  $f(x)$  is the standard normal probability density function  $z(x) = (\sqrt{2\pi})^{-1} \exp(-\frac{1}{2}x^2)$ . Denote  $Z_q$  as the known standard normal quantile with cumulative probability  $q$  and  $x_q$  as the corresponding quantile for  $g$ . One can then consider the equation

$$\int_{-\infty}^{x_q} g(x) dx = \int_{-\infty}^{Z_q} z(x) dx \quad . \quad (3.3.2)$$

Hermite polynomials (see Cramer, 1970) allow us to conveniently express the derivatives of the standard normal p.d.f. as follows:  $D^j z(x) = (-1)^j H_j(x) z(x)$



where  $H_0(x) = 1$ ,  $H_1(x) = x$  and  $H_i(x) = xH_{i-1}(x) - (i-1)H_{i-2}(x)$  for  $i \geq 2$ . The right hand side of (3.3.2) is the standard normal cumulative distribution function evaluated at  $Z_q$ , denoted as  $\Phi(Z_q)$ . A Taylor expansion for  $\Phi(Z_q) = \Phi(x_q + Z_q - x_q)$  gives

$$\Phi(Z_q) = \Phi(x_q) - \sum_{i=1}^{\infty} \frac{(x_q - Z_q)^i}{i!} H_{i-1}(x_q) z(x_q) \quad . \quad (3.3.3)$$

Equation (3.3.2) can be written as  $G(x_q) = \Phi(Z_q)$ . Substituting (3.3.3) and (3.3.1) into this equation and performing extensive manipulation allows one to express  $x_q$  as a function of its own moments and  $Z_q$ . This expression is usually referred to as the inverse Cornish-Fisher expansion. Most references assume standardization of the first two moments of  $g(x)$  but this simplification is not assumed here. From Kendall and Stuart (1957), the terms of the inverse Cornish-Fisher expansion which include moments of  $g(x)$  up to and including the fourth are

$$\begin{aligned} x_q = & Z_q + \kappa_1 + \frac{1}{2}Z_q\kappa_2 + \frac{1}{6}(Z_q^2 - 1)\kappa_3 \\ & + \frac{1}{24}(Z_q^3 - 3Z_q)\kappa_4 - \frac{1}{36}(2Z_q^3 - 5Z_q)\kappa_3^2 \\ & - \frac{1}{24}(Z_q^4 - 5Z_q^2 + 2)\kappa_3\kappa_4 + \frac{1}{324}(12Z_q^4 - 53Z_q^2 + 17)\kappa_3^3 \\ & - \frac{1}{384}(3Z_q^5 - 24Z_q^3 + 29Z_q)\kappa_4^2 + \frac{1}{288}(14Z_q^5 - 103Z_q^3 + 107Z_q)\kappa_3^2\kappa_4 \\ & - \frac{1}{7776}(252Z_q^5 - 1688Z_q^3 + 1511Z_q)\kappa_3^4 \quad . \quad (3.3.4) \end{aligned}$$

This expansion was applied to the reduced test statistic  $Z_{\lambda, red, i}$  in an investigation of its quantiles under  $H_0: \gamma_i = 1$ . Moments of  $Z_{\lambda, red, i}$  beyond the fourth were

considered intractable. As a result, terms involving the fifth and sixth moments are not included. The cumulants listed in the expansion are moments about the mean. In general, if  $\kappa_r$  represents the  $r$ th cumulant and  $\mu_r$  represents the  $r$ th moment about the mean then  $\kappa_1 = \mu_1 = \mu$ ,  $\kappa_2 = \mu_2 = \sigma^2$ ,  $\kappa_3 = \mu_3$ , and  $\kappa_4 = \mu_4 - 3\sigma^2$ .

The moments of  $Z_{\lambda,red,i}$  were required. As given by (3.12), the reduced test statistic is defined as

$$Z_{\lambda,red,i} = \hat{s}_{\lambda,red,i} Z_{ml,i} + \frac{\hat{l}_{red,i} \lambda}{v_{\sigma,i}^{\frac{1}{2}}} \quad (3.3.5)$$

where

$$\hat{s}_{\lambda,red,i} = 1 - 0.5 \hat{v}_{1,red,i} / v_{\sigma,i} \quad .$$

A necessary simplification was to consider  $\hat{s}_{\lambda,red,i}$  fixed (i.e. non-random) for fixed  $\lambda$ . Empirical investigation showed this to be a reasonable assumption. Moments up to and including the fourth were required so powers up to four of  $\hat{s}_{\lambda,red,i}$  were also needed. The approximation for  $\hat{\gamma}_{\lambda,i}$  and  $\hat{V}ar(\hat{\gamma}_{\lambda,i})$  are linear in  $\lambda$ , and this same convention was followed for the powers of  $\hat{s}_{\lambda,red,i}$ .

Higher moments of  $Z_{\lambda,red,i}$  were found by expanding powers of the right hand side of (3.3.5). The higher moments of  $Z_{ml,i}$  were found under  $H_{\sigma,i}$ . Via direct calculation of the third and fourth moments of  $Z_{\lambda,red,i}$ , it was found that many terms of order  $\lambda^2$  or greater were dropped to maintain the convention of keeping expressions to linear in  $\lambda$ . Since  $Z_{\lambda,red,i} = f(\hat{\gamma}_{ml})$ , the delta method for calculating the moments of a function of a random variable can be used. In empirical investigation, this method gave approximate third and fourth moments that were closer to the observed third and fourth moments of  $Z_{\lambda,red,i}$  than the approximate moments derived directly. However, critical values  $x_q$  calculated according to (3.3.4) were very conservative under  $H_{\sigma}$ . This resulted in a low detection ability under  $H_a$  and the conditional approach described in Section 3.3 was chosen for the determination of the critical value of  $Z_{\lambda,red,i}$ .

### Appendix 3.4

#### Conditional moments of $w_i$ given $\hat{\gamma}_{ml,i}$

In Section 3.3.3, the first two conditional moments of  $w_i$  given  $\hat{\gamma}_{ml,i}$  were given as

$$\begin{aligned}\mu_{w_i|\hat{\gamma}_i} &= \mu_{w_i} + \frac{\text{Cov}(w_i, \hat{\gamma}_i)(\hat{\gamma}_i - \mu_{\hat{\gamma}_i})}{\sigma_{\hat{\gamma}_i}^2} \\ \sigma_{w_i|\hat{\gamma}_i}^2 &= \sigma_{w_i}^2 - \frac{[\text{Cov}(w_i, \hat{\gamma}_i)]^2}{\sigma_{\hat{\gamma}_i}^2} .\end{aligned}$$

Denoting  $\hat{\gamma}_{ml,i}$  as  $\hat{\gamma}_i$ , the unconditional mean and variance of  $w_i$  are

$$\begin{aligned}\mu_{w_i} &= E\left(\sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j\right) \\ &= \sum_{j \neq i} \epsilon_{ij} \gamma_j .\end{aligned}$$

and

$$\sigma_{w_i}^2 = \mathbf{e}'_{(i)} \Sigma_{\hat{\gamma}_{(i)}} \mathbf{e}_{(i)}$$

which simplifies to

$$\sigma_{w_i}^2 = \sum_{j \neq i} \epsilon_{ij}^2 \frac{1}{m} \frac{N_+}{N_j} \gamma_j \left(1 - \frac{N_j}{N_+} \gamma_j\right) - \frac{1}{m} \sum_{j \neq i} \sum_{q \neq j, i} \epsilon_{ij} \epsilon_{iq} \hat{\gamma}_j \hat{\gamma}_q .$$

Also,

$$\begin{aligned}\text{Cov}(w_i, \hat{\gamma}_i) &= \text{Cov}\left(\sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j, \hat{\gamma}_i\right) \\ &= -\frac{1}{m} \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_i \hat{\gamma}_j\end{aligned}$$

which becomes  $-(1/m) \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j$  under  $H_0 : \gamma_i = 1$ .  
Finally,  $\sigma_{\hat{\gamma}_i}^2 = (N_+/m.N_i)(1 - N_+/N_i) = v_{o,i}$  under the null hypothesis.

The parameter  $\gamma_j$  ( $j \neq i$ ) is not specified under  $H_0 : \gamma_i = 1$ , so it must be estimated under this constraint. It was shown in Appendix 3.1 that

$$\hat{\gamma}_{j.(i)} = \frac{y_j/N_j}{(m - y_i)/(N_+ - N_i)}$$

Substituting this estimate into the expressions for  $\mu_{w_i}$ ,  $\text{Cov}(w_i, \hat{\gamma}_i)$ ,  $\sigma_{\hat{\gamma}_i}^2$  and subsequently substituting into the expression for  $\mu_{w_i|\hat{\gamma}_i}$  gives

$$\begin{aligned}\hat{\mu}_{w_i|\hat{\gamma}_i} &= \left[ \sum_{j \neq i} \epsilon_{ij} \frac{\left(\frac{y_j}{N_j}\right)}{\left(\frac{m - y_i}{N_+ - N_i}\right)} \right] \left[ 1 + \left(\frac{1}{m}\right) \frac{(\hat{\gamma}_i - 1)}{\left(\frac{N_+ - N_i}{m.N_i}\right)} \right] \\ &= \left(\frac{N_+ - N_i}{m - y_i}\right) \sum_{j \neq i} \epsilon_{ij} \left(\frac{y_j}{N_j}\right) \left[ 1 + \left(\frac{N_i}{N_+ - N_i}\right) (\hat{\gamma}_i - 1) \right]\end{aligned}$$

As expected for a conditional mean, each  $\hat{\mu}_{w_i|\hat{\gamma}_i}$  involved in (3.24) depends on the  $\hat{\gamma}_i$  on which we are conditioning. Further simplification of  $\hat{\gamma}_i$  as  $(y_i/N_i)/(m/N_+)$  could express  $\hat{\mu}_{w_i|\hat{\gamma}_i}$  as a function of  $\mathbf{y}$  and  $\mathbf{N}$ . Replacing  $\text{Cov}(w_i, \hat{\gamma}_i)$  by its estimate  $\hat{\text{Cov}}(w_i, \hat{\gamma}_i)$  gives the following expression

$$\hat{\sigma}_{w_i|\hat{\gamma}_i}^2 = \sum_{j \neq i} \epsilon_{ij}^2 \frac{1}{m} \frac{N_+}{N_j} \hat{\gamma}_{ij} \left(1 - \frac{N_j}{N_+} \hat{\gamma}_{ij}\right) - \frac{1}{m} \sum_{j \neq i} \sum_{q \neq j, i} \epsilon_{ij} \epsilon_{iq} \hat{\gamma}_j \hat{\gamma}_q - \frac{[\hat{\text{Cov}}(w_i, \hat{\gamma}_i)]^2}{\sigma_{\hat{\gamma}_i}^2}$$

The third term of this expression becomes

$$\begin{aligned} \frac{\left[ \sum_{j \neq i} \epsilon_{ij} \text{Cov}(\hat{\gamma}_j, \hat{\gamma}_i) \right]^2}{\sigma_{\hat{\gamma}_i}^2} &= \frac{\left[ \sum_{j \neq i} \epsilon_{ij} \left( \frac{-1}{m} \right) \hat{\gamma}_i \hat{\gamma}_j \right]^2}{\sigma_{\hat{\gamma}_i}^2} \\ &= \frac{\hat{\gamma}_i^2}{m^2 \sigma_{\hat{\gamma}_i}^2} \left[ \sum_{j \neq i} \epsilon_{ij}^2 \hat{\gamma}_j^2 + \sum_{j \neq i} \sum_{q \neq j, i} \epsilon_{ij} \epsilon_{iq} \hat{\gamma}_j \hat{\gamma}_q \right]. \end{aligned}$$

Substituting  $\sigma_{\hat{\gamma}_i}^2 = (N_+ - N_i)/m.N_i$  and setting  $\gamma_i = 1$  in the previous line.

$$\begin{aligned} \frac{[\text{Cov}(w_i, \hat{\gamma}_i)]^2}{\sigma_{\hat{\gamma}_i}^2} &= \frac{\frac{1}{m^2} \left[ \sum_{j \neq i} \epsilon_{ij}^2 \hat{\gamma}_j^2 + \sum_{j \neq i} \sum_{q \neq j, i} \epsilon_{ij} \epsilon_{iq} \hat{\gamma}_j \hat{\gamma}_q \right]}{\frac{N_+ - N_i}{m.N_i}} \\ &= \frac{N_i}{m(N_+ - N_i)} \left[ \sum_{j \neq i} \epsilon_{ij}^2 \hat{\gamma}_j^2 + \sum_{j \neq i} \sum_{q \neq j, i} \epsilon_{ij} \epsilon_{iq} \hat{\gamma}_j \hat{\gamma}_q \right]. \end{aligned}$$

Inserting this expression into the third term of the expression for  $\hat{\sigma}_{w_i|\hat{\gamma}_i}^2$ , gives

$$\begin{aligned} \hat{\sigma}_{w_i|\hat{\gamma}_i}^2 &= \sum_{j \neq i} \epsilon_{ij}^2 \frac{1}{m} \frac{N_+}{N_j} \hat{\gamma}_j \left( 1 - \frac{N_j}{N_+} \hat{\gamma}_j \right) - \frac{1}{m} \sum_{j \neq i} \sum_{q \neq j, i} \epsilon_{ij} \epsilon_{iq} \hat{\gamma}_j \hat{\gamma}_q \\ &\quad - \frac{N_i}{m(N_+ - N_i)} \left[ \sum_{j \neq i} \epsilon_{ij}^2 \hat{\gamma}_j^2 + \sum_{j \neq i} \sum_{q \neq j, i} \epsilon_{ij} \epsilon_{iq} \hat{\gamma}_j \hat{\gamma}_q \right] \end{aligned}$$

Substituting  $\hat{\gamma}_{j,(i)} = (y_j/N_j)/[(m - y_i)/(N_+ - N_i)]$  and rearranging gives

$$\hat{\sigma}_{w_i|\hat{\gamma}_i}^2 = \frac{N_+(N_+ - N_i)}{m(m - y_i)} \sum_{j \neq i} \epsilon_{ij}^2 \left( \frac{y_j}{N_j^2} \right) \left[ 1 - \left( \frac{N_+ - N_i}{m - y_i} \right) \frac{y_j}{N_+} \right]$$

$$\begin{aligned}
& -\frac{1}{m} \left( \frac{.N_+ - .N_i}{m - y_i} \right)^2 \left[ \sum_{j \neq i} \sum_{q \neq j, i} \epsilon_{ij} \epsilon_{iq} \frac{y_j y_q}{.N_j .N_q} \right] \left[ 1 - \frac{.N_i}{.N_+ - .N_i} \right] \\
& - \frac{.N_i .N_+^2}{m^2 (.N_+ - .N_i)} \left[ \sum_{j \neq i} \epsilon_{ij}^2 \frac{y_j^2}{.N_j^2} \right]
\end{aligned}$$

which is given in Section 3.3.3.

The usual conditional variance  $X_2|X_1$  for a bivariate normal vector  $(X_1, X_2)$  does not depend on the conditioning variable  $X_1$  and that is still the case for  $\sigma_{w_i|\hat{\tau}_i}^2$ . However, in the re-estimation of  $\hat{\tau}_j$  under  $H_0$  given above, the estimate  $\hat{\sigma}_{w_i|\hat{\tau}_i}^2$  does indeed depend on  $y_i$  and  $.N_i$ . Hence in (3.24)  $\hat{\sigma}_{w_i|\hat{\tau}_i}^2$  changes for each  $y_i$  in the sum.

### Appendix 3.5

#### Additional Results: Complete Homogeneity, $k = 40$ sites

This appendix gives results for complete null simulations for  $k = 40$  sites arranged on a circle of radius one. Given below the table are the probability limits for the acceptable fraction of rejections for the site and overall levels for a simulation of this size.

Circle configuration								
$k = 40, m = 400$					$k = 40, m = 800$			
N's equal		1 $N_j$ different			N's equal		1 $N_j$ different	
$\lambda$	max site rej.	overall rej.	max site rej.	overall rej.	max site rej.	overall rej.	max site rej.	overall rej.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0	.0010	.0252	.0009	.0286	.0010	.0267	.0012	.0335
$\lambda_1$	.0018	.0528	.0014	.0378	.0016	.0513	.0017	.0390
$\lambda_2$	.0018	.0526	.0016	.0503	.0017	.0499	.0017	.0431
$\lambda_3$	.0018	.0536	.0018	.0514	.0018	.0533	.0018	.0511
$\lambda_4$	.0017	.0522	.0018	.0494	.0017	.0497	.0018	.0508
$\lambda_5$	.0018	.0530	.0017	.0502	.0018	.0494	.0019	.0518

max site rej. = maximim fraction of repetitions rejected at site level  
 overall rej. = fraction of repetitions with at least one test rejected  
 (1) through (8) refer to upper probability limits given below

#### Upper probability limit for acceptable fraction rejected

$\lambda > 0$

site level .0020 . overall level .0542

$\lambda = 0$

(1) .0011 (2) .0267 (3) .0012 (4) .0300 (5) .0012 (6) .0297 (7) .0014 (8) .0366

## Appendix 3.6

### Additional results: Single elevated rate

The following three tables present additional results for simulations involving a single elevated rate. The first and second tables involve the same levels of  $\gamma$  and  $\mathbf{N}$  at each of the two spatial configurations (as in Tables 3.3 and 3.4 respectively). However  $m$  is now 400. The third table presents results for  $k = 40$  and  $m = 400$  for the circle configuration. As in the tables of Section 3.6 (\*\*\*) indicates the initial significant increase from  $\lambda = 0$  while (\*) indicates significantly higher from the observed fraction at  $\lambda = 0$  but not the previous  $\lambda$ . The increases in detection ability observed in the third table with  $k = 40$  and  $m = 400$  are very similar to Table 3.3. In the simulation of Table 3.3,  $m/k$  was also 10.



## Ability to Detect Single Elevated Rate

Circle Configuration. 20 sites and 400 cases

$\gamma_{el}$	N's equal		1 $N_j$ different	
	1.48	1.95	1.48	1.95
	frej.	frej.	frej.	frej.
	ftcor.	ftcor.	ftcor.	ftcor.
	fincl.	fincl.	fincl.	fincl.
0	.2210 .1978 .1998	.7860 .7662 .7806	.1048 .0816 .0834	.4510 .4302 .4360
$\lambda_1$	.2248 .1990 .2008	.7782 .7600 .7773	.1174 .0866 .0886	.4554 .4292 .4384
$\lambda_2$	.2298 .2060 .2074	.7870 .7684 .7806	.1194** .0860 .0870	.4496 .4306 .4386
$\lambda_3$	.2332 .2064 .2092	.7882 .7686 .7798	.1210 .0810 .0822	.4602 .4270 .4378
$\lambda_4$	.2206 .1960 .1984	.7870 .7700 .7794	.1296 .0814 .0834	.4710** .4360 .4474
$\lambda_5$	.2232 .1958 .1988	.7858 .7664 .7788	.1290 .0808 .0826	.4652 .4310 .4442

frej. = fraction of repetitions with at least one test rejected

ftcorr. = fraction of repetitions in which only elevated site is declared elevated

fincl. = fraction of repetitions in which the elevated site is among those declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

## Ability to Detect Single Elevated Rate

Circle + 1 Configuration. 20 sites and 400 cases

	N's equal		1 $N_j$ different	
	$\gamma_{el}$			
	1.46	1.90	1.48	1.95
	frej.	frej.	frej.	frej.
	ftcor.	ftcor.	ftcor.	ftcor.
	fincl.	fincl.	fincl.	fincl.
0	.2312	.7774	.1140	.4616
	.2054	.7652	.0894	.4456
	.2086	.7722	.0904	.4472
$\lambda_1$	.2198	.7896	.1292	.4596
	.1926	.7758	.0900	.4338
	.1946	.7824	.0914	.4408
$\lambda_2$	.2354	.7790	.1380**	.4802
	.2108	.7704	.0906	.4470
	.2340	.7752	.0924	.4546
$\lambda_3$	.2354	.7832	.1314**	.4644
	.2096	.7732	.0902	.4346
	.2122	.7794	.0918	.4420
$\lambda_4$	.2258	.7916	.1324**	.4642
	.2076	.7820	.0904	.4340
	.2090	.7874	.0916	.4396
$\lambda_5$	.2320	.7940	.1280	.4692
	.2134	.7832	.0880	.4410
	.2156	.7912	.0896	.4478

frej. = fraction of repetitions with at least one test rejected

ftcorr. = fraction of repetitions in which only elevated site is declared elevated

fincl. = fraction of repetitions in which the elevated site is among those declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

## Ability to Detect Single Elevated Rate

Circle Configuration. 40 sites and 400 cases

$\gamma_{el}$	N's equal		1 $N_j$ different	
	1.48	1.95	1.49	1.975
	frej.	frej.	frej.	frej.
	ftcor.	ftcor.	ftcor.	ftcor.
	fincl.	fincl.	fincl.	fincl.
0	.0612	.3164	.0472	.1530
	.0450	.3012	.0224	.1310
	.0450	.3040	.0228	.1328
$\lambda_1$	.1164**	.4118**	.0538	.1596
	.0698**	.3738**	.0256	.1334
	.0708**	.3824**	.0260	.1352
$\lambda_2$	.1150*	.4272*	.0600	.1532
	.0678*	.3892*	.0292	.1246
	.0694*	.3968*	.0324	.1278
$\lambda_3$	.1168*	.4230*	.0572	.1662
	.0680*	.3816*	.0260	.1262
	.0704*	.3922*	.0274	.1306
$\lambda_4$	.1232*	.4198*	.0570	.1704
	.0776*	.3768*	.0260	.1348
	.0798*	.3896*	.0268	.1378
$\lambda_5$	.1192*	.4262*	.0542	.1668
	.0714*	.3936*	.0258	.1230
	.0732*	.4048*	.0262	.1242

frej. = fraction of repetitions with at least one test rejected

ftcor. = fraction of repetitions in which only elevated site is declared elevated

fincl. = fraction of repetitions in which the elevated site is among those declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

\* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$

### Appendix 3.7

#### Locations and Contiguities: Main Spatial Configurations

The following table lists the locations for  $k = 20$  for the two spatial configurations used in the simulations with fixed  $\lambda$  in Chapter 3 and also in the simulations with data-based choice of  $\lambda$  in Chapter 4.

Spatial configurations, $k = 20$				
	Circle		Circle + 1	
Site	(x,y) Loc	Contig Sites	(x,y) Loc	Contig Sites
1	( 1.95, 1.30 )	2.20	( 1.94, 1.32 )	2.19.20
2	( 1.80, 1.58 )	1.3	( 1.78, 1.61 )	1.3.20
3	( 1.58, 1.80 )	2.4	( 1.54, 1.83 )	2.4.20
4	( 1.30, 1.95 )	3.5	( 1.24, 1.96 )	3.5.20
5	( 1.2 )	4.6	( 0.91, 1.99 )	4.6.20
6	( 0.69, 1.95 )	5.7	( 0.59, 1.91 )	5.7.20
7	( 0.41, 1.80 )	6.8	( 0.32, 1.73 )	6.8.20
8	( 0.19, 1.58 )	7.9	( 0.12, 1.47 )	7.9.20
9	( 0.04, 1.30 )	8.10	( 0.01, 1.16 )	8.10.20
10	( 0, 1 )	9.11	( 0.01, 0.83 )	9.11
11	( 0.04, 0.69 )	10.12	( 0.12, 0.52 )	10.12
12	( 0.19, 0.41 )	11.13	( 0.32, 0.26 )	11.13
13	( 0.41, 0.19 )	12.14	( 0.59, 0.08 )	12.14
14	( 0.69, 0.04 )	13.15	( 0.91, 0.00 )	13.15
15	( 1.0 )	14.16	( 1.24, 0.03 )	14.16
16	( 1.30, 0.04 )	15.17	( 1.54, 0.16 )	15.17
17	( 1.58, 0.19 )	16.18	( 1.78, 0.38 )	16.18
18	( 1.80, 0.41 )	17.19	( 1.94, 0.675 )	17.19
19	( 1.95, 0.69 )	18.20	( 2.1 )	18.20
20	( 2, 1 )	19.1	( 1.5 )	1 to 9.19

## Appendix 3.8

### Additional Results: Two elevated rates

The following tables give additional inference results for two elevated rates.

Table 1 presents results for two noncontiguous elevated rates for  $k = 20$ ,  $m = 200$  in the 'circle + 1' configuration. The magnitude of the elevated rate ratios are the same as the simulation results presented in Table 3.6 and are given again below.

#### equal N

$$A: \gamma_{el1} = \gamma_{el2} = 1.43$$

$$B: \gamma_{el1} = 1.86, \gamma_{el2} = 1.40$$

$$C: \gamma_{el1} = 1.40, \gamma_{el2} = 1.86$$

#### 1 N different

$$D: \gamma_{el1} = \gamma_{el2} = 1.44$$

$$E: \gamma_{el1} = 1.90, \gamma_{el2} = 1.43$$

$$F: \gamma_{el1} = 1.43, \gamma_{el2} = 1.90$$

Table 2 presents results for two elevated rates (contiguous and noncontiguous) in the random population/location configuration with  $k = 20$  and  $m = 200$ . The elevated contiguous incidence rates are sites 3 and 20 while the elevated noncontiguous incidence rates are sites 12 and 20 (see Appendix 3.9 for locations). The elevated rate ratios for Table 2 are as follows

$$A: \gamma_{el1} = \gamma_{el2} = 1.44$$

$$B: \gamma_{el1} = 1.88, \gamma_{el2} = 1.41$$

$$C: \gamma_{el1} = 1.41, \gamma_{el2} = 1.88$$

Table 3 presents results for two elevated rates at contiguous sites on a circle for  $k = 40$  and  $m = 400$ . Table 4 presents results for two elevated noncontiguous rates in the 'circle + 1' configuration for  $k = 40$  and  $m = 400$ . In simulations A through C of Tables 3 and 4, all 40 sites have a population of 1000 while one site has a population of 500 in simulations D through F. One of the two elevated sites has this low population in simulations D through F. The magnitude of the elevated rate ratios given below has been rounded to two decimal places but it is interesting to note that with 40 sites, the overall  $\bar{p}$  is less affected by the low population and the  $\gamma$  vector is now approximately the same in simulations B and E as well as C and F. The elevated rate ratios are

#### equal N

$$A: \gamma_{el1} = \gamma_{el2} = 1.46$$

$$B: \gamma_{el1} = 1.93, \gamma_{el2} = 1.45$$

$$C: \gamma_{el1} = 1.45, \gamma_{el2} = 1.93$$

#### 1 N different

$$D: \gamma_{el1} = \gamma_{el2} = 1.47$$

$$E: \gamma_{el1} = 1.93, \gamma_{el2} = 1.45$$

$$F: \gamma_{el1} = 1.45, \gamma_{el2} = 1.93$$

Table entries and significance markers are as denoted previously.

The results in Table 1 indicate that the initial impact of smoothing on inference is not as large as in these simulations as when the two elevated rates are equal (simulations A and D in Table 3.6). This implies that less smoothing is required when the two elevated rates are very close together and the same. A (\*\*x) in Table 1 such as for  $\lambda_4$  of simulation B, indicates that the observed fraction is the first  $\lambda$  in the set which is significantly different from the observed fraction at  $\lambda = 0$  but it is also

significantly different from the observed fraction at the previous  $\lambda$ . These significant increases for changes in  $\lambda$  other than from  $\lambda = 0$  to  $\lambda = \lambda_1$  are also different from the contiguous rates for this configuration presented in Table 3.6.

The results presented in Table 2 show that the observed rejection and inclusion rates are slightly lower for the noncontiguous rates for the same  $\gamma$ 's. The position of the second elevated site appears to play a role for the unequal elevated rates. Also, the observed fraction totally correct at some positive  $\lambda$  becomes significantly different from the observed fraction totally correct at  $\lambda = 0$  prior to significant increases in other observed fractions in simulation C.

The results for  $k = 40$  in Tables 3 and 4 show that there are slight drops in detection ability with the increase in  $k$  in some situations.

2 Noncontiguous Elevated Rates						
Circle + 1. $k = 20, m = 200$						
$\lambda$	A	B	C	D	E	F
	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.
0	.1060 .0472 .0488 .0012	.3416 .3054 .0398 .0106	.3438 .0394 .3080 .0096	.1160 .0624 .0318 .0014	.4074 .3762 .0312 .0094	.2476 .0580 .1828 .0084
$\lambda_1$	.1544** .0754** .0518 .0026**	.4460** .3972** .0508** .0162	.3750** .0644** .3100 .0128	.1346** .0750** .0420** .0024**	.4068 .3754 .0322 .0120	.2554 .0604 .1872 .0086
$\lambda_2$	.1576* .0828* .0542 .0018*	.4470* .3982* .0454* .0110	.3862* .0674* .3210 .0176**	.1312* .0714* .0416* .0018	.4076 .3792 .0320 .0122	.2570 .0692 .1874 .0084
$\lambda_3$	.1484* .0800* .0452 .0024*	.4518* .3968* .0514* .0148	.4178*x .0678* .3564*x .0192*	.1316* .0800* .0358 .0039*	.4312** .3994** .0326 .0106	.2532 .0736 .1772 .0116
$\lambda_4$	.1626* .0798* .0620** .0026	.4692* .4144* .0658* .0238**	.4622*x .0732* .3984*x .0224*	.1504*x .0836* .0398** .0034*	.4418* .4104* .0310 .0010	.2756 .0794 .1896 .0126
$\lambda_5$	.1806*x .0792* .0818*x .0050*x	.4462* .3900* .0660* .0226*	.4560* .0596* .4042* .0192*	.1586* .0946* .0386 .0032*	.4470 .4114 .0334 .0110	.2640 .0764 .1766 .0122

frej. = fraction of repetitions with at least one test rejected

f(1)in. = fraction of repetitions in which first elevated site is among those declared elevated

f(2)in. = fraction of repetitions in which second elevated site is among those declared elevated

ftcor. = fraction of repetitions in which only elevated sites are declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

\* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$

\*x indicates significant difference from observed fraction at  $\lambda = 0$  and previous  $\lambda$



2 Elevated Rates Random populations/locations						
	contiguous			noncontiguous		
$\lambda$	A	B	C	A	B	C
	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.
0	.1192 .0528 .0502 .0018	.3496 .3048 .0424 .0098	.3372 .0460 .2878 .0092	.1054 .0420 .0514 .0020	.3370 .2946 .0430 .0102	.3250 .0390 .2798 .0106
$\lambda_1$	.1366** .0704** .0409 .0024	.3958** .3522** .0428 .0124	.3454 .0602 .2884 .0128	.1348** .0700** .0380 .0030	.4288** .3794** .0440 .0114	.3270 .0606** .2782 .0128
$\lambda_2$	.1444* .0742* .0546 .0026	.4110** .3664** .0440 .0126	.3388 .0602** .2794 .0132	.1334* .0700* .0402 .0022	.4176* .3688* .0452 .0128	.3306 .0608* .2786 .0134
$\lambda_3$	.1624*x .0894*x .0542 .0032**	.4536* .4100*x .0446 .0156**	.3392 .0734* .2820 .0168**	.1340* .0690* .0400 .0026	.4190* .0792* .0396 .0124	.3282 .0602* .2760 .0132
$\lambda_4$	.1528* .0840* .0528 .0034*	.4450* .4024* .0428 .0138*	.3548** .0756* .2844 .0170*	.1362* .0704* .0410 .0028	.4184* .3762* .0452 .0150	.3284 .0602* .2762 .0130
$\lambda_5$	.1588 .0910 .0544 .0046	.4490 .4080 .0448 .0178	.3518** .0758* .2844 .0178*	.1384* .0698* .0440 .0028	.4248* .3866* .0462 .0158	.3346 .0620* .2780 .0120

frej. = fraction of repetitions with at least one test rejected

f(1)in. = fraction of repetitions in which first elevated site is among those declared elevated

f(2)in. = fraction of repetitions in which second elevated site is among those declared elevated

ftcor. = fraction of repetitions in which only elevated sites are declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

\* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$

\*x indicates significant difference from observed fraction at  $\lambda = 0$  and previous  $\lambda$

2 elevated contiguous rates						
Circle. $k = 40$ . $m = 400$						
$\lambda$	A	B	C	D	E	F
	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.
0	.0962 .0420 .0424 .0018	.3282 .2990 .0278 .0088	.3348 .0336 .2988 .0072	.0938 .0512 .0234 .0006	.3456 .3168 .0184 .0036	.1816 .0446 .1216 .0032
$\lambda_1$	.1478** .0640** .0576** .0024	.4158** .3654** .0540** .0180**	.4132** .0554** .3538** .0152**	.0932 .0526 .0216 .0010	.3710** .3390** .0188 .0060**	.1758 .0452 .1212 .0056
$\lambda_2$	.1606* .0650* .0618* .0026	.4174* .3578* .0620* .0218*	.4118* .0548* .3514* .0180*	.0992 .0568 .0190 .0012	.3848* .3540* .0162 .0058*	.1904 .0476 .1296 .0068
$\lambda_3$	.1658* .0694* .0672* .0030	.4334* .3704* .0604* .0210*	.4310* .0550* .3718* .0188*	.1188** .0670** .0194 .0012	.4140*x .3812*x .0174 .0068	.1888 .0562 .1176 .0058
$\lambda_4$	.1554* .0632* .0644* .0030	.4372* .3790* .0568* .0218*	.4448* .0578* .3856* .0190*	.1300* .0716* .0210 .0018	.4464*x .4106*x .0178 .0064	.2054** .0566 .1294 .0074**
$\lambda_5$	.1562* .0622* .0632* .0026	.4274* .3698* .0566* .0196*	.4290* .0596* .3734* .0216*	.1402* .0792* .0192 .0012	.4430* .4050* .0168 .0064	.2256* .0686** .1308 .0080*

frej. = fraction of repetitions with at least one test rejected

f(1)in. = fraction of repetitions in which first elevated site is among those declared elevated

f(2)in. = fraction of repetitions in which second elevated site is among those declared elevated

ftcor. = fraction of repetitions in which only elevated sites are declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

\* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$

\*x indicates significant difference from observed fraction at  $\lambda = 0$  and previous  $\lambda$

2 elevated noncontiguous rates Circle + 1. k = 40. m = 400						
$\lambda$	A	B	C	D	E	F
	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.	frej. f(1)in. f(2)in. ftcor.
0	.0900 .0420 .0450 .0020	.3288 .2906 .0348 .0064	.3426 .0330 .3082 .0084	.0894 .0478 .0208 .0008	.3430 .3186 .0142 .0036	.1764 .0474 .1188 .0054
$\lambda_1$	.1348** .0644** .0390** .0026	.4118** .3648** .0382 .0100**	.3494 .0552 .3074 .0186	.1124** .0584** .0226 .0010	.3640** .3368** .0170 .0034	.1964** .0536 .1252 .0064
$\lambda_2$	.1376* .0610* .0410* .0020	.4048* .3574* .0354 .0104*	.3544 .0580 .3076 .0172	.1270*x .0744*x .0202 .0016	.4050*x .3706*x .0198** .0070**	.2146*x .0602** .1280 .0058
$\lambda_3$	.1334* .0570* .0422 .0016	.4300*x .3822*x .0364 .0096*	.3610 .0564 .2994 .0174	.1384* .0722* .0218 .0018	.4282*x .3914*x .0174* .0070	.2262* .0706*x .1278 .0082
$\lambda_4$	.1490*x .0704*x .0434 .0020	.4222* .3732* .0430** .0124*	.3576 .0484 .2972 .0168	.1342* .0726* .0234 .0012	.4400*x .3994* .0210* .0068*	.2234* .0668* .1284 .0064
$\lambda_5$	.1314* .0524 .0476 .0024	.4162* .3656* .0496* .0124*	.3540 .0504 .3024 .0164	.1366* .0732* .0216 .0010	.4502* .4132* .0206* .0070*	.2276* .0684* .1358** .0072

frej. = fraction of repetitions with at least one test rejected

f(1)in. = fraction of repetitions in which first elevated site is among those declared elevated

f(2)in. = fraction of repetitions in which second elevated site is among those declared elevated

ftcor. = fraction of repetitions in which only elevated sites are declared elevated

\*\* indicates significant difference from observed fraction at  $\lambda = 0$

\* indicates significant difference from observed fraction at  $\lambda = 0$  but not previous  $\lambda$

\*x indicates significant difference from observed fraction at  $\lambda = 0$  and previous  $\lambda$

### Appendix 3.9

#### Locations and Contiguities: 'Random' Configuration

The following table lists the locations for  $k = 20$  for the random population and location configuration used in simulations of Chapter 3 and Chapter 4.

Random Populations/Locations. $k = 20$ Spatial Configuration and Contiguity			
Site	Population	(x,y) Loc	Contig Sites
1	897	( 5.54. 3.73 )	7.11.14
2	928	( 0.22. 7.53 )	7.10.13
3	722	( 7.34. 3.86 )	5.7.11.20
4	725	( 9.08. 0.37 )	5.20
5	896	( 6.47. 0.42 )	11.15
6	697	( 8.63. 8.89 )	10
7	594	( 4.19. 5.86 )	10.11.13.14.16.19.20
8	598	( 0.60. 5.13 )	13.18.19
9	513	( 1.50. 2.60 )	12.16.17.18.19
10	852	( 8.60. 8.90 )	2.6.7
11	960	( 7.01. 3.39 )	1.3.5.7
12	793	( 0.19. 0.01 )	9.16
13	553	( 0.82. 5.38 )	2.7.8
14	926	( 4.24. 2.47 )	1.7.16
15	556	( 2.83. 1.38 )	5
16	955	( 2.73. 3.17 )	7.9.11.14
17	806	( 0.95. 0.25 )	19
18	906	( 0.37. 4.65 )	9
19	873	( 2.14. 3.85 )	7.9
20	586	( 8.13. 3.65 )	3.7

# Appendix C

## Chapter 4 Appendices

4.1	$\text{Var}(\hat{l}_{red,i})$ . . . . .	181
4.2	Further investigation of CVPL( $\lambda$ ). . . . .	182
4.3	CVPL for Convex Smoothing . . . . .	185

### Appendix 4.1 $\text{Var}(\hat{l}_{red,i})$

The details below for  $\text{Var}(\hat{l}_{red,i})$  are used in the expression for TOTMSE and the solution for  $\lambda_{\min(MSE)}$  given in Section 4.2.

It is more computationally convenient to work with  $\hat{l}_{red,i}$  (first presented in Section 2.9) which ignores higher powers in  $\hat{\gamma}_{ml,i}$ . Further, empirical investigation also has shown a lower TOTMSE for  $\hat{l}_{red,i}$ .  $\text{Var}(\hat{l}_{red,i})$  is of order four in  $\hat{\gamma}_{ml,i}$ . Denoting  $\hat{\gamma}_{ml,i}$  simply as  $\hat{\gamma}_i$ , the expression for  $\text{Var}(\hat{l}_{red,i})$  given in (4.2) is

$$\begin{aligned} \text{Var}(\hat{l}_{red,i}) &= \text{Var} \left[ t_i \left( -b_i \hat{\gamma}_i^2 + \hat{\gamma}_i \sum_{j \neq i} \epsilon_{ij} \hat{\gamma}_j \right) \right] \\ &= t_i^2 \left[ b_i^2 \text{Var}(\hat{\gamma}_i^2) + \sum_{j \neq i} \epsilon_{ij}^2 \text{Var}(\hat{\gamma}_i \hat{\gamma}_j) - 2b_i \sum_{j \neq i} \epsilon_{ij} \text{Cov}(\hat{\gamma}_i^3, \hat{\gamma}_j) \right] . \end{aligned}$$

where

$$t_i = \frac{4}{m/N_+}.$$

$$\text{Var}(\hat{\gamma}_i^2) = E(\hat{\gamma}_i^4) - [E(\hat{\gamma}_i^2)]^2.$$

$$\text{Var}(\hat{\gamma}_i \hat{\gamma}_j) = E(\hat{\gamma}_i^2 \hat{\gamma}_j^2) - [E(\hat{\gamma}_i \hat{\gamma}_j)]^2$$

and

$$\text{Cov}(\hat{\gamma}_i^3, \hat{\gamma}_j) = E(\hat{\gamma}_i^3 \hat{\gamma}_j) - E(\hat{\gamma}_i^3)E(\hat{\gamma}_j).$$

## Appendix 4.2

### Further investigation of $CVPL(\lambda)$

The function  $CVPL(\lambda)$  can be better understood by considering its limits at zero and infinity. As an aid to the investigation of this function, the case vector  $\mathbf{y}$  consisting of  $m$  cases among  $k$  cells can be rearranged such that the first  $y_1$  rows correspond to the cases falling into cell 1 and so on. In Table 4.1.1 below, three cases fell into cell 1 while four fell into cell  $k$ .

**Table 4.1.1**

Alternate arrangement of  $m$  indicator vectors

$i/j$	1	2	.	.	.	$k$
1	1	0	.	.	.	0
2	1	0	.	.	.	0
3	1	0	.	.	.	.
4	.	1	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
$m-4$	.	.	.	.	1	0
$m-3$	.	.	.	.	.	1
$m-2$	.	.	.	.	.	1
$m-1$	.	.	.	.	.	1
$m$	.	.	.	.	.	1

Without smoothing, the cell estimate  $\hat{\theta}_{\lambda,j,(i)}$  in (4.5) is the maximum likelihood

estimate with the deletion of a case from cell  $i$ . For  $\lambda$  at infinity,  $\hat{\tau}_{\lambda,j,(i)} = 1$  and  $\hat{\theta}_{j,(i),\lambda} = N_j/N_+$ . For  $\lambda = 0$ , (4.6) becomes

$$\text{CVPL}(0) = \frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k (\delta_{ij} - \hat{\theta}_{j,(i)})^2. \quad (4.1.1)$$

where  $\hat{\theta}_{j,(i)}$  is the maximum likelihood estimate under a case deletion from cell  $i$  and  $\delta_{ij}$  is the  $j$ th element of an indicator vector such as row  $i$  of Table 4.1.1.

The order of summation in (4.1.1) can now be reversed so that summation is down a column of a table formatted as Table 4.1.1. If one substitutes  $\hat{\theta}_{j,(i)} = (y_j - 1)/(m - 1)$  for table entries with  $j = i$  and  $\hat{\theta}_{j,(i)} = y_j/(m - 1)$  for  $j \neq i$ ,

$$\text{CVPL}(0) = \frac{1}{m} \sum_{r=1}^k \left[ y_r \left( 1 - \frac{y_r - 1}{m - 1} \right)^2 + (m - y_r) \left( \frac{-y_r}{m - 1} \right)^2 \right].$$

Some manipulation shows that

$$\text{CVPL}(0) = \frac{1}{(m - 1)^2} \left( m^2 - \sum_{r=1}^k y_r^2 \right)$$

The relation  $m^2 = (\sum y_i)^2 > \sum y_i^2$  ensures that  $\text{CVPL}(0)$  is positive.

Taking  $\lambda$  to infinity gives  $\hat{\theta}_{j,(i)} = N_j/N_+$  regardless of whether  $j = i$ . Reversing the order of summation as before, then as  $\lambda$  goes to infinity, CVPL tends to

$$\frac{1}{m} \sum_{r=1}^k \left[ y_r \left( 1 - \frac{N_r}{N_+} \right)^2 + \sum_{i=1}^k (m - y_i) \left( -\frac{N_r}{N_+} \right)^2 \right].$$



Some further rearrangement shows that as  $\lambda$  goes to infinity, CVPL tends to

$$\frac{1}{m} \left[ m - \frac{2}{N_+} \sum_{j=1}^k y_j \cdot N_j + m \sum_{j=1}^k N_j^2 / N_+ \right].$$

Assuming equal populations, this limit for CVPL reduces to  $(1 - 1/k)$ . This investigation of  $\text{CVPL}(\lambda)$  prompted further work on predictive loss using the simple convex smoothing prescription introduced in Section 1.4. This material is presented in Appendix 4.3.

### Appendix 4.3

#### CVPL for Convex Smoothing

In this appendix, predictive loss through simple convex smoothing is investigated.

The following function was considered in Section 4.3 as a possible criterion for the choice of smoothing parameter. Including the  $y_i$  case deletions at each site, the predictive loss criterion written in terms of the smoothing parameter  $\lambda$  is

$$\text{CVPL}(\lambda) = \frac{1}{m} \sum_{i=1}^k y_i \sum_{j=1}^k (\delta_{ij} - \hat{\theta}_{j,(i),\lambda})^2.$$

The function can be reconsidered with the simple convex smoothing prescription described in Sections 1.4, 4.3 and Appendix 4.2. For  $j = i$ ,  $\hat{\theta}_{j,(i),a} = (y_j - 1)(1 - a)/(m - 1) + a.N_j/N_+$  while for  $j \neq i$ ,  $\hat{\theta}_{j,(i),a} = y_j(1 - a)/(m - 1) + a.N_j/N_+$ . Substituting in these expressions, the predictive loss function becomes

$$\begin{aligned} \text{CVPL}(a) = \frac{1}{m} \sum_{i=1}^k \left[ y_i \sum_{i=1}^k \left( 1 - (1 - a) \frac{y_i - 1}{m - 1} - a \frac{N_i}{N_+} \right)^2 \right. \\ \left. + \sum_{j \neq i} \left( (1 - a) \frac{y_j}{m - 1} + a \frac{N_j}{N_+} \right)^2 \right]. \end{aligned}$$

This equation can be re-expressed as the quadratic  $\text{CVPL}(a) = Aa^2 + Ba + C$ . The limits of CVPL were once again considered by evaluating this quadratic at  $a = 0$  and  $a = 1$ . These analytic results for convex smoothing were compared to those found for the criterion involving the smoothing parameter  $\lambda$ .

The value of the function at  $a = 0$  is the predictive loss function with no smoothing. The estimate  $\hat{\theta}_{j,(i)}$  is the maximum likelihood estimate with case  $i$  deleted from the observations. We obtain the value  $C$  of the quadratic as

$$C = \frac{1}{m} \sum_{i=1}^k \left[ y_i \left\{ \left( 1 - \frac{y_i - 1}{m - 1} \right)^2 + \sum_{j \neq i} \left( \frac{y_j}{m - 1} \right)^2 \right\} \right].$$

which reduces to

$$\frac{1}{(m - 1)^2} \left( m^2 - \sum_{i=1}^k y_i^2 \right).$$

As expected, this is the value of the predictive loss function obtained previously in Appendix 4.2 for  $\lambda = 0$ .

The loss function can be written as  $CVPL(a) = Aa^2 + Ba + C$ , the first derivative  $2Aa + B$  evaluated at  $a = 0$  will indicate whether the function is increasing or decreasing at  $a = 0$ . The derivative is

$$\begin{aligned} CVPL'(a) &= \frac{2}{m} \sum_{i=1}^k y_i \left( 1 - (1 - a) \frac{y_i - 1}{m - 1} - a \frac{N_i}{N_+} \right) \left( \frac{y_i - 1}{m - 1} - \frac{N_i}{N_+} \right) \\ &+ \frac{2}{m} \sum_{i=1}^k y_i \left[ \sum_{j \neq i} \left( (1 - a) \frac{y_j}{m - 1} + a \frac{N_j}{N_+} \right) \left( \frac{N_j}{N_+} - \frac{y_j}{m - 1} \right) \right] \end{aligned}$$

and  $B = CVPL'(0)$  is

$$\begin{aligned} &\frac{2}{m} \sum_{i=1}^k y_i \left( 1 - \frac{y_i - 1}{m - 1} \right) \left( \frac{y_i - 1}{m - 1} - \frac{N_i}{N_+} \right) \\ &+ \frac{2}{m} \sum_{i=1}^k y_i \left[ \sum_{j \neq i} \left( \frac{y_j}{m - 1} + \right) \left( \frac{N_j}{N_+} - \frac{y_j}{m - 1} \right) \right] \end{aligned}$$

which reduces to

$$B = \frac{2 \left( \sum_{i=1}^k y_i^2 - m^2 \right)}{m(m-1)^2}.$$

The relationship  $(\sum_{i=1}^k y_i)^2 \geq \sum_{i=1}^k y_i^2$  implies that  $CVPL'(0) \leq 0$ . Hence  $CVPL(a)$  is decreasing at zero unless  $\sum_{i=1}^k y_i^2 = m^2$  (which occurs only if all  $m$  cases occur at one site). This result implies that smoothing will almost always be beneficial with respect to predictive loss. Taking the derivative once again gives

$$CVPL''(a) = \frac{2}{m} \sum_{i=1}^k y_i \left( \frac{y_i - 1}{m - 1} - \frac{N_i}{N_+} \right)^2 + \frac{2}{m} \sum_{i=1}^k y_i \left[ \sum_{j \neq i} \left( \frac{N_j}{N_+} - \frac{y_j}{m - 1} \right)^2 \right].$$

Some extensive manipulation gives

$$CVPL''(0) = 2A = \frac{2}{m(m-1)^2 N_+} \left[ m N_+ + 2(m-1)^2 \sum_{i=1}^k y_i N_i - (m-2) N_+ \sum_{i=1}^k y_i^2 + \frac{m(m-1)^2}{N_+} \sum_{i=1}^k N_i^2 \right]$$

Adding the constants A, B and C gives the value of CVPL at  $a = 1$ , the upper limit for  $a$ .

$$CVPL(1) = \frac{1}{(m-1)^2} \left[ \frac{1}{m} \left( m - 2 \sum_{i=1}^k y_i N_i (m-1)^2 / N_+ + (m-2) \sum_{i=1}^k y_i^2 + m(m-1)^2 \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right) \right]$$

$$\begin{aligned}
& + \frac{1}{(m-1)^2} \left[ \frac{2}{m} \left( \sum_{i=1}^k y_i^2 - m^2 \right) \right] \\
& + \frac{1}{(m-1)^2} \left[ m^2 - \sum_{i=1}^k y_i^2 \right].
\end{aligned}$$

Rearrangement and simplification gives

$$\text{CVPL}(1) = 1 + \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 - \frac{2}{m} \frac{\sum_{i=1}^k y_i \cdot N_i}{N_+} .$$

which is the same solution as that obtained for CVPL as  $\lambda$  tends to infinity.

The expectation of  $\text{CVPL}(a) = Aa^2 + Ba + C$  (where A, B, C have been given above) can also be considered as an aid to understanding the predictive loss function. Under complete homogeneity ( $\gamma_i = 1$  for  $i = 1, \dots, k$ ) it can be shown that

$$\begin{aligned}
\mathbf{E}_{H_0}[\text{CVPL}(a)] & = \frac{\left( 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right)}{(m-1)} a^2 + \frac{2}{(m-1)^2} \left[ 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 - m \right] a \\
& + \frac{m}{m-1} \left( 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right) .
\end{aligned}$$

Hence  $\mathbf{E}_{H_0}[\text{CVPL}(a)]$  is the quadratic function  $Da^2 + Ga + H$  where

$$\begin{aligned}
D = \mathbf{E}_{H_0}(A) & = \frac{1}{m-1} \left( 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right) . \\
G = \mathbf{E}_{H_0}(B) & = \frac{2}{(m-1)^2} \left[ 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 - m \right] .
\end{aligned}$$

and

$$H = \mathbf{E}_{H_o}(C) = \frac{m}{m-1} \left( 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right).$$

Denoting  $\mathbf{E}[\text{CVPL}(a)]$  under homogeneity as simply  $\mathbf{E}_o(a)$ , the limits at  $a = 0$  and  $a = 1$  are

$$\mathbf{E}_o(0) = \frac{m}{m-1} \left( 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right)$$

and

$$\mathbf{E}_o(1) = \frac{m}{m-1} \left( 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right) + \frac{\left( 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right)}{m-1} \left( \frac{2}{m-1} + 1 \right) - \frac{2m}{(m-1)^2}.$$

The two limits are equal as  $m$  goes to infinity.

The first derivative of  $\mathbf{E}_o(a)$  at the extremes for  $a$  reveals whether it is increasing or decreasing at these limits.  $\mathbf{E}'_o(a) = 2Da + G$  so

$$\begin{aligned} \mathbf{E}'_o(0) &= G \\ &= \frac{2}{(m-1)^2} \left[ 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 - m \right]. \end{aligned}$$

which is negative because the term inside the square brackets is negative. Hence the expected value of the predictive loss function is decreasing at  $a = 0$ . This result is not unexpected because  $\text{CVPL}(a)$  was shown to be decreasing at  $a = 0$  in general.

$$\begin{aligned} \mathbf{E}'_o(1) &= 2D + G \\ &= \frac{2}{m-1} \left( 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right) \left( 1 + \frac{1}{m-1} \right) - \frac{2m}{(m-1)^2} \\ &= \mathbf{E}'_o(0) + \frac{2}{m-1} \left( 1 - \sum_{i=1}^k \left( \frac{N_i}{N_+} \right)^2 \right). \end{aligned}$$

Under the assumption of equal populations  $E'_\sigma(1)$  is

$$\begin{aligned} & \frac{2}{m-1} \left(1 - \frac{1}{k}\right) \left(1 + \frac{1}{m-1}\right) - \frac{2m}{(m-1)^2} \\ &= \frac{2m}{(m-1)^2} \frac{k-1}{k} - \frac{2m}{(m-1)^2} \\ &= \frac{2m}{(m-1)^2} \left(\frac{-1}{k}\right) \quad . \end{aligned}$$

which is negative. Hence the expectation of the predictive loss criterion under homogeneity and equal populations is decreasing as  $a$  approaches 1. This result supports empirical evidence which revealed that the CVPL function was decreasing under homogeneity and small departures.

# Appendix D

## Chapter 5 Appendices

5.1 Nova Scotia Municipalities: Co-ordinates, Names and Contiguous Sites 192



## Appendix 5.1

### Nova Scotia Municipalities

Co-ordinates, Names and Contiguous Sites

Site	Latitude (x)	Longitude (y)	Municipality Name	Contiguous Sites
1	-64.008	45.864	AMHERST	45.17.37
2	-65.2	44.95	ANNAP.MUN	3.9.34.41.27.18
3	-65.517	44.75	ANNAP.ROYAL	9.34.2
4	-61.75	45.533	ANTIGONIGH.MUN	5.38.40.20
5	-61.917	45.617	ANTIGONISH	4
6	-65.825	43.75	ARGYLE.MUN	54.7.18
7	-65.565	43.575	BARRINGTON.MUN	15.43.6
8	-64.75	45.045	BERWICK	26
9	-65.333	44.833	BRIDGETOWN	34.2
10	-64.5	44.364	BRIDGEWATER	33.31
11	-61	45.364	CANSO	20
12	-60.48	46.05	C.B.MUN	30.56.58.48.50.42
13	-64.25	44.65	CHESTER	31.21
14	-66.167	44.25	CLARE.MUN	18.53
15	-65.625	43.483	CLARKS.HR.	7
16	-63.55	45.909	COLCHESTER.MUN	46.49.17.21.38.23
17	-63.75	45.9	CUMBERLAND.MUN	16.1.36.37.45
18	-65.925	44.5	DIGBY.MUN.	14.19.2.6
19	-65.75	44.636	DIGBY	18
20	-61.5	45.318	GUYSBOROUGH.MUN	4.11.35.47
21	-63.357	44.7	HALIFAX.MUN	55.47.23.31.6
22	-63.75	45.1	HANTS.EAST.MUN	21.23
23	-64.033	45.035	HANTS.WEST.MUN	21.22.24.51.16
24	-64.188	45.068	HANTSPORT	23
25	-61.375	45.591	PORT.HAWKESBURY	40.35.42
26	-64.5	45.091	KENTVILLE	52.27
27	-64.45	45.227	KINGS.MUN	8.26.52.2.23.31
28	-64.75	44.045	LIVERPOOL	41
29	-65.125	43.704	LOCKEPORT	12
30	-60	45.833	LOUISBOURG	10.32.33.13.2.41

Site	Latitude (x)	Longitude (y)	Municipality Name	Contiguous Sites
31	-64.42	44.25	LUNENBURG.MUN	32
32	-64.3	44.3	LUNENBURG	31.33
33	-64.375	44.425	MAHONE.BAY	10.32.32.13.41
34	-65.067	44.917	MIDDLETON	9.3.2
35	-61.375	45.614	MULGRAVE	20
36	-63.84	45.75	OXFORD	45.17
37	-64.313	45.4	PARRSBORO	1.17
38	-62.55	45.515	PICTOU.MUN	39.57.4.20.16
39	-62.688	45.682	PICTOU	38
40	-61.501	46.065	INVERNESS.MUN.	42.12.4
41	-64.6	44.125	QUEENS.MUN	28.31.2.18.43
42	-60.85	45.65	RICHMOND.MUN	25.40.12.35.20
43	-65.4	43.6	SHELBOURNE.MUN	41.42.53
44	-65.313	43.75	SHELBOURNE	29
45	-64.075	45.636	SPRINGHILL	1.36.17
46	-63.344	45.136	STEWIACKE	16
47	-61.9	45.136	STMARYS.MUN	35.21.38
48	-60.333	46.167	SYDNEY	56.57.12
49	-63.281	45.364	TRURO	16
50	-60.75	46.15	VICTORIA.MUN	12.40
51	-64.125	45.167	WINDSOR	23
52	-64.375	45.05	WOLFVILLE	26.27
53	-66.063	43.92	YARMOUTH.MUN	54.6.43.14
54	-66.125	43.84	YARMOUTH	53
55	-63.6588	44.636	HFX.METRO (*)	21
56	-60	46.25	GLACE.BAY (**)	48.12
57	-62.656	45.591	NEW.GLASGOW(***)	38.18
58	-60.5	46.25	NORTH.SYDNEY(****)	48.12

\* Site 55 is an amalgamation of Halifax, Dartmouth, and Bedford

\*\* Site 56 is an amalgamation of Glace Bay, New Waterford, and Dominion

\*\*\* Site 57 is an amalgamation of New Glasgow, Stellarton, Westville and Trenton

\*\*\*\* Site 58 is an amalgamation of North Sydney and Sydney Mines

# Bibliography

- [1] Bernardinelli. L. and Montomoli. C. (1992). Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine* **11**. 983-1007
- [2] Besag. J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B* **36**. 192-236
- [3] Besag. J. and Newell. J. (1991). The detection of clusters in rare diseases. *Journal of the Royal Statistical Society. Series A* **154**. 143-155
- [4] Bishop. Y.M.M., Fienberg. S.E., and Holland P.W. (1975). *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, Massachusetts : MIT Press.
- [5] Bithell. J.F. and Stone R.A. (1989). On statistical methods for analysing the geographical distribution of cancer cases near nuclear installations. *Journal of Epidemiology and Community Health*. **43**. 79-85.
- [6] Clayton. D. (1990). Penalized likelihood methods for mapping disease incidence. Technical Report. University of Leicester.
- [7] Clayton. D. and Bernardinelli. L. (1992). Bayesian methods for mapping disease risk *Geographic and Environmental Epidemiology. Methods for Small Area Studies*. Oxford: Oxford University Press.
- [8] Cliff. A.D. and Ord. J.K. (1981). *Spatial Processes : Models and Applications*. London: Pion.

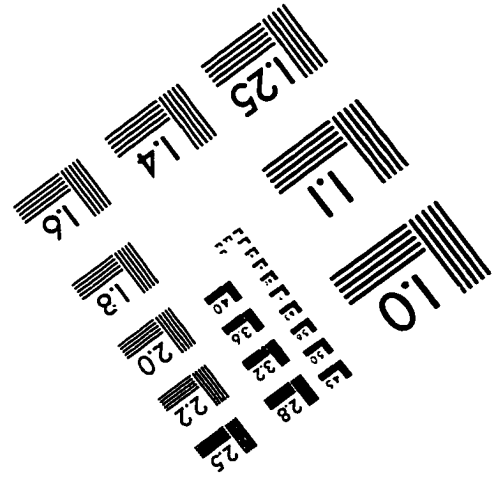
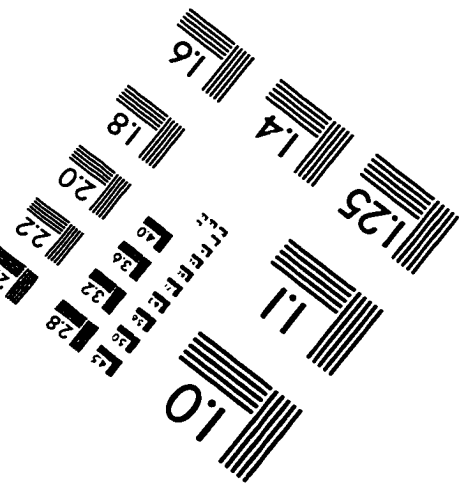
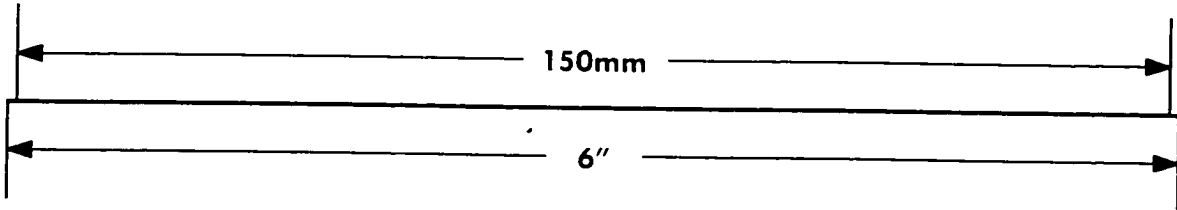
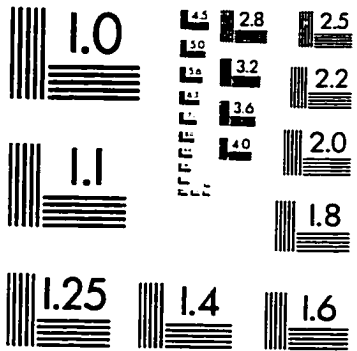
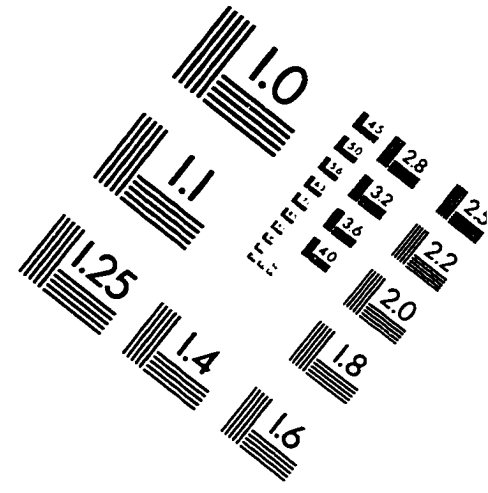
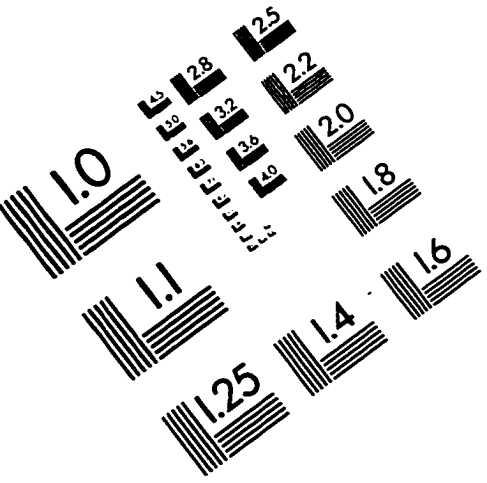
- [9] Cook. D.G. and Pocock. S.J. (1983). Multiple regression in geographical mortality studies. with allowance for spatially correlated errors. *Biometrics* **38**. 361-371.
- [10] Cramer. H. (1970). *Random Variables and Probability Distributions*. London: Cambridge University Press.
- [11] Cressie. N. (1991). *Spatial Analysis*. New York: Wiley.
- [12] Cressie. N. and Chan. N.H. (1989). Spatial modelling of regional variables. *Journal of the American Statistical Association* **84**. 393-401
- [13] Cressie. N. and Read. T.R.C. (1989). Spatial data analysis of regional counts. *Biometrical Journal* **31**. 699-719.
- [14] Cuzick. J. and Edwards. R. (1990). Spatial clustering for inhomogeneous populations. *Journal of the Royal Statistical Society. Series B* **52**. 73-104
- [15] Devine. O., Louis. T.A. and Halloran. M.E. (1994). Empirical Bayes methods for stabilizing incidence rates using empirical Bayes estimators. *Epidemiology* **5**. 622-630.
- [16] Diggle. P.J. (1983). *Statistical analysis of spatial point patterns* New York: Academic Press.
- [17] Downer. R.G. (1996). An introduction to smoothing incidence rates by penalized likelihood. *Statistics in Medicine* **15**. 907-917.
- [18] Ederer. F., Myers. M.H. and Mantel (1964). A statistical problem in space and time: do leukemia cases come in clusters ? *Biometrics* **20**. 626-638.
- [19] Fienberg. S.E. and Holland. P.W. (1973). Simultaneous Estimation of Cell Probabilities. *Journal of the American Statistical Association* **68**. 683-691.
- [20] Geman. S. and Geman D. (1984). Stochastic relaxation. Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**. 721-741

- [21] Gill, P.E. and Murray W. (1976). Minimization subject to bounds on the variables. National Physical Laboratory Report NAC 72.
- [22] Good, I.J. (1965). *The Estimation of Probabilities*. Cambridge, Massachusetts: MIT Press.
- [23] Hills, M. and Alexander, F. (1989). Statistical methods used in assessing the risk of disease near a source of possible environmental pollution: a review. *Journal of the Royal Statistical Society, Series A* **152**, 353-363.
- [24] Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* **55**, 800-802.
- [25] Johnson, N.L. and Kotz, S. (1969). *Discrete Distributions*. New York: Houghton-Mifflin.
- [26] Johnson, N.L. and Kotz, S. (1970). *Continuous Univariate Distributions, Volume 1*. New York: Houghton-Mifflin.
- [27] Kendall, M.G. and Stuart, A. (1957). *The Advanced Theory of Statistics*. New York: Hafner.
- [28] Lehmann, E.L. (1986). *Testing Statistical Hypotheses*. New York: Wiley.
- [29] Leonard, T. (1977). A Bayesian approach to some multinomial and pretesting problems *Journal of the American Statistical Association*, **72**, 869-874
- [30] Mantel, N. (1967). The detection of disease clustering and a generalised regression approach. *Cancer Research* **27**, 209-220.
- [31] Marshall, R.J. (1991). A review of methods for the statistical analysis of spatial patterns of disease *Journal of the Royal Statistical Society, Series A* **154**, 421-441.
-

- [32] Martin. R.L. and Oeppen. J.E. (1975). The identification of regional forecasting models using space-time correlation functions. *Trans. Inst. British Geographers* **66** . 95-118.
- [33] McCullagh. P. and Nelder. J.A. (1989). *Generalized Linear Models* New York: Chapman and Hall.
- [34] Molenaar. W. (1970). *Approximations to the poisson, binomial and hypergeometric distribution functions*. Amsterdam: Mathematical Centre Tracts.
- [35] Naus. J.I. (1965). The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association*. **60**. 532-538.
- [36] Openshaw. S., Craft. A.W., Charlton. M.G. and Birch. J.M. (1988). Investigation of leukemia clusters by use of a geographical analysis machine. *Lancet*, 272-273.
- [37] Peizer. D.B. and Pratt. J.W. (1968). A normal approximation for binomial, F, beta and other common related tail probabilities. *Journal of the American Statistical Association*. **63** 1417-1456.
- [38] Pfeifer P.E. and Deutsch S.J. (1980). A three-stage iterative procedure for space-time modeling. *Technometrics* **22**. 35-47
- [39] Pitman. E.J.G. (1937). The "closest estimates" of statistical parameters. *Proc. Cambridge Philos. Soc.* **33** . 212-222.
- [40] Raubertas. R.F. (1988). Spatial and temporal analysis of disease occurrence for detection of clustering. *Biometrics* **44**. 1121-1129
- [41] Ripley. B.D. (1981). *Spatial Statistics*. New York: Wiley.
- [42] Roberson (1990). Controlling for time-varying population distributions in disease clustering studies. *American Journal of Epidemiology* **132** 131-135.
- [43] Stone. M. (1974). Cross-validation and multinomial prediction. *Biometrika* **61**. 509-515.

- [44] Tango. T. (1984). The detection of disease clustering in time. *Biometrics* **40**.15-26.
- [45] Titterington. D.M. (1985). Common structure of smoothing techniques in statistics. *International Statistical Review* **53** 141-170.
- [46] Tukey. J.W. (1977). *Exploratory Data Analysis* Reading, Massachusetts: Addison-Wesley.
- [47] Wallenstein. S (1980). A test for detection of clustering over time. *American Journal of Epidemiology* **111** 367-372.
- [48] Walter. S.D. (1992). The analysis of regional patterns in health data. *American Journal of Epidemiology* **136** 730-759.
- [49] Walter. S.D. and Birnie. S.E. (1991). Mapping mortality and morbidity patterns: an international comparison. *International Journal of Epidemiology* **20**. 678-689.
- [50] Whittemore. A.S., Friend. N., Brown. B.W. and Holly. E.A. (1987). A test to detect clusters of disease. *Biometrika* **74**. 631-635.

# IMAGE EVALUATION TEST TARGET (QA-3)



**APPLIED IMAGE, Inc**  
 1653 East Main Street  
 Rochester, NY 14609 USA  
 Phone: 716/482-0300  
 Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved