May 2016

Parameter Estimation for the Spatial Ornstein-Uhlenbeck Process with Missing Observations

Sami Cheong
University of Wisconsin-Milwaukee

Parameter Estimation for the Spatial Ornstein-Uhlenbeck process with missing observations

by

Sami Cheong

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Mathematics

at

The University of Wisconsin–Milwaukee

May 2016

ABSTRACT

Parameter Estimation for the Spatial Ornstein-Uhlenbeck process with missing observations

by

Sami Cheong

The University of Wisconsin-Milwaukee, 2016
Under the Supervision of Professor Jugal Ghorai

Suppose we are collecting a set of data on a rectangular sampling grid, it is reasonable to assume that observations (e.g. data that arise in weather forecasting, public health and agriculture) made on each sampling site are spatially correlated. Therefore, when building a model for this type of data, we often pair it with an underlying Gaussian process that contains parameters that correspond to the spatial dependency of the data. Here, we assume that the Gaussian process is characterized by the Ornstein-Uhlenbeck covariance function, which has the property of being both stationary and Markov under the assumption that no observations are missing. However, in reality, the full data assumption may not be a practical one.

In this work, we consider two different scenarios where some observations are missing: 1) a block of observations is missing from the grid and 2) missing observations occur randomly throughout the sampling grid. In each case, we propose an approximate likelihood method to estimate the parameters for the covariance structure. We show that, either by an analytical or a numerical approach, the parameter

estimates from the approximate method have similar properties to those obtained under the full data likelihood function. In particular, we show that the parameter estimators in the missing block case are strongly consistent and asymptotically normal under certain regularity condition, and conclude our work by comparing the results from implementing our methods with simulated data.

Table of Contents

List of Figures

List of Tables

ACKNOWLEDGMENTS

First of all, I would like to express my most sincere gratitiude to Professor Jugal Ghorai. Thank you for being my advisor and offering constant guidance. Your insightful ideas are undoubtedly the most important contribution to the completion of this thesis. Also, thank you Professor Jay Beder, Professor Vytaras Brazauskas, Professor Gabriella Pinter and Professor Chao Zhu for your willingness to serve on my dissertation committee and offer your time and attention on this work.

I would like to thank the Mathematics department at UWM for providing me financial support, a friendly environment and many exceptional opportunities throughout my academic career here.

Finally, thank you to my family and friends for your kindness, encouragement and incredible moral support. You have been there for me all along, and I am forever grateful for that.

Chapter 1

Introduction

1.1 A brief look at spatial statistics

In 1970, geographer Waldo Tobler [18] introduced the concept that "everything is related to everything else, but near things are more related to distant things." This observation, coined the Tobler's First Law of Geography, has taken an important role in the development of spatial analysis, where quantifying the spatial patterns of observations is key to statistical procedures such as experimental design, estimation and prediction. [1] With its roots originating in the mining industry, a spatial model is a stochastic process whose mean and covariance structure are characterized by the distance between observations. Matheron, and later Cressie [2] were among the first to develop the theoretical foundation of spatial statistics. Since then, statistical tools for analyzing and modeling spatially dependant data have been

[1] Although awareness of spatially dependent observations can be traced back as far as the late 17th century, when English astronomer Edmond Halley attempted to map the directions of trade winds and monsoons for voyagers. [2]

generating more interests than ever, thanks to the rise of powerful computing and data storage capabilities. By taking into account the underlying spatial patterns, we can grasp a more accurate description of reality, which in turn allows us to make better decisions using limited data or models.

Example in agriculture

In agriculture, the ability to understand soil properties in a field is an important factor to planting strategies, such as placement of irrigation systems, seed allocations and fertilizer applications, all of which are key to managing yield and quality control of the crops.

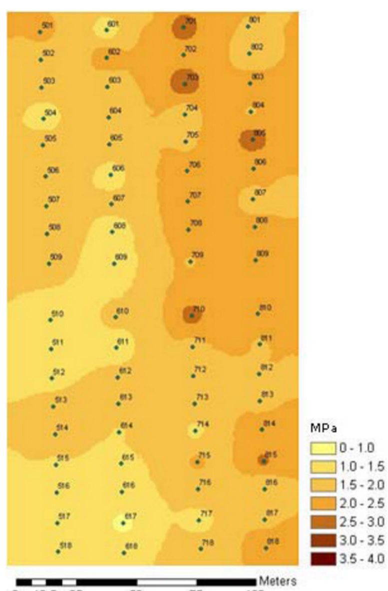


Figure 1.1: An example of spatial data : contour map showing water potential of the soil within a field along with a grid of sampling sites (source of figure: usda.gov)

To understand what pertains to the properties of soil, one can use different soil sampling schemes to analyze the chemical and physical components within a field.

However, data collection can often be time-consuming, expensive and sometimes not possible due to restriction by weather and landscape. Alternatively, one can develop a regression model to predict soil attributes such as moisture content and salinity within a field,

$$\mathbf{Y}(\mathbf{s}) = \mathbf{X}(\mathbf{s})\mathbf{B} + \mathbf{Z}(\mathbf{s}) \tag{1.1}$$

where

- $\mathbf{s}$ is a sampling site of interest
- $\mathbf{Y}(\cdot) \in \mathbb{R}^{n\times 1}$ is a vector of dependent variables
- $\mathbf{X}(\cdot) \in \mathbb{R}^{n\times p}$ is the design matrix containing the independent variables
- $\mathbf{B} \in \mathbb{R}^{p\times 1}$ is a vector of parameters
- $\mathbf{Z}(\cdot) \in \mathbb{R}^{n\times 1}$ is a vector of unobserved, spatially correlated errors affecting the predictions

To address the spatial correlation, $\mathbf{Z}(\cdot)$ is often modeled as a realization of a zero-mean Gaussian process with the covariance matrix being a function of the distance between two samples.

Example in computer experiments

This example serves as a prelude to the main interest of this paper, where lattice data is used in the implementation of computer experiments introduced by Sacks, Schiller and Welch [13] and Sacks, Welch, Mitchell and Wynn [14]. In their experiments, a set of responses from an input grid is modeled as a realization of a stochastic process. Let $S = \{\mathbf{s}_1, \dots \mathbf{s}_N\} \subset \mathbb{R}^d$ be the sample space of all possible computer inputs. Let $X(\mathbf{s})$ be the computer response at the input point $\mathbf{s} \in S$. The set of responses, $\{X(\mathbf{s})\}_{\mathbf{s}\in S}$ is assumed to be a Gaussian random field with the Ornstein-Uhlenbeck covariance function

$$V_q(\sigma^2, \mu, \mathbf{t}, \mathbf{s}) := \sigma^2 \exp\left\{-\sum_{i=1}^{d} \mu_i |t_i - s_i|^q\right\} \tag{1.2}$$

where

- $t = (t_1, \dots t_d)^T, s = (s_1, \dots s_d)^T \in S$ are any two sampled inputs
- $\sigma^2 > 0, \mu = (\mu_1, \dots \mu_d) \in (0, \infty)$ are unknown parameters, and
- $q \in (0, 2]$ is the fixed smoothness parameter of the process $X(\mathbf{s})$

An example of the application of this model is to predict at un-sampled points, which requires estimation of the unknown parameters such as the mean and covariance function.

1.2 The Ornstein-Uhlenbeck (O-U) process

To simplify the problem in the computer experiment example, Ying [24] considered a zero-mean process with dimension $d = 2$, and sampling space $U = [0, 1]^2$. In this model, U is partitioned into an m-by-n grid, with each set of input points being increasing sequences $\{u_j : j = 1, \dots, m\}$ and $\{v_k : k = 1 \dots n\}$ (see Figure 1.2). Let X denote the set of outputs, i.e.

$$X := \{X(u_i, v_j) : i = 1, \dots, m, j = 1, \dots, n\}, \tag{1.3}$$

with $\mathbb{E}\left[X(u_j, v_k)\right] = 0$, and

$$\mathrm{Cov}\left(X\left(u_j, v_k\right), X\left(u_{j'}, v_{k'}\right)\right) = \sigma^2 e^{-\mu\left|u_j - u_{j'}\right| - \lambda\left|v_k - v_{k'}\right|}, \tag{1.4}$$

where $\sigma^2 > 0$, and $(\lambda, \mu) \in [a, b]^2 \subset (0, \infty)^2$. Then X is a two-dimensional Ornstein-Uhlenbeck process with parameters λ, μ and σ^2. The O-U process was originally derived in 1930 as a stochastic process that describes the velocity of a Brownian motion. It is the only Gaussian process that satisfies both the Markov property and stationarity, as shown by Doob in his 1942 paper [4]. Interestingly, the dimension

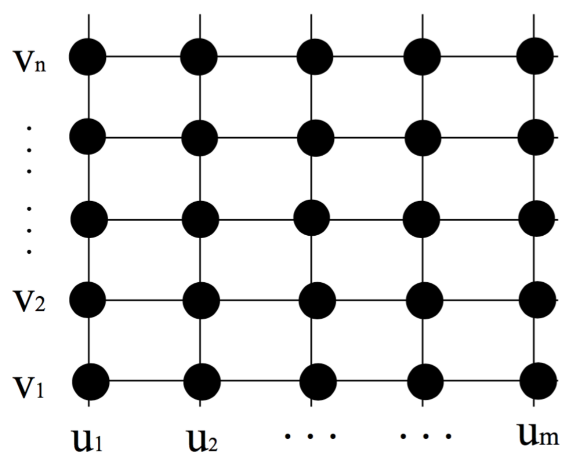


Figure 1.2: An example of a sampling space defined on a rectangular grid (lattice).

of the process plays an important role in the identifiability of the parameters for its covariance structure. In the one-dimensional case, where we have $X(u)$ instead of $X(u, v)$, the probability measure induced by $\sigma_1^2\lambda_1$ is equivalent to that induced by $\sigma_2^2\lambda_2$ if $\sigma_1^2\lambda_1 = \sigma_2^2\lambda_2$. This characteristic of the one-dimensional process raises the issue of identifiabiliy of λ and σ^2, when neither of the parameters is known. In contrast, when the O-U process is at least two-dimensional, as in (1.4), the parameters are all identifiable, as asserted by Ying [24]. Moreover, the Markovian property of X provides an important advantage, in the form of dimension reduction, to derive the asymptotics for the maximum likelihood estimator (MLE) for λ, μ and σ^2.

1.2.1 Parameter estimation using maximum likelihood approach

Recall the random field defined in (1.3). Now, for $j = 1, \ldots, m$ and $k = 1, \ldots, n$, define the following

- $\xi_j = |u_j - u_{j-1}|$ and $\zeta_k = |v_k - v_{k-1}|$
- $a_j = e^{-\lambda \xi_j}$ and $b_k = e^{-\mu \zeta_k}$
- $\underset{\sim}{X}_j = \begin{bmatrix} X(u_j, v_1) \\ X(u_j, v_2) \\ \vdots \\ X(u_j, v_n) \end{bmatrix}, \underset{\sim}{X} = \begin{bmatrix} \underset{\sim}{X}_1 \\ \underset{\sim}{X}_2 \\ \vdots \\ \underset{\sim}{X}_m \end{bmatrix}$

In here, $\underset{\sim}{X}$ is the 'stacked' version of the random field X. As a result,

$$\underset{\sim}{X} \in \mathbb{R}^{mn \times 1} \sim N(0, \sigma^2 A(\lambda) \otimes B(\mu)),$$

where

$$A(\lambda) = \begin{bmatrix} 1 & a_2 & a_2a_3 & \dots & a_2a_3\cdots a_M \\ a_2 & 1 & a_3 & \dots & a_3\cdots a_m \\ \vdots & & \ddots & & \vdots \\ a_2a_3\cdots a_M & a_3a_4\cdots a_M & a_4\cdots a_M & \dots & 1 \end{bmatrix} \tag{1.5}$$

and

$$B(\mu) = \begin{bmatrix} 1 & b_2 & b_2b_3 & \dots & b_2b_3\cdots b_n \\ b_2 & 1 & b_3 & \dots & b_3\cdots b_N \\ \vdots & & \ddots & & \vdots \\ b_2b_3\cdots b_N & b_3b_4\cdots b_N & b_4\cdots b_n & \dots & 1 \end{bmatrix}. \tag{1.6}$$

Notice that, the arrangement of $\underset{\sim}{X}$ at each sampling site (u_j, v_k) can be expressed as a set of observations made on a lattice, shown in Table 1.1.

v_1	$X(u_1, v_1)$	$\cdots$	$X(u_i, v_1)$	$\cdots$	$X(u_j, v_1)$	$\cdots$	$X(u_m, v_1)$
.	.	$\cdots$	.	$\cdots$	.	$\cdots$	.
.	.	$\cdots$	.	$\cdots$	.	$\cdots$	.
.	.	$\cdots$	.	$\cdots$	.	$\cdots$	.
v_k	$X(u_1, v_k)$	$\cdots$	$X(u_i, v_k)$	$\cdots$	$X(u_j, v_k)$	$\cdots$	$X(u_m, v_k)$
.	.	$\cdots$	.	$\cdots$	.	$\cdots$	.
.	.	$\cdots$	.	$\cdots$	.	$\cdots$	.
.	.	$\cdots$	.	$\cdots$	.	$\cdots$	.
v_l	$X(u_1, v_l)$	$\cdots$	$X(u_i, v_l)$	$\cdots$	$X(u_j, v_l)$	$\cdots$	$X(u_m, v_l)$
.	.	$\cdots$	.	$\cdots$	.	$\cdots$	.
.	.	$\cdots$	.	$\cdots$	.	$\cdots$	.
.	.	$\cdots$	.	$\cdots$	.	$\cdots$	.
v_n	$X(u_1, v_n)$	$\cdots$	$X(u_i, v_n)$	$\cdots$	$X(u_j, v_n)$	$\cdots$	$X(u_m, v_n)$
	u_1	$\cdots$	u_i	$\cdots$	u_j	$\cdots$	u_m
	$\underset{\sim}{X}_1$	$\cdots$	$\underset{\sim}{X}_i$	$\cdots$	$\underset{\sim}{X}_j$	$\cdots$	$\underset{\sim}{X}_m$

Table 1.1: A tabular representation of a realization of the OU field with complete data.

Since the covariance matrix of $\underset{\sim}{X}$ is a kronecker product, we have

$$(A(\lambda) \otimes B(\mu)) = A(\lambda)^{-1} \otimes B^{-1}(\mu).$$

Moreover, due to the multiplicative properties of the covariance function, $A(\lambda)^{-1}$ and $B(\mu)^{-1}$ are both tridiagonal, which allows us to express the log likelihood function explicitly in terms of the parameters λ, μ and σ^2. Below we provide a lemma from [24], which can be used to obtain the exact form of $(A(\lambda) \otimes B(\mu))$.

Lemma 1.2.1 (Ying 1993). *Let $\theta > 0$ and $-\infty < s_1 < \cdots < s_r < \infty$. Define the $r \times r$ matrix*

$$G := \begin{bmatrix} 1 & e^{-\theta|s_1-s_2|} & \cdots & e^{-\theta|s_1-s_r|} \\ e^{-\theta|s_2-s_1|} & 1 & \cdot & e^{-\theta|s_2-s_r|} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-\theta|s_r-2_1|} & e^{-\theta|s_r-s_2|} & \cdots & 1 \end{bmatrix},$$

the $s \times 1$ vector

$$g(s) := \begin{bmatrix} e^{-\theta|s-s_1|} \\ \vdots \\ e^{-\theta|s-s_r|} \end{bmatrix}, \text{ where } s \leq s_r,$$

and the $rk \times 1$ vectors

$$f = \begin{bmatrix} f_1 \\ \vdots \\ f_r \end{bmatrix}, \quad h = \begin{bmatrix} h_1 \\ \vdots \\ h_r \end{bmatrix},$$

where for $i = 1, \ldots, r$, f_i and h_i are $k \times 1$ vectors. Then for any $k \times k$ matrix H,

1. $G^{-1}g(s) = \begin{bmatrix} 0 & 0 & \ldots & e^{-\theta(s-s_r)} \end{bmatrix}'$

2. $f'(G \otimes H)^{-1}h = f_1'H^{-1}h_1 + \sum_{i=2}^{r} \frac{\left(f_i - e^{-\theta(s_i-s_{i-1})}f_{i-1}\right)' H^{-1} \left(h_i - e^{-\theta(s_i-s_{i-1})}h_{i-1}\right)'}{1 - e^{-2\theta(s_i-s_{i-1})}}$

3. $\det G = \prod_{i=2}^{r} \left(1 - e^{-2\theta(s_i-s_{i-1})}\right)$

The likelihood function for the complete data is:

$$L(\lambda,\mu,\sigma^2|\underset{\sim}{X}) = (2\pi\sigma^2)^{-mn/2}\left[\det(A(\lambda)\otimes B(\mu))\right]^{1/2}\exp\left\{\frac{-1}{2\sigma^2}\underset{\sim}{X}'(A(\lambda)\otimes B(\mu))^{-1}\underset{\sim}{X}\right\}. \tag{1.7}$$

Let $l(\lambda,\mu,\sigma^2|\underset{\sim}{X}) = -2\ln L(\lambda,\mu,\sigma^2|\underset{\sim}{X})$, then the log-likelihood becomes

$$l(\lambda,\mu,\sigma^2|\underset{\sim}{X}) = mn\ln(2\pi\sigma^2) + \ln[\det(A(\lambda)\otimes B(\mu))] + \frac{1}{\sigma^2}\underset{\sim}{X}'(A(\lambda)\otimes B(\mu))^{-1}\underset{\sim}{X}. \tag{1.8}$$

Let $\bar{a}_{ij}$ be the ij^{th} element of $A(\lambda)^{-1}$, and $\bar{b}_{ij}$ be the ij^{th} element of $B^{-1}(\mu)$. Recall that since $A(\lambda)^{-1}$ and $B(\mu)^{-1}$ are tridiagonal, we have by lemma (1.2.1) :

- $\bar{a}_{11} = \dfrac{1}{1-a_2^2}, \bar{a}_{mm} = \dfrac{1}{1-a_m^2}; \bar{b}_{11} = \dfrac{1}{1-b_2^2}, \bar{b}_{nn} = \dfrac{1}{1-b_n^2}$

  For $j = 2,\ldots,m; k = 2,\ldots,n$:

- $\bar{a}_{jj} = \dfrac{1}{1-a_j^2} + \dfrac{1}{1-a_{j+1}^2} - 1, \bar{b}_{kk} = \dfrac{1}{1-b_k^2} + \dfrac{1}{1-b_{k+1}^2} - 1$
- $\bar{a}_{jj-1} = \dfrac{-a_j}{1-a_j^2}; \bar{b}_{kk-1} = \dfrac{-b_k}{1-b_j^2}$
- $\bar{a}_{ij} = 0$ and $\bar{b}_{ij} = 0$ if $|i-j| > 1$

On the other hand, since the O-U process satisfies the Markov property, we can express the joint distribution of $\underset{\sim}{X}$ as

$$f\left(\underset{\sim}{X}\right) = f\left(\underset{\sim}{X}_1\right)\prod_{j=2}^{m} f\left(\underset{\sim}{X}_j|\underset{\sim}{X}_{j-1}\right). \tag{1.9}$$

By direct calculation, we have

$$\underset{\sim}{X}_1 \sim N_n\left(\mathbf{0}, \sigma^2 B(\mu)\right), \text{ and } \underset{\sim}{X}_j|\underset{\sim}{X}_{j-1} \sim N_n\left(e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}, \sigma^2(1-e^{-2\lambda\xi_j})B(\mu)\right), \tag{1.10}$$

from which we can derive a representation of (1.8) as

$$\begin{aligned} l(\lambda, \mu, \sigma^2|\underset{\sim}{X}) = {} & mn\ln(2\pi\sigma^2) \\ & + n\sum_{j=2}^{m}\ln(1-e^{-2\lambda\xi_j}) + m\sum_{k=2}^{n}\ln(1-e^{-2\mu\eta_k}) \\ & + \frac{1}{\sigma^2}\left[\underset{\sim}{X}_1' B^{-1}(\mu)\underset{\sim}{X}_1 + \sum_{j=2}^{m}\frac{(\underset{\sim}{X}_j - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1})' B^{-1}(\mu)(\underset{\sim}{X}_j - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1})}{1-e^{-2\lambda\xi_j}}\right]. \end{aligned} \tag{1.11}$$

1.2.2 Properties of the MLE given complete observations

Ying [24] has shown that the MLE's for λ, μ and σ^2 derived from (1.11) are strongly consistent, that is, if λ_0, μ_0 and σ_0^2 are the true parameters for the random field X, then the MLE's $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\sigma}^2$ will converge to the true values almost surely, as $m, n \to \infty$. In particular, when the spacing of the sampling grid follows a certain regularity condition, the MLE's are asymptotically normal. Also, under the same regularity assumption, we have that, asymptotically, $\hat{\lambda} - \lambda_0$ and $\hat{\mu} - \mu_0$ are uncorrelated, which implies independece in the normal case. As a result, the vertical partition can be arragned independently of the horizontal partition, allowing freedom in designing the sampling scheme.

However, in the practical point of view, missing observations are often unavoidable due to many factors, from physical constraint to human error. Therefore, we are interested in investigating the properties of the estimates under the assumption that some observations are missing. In the rest of this paper, we will investigate methods to estimate the parameters λ, μ and σ^2 when we no longer have the complete lattice assumption that was made in [24]. We begin with defining the patterns of the missing observations. Then, we present our proposed methods to estimate the parameters, followed by an investigation on the properties of the estimates as well as implementation based on numerical examples using simulated data.

Chapter 2

Parameter Estimation in the Presence of Missing Data

We live in a time where data is ubiquitous. From personal exercise records generated from mobile devices to coffee preferences arranged by zipcodes, data availability is a trend that continues to spread as tools for collecting, storing and visualizing data are more accessible and cost-effective than ever. Consequently, situations where one has to deal with missing data are occuring more frequently.

A large body of litereture has been developed on statistical inference with respect to missing data, with some of the most notable methods being the expectation-maximization (EM) algorithm [3], which is an iterative approach that repeatedly updates the parameter estimates of a model based on its conditional likelihood given the available data, and multiple-imputation (MI), which employs many different modeling procedures (such as regression) to produce multiple values to fill in the missing observations, after which the full data estimation method can be applied.

The advantages of these inference methods are the ease of implementation, and, in the case of the EM algorithm, a guarentee of convergence to the MLE. However, convergence of the EM estimates can be extremely slow, and, on the other hand, MI may introduce unwanted bias due to the variance from drawing multiple simulations. In here, we focus on building parameter estimation methods that do not rely on iterative steps or imputations, as in the EM algorithm and the MI approach. Rather, we seek to use the available information we have and define inference functions where parameter estimates can be derived in a similar manner as the MLE in the full-data case.

2.1 Patterns of missing data : missing data block and randomly missing data

There are many scenarios in which a block of observations can be absent in a dataset. For example, suppose we want to model sea surface temperature based on data collected in a sea area where there is an island. This island serves as an origin of a missing block in the resulting dataset, as sea surface temperature is clearly inacessible when the sampling site is away from water. Another example would be data collected from air monitoring stations in a metro area. In this case, a missing block can occur due to a power outage in a small region, which prevents the equipments from recording the observations. On the other hand, suppose we wish to collect income data from each house in a geographical area, then the missing observations may follow a random pattern due to factors such as non-response or human error. In either of these examples, properties of the MLE's for λ, μ and σ^2 as defined by [24] in the complete data case may no longer be valid, as the missing sampling sites will impose many new restrictions in computing the likelihood function.

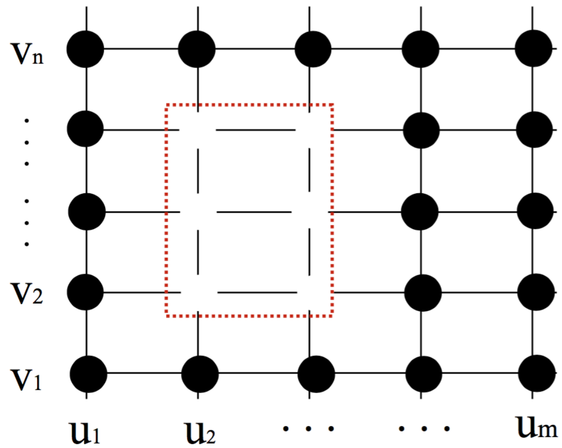


Figure 2.1: An illustration of a missing block in a rectangular sampling grid.

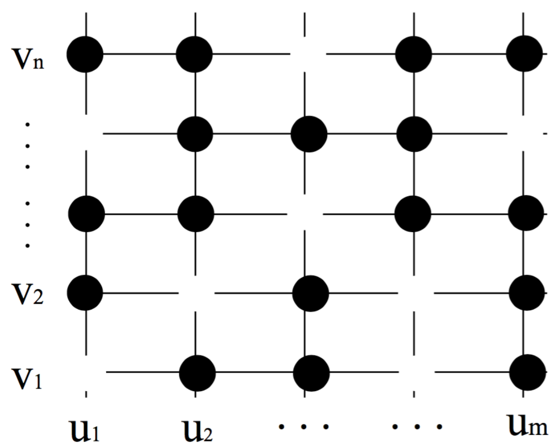


Figure 2.2: An illustration of randomly missing sampling sites in a rectangular sampling grid.

In this chapter, assuming the data is modeled as an O-U process with covariance parameters λ, μ and σ^2. We propose, for each missing data scenario, a method of estimating the parameters via an approximation of the likelihood function. This idea is inspired by the work of Besag [1] and Vecchia [20], who suggested approximating the likelihood function by conditioning on a selected number of neighboring observations. This means that, if $Z \sim N(\mu(\theta), \Sigma(\theta))$ is a stochastic process observed on sites $l = 1, \ldots, K$, then the full likelihood can be approximated by

$$l(\theta) = f(Z_1|\theta) \prod_{l=2}^{K} f(Z_l|Z_{l-1}; \theta) \approx f(Z_1|\theta) \prod_{l=2}^{K} f(Z_l|Z_{l-1}^{*}; \theta),$$

where Z_l^*, known as the conditioning vector, is a subvector of Z_l chosen to simplify the computation of the likelihood function.

2.2 Approximated likelihood estimation for missing data block

Suppose some observations in a realization of an O-U process are missing in a rectangular grid formed by columns $\{m_1, m_1+1, \ldots, m_2-1, m_2\}$ and rows $\{n_1, n_1+1, \ldots, n_2-1, n_2\}$, where $1 < m_1 < m_2 < m$ and $1 < n_1 < n_2 < n$. Then for $j = m_1, \ldots, m_2$, we express the observation columns in three parts:

$$\underset{\sim}{X}_j = \begin{bmatrix} \underset{\sim}{X}_j^{(1)} \\ \underset{\sim}{X}_j^{(2)} \\ \underset{\sim}{X}_j^{(3)} \end{bmatrix},$$

$$\text{with } \underset{\sim}{X}_j^{(1)} = \begin{bmatrix} X(u_j, v_1) \\ \vdots \\ X(u_j, v_{n_1-1}) \end{bmatrix}; \underset{\sim}{X}_j^{(2)} = \underbrace{\begin{bmatrix} X(u_j, v_{n_1}) \\ \vdots \\ X(u_j, v_{n_2}) \end{bmatrix}}_{\text{missing data}}; \underset{\sim}{X}_j^{(3)} = \begin{bmatrix} X(u_j, v_{n_2+1}) \\ \vdots \\ X(u_j, v_n) \end{bmatrix}.$$

Notice that the partition of an observation vector corresponds to the covariance matrix for $\underset{\sim}{X}_j$, in particular, let M be an arbitrary matrix, and denote $M[a:b, c:d]$ to be a submatrix of M formed by rows a to b and columns c to d. Let

- $B_{11} = B\left[1 : n_1 - 1, 1 : n_1 - 1\right], B_{22} = B[n_1 : n_2, n_1 : n_2]$,
  $B_{33} = B[n_2 + 1 : n, n_2 + 1 : n]$
- $B_{12} = B[1 : n_1 - 1, n_1 : n_2], B_{13} = B[1 : n_1 - 1, n_2 + 1 : n]$
- $B_{23} = B[n_1 : n_2, n_2 + 1 : n]$
- $B_{31} = B'_{13}, B_{32} = B'_{23}$,

so that

$$B(\mu) = \left[\begin{array}{c|c|c} B_{11} & B_{12} & B_{13} \\ \hline B_{21} & B_{22} & B_{23} \\ \hline B_{31} & B_{32} & B_{33} \end{array}\right].$$

In other words, each partition $B_{s,s'}$ is expressed by $\text{Cov}\left(\underset{\sim}{X}_j^{(s)} \underset{\sim}{X}_j^{(s')}\right)$. Moreover, $B_{s,s'}^{-1}$ has a tridiagonal form if $s = s'$.

	u_1	$\cdot\cdot$	u_{m_1-1}	$u_{m_1} \cdot\cdot u_{m_2}$	u_{m_2+1}	$\cdot\cdot$	u_m
v_1	$X(u_1, v_1)$	$\cdot\cdot$	$X(u_{m_1-1}, v_1)$	$X(u_{m_1}, v_1) \cdot\cdot X(u_{m_2}, v_1)$	$X(u_{m_2+1}, v_1)$	$\cdot\cdot$	$X(u_m, v_1)$
.	.	$\cdot\cdot$	.	$\cdot\cdot$	.	$\cdot\cdot$	.
.	.	$\cdot\cdot$	.	$\cdot\cdot$	.	$\cdot\cdot$	.
.	.	$\cdot\cdot$	.	$\cdot\cdot$	.	$\cdot\cdot$	.
v_{n_1-1}	$X(u_1, v_{n_1-1})$	$\cdot\cdot$	$X(u_{m_1-1}, v_{n_1-1})$	$X(u_{m_1}, v_{n_1-1}) \cdot\cdot X(u_{m_2}, v_{n_1-1})$	$X(u_{m_2+1}, v_{n_1-1})$	$\cdot\cdot$	$X(u_m, v_{n_1-1})$
v_{n_1}	$X(u_1, v_{n_1})$	$\cdot\cdot$	$X(u_{m_1-1}, v_{n_1})$	$X(u_{m_1}, v_{n_1}) \cdot\cdot X(u_{m_2}, v_{n_1})$	$X(u_{m_2+1}, v_{n_1})$	$\cdot\cdot$	$X(u_m, v_{n_1})$
.	.	$\cdot\cdot$	.	$\cdot\cdot$	.	$\cdot\cdot$	.
.	.	$\cdot\cdot$	.	$\cdot\cdot$	.	$\cdot\cdot$	.
.	.	$\cdot\cdot$	.	$\cdot\cdot$	.	$\cdot\cdot$	.
v_{n_2}	$X(u_1, v_{n_2})$	$\cdot\cdot$	$X(u_{m_1-1}, v_{n_2})$	$X(u_{m_1}, v_{n_2}) \cdot\cdot X(u_{m_2}, v_{n_2})$	$X(u_{m_2+1}, v_{n_2})$	$\cdot\cdot$	$X(u_m, v_{n_2})$
v_{n_2+1}	$X(u_1, v_{n_2+1})$	$\cdot\cdot$	$X(u_{m_1-1}, v_{n_2+1})$	$X(u_{m_1}, v_{n_2+1}) \cdot\cdot X(u_{m_2}, v_{n_2+1})$	$X(u_{m_2+1}, v_{n_2+1})$	$\cdot\cdot$	$X(u_m, v_{n_2+1})$
.	.	$\cdot\cdot$	.	$\cdot\cdot$	.	$\cdot\cdot$	.
.	.	$\cdot\cdot$	.	$\cdot\cdot$	.	$\cdot\cdot$	.
.	.	$\cdot\cdot$	.	$\cdot\cdot$	.	$\cdot\cdot$	.
v_n	$X(u_1, v_n)$	$\cdot\cdot$	$X(u_{m_1-1}, v_n)$	$X(u_{m_1}, v_n) \cdot\cdot X(u_{m_2}, v_n)$	$X(u_{m_2+1}, v_n)$	$\cdot\cdot$	$X(u_m, v_n)$
	u_1	$\cdot\cdot$	u_{m_1-1}	$u_{m_1} \cdot\cdot u_{m_2}$	u_{m_2+1}	$\cdot\cdot$	u_m
	$\underset{\sim}{X}_1$	$\cdot\cdot$	$\underset{\sim}{X}_{m_1-1}$	$\underset{\sim}{X}_{m_1} \cdot\cdot \underset{\sim}{X}_{m_2}$	$\underset{\sim}{X}_{m_2+1}$	$\cdot\cdot$	$\underset{\sim}{X}_m$

Table 2.1: A tabular representation of $X^{(o)}$ with missing observations (indicated in red) in a rectangular pattern.

For $j = m_1 - 1, \dots, m_2 + 1$, define

$$\underset{\sim}{X}_j^* = \begin{bmatrix} \underset{\sim}{X}_j^{(1)} \\ \underset{\sim}{X}_j^{(3)} \end{bmatrix} \tag{2.1}$$

to be the vector of remaining observations from each column. Then $\underset{\sim}{X}_j^* \sim N(\mathbf{0}, \sigma^2 B^*(\mu))$, where

$$B^*(\mu) = \begin{bmatrix} B_{11} & B_{13} \\ B_{31} & B_{33} \end{bmatrix}, \tag{2.2}$$

and $(B^*(\mu))^{-1}$ is a tridiagonal matrix with entries in similar form as those in $B^{-1}(\mu)$, except for the $n_1 - 1^{th}$ row and the row immediately after that. By direct calculation, we get $\underset{\sim}{X}_j^* | \underset{\sim}{X}_{j-1}^* \sim N\left(e^{-\lambda \xi_j} \underset{\sim}{X}_{j-1}^*, \sigma^2 (1 - e^{-2\lambda \xi_j}) B^*(\mu)\right)$. On the other hand, define

- $\begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{13} \\ \tilde{B}_{31} & \tilde{B}_{33} \end{bmatrix} = \begin{bmatrix} B_{11} & B_{13} \\ B_{31} & B_{33} \end{bmatrix}^{-1}$

- $H = B_{21}\tilde{B}_{11} + B_{23}\tilde{B}_{31}$

- $J = B_{21}\tilde{B}_{13} + B_{23}\tilde{B}_{33}$

Then, for $\underset{\sim}{X}_j|\underset{\sim}{X}^*_{j-1}$, where $\underset{\sim}{X}_j$ and $\underset{\sim}{X}^*_{j-1}$ are column observations bordering the missing block, we have

$$\begin{aligned} \mathbb{E}[\underset{\sim}{X}_j|\underset{\sim}{X}^*_{j-1}] &= e^{-\lambda\xi_j} \begin{bmatrix} B_{11} & B_{13} \\ B_{21} & B_{23} \\ B_{31} & B_{33} \end{bmatrix} \cdot \begin{bmatrix} \tilde{B}_{11} & \tilde{B}_{13} \\ \tilde{B}_{31} & \tilde{B}_{33} \end{bmatrix} \\ &= e^{-\lambda\xi_j} \begin{bmatrix} I_{n_1-1} & 0_{n_1-1\times n-n_2} \\ H & J \\ 0_{n-n_2\times n_1-1} & I_{n-n_2} \end{bmatrix} \cdot \underset{\sim}{X}^*_j \\ &= e^{-\lambda\xi_j} \begin{bmatrix} \underset{\sim}{X}^{(1)}_j \\ H\underset{\sim}{X}^{(1)}_j + J\underset{\sim}{X}^{(3)}_j \\ \underset{\sim}{X}^{(3)}_j \end{bmatrix}, \end{aligned} \tag{2.3}$$

and

$$
\operatorname{Cov}[\underset{\sim}{X}_j|\underset{\sim}{X}^*_{j-1}] = \sigma^2 \left(B(\mu) - e^{-2\lambda\xi_j} \begin{bmatrix} I_{n_1-1} & 0_{n_1-1\times n-n_2} \\ H & J \\ 0_{n-n_2\times n_1-1} & I_{n-n_2} \end{bmatrix} \begin{bmatrix} B_{11} & B'_{21} & B'_{31} \\ B'_{13} & B'_{23} & B_{33} \end{bmatrix} \right)
$$

$$
= \sigma^2 \left(B(\mu) - e^{-2\lambda\xi_j} \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ HB_{11}+JB_{31} & HB_{12}+JB_{32} & HB_{13}+JB_{33} \\ B_{31} & B_{32} & B_{33} \end{bmatrix} \right). \tag{2.4}
$$

From (2.4), we see that, unlike $B^{-1}(\mu)$, the inverse matrix for $\operatorname{Cov}[\underset{\sim}{X}_i|\underset{\sim}{X}^*_j]$ can no longer be expressed explicitly. As a result, we consider only conditional variables in the form of $\underset{\sim}{X}_i|\underset{\sim}{X}_{i-1}$ and $\underset{\sim}{X}^*_j|\underset{\sim}{X}^*_{j-1}$ when computing the approximated likelihood function. Denote

$$
X^{(o)} := \{\underset{\sim}{X}_1, \ldots, \underset{\sim}{X}_{m_1-1}, \underset{\sim}{X}^*_{m_1}, \ldots, \underset{\sim}{X}^*_{m_2+1}, \underset{\sim}{X}_{m_2}, \ldots \underset{\sim}{X}_m\} \tag{2.5}
$$

to be the set of available observations from the O-U field. Let

- $K_o = \{2, 3, \ldots n_1 - 1, n_2 + 1, \ldots, n\}$,
- $J_o = \{2, 3, \ldots, m_1 - 1, m_2 + 2, \ldots, m\}$,
- $\zeta^*_{n_2+1} = |v_{n_2+1} - v_{n_1-1}|$
- $\xi^*_{m_2+2} = |u_{m_2+2} - u_{m_1-1}|$

and define

$$L(\lambda,\mu,\sigma^2|X^{(o)}) = f\left(\underset{\sim}{X}_1\right)\prod_{j=2}^{m_1-1} f\left(\underset{\sim}{X}_j|\underset{\sim}{X}_{j-1}\right)\prod_{j=m_1}^{m_2+1} f\left(\underset{\sim}{X}_j^*|\underset{\sim}{X}_{j-1}^*\right)\prod_{j=m_2+2}^{m} f\left(\underset{\sim}{X}_j|\underset{\sim}{X}_{j-1}\right).$$

Notice that

$$-2\ln f\left(\underset{\sim}{X}_1\right) = n\ln\left(2\pi\sigma^2\right) + \ln\det\left(B(\mu)\right) + \frac{1}{\sigma^2}\underset{\sim}{X}_1' B^{-1}\left(\mu\right)\underset{\sim}{X}_1,$$

for $j = 2,\ldots,m_1-1, m_2+1,\ldots,m,$

$$-2\ln f(\underset{\sim}{X}_j|\underset{\sim}{X}_{j-1}) = n\ln\left(2\pi\sigma^2\right) + \ln\left(1-e^{-2\lambda\xi_j}\right) + \sum_{k=2}^{n}\ln\left(1-e^{-2\mu\zeta_k}\right)$$

$$+\frac{(\underset{\sim}{X}_j - e^{-2\lambda\xi_j}\underset{\sim}{X}_{j-1})' B^{-1}\left(\mu\right)(\underset{\sim}{X}_j - e^{-2\lambda\xi_j}\underset{\sim}{X}_{j-1})}{\sigma^2(1-e^{2-\lambda\xi_j})},$$

and for $j = m_1, m_1+1,\ldots m_2+1,$

$$-2\ln f(\underset{\sim}{X}_j^*|\underset{\sim}{X}_{j-1}^*) = (n-(n_2-n_1+1))(\ln(2\pi\sigma^2)+\ln(1-e^{-2\lambda\xi_j}))$$

$$+\sum_{k\in K_o}\ln(1-e^{-2\mu\zeta_k}) + \ln(1-e^{-2\mu\zeta^*_{n_2+1}})$$

$$+\frac{(\underset{\sim}{X}_j - e^{-2\lambda\xi_j}\underset{\sim}{X}_{j-1})'(B^*(\mu))^{-1}(\underset{\sim}{X}_j^* - e^{-2\lambda\xi_j}\underset{\sim}{X}_{j-1}^*)}{\sigma^2(1-e^{2-\lambda\xi_j})}.$$

Then the approximated likelihood function has the form

$$l\left(\lambda,\mu,\sigma^2|X^{(o)}\right) = -2\ln L\left(\lambda,\mu,\sigma^2|X^{(o)}\right) \tag{2.6}$$

$$= [nm-(m_2-m_1+2)(n_2-n_1+1)]\ln\left(2\pi\sigma^2\right)$$

$$+ n\sum_{j\in J_o}\ln\left(1-e^{-2\lambda\xi_j}\right) + [n-(n_2-n_1+1)]\sum_{j=m_1}^{m_2+1}\ln\left(1-e^{-2\lambda\xi_j}\right)$$

$$+ [m-(m_2-m_1+2)]\sum_{k=2}^{n}\ln\left(1-e^{-2\mu\zeta_k}\right)$$

$$+ (m_2-m_1+2)\left[\sum_{k\in K_o}\ln\left(1-e^{-2\mu\zeta_k}\right)+\ln\left(1-e^{-2\mu\zeta^*_{n_2+1}}\right)\right]$$

$$+ \frac{1}{\sigma^2}\left[\underset{\sim}{X}'_1 B^{-1}(\mu)X_1\right.$$

$$+ \sum_{j\in J_o}\frac{\left(\underset{\sim}{X}_j-e^{-2\lambda\xi_j}\underset{\sim}{X}_{j-1}\right)' B^{-1}(\mu)\left(\underset{\sim}{X}_j-e^{-2\lambda\xi_j}\underset{\sim}{X}_{j-1}\right)}{1-e^{-2\lambda\xi_j}}$$

$$\left. + \sum_{j=m_1}^{m_2+1}\frac{\left(\underset{\sim}{X}^*_j-e^{-2\lambda\xi_j}\underset{\sim}{X}^*_{j-1}\right)'(B^*(\mu))^{-1}\left(\underset{\sim}{X}^*_j-e^{-2\lambda\xi_j}\underset{\sim}{X}^*_{j-1}\right)}{1-e^{-2\lambda\xi_j}}\right]. \tag{2.7}$$

Consequently, the maximum likelihood estimators $\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2$ for $L((\lambda, \mu, \sigma^2|X^{(o)})$ are defined as the solution to

$$\frac{\partial}{\partial\lambda} l(\lambda, \mu, \sigma^2|X^{(o)}) = 0$$
$$\frac{\partial}{\partial\mu} l(\lambda, \mu, \sigma^2|X^{(o)}) = 0$$
$$\frac{\partial}{\partial\sigma^2} l(\lambda, \mu, \sigma^2|X^{(o)}) = 0.$$

Finding the ML estimators for $\hat{\lambda}$ and $\hat{\mu}$ explicitly may not be possible due to the non-linearity of the equations. In order to study the asymptotic property of the estimators, we rely on approximation techniques that utilize transformation of correlated random variables to independent ones. By expressing the quadratic forms in (2.6) using combinations of independent standard normal and chi-squared random variables, we can draw some conclusion on the consistency of the estimator through the behavior of

$$l(\lambda, \mu, \sigma^2|X^{(o)}) - l(\lambda_0, \mu_0, \sigma_0^2|X^{(o)})$$

as $m, n \to \infty$.

2.2.1 Asymptotic results

In this section we present the main theorems that describe the asymptotic properties of $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ estimated from (2.6).

Theorem 2.2.1 (Strong Consistency)**.** *Let $C \subset \mathbb{R}^2_+$ be a compact subspace, and let λ_0, μ_0 and σ_0^2 denote the true parameters for the random field X with joint density defined in* (1.9) *and* (1.10)*, where $(\lambda_0, \mu_0) \in C$ and $\sigma_0^2 > 0$. Let $X^{(o)}$ be as defined in* (2.5)*. If $(m_2 - m_1)(n_2 - n_1) = o(mn)$, then $(\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2)$, the approximated likelihood estimator that maximizes* (2.6) *over $C \times \mathbb{R}^+$, is strongly consistent, i.e.*

$$(\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2) \to (\lambda_0, \mu_0, \sigma_0^2) \tag{2.8}$$

almost surely in $C \times \mathbb{R}_+$.

Theorem 2.2.2 (Asymptotic Normality)**.** *Assume the same notations in theorem* 2.2.1*. Let $\xi_j = |u_j - u_{j-1}|$, $\zeta_k = |v_k - v_{k-1}|$, and suppose the following holds:*

- $\xi_j, \zeta_k < o(n^{1/2})$
- $(m_2 - m_1)(n_2 - n_1) \leq O(n^{1-\epsilon_0})$*, where* $0 < \epsilon_0 < 1$

then

$$\begin{bmatrix} \sqrt{n}(\hat{\lambda} - \lambda_0) \\ \sqrt{m}(\hat{\mu} - \mu_0) \end{bmatrix} \to_{\mathcal{D}} N\left(\mathbf{0}, \quad \mathbf{\Sigma_1}\right) \tag{2.9}$$

where

$$\mathbf{\Sigma_1} = \begin{bmatrix} \frac{2{\lambda_0}^2}{1+\lambda_0} & 0 \\ 0 & \frac{2{\mu_0}^2}{1+\mu_0} \end{bmatrix}.$$

Furthermore, suppose $\frac{n}{m} \to \rho$, *where* ρ *is a positive constant, then*

$$\begin{bmatrix} \sqrt{n}\left(\hat{\lambda}-\lambda_0\right) \\ \sqrt{n}\left(\hat{\mu}-\mu_0\right) \\ \sqrt{n}\left(\hat{\sigma}^2-\sigma_0^2\right) \end{bmatrix} \to_D N\left(\mathbf{0}, \begin{bmatrix} 0 & \frac{{\lambda_0}^2}{1+\lambda_0} & \frac{-2\lambda_0{\sigma_0^2}^2}{1+\lambda_0} \\ \frac{{\mu_0}^2}{1+\mu_0} & 0 & \frac{-2\mu_0\sigma_0^2}{1+\mu_0} \\ \frac{-2\mu_0\sigma_0^2}{1+\mu_0} & \frac{-2\lambda_0\sigma_0^2}{1+\lambda_0} & 2\sigma_0^4\left[\frac{1}{1+\lambda_0}+\frac{\rho}{1+\mu_0}\right] \end{bmatrix}\right). \tag{2.10}$$

Proofs of theorems (2.2.1) and (2.2.2) will be provided in the upcoming sections. To this end, the main message in here is that the approximated likelihood estimator shows similar asymptotic result as the MLE under the full data case, as long as the size of the missing block can be controlled by $O(n^{1-\epsilon_0})$, where $0 < \epsilon_0 < 1$.

2.2.2 Variable transformation

In this section, we introduce a set of new variables that will be utilized in the approximation process. This is based on the general idea that a normal random variable is a linear combination of *i.i.d.* standard normal random variables. That is, if $Y = MZ$, where $Y \sim N(0, \Sigma)$ and $Z \sim N(0, I)$, then we can study the asymptotic properties of Y by investigating the behavior of the matrix M. Similarly, in our case, we seek to express the quadratic forms in the likelihood function of the O-U process as a linear combination of independent variables, which can then simplify the analysis of the asymptotic properties of the estimators.

Lemma 2.2.3. *For* $j = 2, \ldots, m$, $k = 2, \ldots, n$, *let* $\xi_j = |u_j - u_{j-1}|$, $\zeta_k = |v_k - v_{k-1}|$, *and* $A(\lambda_0)$, $B(\mu_0)$ *as defined in* (1.5) *and* (1.6). *Let*

- $\eta_{j,k} = \dfrac{x_{j,k} - e^{-\lambda_0\xi_j}x_{j-1,k}}{\sigma_o\sqrt{1-e^{-2\lambda_0\xi_j}}}$

- $\gamma_{j,k} = \dfrac{x_{j,k} - e^{-\mu_0 \xi_j} x_{j,k-1}}{\sigma_o \sqrt{1 - e^{-2\mu_0 \zeta_k}}}$
- $w_{jk} = \dfrac{\eta_{j,k} - e^{-\mu_0 \zeta_k} \eta_{j,k-1}}{\sqrt{1 - e^{-2\mu_0 \zeta_k}}}$

then we have the following

1. $\eta_{j_1,\cdot}$ *is independent of* $x_{j_2,\cdot}$ *for* $j_1 \geq j_2$
2. $\gamma_{\cdot,k_1}$ *is independent of* $x_{\cdot,k_2}$ *for* $k_1 \geq k_2$
3. η_{j_1,k_1} *is independent of* γ_{j_2,k_2} *if either* $j_1 \geq j_2$ *or* $k_1 \geq k_2$
4. *For a fixed* k*,*$\{\eta_{j,k}\}$ *is a sequence of i.i.d normal random variables in* j*, with distribution* $N(0, B(\mu_0))$*. For a fixed* j*,* $\{\gamma_{j,k}\}$ *is a sequence of i.i.d standard normal in* k*, with distribution* $N(0, A(\lambda_0))$ *, and* $\{w_{j,k}\}$ *is a sequence of i.i.d. standard normal ranomd variables in both* j *and* k*.*

Proof of lemma 2.2.3. To show 1 (and similarly for 2), notice that

$$\mathbb{E}\left[\eta_{j_1,\cdot} \underset{\sim}{X}'_{j_2}\right] = \frac{1}{\sigma_0 \sqrt{1 - e^{-2\lambda_0 \xi_{j_1}}}} \mathbb{E}\left[\underset{\sim}{X}_{j_1} \underset{\sim}{X}'_{j_2} - e^{-\lambda_0 \xi_{j_1}} \underset{\sim}{X}_{j_1-1} \underset{\sim}{X}'_{j_2}\right].$$

Then,

$$\mathbb{E}\left[\underset{\sim}{X}_{j_1} \underset{\sim}{X}'_{j_2} - e^{-\lambda_0 \xi_{j_1}} \underset{\sim}{X}_{j_1-1} \underset{\sim}{X}'_{j_2}\right] = \left(e^{-\lambda_0 |u_{j_1} - u_{j_2}|} - e^{-\lambda_0 \xi_{j_1} - \lambda_0 |u_{j_1 -1} - u_{j_2}|}\right) B(\mu_0).$$

Now, if $j_1 > j_2$, then $\xi_{j_1} + u_{j_1-1} - u_{j_2} = u_{j_1} - u_{j_2}$, therefore we have $\mathbb{E}\left[\eta_{j_1,\cdot} \underset{\sim}{X}'_{j_2}\right] = 0$. For 3, since $\gamma_{j,k}$ is a function of $x_{j,k}$ and $x_{j,k-1}$, and similarly, $\eta_{j,k}$ is a function of $x_{j,k}$ and $x_{j-1,k}$, the result stated in part 3 follows from part 1 and 2 of the lemma.

To show 4, first notice that for a fixed column j, we have

$$
\begin{aligned}
\mathbb{E}\left[\eta_{j,k_2}\eta_{j,k_2}\right] &= \frac{\mathbb{E}\left(x_{j,k_1} - e^{-\lambda_0\xi_j}x_{j-1,k_1}\right)\left(x_{j,k_2} - e^{-\lambda_0\xi_j}x_{j-1,k_2}\right)}{\sigma_0^2\left(1-e^{-2\lambda_0\xi_j}\right)} \\
&= \frac{\sigma_0^2\left(1-e^{-2\lambda_0\xi_j}\right)e^{-\mu_0|v_{k_1}-v_{k_2}|}}{\sigma_0^2\left(1-e^{-2\lambda_0\xi_j}\right)} \\
&= e^{-\mu_0|v_{k_1}-v_{k_2}|},
\end{aligned}
$$

so $\eta_{j,\cdot} \sim N\left(0, B(\mu_0)\right)$ for $j = 2, \ldots, m$.

On the other hand, without loss of generality, let k be fixed and $j_1 > j_2$, consider

$$
\begin{aligned}
\mathbb{E}\left[\eta_{j_1,k}\eta'_{j_2,k}\right] &= \frac{\mathbb{E}\left[x_{j_1,k}x_{j_2,k}\right] - e^{-\lambda_0\xi_{j_1}}\mathbb{E}\left[x_{j_1-1,k}x_{j_2,k}\right] - e^{-\lambda_0\xi_{j_2}}\mathbb{E}\left[x_{j_1,k}x_{j_2-1,k}\right]}{\sigma_0^2\sqrt{\left(1-e^{-2\lambda_0\xi_{j_1}}\right)\left(1-e^{-2\lambda_0\xi_{j_2}}\right)}} \\
&\quad + \frac{e^{-\lambda_0(\xi_{j_1}+\xi_{j_2})}\mathbb{E}\left[x_{j_1-1,k}x_{j_2-1,k}\right]}{\sigma_0^2\sqrt{\left(1-e^{-2\lambda_0\xi_{j_1}}\right)\left(1-e^{-2\lambda_0\xi_{j_2}}\right)}} \\
&= \frac{e^{-\lambda_0(u_{j_1}-u_{j_2})} - e^{-\lambda_0(u_{j_1}-u_{j_2})} - e^{\lambda_0(u_{j_1}-u_{j_2}-2u_{j_2})}}{\sqrt{\left(1-e^{-2\lambda_0\xi_{j_1}}\right)\left(1-e^{-2\lambda_0\xi_{j_2}}\right)}} \\
&\quad + \frac{e^{-\lambda_0(u_{j_1}+u_{j_2}-2u_{j_2-1})}}{\sqrt{\left(1-e^{-2\lambda_0\xi_{j_1}}\right)\left(1-e^{-2\lambda_0\xi_{j_2}}\right)}} \\
&= 0,
\end{aligned}
$$

which implies that $\eta_{j,\cdot}$'s are indenpendent, and the same argument can be used for $\gamma_{j,k}$ to show that $\gamma_{\cdot,k} \overset{i.i.d}{\sim} N\left(0, A(\lambda_0)\right)$ as well. Finally, for $w_{j,k}$, notice that, clearly,

$\mathbb{E}[w_{j,k}] = 0$, and $\mathbb{E}[w_{j_1,k_1} w_{j_2,k_2}] = 0$, if $j_1 \neq j_2$, since $\eta'_{j,\cdot}s$ are *i.i.d.* in j. Now,

$$\mathbb{E}\left[w_{j,k}^2\right] = \frac{\mathbb{E}\left[\eta_{j,k}^2\right] - 2e^{-\mu_0\zeta_k}\mathbb{E}\left[\eta_{j,k}\eta_{j,k-1}\right] + e^{-2\mu_0\zeta_k}\mathbb{E}\left[\eta_{j,k-1}^2\right]}{1 - e^{-2\mu_0\zeta_k}}$$
$$= 1,$$

thus $w_{j,k}$'s are a sequence of *i.i.d.* standard normal random variables. □

Lemma 2.2.4. *For $j = 1, \dots, m$ and $k = 1, \dots, n$, let $\eta_{j,k}, \gamma_{j,k}$ and $w_{j,k}$ be the same as previously defined. Let J, K be subsets of the indices $\{1, \dots, m\}$, and $\{1, \dots, n\}$ respectively, with $|J|$ and $|K|$ denote the cardinality of J and K. Also, assume $|J|, |K| \leq O(n^q)$, where $0 < q < 1$, and let $\xi_j = |u_j - u_{j-1}|, \zeta_k = |v_k - v_{k-1}|$ and suppose that $\sum_{j \in J} \xi_j \leq 1$ and $\sum_{k \in K} \zeta_k \leq 1$. Then we have*

1. $\sum_{j \in J} \sum_{k \in K} (w_{j,k}^2 - 1) = o(mn)$ *a.s.*
2. $\sum_{j \in J} (\eta_{j,k}^2 - 1) = o(n)$ *a.s.*
3. $\sum_{k \in K} (\gamma_{j,k}^2 - 1) = o(n)$ *a.s.*
4. *Furthermore, if $\max_{j \in J} \xi_j = \max_{k \in K} \zeta_k \leq o(n^{-1/2})$, then each of*

$$\sum_{j \in J} \sum_{k \in K} \xi_j (w_{j,k}^2 - 1), \sum_{j \in J} \sum_{k \in K} \zeta_k (w_{j,k}^2 - 1), \sum_{j \in J} \sum_{k \in K} \xi_j (\gamma_{j,k}^2 - 1), \sum_{j \in J} \sum_{k \in K} \zeta_k (\eta_{j,k}^2 - 1)$$

is $o(n)$ a.s., or $o_p\left(n^{1/2}\right)$.

Proof of lemma (2.2.4). To show (1), notice that since $w_{j,k}^2$ is a sequence of *i.i.d.* standard normal random variables, $\sum_{j \in J} \sum_{k \in K} (w_{j,k}^2 - 1)$ is then a centered chi-square random variable with $|J||K| - 1$ degrees of freedom. By Chebychev's inequality

$$\begin{aligned} \mathbb{P}\left(\sum_{j \in J} \sum_{k \in K} \left(w_{j,k}^2 - 1\right) > mn \right) &\leq \frac{\mathbb{E}\left[\left(\sum_{j \in J} \sum_{k \in K} (w_{j,k}^2 - 1) \right)^2 \right]}{m^2 n^2} \\ &= \frac{2\left(|J||K| - 1\right)}{m^2 n^2} \leq O\left(n^{-2}\right) \end{aligned}$$

where $2(|J||K|-1)$ is the variance of $\sum_{j\in J}\sum_{k\in K}(w_{j,k}^2-1)$, and the last inequality is due to the assumptions that $|J||K| = O(n^{2q}) = O(n^{2-\epsilon_0})$, and $m = O(n)$. This implies that $\mathbb{P}\left(\sum_{j\in J}\sum_{k\in K}\left(w_{j,k}^2-1\right) > mn\right)$ is summable, and by the Borel-Cantelli lemma,

$$\sum_{j\in J}\sum_{k\in K}(w_{j,k}^2-1) = o(mn) \text{ a.s. } .$$

Similarly to (1), since $\eta_{j,k}$ is a sequence of $i.i.d$ standard normal in j, $\sum_{j\in J}(\eta_{j,k}^2-1) \sim \mathcal{X}^2(|J|-1)$, and therefore

$$\mathbb{P}\left(\sum_{j\in J}(\eta_{j,k}^2-1) > n\right) \leq \frac{2\left(|J|-1\right)}{n^2} = O\left(n^{-2+q}\right).$$

Since $|J| = O(n^q), 0 < q < 1$, $\mathbb{P}\left(\sum_{j\in J}(\eta_{j,k}^2-1) > n\right)$ is also summable,therefore

$$\sum_{j\in J}\left(\eta_{j,k}^2-1\right) = o\left(n\right) \text{a.s.}.$$

Moreover, it can be shown that, by choosing $q < \frac{5}{2}$, we have

$$\sum_{j\in J}\left(\eta_{j,k}^2-1\right) = o_p\left(n^{1/2}\right)$$

as well.

By the same argument, we have

$$\sum_{k\in K}\left(\gamma_{j,k}^2-1\right) = o(n) \text{ a.s., or } o_p\left(n^{1/2}\right)$$

as well. To show (4), notice that since, $\xi_j < o(n^{1/2})$,

$$\mathbb{P}\left(\sum_{j\in J}\sum_{k\in K}\xi_j(w_{j,k}^2-1)>n\right) \leq \frac{\mathbb{E}\left[\left(\sum_{j\in J}\sum_{k\in K}\xi_j(w_{j,k}^2-1)\right)^2\right]}{n^2}$$

$$\leq \frac{\mathbb{E}\left[\left(\sum_{j\in J}\sum_{k\in K}(w_{j,k}^2-1)\right)^2\right]}{n^3}$$

$$= \frac{2|J||K|}{n^3} = O\left(n^{-(1+\epsilon_0)}\right),$$

recalling that $|J||K| = O(n^{2q}) = O(n^{2-\epsilon_0})$, since $0 < q < 1$. Therefore, this is summable and we have

$$\sum_{j\in J}\sum_{k\in K}\xi_j\left(w_{j,k}^2-1\right) = o(n) \text{ a.s..}$$

On the other hand, since $\sum_{k\in K}\xi_k\left(\gamma_{j,k}^2-1\right) = o(n)$, a.s., and $\sum_{j\in J}\xi_j \leq 1$,

$$\sum_{j\in J}\sum_{k\in K}\xi_j\left(\gamma_{j,k}^2-1\right) = \sum_{j\in J}\xi_j\left(\sum_{k\in K}(\gamma_{j,k}^2-1)\right) = \left(\sum_{j\in J}\xi_j\right)o(n) = o(n) \text{ a.s..}$$

The same arguments apply to $\sum_{j\in J}\sum_{k\in K}\zeta_k\left(w_{j,k}^2-1\right)$ and $\sum_{j\in J}\sum_{k\in K}\zeta_k\left(\eta_{j,k}^2-1\right)$ as well.

□

The following lemma, which is a more detailed version of lemma 3 in [24], provides approximations for the coefficients of the linear combination of the transformed variables.

Lemma 2.2.5. *Let* $0 < \delta_j < 1$*, and* $\lim_{j\to\infty} \delta_j = 0$*. Let* θ_0 *and* θ *be parameters such that each of them is in* $(0, c]$*, where* $c > 0$ *is finite. Then*

1. $1 - e^{-2\theta\delta_j} \le 2\theta\delta_j$

2. $\left| \dfrac{1}{1 - e^{-2\theta\delta_j}} - \dfrac{1}{2\theta\delta_j} - \dfrac{1}{2} \right| \le M\delta_j$

3. $\dfrac{1 - e^{-2\theta_0\delta_j}}{1 - e^{-2\theta\delta_j}} = \dfrac{\theta_0}{\theta} + \dfrac{\theta_0(\theta - \theta_0)}{\theta}\delta_j + o(1)$

4. $\dfrac{\left(e^{-\theta_0\delta_j} - e^{-\theta\delta_j}\right)\left(1 - e^{-2\theta_0\delta_j}\right)^{1/2}}{1 - e^{-2\theta\delta_j}} = O({\delta_j}^{1/2})$

5. $\dfrac{\left(e^{-\theta_0\delta_j} - e^{-\theta\delta_j}\right)^2}{1 - e^{-2\theta\delta_j}} = \dfrac{(\theta - \theta_0)^2}{2\theta}\delta_j + o(1)$

6. $\ln \dfrac{1 - e^{-2\theta\delta_i}}{1 - e^{-2\theta_0\delta_i}} = \ln \dfrac{\theta}{\theta_0} + (\theta_0 - \theta)\delta_i + o(1)$

7. $\dfrac{\delta_j e^{-2\theta\delta_j}}{1 - e^{-2\theta\delta_j}} = \dfrac{1}{2\theta} - \dfrac{\delta_j}{2} + O(\delta_j^2)$

8. $\dfrac{\delta_j e^{-\theta\delta_j}}{1 - e^{-2\theta\delta_j}} = \dfrac{1}{2\theta} + O(\delta_j^2)$

Proof of lemma (2.2.5). For each of the following proofs, we use the fact that

$$e^x = T_n(x) + R_n(x),$$

where

- $T_n(x) = \sum_{k=0}^{n} \frac{x^k}{k!}$ is the n^{th} order Taylor approximation of e^x
- $R_n(x) = \frac{e^c}{(n+1)!}x^n$ is its lagrange remainder, with $0 < c < x$.

(1) is clear by letting $0 < c < \delta_j$ and writing

$$1 - e^{-2\theta\delta_j} = 2\theta\delta_j e^{-2\theta c} \leq 2\theta\delta.$$

For (2), notice that

$$\begin{aligned}
\left|\frac{1}{1-e^{-2\theta\delta_j}} - \frac{1}{2\theta\delta_j} - \frac{1}{2}\right| &= \left|\frac{2\theta\delta_j - \left(1-e^{-2\theta\delta_j}\right) - \theta\delta_j\left(1-e^{-2\theta\delta_j}\right)}{2\theta\delta_j\left(1-e^{-2\theta\delta_j}\right)}\right| \\
&= \left|\frac{2\theta\delta_j - (1+\theta\delta_j)\left(1-e^{-2\theta\delta_j}\right)}{(2\theta\delta)^2 e^{-2\theta c_1}}\right| \\
&= \left|\frac{2\theta\delta_j - (1+\theta\delta_j)\left(2\theta\delta_j - 2\theta^2\delta_j^2 + \frac{4}{3}\theta^3\delta_j^2 e^{-2\theta c_2}\right)}{(2\theta\delta)^2 e^{-2\theta c_1}}\right| \\
&= \left|\frac{\theta\delta_j\left(1 - \frac{2}{3}e^{-2\theta c_2} - \frac{2}{3}\theta\delta_j e^{-2\theta c_2}\right)}{e^{-2\theta c_1}}\right|
\end{aligned}$$

and by the triangular inequality,

$$\left|\frac{1}{1-e^{-2\theta\delta_j}} - \frac{1}{2\theta\delta_j} - \frac{1}{2}\right| \le \delta_j \left|\frac{5\theta + 2\theta^2}{2e^{-2\theta c_1}}\right| = M\delta_j.$$

For (3), notice that by (1) and (2), we have

$$\left|1 - e^{-2\theta_0\delta_j}\right| \left|\frac{1}{1-e^{-2\theta\delta_j}} - \frac{1}{2\theta\delta_j} - \frac{1}{2}\right| \le M\delta_j^2,$$

which implies that

$$\underbrace{\left|\frac{1-e^{-2\theta_0\delta_j}}{1-e^{-2\theta\delta_j}} - \frac{2\theta_0\delta_j - 2{\theta_0}^2\delta_j^2 + \frac{4}{3}{\theta_0}^3{\delta_j}^3 e^{-2\theta_0 c_1}}{2\theta\delta_j} - \frac{1}{2}(2\theta_0\delta_j - 2{\theta_0}^2\delta_j^2 + \frac{4}{3}{\theta_0}^3\delta_j^3 e^{-2\theta_0 c_1})\right|}_{(*)}$$

$$\le M{\delta_j}^2,$$

but since

$$\begin{aligned}(*) &= \left|\frac{1-e^{-2\theta_0\delta_j}}{1-e^{-2\theta\delta_j}} - \frac{\theta_0}{\theta} - \frac{\theta_0(\theta-\theta_0)}{\theta}\delta_j - {\theta_0}^2{\delta_j}^2 + \frac{2}{3}\frac{\theta^3}{\theta}e^{-2\theta_0 c_1}({\delta_j}^2 + \theta{\delta_j}^3)\right| \\ &= \left|\frac{1-e^{-2\theta_0\delta_j}}{1-e^{-2\theta\delta_j}} - \frac{\theta_0}{\theta} - \frac{\theta_0(\theta-\theta_0)}{\theta}\delta_j\right| + o(\delta_j),\end{aligned}$$

therefore, we have

$$\frac{1-e^{-2\theta_0\delta_j}}{1-e^{-2\theta\delta_j}} = \frac{\theta_0}{\theta} + \frac{\theta_0(\theta-\theta_0)}{\theta}\delta_j + o(\delta_j).$$

To show (4), notice that

$$
\begin{aligned}
\frac{\left(e^{-\theta_0\delta_j} - e^{-\theta\delta_j}\right)\left(1 - e^{-2\theta_0\delta_j}\right)^{1/2}}{1 - e^{-2\theta\delta_j}} &\leq \sqrt{2\theta\delta_j}\frac{\left(1 - e^{-(\theta_0-\theta)\delta_j}\right)}{1 - e^{-2\theta\delta_j}} \\
&= {\delta_j}^{1/2}\frac{\sqrt{2\theta}(\theta_0 - \theta)}{2\theta}\left(1 + (\theta_0 + \theta)\delta_j + o(1)\right) \\
&\leq {\delta_j}^{1/2}\frac{\sqrt{2\theta}(\theta_0 - \theta)}{2\theta}\left(1 + \theta_0 + \theta + o(1)\right) \\
&= O({\delta_j}^{1/2}).
\end{aligned}
$$

Similarly, for (5)

$$
\begin{aligned}
\frac{(e^{-\theta_0\delta_j} - e^{-\theta\delta_j})^2}{1 - e^{-2\theta\delta_j}} &\leq (\theta - \theta_0)\delta_j\frac{1 - e^{-(\theta-\theta_0)\delta_j}}{1 - e^{-2\theta\delta_j}} \\
&= \frac{(\theta - \theta_0)^2}{2\theta}\delta_j(1 + \frac{\theta_0 + \theta}{2\theta}\delta_j + o(1)) \\
&= \frac{(\theta - \theta_0)^2}{2\theta}\delta_j + O({\delta_j}^2) \\
&= \frac{(\theta - \theta_0)^2}{2\theta}\delta_j + o(1).
\end{aligned}
$$

For (6), we can use (3) and second order Taylor expansion to get

$$\ln\left(\frac{1-e^{-2\theta_0\delta_j}}{1-e^{-2\theta\delta_j}}\right) = \ln\left(\frac{\theta_0}{\theta} + \frac{\theta_0(\theta-\theta_0)}{\theta}\delta_j + o(1)\delta_j\right)$$

$$= \ln\frac{\theta_0}{\theta} + \ln(1 + (\theta-\theta_0)\delta_j + o(1))$$

$$= \ln\frac{\theta_0}{\theta} + (\theta-\theta_0)\delta_j + \frac{(\theta-\theta_0)^2}{2}{\delta_j}^2 + o(1)$$

$$= \ln\frac{\theta_0}{\theta} + (\theta-\theta_0)\delta_j + o(1).$$

To prove (7), notice that

$$\frac{\delta_j e^{-2\theta\delta_j}}{1-e^{-2\theta\delta_j}} - \frac{1}{2\theta} + \frac{\delta_j}{2} = \delta_j e^{-2\theta\delta_j}\left(\frac{1}{1-e^{-2\theta\delta_j}} - \frac{1}{2\theta\delta_j e^{-2\theta\delta_j}} + \frac{1}{2e^{-2\theta\delta_j}}\right),$$

and

$$\frac{1}{1-e^{-2\theta\delta_j}} - \frac{1}{2\theta\delta_j e^{-2\theta\delta_j}} + \frac{1}{2e^{-2\theta\delta_j}}$$

$$= \frac{2\theta\delta_j e^{-2\theta\delta_j} - \left(1-e^{-2\theta\delta_j}\right)(1-\theta\delta_j)}{2\theta\delta_j e^{-2\theta\delta_j}\left(1-e^{-2\theta\delta_j}\right)}$$

$$= \frac{2\theta\delta_j\left(1-2\theta\delta_j+2\theta^2{\delta_j}^2 e^{-2\theta c_2}\right) - (1-\theta\delta_j)\left(2\theta\delta_j - 2\theta^2{\delta_j}^2 + \frac{4}{3}\theta^3{\delta_j}^3 e^{-2\theta c_3}\right)}{4\theta^2{\delta_j}^2 e^{-2\theta(\delta_j+c_1)}}$$

$$= \frac{{\delta_j}^3(2\theta^3(2e^{-2\theta c_2}-1) + \frac{4}{3}\theta^3 e^{-2\theta c_3}(\theta\delta_j - 1))}{4{\delta_j}^2\left(\theta^2 e^{-2\theta(\delta_j+c_1)}\right)}$$

$$= O(\delta_j),$$

which implies that

$$\frac{\delta_j e^{-2\theta\delta_j}}{1-e^{-2\theta\delta_j}} - \frac{1}{2\theta} + \frac{\delta_j}{2} = 2\theta\delta_j e^{-2\theta c} O(\delta_j)$$
$$= O({\delta_j}^2).$$

Similarly, for (8), we have

$$\frac{2\theta\delta_j - (1-e^{-2\theta\delta_j})(1-\theta\delta_j)}{2\theta\delta_j e^{-\theta\delta_j}(1-e^{-2\theta\delta_j})} = \frac{{\delta_j}^3\theta^3(e^{-\theta c_2} + \frac{4}{3}e^{-2\theta c_3})}{4\theta^2{\delta_j}^2 e^{-\theta(\delta_j + c_1)}}$$
$$= O(\delta_j),$$

thus

$$\frac{\delta_j e^{-\theta\delta_j}}{1-e^{-2\theta\delta_j}} - \frac{1}{2\theta} = O({\delta_j}^2).$$

□

Lemma 2.2.6 (Ying [24])**.** *Let $\theta > 0$. For any constant $\epsilon > 0$, there exists a $\delta > 0$ such that*

$$\inf_{|\theta-1|\geq\epsilon}(\theta - 1 - \ln\theta) \geq \delta.$$

Proof of lemma (2.2.6)*.* Let

$$f(\theta) = \theta - 1 - \ln\theta$$

Then, we have $f'(\theta) = 1 - \dfrac{1}{\theta}$. Notice that, from its derivative we can see that f is continuously decresing on $(0, 1)$ and continuously increasing on $\theta \in [1, \infty)$. As a result, we can choose δ to be $f(1 + \epsilon)$. □

Lemma 2.2.6 provides a lower bound for $l(\lambda, \mu, \sigma^2) - l(\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2)$, which will be useful later for showing strong consistency of the estimators.

2.2.3 Expanding and approximating the quadratic forms

The idea in here is to express the quadratic forms as a linear combination of either chi-squared random variables, or product of two random variables that are independent in at least one direction. Combining with previous lemmas, we can utilize these expressions to control the magnitude of the quadratic forms, which will be useful for the asymptotic studies in later sections.

To this end, write $\underset{\sim}{X}_j - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}$ as $\underset{\sim}{X}_j - e^{-\lambda_0\xi_j}\underset{\sim}{X}_{j-1} + (e^{-\lambda_0\xi_j} - e^{-\lambda\xi_j})\underset{\sim}{X}_{j-1}$ (and similarly for $\underset{\sim}{X}^*_j - e^{-\lambda\xi_j}\underset{\sim}{X}^*_{j-1}$). Then, using variable tranformations, we can rewrite the quadratic forms from (2.6) in the following way. For the columns with complete observations, we have

$$\sum_{j\in J_o}\frac{\left(\underset{\sim}{X}_j-e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}\right)'B^{-1}(\mu)\left(\underset{\sim}{X}_j-e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}\right)}{1-e^{-2\lambda\xi_j}}$$

$$=\sigma_0^2\sum_{j\in J_o}\frac{1-e^{-2\lambda_0\xi_j}}{1-e^{-2\lambda\xi_j}}\eta_j'B^{-1}(\mu)\eta_j$$

$$+2\sigma_o\sum_{j\in J_o}\frac{\left(e^{-\lambda_0\xi_j}-e^{-\lambda\xi_j}\right)\left(1-e^{-2\lambda_0\xi_j}\right)^{1/2}}{1-e^{-2\lambda\xi_j}}\eta_j'B^{-1}(\mu)\underset{\sim}{X}_{j-1}$$

$$+\sum_{j\in J_o}\frac{\left(e^{-\lambda_0\xi_j}-e^{-\lambda\xi_j}\right)^2}{1-e^{-2\lambda\xi_j}}\underset{\sim}{X}_{j-1}'B^{-1}(\mu)\underset{\sim}{X}_{j-1}$$

$$=Q_1+Q_2+Q_3, \tag{2.11}$$

where

$$
\begin{aligned}
Q_1 &= \sigma_0^2 \sum_{j \in J_o} \frac{1-e^{-2\lambda_0\xi_j}}{1-e^{-2\lambda\xi_j}} \eta_{j,1}^2 \\
&+ \sigma_0^2 \sum_{j \in J_o} \frac{1-e^{-2\lambda_0\xi_j}}{1-e^{-2\lambda\xi_j}} \sum_{k=2}^{n} \frac{1-e^{-2\mu_0\zeta_k}}{1-e^{-2\mu\zeta_k}} w_{j,k}^2 \\
&+ 2\sigma_0^2 \sum_{j \in J_o} \frac{1-e^{-2\lambda_0\xi_j}}{1-e^{-2\lambda\xi_j}} \sum_{k=2}^{n} \frac{\left(e^{-\mu_0\zeta_k}-e^{-\mu\zeta_k}\right)\left(1-e^{-2\mu\zeta_k}\right)^{1/2}}{1-2^{-2\mu\zeta_k}} \eta_{j,k-1} w_{j,k} \\
&= \sigma^2 \left(\frac{\lambda_0}{\lambda} + o(1)\right) \sum_{j \in J_o} \eta_{j,1}^2 \\
&+ \frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu} \left(\sum_{j \in J_o}\sum_{k=2}^{n} w_{j,k}^2 + (\lambda-\lambda_0) \sum_{j \in J_o}\sum_{k=2}^{n} \xi_j w_{j,k}^2 + (\mu-\mu_0) \sum_{j \in J_o}\sum_{k=2}^{n} \zeta_k w_{j,k}^2 \right) \\
&+ \frac{\lambda_0\sigma_0^2(\mu-\mu_0)^2}{2\lambda\mu} \sum_{j \in J_o}\sum_{k=2}^{n} \zeta_k \eta_{j,k-1}^2 + \sum_{j \in J_o}\sum_{k=2}^{n} \xi_j \zeta_k \eta_{j,k-1}^2, \qquad (2.12)
\end{aligned}
$$

$$
\begin{aligned}
Q_2 &= 2\sigma_o \sum_{j \in J_o} \frac{\left(e^{-\lambda_0 \xi_j} - e^{-\lambda \xi_j}\right)\left(1 - e^{-2\lambda_0 \xi_j}\right)^{1/2}}{1 - e^{e^{-2\lambda \xi_j}}} \underset{\sim}{X}_{j-1} B^{-1} \eta_j \\
&= 2\sigma_o \Bigg[\sum_{j \in J_o} \sum_{k=2}^{n} \frac{(e^{-\lambda_0 \xi_j} - e^{-\lambda \xi_j})(1 - e^{-2\lambda_0 \xi_j})^{1/2}}{1 - e^{e^{-2\lambda \xi_j}}} \frac{1 - e^{-2\mu_0 \zeta_k}}{1 - e^{-2\mu \zeta_k}} \gamma_{j-1,k} w_{j,k} \\
&+ \sum_{j \in J_o} \sum_{k=2}^{n} \frac{\left(e^{-\lambda_0 \xi_j} - e^{-\lambda \xi_j}\right)\left(1 - e^{-2\lambda_0 \xi_j}\right)^{1/2}}{1 - e^{e^{-2\lambda \xi_j}}} \frac{(e^{-\mu_0 \zeta_k} - e^{-\mu \zeta_k})^2}{1 - e^{-2\mu \zeta_k}} x_{j-1,k-1} \eta_{j,k-1} \\
&+ \sum_{j \in J_o} \sum_{k=2}^{n} \Bigg(\frac{\left(e^{-\lambda_0 \xi_j} - e^{-\lambda \xi_j}\right)\left(1 - e^{-2\lambda_0 \xi_j}\right)^{1/2}}{1 - e^{e^{-2\lambda \xi_j}}} \\
&\cdot \frac{\left(e^{-\mu_0 \zeta_k} - e^{-\mu \zeta_k}\right)\left(1 - e^{-2\mu_0 \zeta_k}\right)^{1/2}}{1 - e^{-2\mu \zeta_k}} x_{j-1,k} w_{j,k} \Bigg) \\
&+ \sum_{j \in J_o} \sum_{k=2}^{n} \Bigg(\frac{\left(e^{-\lambda_0 \xi_j} - e^{-\lambda \xi_j}\right)\left(1 - e^{-2\lambda_0 \xi_j}\right)^{1/2}}{1 - e^{e^{-2\lambda \xi_j}}} \\
&\cdot \frac{\left(e^{-\mu_0 \zeta_k} - e^{-\mu \zeta_k}\right)\left(1 - e^{-2\mu_0 \zeta_k}\right)^{1/2}}{1 - e^{-2\mu \zeta_k}} \eta_{j,k-1} \gamma_{j-1,k-1} \Bigg) \Bigg] \\
&= 2\sigma_0 \Bigg(\sum_{j \in J_o} \sum_{k=2}^{n} \xi_j^{1/2} \gamma_{j-1,k} w_{j,k} + \sum_{j \in J_o} \sum_{k=2}^{n} \xi_j^{1/2} \zeta_k x_{j-1,k-1} \eta_{j,k-1} \\
&+ \sum_{j \in J_o} \sum_{k=2}^{n} \xi_j^{1/2} \zeta_k^{1/2} \left(x_{j-1,k-1} w_{j,k} + \eta_{j,k-1} \gamma_{j-1,k-1} \right) \Bigg), \qquad (2.13)
\end{aligned}
$$

and

$$Q_3 = \sum_{j\in J_o} \frac{\left(e^{-\lambda_0\xi_j} - e^{-\lambda\xi_j}\right)^2}{1-e^{-2\lambda\xi_j}} \underset{\sim}{X}'_{j-1} B^{-1}(\mu) \underset{\sim}{X}_{j-1}$$

$$= \sum_{j\in J_o} \frac{\left(e^{-\lambda_0\xi_j} - e^{-\lambda\xi_j}\right)^2}{1-e^{-2\lambda\xi_j}} \left[x_{j-1,1}^2 + \sum_{k=2}^{n} \frac{(x_{j-1,k} - e^{-\mu\zeta_k} x_{j-1,k-1})^2}{1-e^{-2\mu\zeta}}\right]$$

$$= \sum_{j\in J_o} \frac{\left(e^{-\lambda_0\xi_j} - e^{-\lambda\xi_j}\right)^2}{1-e^{-2\lambda\xi_j}} \left[\sum_{k=2}^{n} \sigma_0^2 \frac{\left(1-e^{-2\mu_0\zeta_k}\right)}{(1-e^{-2\mu\zeta_k})} \gamma_{j-1,k}^2\right.$$

$$+ \sum_{k=2}^{n} \frac{\left(e^{-\mu_0\zeta_k} - e^{-\mu\zeta_k}\right)^2}{1-e^{-2\mu\zeta_k}} x_{j-1,k-1}^2$$

$$\left. + 2\sum_{k=2}^{n} \frac{\left(e^{-\mu_0\zeta_k} - e^{-\mu\zeta_k}\right)\left(1-e^{-2\mu_0\zeta_k}\right)^{1/2}}{1-e^{-2\mu\zeta_k}} x_{j-1,k-1}\gamma_{j-1,k}\right] + O(1)$$

$$= \frac{\mu_0\sigma_0^2(\lambda-\lambda_0)^2}{2\lambda\mu} \sum_{j\in J_o}\sum_{k=2}^{n} \xi_j \gamma_{j-1,k}^2$$

$$+ \sum_{j\in J_o}\sum_{k=2}^{n} \xi_j\zeta_k x_{j-1,k-1}^2 + 2\sum_{j\in J_o}\sum_{k=2}^{n} \xi_j \zeta_k^{1/2} x_{j-1,k-1}\gamma_{j-1,k} + O(1) \text{ a.s. } . \qquad (2.14)$$

For the columns with partial observations, notice that

$$\underset{\sim}{X}_j^*(B^*(\mu))^{-1}\underset{\sim}{X}_j^* = x_{j,1}^2 + \sum_{k=2}^{n_1-1} \frac{\left(x_{j,k} - e^{-\mu\zeta_k} x_{j,k-1}\right)^2}{1 - e^{-2\mu\zeta_k}}$$

$$+ \frac{\left(x_{j,n_2+1} - e^{-\mu\zeta^*_{n_2+1}} x_{j,n_1-1}\right)^2}{1 - e^{-2\mu\zeta^*_{n_2+1}}} + \sum_{k=n_2+1}^{n} \frac{\left(x_{j,k} - e^{-\mu\zeta_k} x_{j,k-1}\right)^2}{1 - e^{-2\mu\zeta_k}}$$

$$= x_{j,1}^2 + \sum_{k\in K_o} \frac{\left(x_{j,k} - e^{-\mu\zeta_k} x_{j,k-1}\right)^2}{1 - e^{-2\mu\zeta_k}} + \frac{\left(x_{j,n_2+1} - e^{-\mu\zeta^*_{n_2+1}} x_{j,n_1-1}\right)^2}{1 - e^{-2\mu\zeta^*_{n_2+1}}}, \tag{2.15}$$

thus

$$\sum_{j=m_1}^{m_2+1} \frac{\left(\underset{\sim}{X}_j^* - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}^*\right)' (B^*(\mu))^{-1} \left(\underset{\sim}{X}_j^* - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}^*\right)}{1 - e^{-2\lambda\xi_j}} \tag{2.16}$$

$$= Q_1^* + Q_2^* + Q_3^* + Q_4^*. \tag{2.17}$$

Similar to the quadratic form from complete observation columns, we have

$$Q_1^* = \sigma^2 \left(\frac{\lambda_0}{\lambda} + o(1) \right) \sum_{j=m_1}^{m_2+1} \eta_{j,1}^2$$

$$+ \frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu} \left(\sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} w_{j,k}^2 + (\lambda - \lambda_0) \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \xi_j w_{j,k}^2 + (\mu - \mu_0) \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \zeta_k w_{j,k}^2 \right)$$

$$+ \frac{\lambda_0 \sigma_0^2 (\mu - \mu_0)^2}{2\lambda\mu} \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \zeta_k \eta_{j,k-1}^2 + \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \xi_j \zeta_k \eta_{j,k-1}^2, \tag{2.18}$$

$$Q_2^* = \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \xi_j^{1/2} \gamma_{j-1,k} w_{j,k} + \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \xi_j^{1/2} \zeta_k x_{j-1,k-1} \eta_{j,k-1}$$

$$+ \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \xi_j^{1/2} \zeta_k^{1/2} \left(x_{j-1,k-1} w_{j,k} + \eta_{j,k-1} \gamma_{j-1,k-1} \right), \tag{2.19}$$

$$Q_3^* = \frac{\mu_0 \sigma_0^2 (\lambda - \lambda_0)^2}{2\lambda\mu} \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \xi_j \gamma_{j-1,k}^2$$

$$+ \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \xi_j \zeta_k x_{j-1,k-1}^2 + 2 \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \xi_j \zeta_k^{1/2} x_{j-1,k-1} \gamma_{j-1,k} + O(1), \tag{2.20}$$

and

$$Q_4^* = \frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu}\left(\sum_{j=m_1}^{m_2+1} w_{j,n_2+1}^2 + (\lambda-\lambda_0)\sum_{j=m_1}^{m_2+1}\xi_j w_{j,n_2+1}^2 + (\mu-\mu_0)\sum_{j=m_1}^{m_2+1}\zeta_{n_2+1}^* w_{j,n_2+1}^2\right)$$

$$+ \sigma_0^2\Bigg(\sum_{j=m_1}^{m_2+1}\zeta_{n_2+1}^{*\ 1/2}\eta_{j,n_1-1}w_{j,n_2+1} + \frac{\lambda_0(\mu-\mu_0)^2}{2\lambda\mu}\sum_{j=m_1}^{m_2+1}\zeta_{n_2+1}^*\eta_{j,n_1-1}^2$$

$$+ \sum_{j=m_1}^{m_2+1}\xi_j\zeta_{n_2+1}^*\eta_{j,n_1-1}^2\Bigg)$$

$$+ 2\sigma_o\left(\sum_{j=m_1}^{m_2+1}\xi_j^{1/2}\gamma_{j-1,n_2+1}w_{j,n_2+1} + \sum_{j=m_1}^{m_2+1}\xi_j^{1/2}\zeta_{n_2+1}^* x_{j-1,n_1-1}\eta_{j,n_1-1}\right)$$

$$+ 2\sigma_0^2\sum_{j=m_1}^{m_2+1}\xi_j^{1/2}\zeta_{n_2+1}^{*\ 1/2}\left(x_{j-1,n_1-1}w_{j,n_2+1} + \eta_{j,n_1-1}\gamma_{j-1,n_1-1}\right)$$

$$+ \frac{\mu_0(\lambda-\lambda_0)}{2\lambda\mu}\sum_{j=m_1}^{m_2+1}\xi_j y_{j-1,n_2+1}^2 + \sum_{j=m_1}^{m_2+1}\xi_j\zeta_{n_2+1}^* x_{j-1,n_1-1}^2$$

$$+ 2\sum_{j=m_1}^{m_2+1}\xi_j\zeta_{n_2+1}^{*\ 1/2}x_{j-1,n_1-1}\gamma_{j-1,n_2+1}.$$

Claim 2.2.7. *Assume* $\xi_j, \zeta_k \le o(n^{-1/2})$*, and* $(m_2 - m_1), (n_2 - n_1) = O(n^{\frac{1}{2}-\epsilon_0})$*, where* $0 < \epsilon_o < \frac{1}{2}$*. We have,*

$$Q_1 + Q_1^* = \frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu}\left(\sum_{j \in J_o}\sum_{k=2}^{n} w_{j,k}^2 + \sum_{j=m_1}^{m_2+1}\sum_{k \in K_o} w_{j,k}^2\right)$$

$$+ m\left[\frac{\lambda_0 \sigma_0^2}{\lambda} + \frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu}(\mu - \mu_0) + \frac{\lambda_0 \sigma_0^2}{2\lambda \mu}(\mu - \mu_0)^2\right]$$

$$+ n\left[\frac{\lambda_0 \mu_0 \sigma_0^2}{\lambda \mu}(\lambda - \lambda_0)\right] + o(n) \ \ a.s., \tag{2.21}$$

$$Q_3 + Q_3^* = \frac{\mu_0 (\lambda - \lambda_0)^2}{2\lambda \mu} n + o(n) \ \ a.s., \tag{2.22}$$

$$Q_2 + Q_2^* = o(n) \ \ a.s., \tag{2.23}$$

$$Q_4^* = o(n) \ \ a.s., \tag{2.24}$$

and

$$\underset{\sim}{X}_1 B^{-1}(\mu) \underset{\sim}{X}_1 = \frac{\mu_0 \sigma_0^2}{\mu} n + o(n) \ \ a.s.. \tag{2.25}$$

Proof of claim. (2.21) is obvious under lemma (2.2.4). In particular, notice that for Q_1 (and similarly for Q_1^*),

$$\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k\eta_{j,k-1}^2 = \sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k\left(\eta_{j,k-1}^2-1\right)+\sum_{j\in J_o}\xi_j\sum_{k=2}^{n}\zeta_k$$

$$= \sum_{j\in J_o}\xi_j\sum_{k=2}^{n}\zeta_k\left(\eta_{j,k-1}^2-1\right)+\sum_{j\in J_o}\xi_j\sum_{k=2}^{n}\zeta_k$$

$$\leq \sum_{k=2}^{n}\zeta_k\left(\eta_{j,k-1}^2-1\right)+1=o(n)$$

To show (2.22), first notice that lemma (2.2.4) gives

$$\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\gamma_{j-1,k}^2 = n+o(n),$$

and

$$\sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\xi_j\gamma_{j-1,k}^2=(n-(n_2-n_1+1))\left(\sum_{j=m_1}^{m_2+1}\xi_j\right)+o(n).$$

Now,

$$\sum_{j=m_1}^{m_2+1}\xi_j\leq(m_2-m_1)n^{-1/2}<O(n^{-\epsilon}),$$

so we have

$$\sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\xi_j\gamma_{j-1,k}^2=O(n^{1-\epsilon})+o(n)=o(n).$$

On the other hand, notice that since $x_{j,k}^2$ is non-negative, and continuous on $[0,1]^2$, therefore $\sup_{j,k} \mathbb{E}[x_{j,k}^2]$ is bounded, and

$$\begin{aligned}
\mathbb{P}\left(\sum_{j\in J}\sum_{k=2}^{n}\xi_j\zeta_k x_{j-1,k-1}^2 > n^{1+\epsilon_0}\right) &\leq \frac{\sum_{j\in J}\sum_{k=2}^{n}\xi_j\zeta_k \mathbb{E}\left[x_{j-1,k-1}^2\right]}{n^{1+\epsilon_0}} \\
&\leq \frac{\left(\sum_{j\in J}\sum_{k=2}^{n}\xi_j\zeta_k\right)\sup_{j,k}\mathbb{E}\left[x_{j-1,k-1}^2\right]}{n^{1+\epsilon_0}} \\
&\leq \frac{O(1)}{n^{1+\epsilon_0}}.
\end{aligned}$$

Thus, $\sum_{n=1}^{\infty}\mathbb{P}\left(\sum_{j\in J}\sum_{k=2}^{n}\xi_j\zeta_k x_{j-1,k-1}^2 > n^{1+\epsilon_0}\right)$ is finite, and by the Borel-Cantelli lemma

$$\sum_{j\in J}\sum_{k=2}^{n}\xi_j\zeta_k x_{j-1,k-1}^2 = o(n^{1+\epsilon_0})$$

for $\epsilon_0 > 0$. This also holds for $\sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\xi_j\zeta_k x_{j-1,k-1}^2$ as well.

Now, for $\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k^{1/2}x_{j-1,k-1}\gamma_{j-1,k}$, we have

$$\begin{aligned}
\mathbb{P}\left(\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k^{1/2}x_{j-1,k-1}\gamma_{j-1,k} > n\right) &\leq \frac{\mathbb{E}\left[\left(\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k^{1/2}x_{j-1,k-1}\gamma_{j-1,k}\right)^2\right]}{n^2} \\
&= \sum_{j_1\in J_o}\sum_{j_2\in J_o}\sum_{k_1=2}^{n}\sum_{k_2=2}\xi_{j_1}\xi_{j_2}\zeta_{k_1}\zeta_{k_2} \\
&\quad\times\frac{\mathbb{E}\left[x_{j_1-1,k_1-1}\gamma_{j_1-1,k_1}x_{j_2-1,k_2-1}\gamma_{j_2-1,k_2}\right]}{n^2}.
\end{aligned}$$

Recalling from lemma (2.2.3) that γ_{j_1,k_1} is independent of x_{j_2,k_2} for $j_1 \neq j_2$ or $k_1 > k_2$, this implies that $\mathbb{E}[x_{j_1-1,k_1-1}\gamma_{j_1-1,k_1}x_{j_2-1,k_2-1}\gamma_{j_2-1,k_2}] = 0$ unless $j_1 = j_2$ and $k_2 = k_2$. Therefore

$$\begin{aligned}
\mathbb{P}(\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k^{1/2}x_{j-1,k-1}\gamma_{j-1,k} > n) &\leq \frac{\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j^2\zeta_k\mathbb{E}[x_{j-1,k-1}^2]\mathbb{E}[\gamma_{j-1,k}^2]}{n^2} \\
&\leq \sigma^2\frac{\sum_{j\in J_o}\xi_j\sum_{k=2}^{n}\zeta_k}{n^2} \\
&\leq \frac{\sigma^2}{n^2}
\end{aligned}$$

is again summable, and the same argument can be applied to $\sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\xi_j\zeta_k^{1/2}x_{j-1,k-1}\gamma_{j-1,k}$.

To show (2.23), notice that we can apply similar arguments that were used to prove (2.22) on the terms

$$\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j^{1/2}\zeta_k x_{j-1,k-1}\eta_{j,k-1},\ \sum_{j\in J_o}\sum_{k=2}^{n}\xi^{1/2}\zeta_k^{1/2}x_{j-1,k-1}w_{j,k},\ \text{ and }\ \sum_{j\in J_o}\sum_{k=2}^{n}\xi^{1/2}\zeta_k^{1/2}\eta_{j,k-1}\gamma_{j-1,k-1}$$

(and similarly for the same terms summing over $m_1 \dots m_2+1$ and K_o). In particular, by independence we have

$$\begin{aligned}\mathbb{P}(\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j^{1/2}\gamma_{j-1,k}w_{j,k} > n^{1+\epsilon_o}) &\leq \frac{\sum_{j\in J_o}\xi_j\sum_{k=2}^{n}\mathbb{E}[\gamma_{j-1,k}^2]\mathbb{E}[w_{j,k}^2]}{n^{2(1+\epsilon_o)}}\\ &= \frac{\sum_{k=2}^{n}\sum_{j\in J_o}\xi_j}{n^{2(1+\epsilon_o)}}, \epsilon_o > 0\\ &= \frac{1}{n^{1+2\epsilon_o}}\end{aligned}$$

which is summable. Applying the same procedure to $\sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\xi_j^{1/2}\gamma_{j-1,k}w_{j,k}$, we have that $Q_2 + Q_2^* = o(n)$ a.s.. For (2.24), notice that Q_4^* consists of a finite sum of continuous functions on $[0,1]^2$, each of them bounded almost surely. Therefore, assuming $m_2 - m_1 = o(n)$, we have $Q_4^* = o(n)$ as well. Finally, to show (2.25), notice that similar to previous arguments,

$$\underset{\sim}{X}_1 B^{-1}(\mu)\underset{\sim}{X}_1 = x_{1,1}^2 + \sum_{k=2}^{n}\frac{\left(x_1, k - e^{-\mu\zeta_k}x_{1,k-1}\right)^2}{1-e^{-2\mu\zeta}}$$

$$= x_{1,1}^2 + \sigma_0^2\sum_{k=2}^{n}\frac{\left(1-e^{-2\mu_0\zeta_k}\right)}{(1-e^{-2\mu\zeta_k})}\gamma_{1,k}^2 + \sum_{k=2}^{n}\frac{\left(e^{-\mu\zeta_k}-e^{-\mu_0\zeta_k}\right)^2}{(1-e^{-2\mu\zeta_k})}x_{1,k-1}^2$$

$$+ 2\sigma_o\sum_{k=2}^{n}\frac{\left(e^{-\mu\zeta_k}-e^{-\mu_0\zeta_k}\right)\left(1-e^{-2\mu_0\zeta}\right)^{1/2}}{(1-e^{-2\mu_0\zeta_k})}x_{1,k-1}\gamma_{1,k}$$

$$= \left(\frac{\sigma^2\mu_0}{\mu}+o(1)\right)\sum_{k=2}^{n}\gamma_{1,k}^2 + \sum_{k=2}^{n}\zeta_k x_{1,k-1}^2 + \sum_{k=2}^{n}\zeta_k^{1/2}x_{1,k-1}\gamma_{1,k}$$

$$= \left(\frac{\sigma^2\mu_0}{\mu}+o(1)\right)\sum_{k=2}^{n}\gamma_{1,k}^2 + o(n)$$

$$= \frac{\sigma^2\mu_0}{\mu}n + o(n),\ \text{a.s. },\ \text{or } o_p\left(n^{1/2}\right).$$

□

As a result from claim 2.21, we can express (2.6) as

$$
\begin{aligned}
l(\lambda, \mu, \sigma^2 | X^{(o)}) = {} & [nm - (m_2 - m_1 + 2)(n_2 - n_1 + 1)] \ln\left(2\pi\sigma^2\right) \\
& + n \sum_{j \in J_o} \ln\left(1 - e^{-2\lambda\xi_j}\right) + [n - (n_2 - n_1 + 1)] \sum_{j=m_1}^{m_2+1} \ln\left(1 - e^{-2\lambda\xi_j}\right) \\
& + [m - (m_2 - m_1 + 2)] \sum_{k=2}^{n} \ln\left(1 - e^{-2\mu\zeta_k}\right) \\
& + (m_2 - m_1 + 2) \left[\sum_{k \in K_o} \ln\left(1 - e^{-2\mu\zeta_k}\right) + \ln\left(1 - e^{-2\mu\zeta^*_{n_2+1}}\right)\right] \\
& + \frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} \left(\sum_{j \in J_o}\sum_{k=2}^{n} w_{j,k}^2 + \sum_{j=m_1}^{m_2+1}\sum_{k \in K_o} w_{j,k}^2\right) \\
& + m \left[\frac{\lambda_0\sigma_0^2}{\lambda\sigma^2} + \frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2}(\mu - \mu_0) + \frac{\lambda_0\sigma_0^2}{2\lambda\mu\sigma^2}(\mu - \mu_0)^2\right] \\
& + n \left[\frac{\mu_0\sigma_0^2}{\mu\sigma^2} + \frac{\lambda_0\mu_0\sigma^2}{\lambda\mu\sigma^2}(\lambda - \lambda_0) + \frac{\mu_0\sigma_0^2}{2\lambda\mu\sigma^2}(\lambda - \lambda_0)^2\right] + o(n) \text{ a.s..}
\end{aligned}
\tag{2.26}
$$

Notice that,

$$
\begin{aligned}
\sum_{j \in J_o}\sum_{k=2}^{n} w_{j,k}^2 + \sum_{j=m_1}^{m_2+1}\sum_{k \in K_o} w_{j,k}^2 &= (m-1)(n-1) - (n_2 - n_1 + 1)(m_2 - m_1 + 2) + o(mn) \text{ a.s.} \\
&= (m-1)(n-1) + o(mn) \text{ a.s. .}
\end{aligned}
$$

On the other hand, since we assume $m_2 - m_1 = O(n^{1/2-\epsilon_0})$ and $\xi_j < o(n^{-1/2})$, we can apply lemma (2.2.5) part (6) to get

$$
\begin{aligned}
&[nm - (m_2 - m_1 + 2)(n_2 - n_1 + 1)] \ln \frac{\sigma^2}{\sigma_0^2} \\
&+ n \sum_{j \in J_o} \ln \frac{1 - e^{-2\lambda \xi_j}}{1 - e^{-2\lambda_0 \xi_j}} + [n - (n_2 - n_1 + 1)] \sum_{j=m_1}^{m_2+1} \ln \frac{1 - e^{-2\lambda \xi_j}}{1 - e^{-2\lambda \xi_j}} \\
&+ [m - (m_2 - m_1 + 2)] \sum_{k=2}^{n} \ln \frac{1 - e^{-2\mu \zeta_k}}{1 - e^{-2\mu_0 \zeta_k}} \\
&+ (m_2 - m_1 + 2) \left[\sum_{k \in K_o} \ln \frac{1 - e^{-2\mu \zeta_k}}{1 - e^{-2\mu \zeta_k}} + \ln \frac{1 - e^{-2\mu \zeta^*_{n_2+1}}}{1 - e^{-2\mu \zeta^*_{n_2+1}}} \right] \\
&= [(m-1)(n-1) + m + n - 1 - (m_2 - m_1 + 1)(n_2 - n_1 + 1)] \ln \frac{\sigma^2}{\sigma_0^2} \\
&+ [(n-1)m - (m_2 - m_1 + 1)(n_2 - n_1 + 1)] \ln \frac{\mu}{\mu_0} \\
&+ [n(m-1) - (m_2 - m_1 + 1)(n_2 - n_1 + 1)] \ln \frac{\lambda}{\lambda_0} \\
&+ m(\mu_0 - \mu) + n(\lambda_0 - \lambda) + (n - (n_2 - n_1 + 1)) \left(\sum_{j=m_1}^{m_2+1} \xi_j \right) (\lambda_0 - \lambda) \\
&= (m-1)(n-1) \ln \frac{\lambda \mu \sigma^2}{\lambda_0 \mu_0 \sigma_0^2} + (n-1) \ln \frac{\mu \sigma^2}{\mu_0 \sigma_0^2} + (m-1) \ln \frac{\lambda \sigma^2}{\lambda_0 \sigma_0^2} \\
&- (m_2 - m_1 + 2)(n_2 - n_1 + 1) \ln \frac{\lambda \mu \sigma^2}{\lambda_0 \mu_0 \sigma_0^2} + o(n), \qquad (2.27)
\end{aligned}
$$

Putting together the results from (2.21) – (2.25), and (2.27), we have

$$
\begin{aligned}
& l\left(\lambda, \mu, \sigma^2 | X^{(o)}\right) - l\left(\lambda_0, \mu_0, \sigma_0^2 | X^{(o)}\right) \\
& = (n-1)(m-1)\left[\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - \ln\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - 1\right] \\
& + (m-1)\left[\frac{\lambda_0\sigma_0^2}{\lambda\sigma^2} - 1 - \ln\frac{\lambda_0\sigma_0^2}{\lambda\sigma^2} + (\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - 1)(\mu-\mu_0) + \frac{\lambda_0\sigma_0^2(\mu-\mu_0)^2}{2\lambda\mu\sigma^2}\right] \\
& + (n-1)\left[\frac{\mu_0\sigma_0^2}{\mu\sigma^2} - 1 - \ln\frac{\mu_0\sigma_0^2}{\mu\sigma^2} + (\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - 1)(\lambda-\lambda_0) + \frac{\mu_0\sigma_0^2(\lambda-\lambda_0)^2}{2\lambda\mu\sigma^2}\right] \\
& + (n_2-n_1+1)(m_2-m_1+2)\left[\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - \ln\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - 1\right] + o(n) \text{ a.s.}. \qquad (2.28)
\end{aligned}
$$

Assuming $(m_2 - m_1)(n_2 - n_1) = o(mn)$, we have

$$
\begin{aligned}
l\left(\lambda, \mu, \sigma^2 | X^{(o)}\right) - l\left(\lambda_0, \mu_0, \sigma_0^2 | X^{(o)}\right) &= (m-1)(n-1)\left(\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - \ln\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - 1\right) \\
& + o(mn) \text{ a.s.} \qquad (2.29)
\end{aligned}
$$

and

$$
\begin{aligned}
l\left(\lambda, \mu, \sigma^2 | X^{(o)}\right) - l\left(\lambda_0, \mu_0, \frac{\lambda\mu\sigma^2}{\lambda_0\mu_0} | X^{(o)}\right) &= (m-1)\left[\frac{\mu}{\mu_0} - 1 - \ln\frac{\mu}{\mu_0} + \frac{(\mu-\mu_0)^2}{2\mu_0}\right] \\
& + (n-1)\left[\frac{\lambda}{\lambda_0} - 1 - \ln\frac{\lambda}{\lambda_0} + \frac{(\lambda-\lambda_0)^2}{2\lambda_0}\right] \\
& + o(n) \text{ a.s.}. \qquad (2.30)
\end{aligned}
$$

2.2.4 Proof of strong consistency

In here we prove Theorem 2.2.1 by investigating the consistency of $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ using results shown in previous sections. Our first goal is to show that $\hat{\lambda}\hat{\mu}\hat{\sigma}^2$ converges to $\lambda_0\mu_0\sigma_0^2$ as a product, and consequently we can use that result to show $\hat{\lambda} \to \lambda_0, \hat{\mu} \to \mu_0$ and $\hat{\sigma}^2 \to \sigma_0^2$. Consider

$$\inf_{(\lambda,\mu,\sigma^2)\in\bar{V}_\epsilon} \left(l(\lambda, \mu, \sigma^2|X^{(o)}\right) - l\left(\lambda_0, \mu_0, \sigma_0^2|X^{(o)}\right)$$

where for $\epsilon > 0$,

$$V_\epsilon = \left\{\left(\lambda, \mu, \sigma^2\right) : \left|\frac{\lambda\mu\sigma^2}{\lambda_0\mu_0\sigma_0^2} - 1\right| < \epsilon, (\lambda, \mu) \in C \subset \mathbb{R}^2, 0 < \sigma^2 < \infty\right\}, \text{ and}$$

$$\bar{V}_\epsilon = \left\{(\lambda, \mu, \sigma^2) : \left|\frac{\lambda\mu\sigma^2}{\lambda_0\mu_0\sigma_0^2} - 1\right| \geq \epsilon, (\lambda, \mu) \in C \subset \mathbb{R}^2, 0 < \sigma^2 < \infty\right\},$$

with C being a compact set in $\mathbb{R}^2$. Notice that, since $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ are the maximizers of $L(\lambda, \mu, \sigma^2|X^{(o)})$, we have

$$\frac{L\left(\lambda, \mu, \sigma^2|X^{(o)}\right)}{L\left(\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2|X^{(o)}\right)} \leq 1. \tag{2.31}$$

Now, if the maximizers are outside of V_ϵ, then

$$\frac{L\left(\lambda_0, \mu_0, \sigma_0^2|X^{(o)}\right)}{\sup_{(\lambda,\mu,\sigma)\in\bar{V}_\epsilon} L\left(\hat{\lambda}, \hat{\mu}, \hat{\sigma}^2|X^{(o)}\right)} \to \infty \text{ a.s.} \tag{2.32}$$

is a contradiction to (2.31) by definition. In other words, if (2.32) holds almost surely outside of a small neighborhood of $\lambda_0\mu_0\sigma_0^2$, then the maximizers of the approximate likelihood function must be near $\lambda_0\mu_0\sigma_0^2$. Therefore, we can show that $\hat{\lambda}\hat{\mu}\hat{\sigma}^2 \to \lambda_0\mu_0\sigma_0^2$ almost surely with respect to $X(\omega)$ generated from the probability space $(\Omega, \mathcal{A}, P_0)$, where $P_0 \sim N(0, \sigma_0^2 A(\lambda_0) \otimes B(\mu_0))$. In our case, it is equivalent to showing that

$$\inf_{(\lambda,\mu,\sigma^2)\in\bar{V}_\epsilon} \left(l\left(\lambda,\mu,\sigma^2|X^{(o)}\right) - l\left(\lambda_0,\mu_0,\sigma_0^2|X^{(o)}\right)\right) \to \infty \text{ a.s.} \tag{2.33}$$

with respect to P_0.

Now, from (2.29), we have that

$$l\left(\lambda,\mu,\sigma^2|X^{(o)}\right) - l\left(\lambda_0,\mu_0,\sigma_0^2|X^{(o)}\right) = (m-1)(n-1)\left(\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - \ln\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - 1\right) + o(mn) \text{ a.s. } .$$

By lemma (2.2.6), $\left(\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - \ln\frac{\lambda_0\mu_0\sigma_0^2}{\lambda\mu\sigma^2} - 1\right)$ is bounded below by a positive constant for $(\lambda,\mu,\sigma^2) \in \bar{V}_\epsilon$, and any $\epsilon > 0$, therefore, (2.33) holds as $m, n \to \infty$, thus we have

$$\hat{\lambda}\hat{\mu}\hat{\sigma}^2 \to \lambda_0\mu_0\sigma_0^2 \text{ a.s.} \tag{2.34}$$

On the other hand, let

$$U_\epsilon = \left\{(\lambda,\mu) : \left|\frac{\lambda}{\lambda_0} - 1\right| < \epsilon \text{ and } \left|\frac{\mu}{\mu_0} - 1\right| < \epsilon\right\}$$

and

$$\bar{U}_\epsilon = \left\{(\lambda, \mu) : \left|\frac{\lambda}{\lambda_0} - 1\right| \geq \epsilon \text{ or } \left|\frac{\mu}{\mu_0} - 1\right| \geq \epsilon\right\}.$$

Similary, we have from (2.30)

$$\begin{aligned} l\left(\lambda, \mu, \sigma^2 | X^{(o)}\right) - l\left(\lambda_0, \mu_0, \frac{\lambda\mu\sigma^2}{\lambda_0\mu_0} | X^{(o)}\right) &= (m-1)[\frac{\mu}{\mu_0} - 1 - \ln\frac{\mu}{\mu_0} + \frac{(\mu-\mu_0)^2}{2\mu_0}] \\ &+ (n-1)[\frac{\lambda}{\lambda_0} - 1 - \ln\frac{\lambda}{\lambda_0} + \frac{(\lambda-\lambda_0)^2}{2\lambda_0}] \\ &+ o(n) \text{ a.s..} \end{aligned}$$

Thus

$$\inf_{(\lambda,\mu)\in\bar{U}_\epsilon} l(\lambda, \mu, \sigma^2 | X^{(o)}) - l\left(\lambda_0, \mu_0, \frac{\lambda\mu\sigma^2}{\lambda_0\mu_0} | X^{(o)}\right) \to \infty \text{ a.s.} \tag{2.35}$$

with respect to P_0 as $m, n \to \infty$. This implies that $\hat{\lambda} \to \lambda_0$ and $\hat{\mu} \to \mu_0$, together with (2.34), we have, almost surely

$$\hat{\lambda} \to \lambda_0, \hat{\mu} \to \mu_0 \text{ and } \hat{\sigma}^2 \to \sigma_0^2, \tag{2.36}$$

this concludes the consistency of the ML estimators, with the assumption that $(m_2 - m_1)(n_2 - n_1) < o(mn)$, and $\xi_j, \zeta_k < o(n^{-1/2})$. For asymptotic normality, we need $(m_2 - m_1)(n_2 - n_1) = o_p(n)$, but this is also attainable since $(m_2 - m_1)$, and $(n_2 - n_1)$ are both assumed to be $O(n^{1/2} - \epsilon_0)$, which implies that we have $(m_2 - m_1)(n_2 - n_1) = O(n^{1-2\epsilon}) = o_p(n)$.

2.2.5 Proof of asymptotic normality

In this section, we look at the asymptotic behavior of the distributions of $\hat{\lambda}\hat{\mu}\hat{\sigma}^2, \hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ in order to prove Theorem 2.2.2. Since these estimators are based on observations that are correlated, we will use a generalization of Lindeberg's central limit theorem.(For example, see [7] and [9]. In cases with sums of *i.i.d.* random variables, a major requirement in establishing asymptotic normality is the restriction on the magnitude of abnormally large observations, which is achieved by truncation of the random variables. In our case, instead of trying to control large elements in a sequence of random variables, we shift our focus to looking at the information as a martingale-difference array, and seek to control the magnitude of the *expectation* of large elements based on past behavior of the sequence. The condition imposed on the martingale difference sequence provides a version of the Lindeberg's condition, which is an essential characteristic of the central limit theorem.

To this end, let us look at the equations that lead to the MLE of λ, μ and σ^2. Notice that, by taking deriviatve of $l(\lambda, \mu, \sigma^2)$ with respect to each parameter and setting them to zero, we have

$$\begin{aligned}\frac{\partial}{\partial \sigma^2} l\left(\lambda, \mu, \sigma^2\right) = {} & mn - (m_2 - m_1 + 2)(n_2 - n_1 + 1)\sigma^2 - \frac{1}{\sigma^2} \underset{\sim}{X}_1' B^{-1}(\mu) \underset{\sim}{X}_1 \\ & - \frac{1}{\sigma^2}\left[\sum_{j \in J_o} \frac{\left(\underset{\sim}{X}_j - e^{-\lambda \xi_j} \underset{\sim}{X}_{j-1}\right)' B^{-1}(\mu) \left(\underset{\sim}{X}_j - e^{-\lambda \xi_j} \underset{\sim}{X}_{j-1}\right)}{1 - e^{-2\lambda \xi_j}}\right. \\ & \left. + \sum_{j=m_1}^{m_2+1} \frac{\left(\underset{\sim}{X}_j^* - e^{-\lambda \xi_j} \underset{\sim}{X}_{j-1}^*\right)' (B^*(\mu))^{-1} \left(\underset{\sim}{X}_j^* - e^{-\lambda \xi_j} \underset{\sim}{X}_{j-1}^*\right)}{1 - e^{-2\lambda \xi_j}}\right] \qquad (2.37)\end{aligned}$$

$$\frac{\partial}{\partial\lambda}l\left(\lambda,\mu,\sigma^2\right) = 2n\sum_{j\in J_o}\frac{\xi_j e^{-2\lambda\xi_j}}{1-e^{-2\lambda\xi_j}} + 2\left(n-\left(n_2-n_1+1\right)\right)\sum_{j=m_1}^{m_2+1}\frac{\xi_j e^{-2\lambda\xi_j}}{1-e^{-2\lambda\xi_j}}$$

$$+\frac{2}{\sigma^2}\Bigg[\sum_{j\in J_o}\frac{e^{-2\lambda\xi_j}}{1-e^{-2\lambda\xi_j}}\underset{\sim}{X}'_{j-1}B^{-1}\left(\mu\right)\left(\underset{\sim}{X}_j - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}\right)$$

$$+\sum_{j=m_1}^{m_2+1}\frac{e^{-2\lambda\xi_j}}{1-e^{-2\lambda\xi_j}}\left(\underset{\sim}{X}^*_{j-1}\right)'\left(B^*(\mu)\right)^{-1}\left(\underset{\sim}{X}^*_j - e^{-\lambda\xi_j}\underset{\sim}{X}^*_{j-1}\right)\Bigg]$$

$$-\frac{2}{\sigma^2}\Bigg[\sum_{j\in J_o}\frac{\xi_j e^{-2\lambda\xi_j}}{1-e^{-2\lambda\xi_j}}\frac{\left(\underset{\sim}{X}_j - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}\right)'B^{-1}\left(\mu\right)\left(\underset{\sim}{X}_j - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}\right)}{1-e^{-2\lambda\xi_j}}$$

$$+\sum_{j=m_1}^{m_2+1}\frac{\xi_j e^{-2\lambda\xi_j}}{1-e^{-2\lambda\xi_j}}\frac{\left(\underset{\sim}{X}^*_j - e^{-\lambda\xi_j}\underset{\sim}{X}^*_{j-1}\right)'\left(B^*(\mu)\right)^{-1}\left(\underset{\sim}{X}^*_j - e^{-\lambda\xi_j}\underset{\sim}{X}^*_{j-1}\right)}{1-e^{-2\lambda\xi_j}}\Bigg] \tag{2.38}$$

$$\frac{\partial}{\partial\mu}l(\lambda,\mu,\sigma^2) = 2m\sum_{k=2}^{n}\frac{\zeta_k e^{-2\mu\zeta_k}}{1-e^{-2\mu\zeta_k}} + (m_2-m_1+2)\sum_{k=2}^{n}\left[\frac{\zeta_k e^{-2\mu\zeta_k}}{1-e^{-2\mu\zeta_k}} - \frac{\zeta^*_{n_2+1}e^{-\mu\zeta^*_{n_2+1}}}{1-e^{-2\mu\zeta^*_{n_2+1}}}\right]$$

$$+\frac{1}{\sigma^2}\Bigg[\underset{\sim}{X}'_1 D_\mu B^{-1}\left(\mu\right)\underset{\sim}{X}_1$$

$$+\sum_{j\in J_o}\frac{\left(\underset{\sim}{X}_j - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}\right)'D_\mu B^{-1}\left(\mu\right)\left(\underset{\sim}{X}_j - e^{-\lambda\xi_j}\underset{\sim}{X}_{j-1}\right)}{1-e^{-2\lambda\xi_j}}$$

$$+\sum_{j=m_1}^{m_2+1}\frac{\left(\underset{\sim}{X}^*_j - e^{-\lambda\xi_j}\underset{\sim}{X}^*_{j-1}\right)'D_\mu(B^*(\mu))^{-1}\left(\underset{\sim}{X}^*_j - e^{-\lambda\xi_j}\underset{\sim}{X}^*_{j-1}\right)}{1-e^{-2\lambda\xi_j}}\Bigg] \tag{2.39}$$

Setting (2.37) = 0, we have an explicit expression of $\hat{\sigma}^2$ in terms of $\hat{\lambda}$ and $\hat{\mu}$,

$$\hat{\sigma}^2 = \frac{1}{mn - (m_2 - m_1 + 2)(n_2 - n_1 + 1)} \times \left[\underset{\sim}{X}_1' B^{-1}(\hat{\mu}) \underset{\sim}{X}_1 \right.$$

$$+ \sum_{j \in J_o} \frac{\left(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1}\right)' B^{-1}(\hat{\mu}) \left(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1}\right)}{1 - e^{-2\hat{\lambda}\xi_j}}$$

$$\left. + \sum_{j=m_1}^{m_2+1} \frac{\left(\underset{\sim}{X}_j^* - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1}^*\right)' (B^*(\hat{\mu}))^{-1} \left(\underset{\sim}{X}_j^* - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1}^*\right)}{1 - e^{-2\hat{\lambda}\xi_j}} \right] \tag{2.40}$$

which also implies that

$$(mn - (m_2 - m_1 + 2)(n_2 - n_1 + 1))\hat{\sigma}^2 - \underset{\sim}{X}_1' B^{-1}(\hat{\mu}) \underset{\sim}{X}_1$$

$$= \sum_{j \in J_o} \frac{\left(\underset{\sim}{X}_j - e^{-\lambda\xi_j} \underset{\sim}{X}_{j-1}\right)' B^{-1}(\mu) \left(\underset{\sim}{X}_j - e^{-\lambda\xi_j} \underset{\sim}{X}_{j-1}\right)}{1 - e^{-2\lambda\xi_j}}$$

$$+ \sum_{j=m_1}^{m_2+1} \frac{\left(\underset{\sim}{X}_j^* - e^{-\lambda\xi_j} \underset{\sim}{X}_{j-1}^*\right)' (B^*(\mu))^{-1} \left(\underset{\sim}{X}_j^* - e^{-\lambda\xi_j} \underset{\sim}{X}_{j-1}^*\right)}{1 - e^{-2\lambda\xi_j}}. \tag{2.41}$$

From (2.40), we see that $\hat{\sigma}^2$ can be explicitly written as a function of $\hat{\lambda}$ and $\hat{\mu}$. Now, utilizing the consistency result for $\hat{\lambda}$ and $\hat{\mu}$, as well as the approximations derived from (2.21) - (2.24), we have

$$\hat{\sigma}^2 = \frac{1}{mn - m^*n^*}\left[\frac{\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}}\left(\sum_{j\in J_o}\sum_{k=2}^{n} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\right)\left(w_{j,k}^2 - 1\right) + O(n)\right] + \frac{\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}}$$

$$= \frac{1}{mn - m^*n^*}\left[\frac{\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}}\left(\sum_{j\in J_o}\sum_{k=2}^{n} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\right)\left(w_{j,k}^2 - 1\right)\right] + \frac{\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}} + o_p(1),$$

which implies that

$$\sqrt{mn - m^*n^*}(\hat{\lambda}\hat{\mu}\hat{\sigma}^2 - \lambda_0\mu_0\sigma_0^2) = \frac{\lambda_0\mu_0\sigma_0^2\left(\sum_{j\in J_o}\sum_{k=2}^{n} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\right)\left(w_{j,k}^2 - 1\right)}{\sqrt{mn - m^*n^*}} + o_p(1) \tag{2.42}$$

Now, since $\{w_{j,k}^2 - 1\}$ is a sequence of independent and centered χ_1^2 random variables, we have

$$\text{Var}\left[\frac{\lambda_0\mu_0\sigma_0^2\left(\sum_{j\in J_o}\sum_{k=2}^{n} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\right)\left(w_{j,k}^2 - 1\right)}{\sqrt{mn - m^*n^*}}\right] = 2(\lambda_0\mu_0\sigma_0^2)^2$$

Then, by the central limit theorem, we have

$$\sqrt{mn - m^*n^*}(\hat{\lambda}\hat{\mu}\hat{\sigma}^2 - \lambda_0\mu_0\sigma_0^2) \to_D N\left(0, 2(\lambda_0\mu_0\sigma_0^2)^2\right),$$

where $m^* = (m_2 - m_1 + 2)$ and $n^* = (n_2 - n_1 + 2)$. On the other hand, due to their highly nonlinear equations, we cannot express $\hat{\lambda}$ and $\hat{\mu}$ in explicit forms. Thus, to investigate the asymptotic properties of their distribution, we would need to utilize different approximation techniques similar to those used to prove their consistency.

2.2.6 Approximating $\frac{\partial}{\partial\lambda} l(\lambda, \mu, \sigma^2) = 0$

Claim 2.2.8. $\dfrac{\partial}{\partial\lambda} l(\lambda, \mu, \sigma^2) = 0$ *can be expressed as*

$$\begin{aligned}
0 = & \frac{n(m-1)}{\hat{\lambda}} - n\sum_{j=2}^{m} \xi_j - 2(n_2 - n_1 + 1)\sum_{j=m_1}^{m_2+1} \frac{\xi_j e^{-\hat{\lambda}\xi_j}}{1 - e^{-2\hat{\lambda}\xi_j}} \\
& + \frac{mn - (m_2 - m_1 + 2)(n_2 - n_1 + 1)}{\hat{\lambda}} + \frac{\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\sum_{k=2}^{n} \gamma_{1,k}^2 \\
& + \frac{\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\left(\sum_{j\in J_o}\sum_{k=2}^{n} \xi_j w_{j,k}^2 + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o} \xi_j w_{jk}^2\right) \\
& + \frac{\mu_0\sigma_0^2\sqrt{2\lambda_0}}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\left(\sum_{j\in J_o}\sum_{k=2}^{n} {\xi_j}^{1/2} w_{j,k}\gamma_{j-1,k} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o} {\xi_j}^{1/2} w_{j,k}\gamma_{j-1,k}\right) \\
& + \frac{(\hat{\lambda} - \lambda_0)\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\left(\sum_{j\in J_o}\sum_{k=2}^{n} \xi_j \gamma_{j-1,k}^2 + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o} \xi_j \gamma_{j-1,k}^2\right) + o_p\left(n^{1/2}\right)
\end{aligned}$$

Proof of Claim 2.2.8. Write

$$\frac{\partial}{\partial\lambda} l(\lambda, \mu, \sigma^2) = 0 = \mathbf{L_1} - \frac{2}{\hat{\sigma}^2}\mathbf{L_2} + \frac{2}{\hat{\sigma}^2}\mathbf{L_3},$$

where

$$\mathbf{L_1} = 2n \sum_{j \in J_o} \frac{\xi_j e^{-2\hat{\lambda}\xi_j}}{1 - e^{-2\hat{\lambda}\xi_j}} + 2(n - (n_2 - n_1 + 1)) \sum_{j=m_1}^{m_2+1} \frac{\xi_j e^{-2\hat{\lambda}\xi_j}}{1 - e^{-2\hat{\lambda}\xi_j}},$$

$$\mathbf{L_2} = \sum_{j \in J_o} \frac{\xi_j e^{-2\hat{\lambda}\xi_j}}{1 - e^{-2\hat{\lambda}\xi_j}} \frac{\left(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1} \right)' B^{-1}(\hat{\mu}) \left(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1} \right)}{1 - e^{-2\hat{\lambda}\xi_j}}$$

$$+ \sum_{j=m_1}^{m_2+1} \frac{\xi_j e^{-2\hat{\lambda}\xi_j}}{1 - e^{-2\hat{\lambda}\xi_j}} \frac{\left(\underset{\sim}{X}_j^* - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1}^* \right)' (B^*(\hat{\mu}))^{-1} \left(\underset{\sim}{X}_j^* - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1}^* \right)}{1 - e^{-2\hat{\lambda}\xi_j}},$$

$$\mathbf{L_3} = \sum_{j \in J_o} \frac{e^{-2\hat{\lambda}\xi_j}}{1 - e^{-2\hat{\lambda}\xi_j}} \underset{\sim}{X}'_{j-1} B^{-1}(\hat{\mu}) \left(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1} \right),$$

$$+ \sum_{j=m_1}^{m_2+1} \frac{e^{-2\hat{\lambda}\xi_j}}{1 - e^{-2\hat{\lambda}\xi_j}} (\underset{\sim}{X}_{j-1}^*)' (B^*(\hat{\mu}))^{-1} \left(\underset{\sim}{X}_j^* - e^{-\hat{\lambda}\xi_j} \underset{\sim}{X}_{j-1}^* \right).$$

Utilizing the result from (2.41), the approximations from lemma (2.2.5), i.e.

$$\frac{\xi_j e^{-2\hat{\lambda}\xi_j}}{1 - e^{-2\hat{\lambda}\xi_j}} = \frac{1}{2\hat{\lambda}} - \frac{\xi_j}{2} + O({\xi_j}^2),$$

and also

$$\frac{\xi_j}{1 - e^{-2\hat{\lambda}\xi_j}} = \frac{1}{2\hat{\lambda}} + o(1),$$

we have

$$\mathbf{L_2} = \frac{1}{2\hat{\lambda}}\left[mn\hat{\sigma}^2 - (m_2 - m_1 + 2)(n_2 - n_1 + 1)\hat{\sigma}^2 - \underset{\sim}{X}_1' B^{-1}(\hat{\mu})\underset{\sim}{X}_1\right] \tag{2.43}$$

$$- \frac{1}{4\hat{\lambda}}\left[\sum_{j\in J_o}\left(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)' B^{-1}(\hat{\mu})\left(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)\right.$$

$$\left. + \sum_{j=m_1}^{m_2+1}\left(\underset{\sim}{X}_j^* - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}^*\right)' (B^*(\hat{\mu}))^{-1}\left(\underset{\sim}{X}_j^* - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}^*\right)\right] \tag{2.44}$$

$$+ O(1)\left[\sum_{j\in J_o}\xi_j\left(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)' B^{-1}(\hat{\mu})\left(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)\right.$$

$$\left. + \sum_{j=m_1}^{m_2+1}\xi_j\left(\underset{\sim}{X}_j^* - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}^*\right)' (B^*(\hat{\mu}))^{-1}\left(\underset{\sim}{X}_j^* - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}^*\right)\right]. \tag{2.45}$$

Our goal in here is to find a way to approximate (2.43) – (2.45). First, let us focus on (2.44). By repeatedly using the transformations defined in lemma (2.2.3), and expanding the quadratic forms using lemma (1.2.1), and applying approximation techniques to the resulting terms, we have

$$\sum_{j\in J_o}(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1})' B^{-1}(\hat{\mu})(\underset{\sim}{X}_j - e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1})$$

$$= \sum_{j\in J_o}\left[\sigma_0^2(1 - e^{-2\lambda_0\xi_j})\eta_j' B^{-1}(\hat{\mu})\eta_j + (e^{-2\lambda_0\xi_j} - e^{-2\hat{\lambda}\xi_j})^2 \underset{\sim}{X}_{j-1}' B^{-1}(\hat{\mu})\underset{\sim}{X}_{j-1}\right.$$

$$\left. + 2\sigma_o(1 - e^{-2\lambda_0\xi_j})(e^{-\lambda_0\xi_j} - e^{-\lambda\xi_j})\eta_j' B^{-1}(\hat{\mu})\underset{\sim}{X}_{j-1}\right]$$

$$= \sum_{j\in J_o}\left[2\lambda_0\sigma_0^2\eta_j' B^{-1}(\hat{\mu})\eta_j + O(1){\xi_j}^2\underset{\sim}{X}_{j-1}' B^{-1}(\hat{\mu})\underset{\sim}{X}_{j-1} + O(1){\xi_j}^2\eta_j' B^{-1}(\hat{\mu})\underset{\sim}{X}_{j-1}\right]$$

where

$$2\lambda_0\sigma_0^2\eta_j' B^{-1}(\hat{\mu})\eta_j$$

$$= 2\lambda_0\sigma_0^2\sum_{j\in J_o}\xi_j\left[\eta_{j,1}^2 + \sum_{k=2}^{n}\frac{1-e^{-2\mu_0\zeta_k}}{1-e^{-2\hat{\mu}\zeta_k}}w_{j,k}^2 + \sum_{k=2}^{n}\frac{\left(e^{-\mu_0\zeta_k}-e^{-\hat{\mu}\zeta_k}\right)^2}{\left(1-e^{-2\hat{\mu}\zeta_k}\right)}\eta_{j,k-1}^2\right.$$

$$\left.+\sum_{k=2}^{n}\frac{\left(e^{-\mu_0\zeta_k}-e^{\hat{\mu}\zeta_k}\right)\left(1-e^{-2\mu_0\zeta_k}\right)^{1/2}}{1-e^{-2\hat{\mu}\zeta_k}}w_{j,k}\eta_{j,k-1}\right]$$

$$= \frac{2\lambda_0\mu_0\sigma_0^2}{\hat{\mu}}\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j w_{jk}^2$$

$$+ O(1)\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k\eta_{j,k-1}^2 + O(1)\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k^{1/2} + O(1)w_{j,k}\eta_{j,k-1}$$

$$= \frac{2\lambda_0\mu_0\sigma_0^2}{\hat{\mu}}\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j w_{j,k}^2 + o_p\left(n^{1/2}\right), \tag{2.46}$$

since

$$\mathbb{P}\left(\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j{\zeta_k}^{1/2}w_{j,k}\eta_{j,k-1} > n^{\epsilon}\right) \leq \frac{\sum_{j\in J_o}\sum_{k=2}^{n}\zeta_k{\xi_j}^2\mathbb{E}\left[w_{j,k}^2\right]\mathbb{E}\left[\eta_{j,k-1}^2\right]}{n^{2\epsilon}}$$

$$= \frac{(n-1)(m-m2-m_1-2)}{n^{2\epsilon+3/2}} = O\left(n^{\frac{1}{2}-\epsilon_0}\right)$$

where the last equality is a result from the assumption that $\xi_j, \zeta_k \leq o(n^{-1/2})$ and the independence between $w_{j,k}$. Also, the non-negativity of $\eta^2_{j,k-1}$ implies that

$$\begin{aligned}\mathbb{P}(\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k\eta^2_{j,k-1} > n^\epsilon) &\leq \frac{\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\zeta_k\mathbb{E}\left[\eta^2_{j,k-1}\right]}{n^\epsilon} \\ &= \frac{(n-1)(m-m_2-m_1-2)}{n^{1+\epsilon}} = O(n^{1-\epsilon}),\end{aligned}$$

which gives us (2.46). On the other hand, we have

$$\begin{aligned}\sum_{j\in J_o}{\xi_j}^2\underset{\sim}{X}'_{j-1}B^{-1}(\hat{\mu})\underset{\sim}{X}_{j-1} = O(1)\sum_{j\in J_o}{\xi_j}^2\Bigg[x^2_{j-1,1} + \sum_{k=2}^{n}\gamma^2_{j-1,k} \\ + \sum_{k=2}^{n}{\zeta_k}^{1/2}\gamma_{j-1,k}x_{j-1,k-1} + \sum_{k=2}^{n}\zeta_k x^2_{j-1,k-1}\Bigg],\end{aligned}$$

and

$$\begin{aligned}\sum_{j\in J_o}{\xi_j}^2\eta'_jB^{-1}(\hat{\mu})\underset{\sim}{X}_{j-1} = O(1)\sum_{j\in J_o}{\xi_j}^2\Bigg[\sum_{k=2}^{n}w^2_{j,k} + \sum_{k=2}^{n}{\zeta_k}^{1/2}w_{j,k}x_{j-1,k-1} \\ + \sum_{k=2}^{n}{\zeta_k}^{1/2}\eta_{j,k-1}\gamma_{j-1,k-1} + \sum_{k=2}^{n}\zeta_k x^2_{j-1,k-1}\Bigg].\end{aligned}$$

In particular, notice that by choosing $\epsilon = 1/2$, we have (and similarly for $\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j\gamma^2_{j-1,k}$),

$$\mathbb{P}\left(\sum_{j\in J_o}\sum_{k=2}^{n}{\xi_j}^2w^2_{j,k} > n^\epsilon\right) \leq \frac{(n-1)(m-m_2-m1-2)}{n^{1+\epsilon}} = O\left(n^{1-\epsilon}\right) = o_p\left(n^{1/2}\right).$$

Also,

$$
\begin{aligned}
\mathbb{P}\left(\sum_{j\in J_o}\sum_{k=2}^{n}{\xi_j}^2{\zeta_k}^{1/2}w_{j,k}x_{j-1,k-1} > n^{\epsilon}\right) &\leq \frac{\sum_{j\in J_o}\sum_{k=2}^{n}{\xi_j}^4\zeta_k\mathbb{E}[w_{j,k}^2\mathbb{E}[x_{j-1,k-1}^2]]}{n^{2\epsilon}} \\
&= O(1)\frac{(n-1)(m-m_2-m_1-2)}{n^{2\epsilon+3}} \\
&= O\left(n^{-1-2\epsilon}\right) < o_p\left(n^{1/2}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{P}\left(\sum_{j\in J_o}\sum_{k=2}^{n}{\xi_j}^2\zeta_k x_{j-1,k-1}^2\right) &\leq \frac{\sum_{j\in J_o}\sum_{k=2}^{n}\mathbb{E}\left[x_{j-1,k-1}^2\right]}{n^{\epsilon+3/2}} \\
&= O\left(n^{1/2-\epsilon}\right) = o_p\left(n^{1/2}\right).
\end{aligned}
$$

The rest of the terms are approximated similarly. Therefore, we have

$$
{\xi_j}^2\underset{\sim}{X}'_{j-1}B^{-1}(\hat{\mu})\underset{\sim}{X}_{j-1} + \sum_{j\in J_o}{\xi_j}^2\eta_j' B^{-1}(\hat{\mu})\underset{\sim}{X}_{j-1} = o_p\left(n^{1/2}\right). \tag{2.47}
$$

Combing the above results from (2.46), (2.47), and applying the exact same procedue for $j = m_1, \ldots m_2+1$, we have

$$
(2.44) = -\frac{\lambda_0\mu_0\sigma_0^2}{2\hat{\lambda}\hat{\mu}}\left(\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j w_{j,k}^2 + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\xi_j w_{j,k}^2\right) + o_p\left(n^{1/2}\right).
$$

Now, for (2.45), we want to approximate

$$\sum_{j\in J_o}\xi_j\left(\underset{\sim}{X}_j-e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)'B^{-1}(\hat{\mu})\left(\underset{\sim}{X}_j-e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)$$
$$+\sum_{j=m_1}^{m_2+1}\xi_j\left(\underset{\sim}{X}^*_j-e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}^*_{j-1}\right)'\left(B^*(\hat{\mu})\right)^{-1}\left(\underset{\sim}{X}^*_j-e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}^*_{j-1}\right).$$

Similarly to previous terms, since the first and second quadratic forms are essentially the same execpt with different number of coulmn and row elements, we show only the expansion of the quadratic form from the columns with complete observations, which is

$$\sum_{j\in J_o}\xi_j\left(\underset{\sim}{X}_j-e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)'B^{-1}(\hat{\mu})\left(\underset{\sim}{X}_j-e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)$$
$$=O(1)\sum_{j\in J_o}\left[{\xi_j}^2\eta_j'B^{-1}(\hat{\mu})\eta_j+{\xi_j}^3\eta_j'B^{-1}(\hat{\mu})\underset{\sim}{X}_{j-1}+{\xi_j}^3\underset{\sim}{X}_{j-1}B^{-1}(\mu)\underset{\sim}{X}_{j-1}\right].$$

Using what we have already shown for (2.44), we have (2.45) $=o_p\left(n^{1/2}\right)$ as well. Now, focusing on (2.43), notice that

$$\underset{\sim}{X}_1'B^{-1}(\mu)\underset{\sim}{X}_1=x_{1,1}^2+\sum_{k=2}^{n}\frac{\left(x_{1,k}-e^{-\mu\zeta_k}x_{1,k-1}\right)^2}{1-e^{-\hat{\mu}\zeta_k}}$$
$$=x_{1,1}^2+\frac{\mu_0\sigma_0^2}{\hat{\mu}}\sum_{k=2}^{n}\gamma_{1,k}^2+O(1)\sum_{k=2}^{n}\zeta_k x_{1,k-1}^2+O(1)\sum_{k=2}^{n}{\zeta_k}^{1/2}\gamma_{1,k}x_{1,k-1},$$

and since

$$\mathbb{P}\left(\sum_{k=2}^{n}\zeta_k x_{1,k-1}^2 > n^{\epsilon}\right) \leq \frac{\sum_{k=2}^{n}\zeta_k \mathbb{E}\left[x_{1,k-1}^2\right]}{n^{\epsilon}}$$

$$= O(1)\frac{n}{n^{1/2+\epsilon}} = O\left(n^{1/2-\epsilon_0}\right),$$

$$\mathbb{P}\left(\sum_{k=2}^{n}\zeta_k{}^{1/2}\gamma_{1,k}x_{1,k-1}\right) \leq \frac{\sum_{k=2}^{n}\zeta_k \mathbb{E}\left[\gamma_{1,k}^2\right]\mathbb{E}\left[x_{1,k-1}^2\right]}{n^{2\epsilon}}$$

$$= O(1)\frac{n}{n^{2\epsilon+1/2}} = O\left(n^{1/2-\epsilon_0}\right).$$

Therefore,

$$\underset{\sim}{X}_1' B^{-1}(\mu)\underset{\sim}{X}_1 = \frac{\mu_0\sigma_0^2}{\hat{\mu}}\sum_{k=2}^{n}\gamma_{1,k}^2 + o_p\left(n^{1/2}\right),$$

as a result,

$$\mathbf{L_2} = \frac{\hat{\sigma}^2}{2\hat{\lambda}}\left(mn - (m_2 - m_1 + 2)(n_2 - n_1 + 1)\right) - \frac{\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}}\sum_{k=2}^{n}\gamma_{1,k}^2$$

$$- \frac{2\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}}\left[\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j w_{j,k}^2 + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\xi_j w_{j,k}^2\right] + o_p\left(n^{1/2}\right).$$

To approximate **L3**, notice that

$$\sum_{j\in J_o}\frac{\xi_j e^{-\hat{\lambda}\xi_j}}{1-e^{-2\hat{\lambda}\xi_j}}\underset{\sim}{X}'_{j-1}B^{-1}\left(\hat{\mu}\right)\left(\underset{\sim}{X}_j-e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)$$

$$=\frac{1}{2\hat{\lambda}}\sum_{j\in J_o}\underset{\sim}{X}'_{j-1}B^{-1}\left(\hat{\mu}\right)\left(\underset{\sim}{X}_j-e^{-\hat{\lambda}\xi_j}\underset{\sim}{X}_{j-1}\right)$$

$$=\frac{\sigma_0}{2\hat{\lambda}}\sum_{j\in J_o}\left(1-e^{-2\lambda_0\xi_j}\right)\underset{\sim}{X}'_{j-1}B^{-1}\left(\hat{\mu}\right)\eta_j$$

$$+\frac{1}{2\hat{\lambda}}\sum_{j\in J_o}\left(e^{-\lambda_0\xi_j}-e^{-\lambda\xi_j}\right)\underset{\sim}{X}'_{j-1}B^{-1}\left(\hat{\mu}\right)\underset{\sim}{X}_{j-1}$$

$$=\frac{\sigma_0^2\sqrt{2\lambda_0}}{2\hat{\lambda}}\sum_{j\in J_o}\xi_j{}^{1/2}\underset{\sim}{X}'_{j-1}B^{-1}\left(\hat{\mu}\right)\eta_j+\frac{(\lambda_0-\hat{\lambda})}{2\hat{\lambda}}\sum_{j\in J_o}\xi_j\underset{\sim}{X}'_{j-1}B^{-1}\left(\hat{\mu}\right)\underset{\sim}{X}_{j-1}.$$

Now,

$$\sum_{j\in J_o}\xi_j{}^{1/2}\underset{\sim}{X}'_{j-1}B^{-1}\left(\hat{\mu}\right)\eta_j$$

$$=\sum_{j\in J_o}\xi_j{}^{1/2}\left[x_{j-1,1}\eta_{j,1}+\sum_{k=2}^{n}\frac{\left(x_{j-1,k}-e^{-\hat{\mu}\zeta_k}x_{j-1,k-1}\right)\left(\eta_{j,k}-e^{-\hat{\mu}\zeta_k}\eta_{j,k-1}\right)}{1-e^{-2\hat{\mu}\zeta_k}}\right]$$

$$=\sum_{j\in J_o}\xi_j{}^{1/2}\left[\sigma_0^2\frac{\left(1-e^{-2\mu_0\zeta_k}\right)}{\left(1-e^{-2\hat{\mu}\zeta_k}\right)}w_{j,k}\gamma_{j-1,k}+O(1)\text{ a.s.},\right.$$

$$+\sum_{k=2}^{n}\frac{\left(e^{-\mu_0\zeta_k}-e^{-\hat{\mu}\zeta_k}\right)^2}{1-e^{-2\hat{\mu}\zeta_k}}x_{j-1,k-1}\eta_{j,k-1}$$

$$\left.+\sum_{k=2}^{n}\frac{\left(1-e^{-2\mu_0\zeta_k}\right)^{1/2}\left(e^{-\mu_0\zeta_k}-e^{-\hat{\mu}\zeta_k}\right)}{1-e^{-2\hat{\mu}\zeta_k}}\left(x_{j-1,k-1}w_{j,k}+\frac{\gamma_{j-1,k}}{\sigma_0^2}\eta_{j,k-1}\right)\right]$$

$$=\sum_{j\in J_o}\xi_j{}^{1/2}\left[\frac{\mu_0\sigma_0^2}{\hat{\mu}}\sum_{k=2}^{n}w_{j,k}\gamma_{j-1,k}+O(1)\right.$$

$$\left.+O(1)\sum_{k=2}^{n}\zeta_k x_{j-1,k-1}\eta_{j,k-1}+O(1)\sum_{k=2}^{n}\zeta_k{}^{1/2}\left(x_{j-1,k-1}w_{j,k}+\gamma_{j-1,k}\eta_{j,k-1}\right)\right]$$

$$=\frac{\mu_0\sigma_0^2}{\hat{\mu}}\sum_{j\in J_o}\sum_{k=2}^{n}\xi_j{}^{1/2}w_{j,k}\gamma_{j-1,k}+o_p\left(n^{1/2}\right),$$

where the last equality is due to similar arguments used in $\mathbf{L_2}$. Observe that, since

$x_{j-1,k-1}$ is independent of $w_{j,k}$ for j and k, by choosing $\epsilon_0 = \frac{1}{2} + \epsilon$, we have

$$\mathbb{P}(\sum_{j \in J_o} \sum_{k=2}^{n} {\xi_j}^{1/2} {\zeta_k}^{1/2} x_{j-1,k-1} w_{j,k} > n^{\epsilon}) \leq \frac{(n-1)(m - m_2 - m_1 - 2)}{n^{2\epsilon+1}}$$

$$= O(n^{1/2-\epsilon_0}) \text{ a.s.,}$$

$$= o_p(n^{1/2})$$

Similarly, for the second part of **L3**,

$$\sum_{j \in J_o} \xi_j \underset{\sim}{X}_{j-1} B^{-1} (\hat{\mu}) \underset{\sim}{X}_{j-1}$$

$$= \sum_{j \in J_o} \xi_j \left[\frac{\mu_0 \sigma_0^2}{\hat{\mu}} \sum_{k=2}^{n} \gamma_{j-1,k}^2 + O(1) {\zeta_k}^{1/2} x_{j-1,k-1} \gamma_{j-1,k} + O(1) \sum_{k=2}^{n} \zeta_k x_{j-1,k-1}^2 \right],$$

$$= \frac{\mu_0 \sigma_0^2}{\hat{\mu}} \sum_{j \in J_o} \sum_{k=2}^{n} \xi_j \gamma_{j-1,k}^2 + o_p\left(n^{1/2}\right),$$

since we have $\sum_{j \in J_o} \sum_{k=2}^{n} \xi_j {\zeta_k}^{1/2} x_{j-1,k-1} \gamma_{j-1,k} + \sum_{j \in J_o} \sum_{k=2}^{n} \xi_j \zeta x_{j-1,k-1}^2 = o_p\left(n^{1/2}\right)$ from previous arguments that were used to show $\mathbf{L_2}$. Therefore applying the exact same step to the terms involved for $j = m_1, \ldots, m_2 + 1$, we obtain

$$\mathbf{L3} = \frac{\mu_0 \sigma_0^2 \sqrt{2\lambda_0}}{2\hat{\lambda}\hat{\mu}} \left[\sum_{j \in J_o} \sum_{k=2}^{n} {\xi_j}^{1/2} w_{j,k} \gamma_{j-1,k} + \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} {\xi_j}^{1/2} w_{j,k} \gamma_{j-1,k} \right]$$

$$+ \frac{\left(\hat{\lambda} - \lambda_0\right) \mu_0 \sigma_0^2}{2\hat{\lambda}\hat{\mu}} \left[\sum_{j \in J_o} \sum_{k=2}^{n} \xi_j \gamma_{j-1,k}^2 + \sum_{j=m_1}^{m_2+1} \sum_{k \in K_o} \xi_j \gamma_{j-1,k}^2 \right].$$

Putting together $\mathbf{L_1}, \mathbf{L_2}$ and $\mathbf{L_3}$ gives us the desired result. □

So far, we show that expression from claim (2.2.8) can be written as the following

$$
\begin{aligned}
0 =& \frac{-n}{\hat{\lambda}} + (n_2 - n_1 + 1)\sum_{j=m_1}^{m_2+1} \xi_j - n\sum_{j=2}^{m} \xi_j + \frac{\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\left[\sum_{k=2}^{n}(\gamma_{1,k}^2 - 1)\right] + n\frac{\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2} \\
&+ \frac{\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\left(\sum_{j\in J_o}\sum_{k=2}^{n} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\right)\xi_j(w_{j,k}^2 - 1) + (n\sum_{j=2}^{m}\xi_j - (n_2 - n_1 + 1)\sum_{j=m_1}^{m_2+1})\frac{\lambda_0\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2} \\
&+ \frac{\mu_0\sigma_0^2\sqrt{2\lambda_0}}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\left(\sum_{j\in J_o}\sum_{k=2}^{n} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\right){\xi_j}^{1/2} w_{j,k}\gamma_{j-1,k} \\
&+ \frac{(\hat{\lambda} - \lambda_0)\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\left(\sum_{j\in J_o}\sum_{k=2}^{n} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\right)\xi_j(\gamma_{j-1,k}^2 - 1) \\
&+ n\frac{(\hat{\lambda} - \lambda_0)\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2} - (n_2 - n_1 + 1)\frac{(\hat{\lambda} - \lambda_0)\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\sum_{j=m_1}^{m_2+1}\xi_j + o_p\left(n^{1/2}\right).
\end{aligned}
$$

Notice that by lemma (2.2.4) part (4), we have the following approximations

$$
\left(\sum_{j\in J_o}\sum_{k=2}^{n} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\right)\xi_j(w_{j,k}^2 - 1) = o_p\left(n^{1/2}\right), \text{ and}
$$

$$
\left(\sum_{j\in J_o}\sum_{k=2}^{n} + \sum_{j=m_1}^{m_2+1}\sum_{k\in K_o}\right)\xi_j(\gamma_{j-1,k}^2 - 1) = o_p\left(n^{1/2}\right).
$$

Therefore,

$$0 = n\left[\frac{\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2} - \frac{1}{\hat{\lambda}} + \frac{(\hat{\lambda}-\lambda_0)\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\right] + \frac{\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\sum_{k=2}^{n}\left[\sqrt{2\lambda_0}\sum_{j=2}^{m}{\xi_j}^{1/2}w_{j,k}\gamma_{j-1,k} + \gamma_{1,k-1}^2\right]$$

$$- \frac{\mu_0\sigma_0^2}{\hat{\lambda}\hat{\mu}\hat{\sigma}^2}\sum_{j=m_1}^{m_2+1}\sum_{k=2}^{n_1-1}{\xi_j}^{1/2}w_{j,k}\gamma_{j-1,k} - (n_2-n_1+1)\frac{\hat{\lambda}-\lambda_0}{\mu_0\sigma_0^2}\hat{\lambda}\hat{\mu}\hat{\sigma}^2\sum_{j=m_1}^{m_2+1}\xi_j + o_p\left(n^{1/2}\right) \tag{2.48}$$

Now, recall that $(m_2 - m_1), (n_2 - n_1) < O\left(n^{1/2-\epsilon}\right)$, $0 < \epsilon < 1/2$. This implies that

$$(m_2 - m_1)(n_2 - n - 1) = O\left(n^{1-\epsilon_0}\right), 0 < \epsilon_0 < 1,$$

so

$$(n_2 - n_1 + 1)\sum_{j=m_1}^{m_2+1}\xi_j = O\left(n^{1-\epsilon_0-\frac{1}{2}}\right) = o\left(n^{1/2}\right) = o_p\left(n^{1/2}\right).$$

Similarly,

$$\begin{aligned}\mathbb{P}\left(\sum_{j=m_1}^{m_2+1}\sum_{k=n_1}^{n_2}{\xi_j}^{1/2}w_{j,k}\gamma_{j-1,k} > n^{\epsilon}\right) &\leq \frac{\sum_{j=m_1}^{m_2+1}\sum_{k=n_1}^{n_2}\xi_j\mathbb{E}\left[w_{j,k}^2\right]\mathbb{E}\left[\gamma_{j-1,k}^2\right]}{n^{2\epsilon}} \\ &\leq \frac{(m_2-m_1)(n_2-n_1)}{n^{2\epsilon+\frac{1}{2}}} \\ &= o_p\left(n^{1/2}\right).\end{aligned}$$

Together with the consistency results of $\hat{\lambda}\hat{\mu}\hat{\sigma}^2$, $\hat{\lambda}, \hat{\mu}$, and $\hat{\sigma}^2$ from (2.34) and (2.36), we can express (2.48) as

$$0 = \frac{n}{\lambda_0}\left(\frac{1+\lambda_0}{\lambda_0}\right)+\frac{1}{\lambda_0}\sum_{k=2}^{n}\left[\sqrt{2\lambda_0}\sum_{j=2}^{m}\xi_j{}^{1/2}w_{j,k}\gamma_{j-1,k}+\gamma_{1,k-1}^2-1\right]+o_p\left(n^{1/2}\right)$$

which gives us

$$\sqrt{n}(\hat{\lambda}-\lambda_0) = \frac{-\lambda_0}{\sqrt{n}(1+\lambda_0)}\times\sum_{k=2}^{n}\left[\sqrt{2\lambda_0}\sum_{j=2}^{m}\xi_j{}^{1/2}w_{j,k}\gamma_{j-1,k}+\gamma_{1,k-1}^2-1\right]+o_p\left(n^{1/2}\right), \tag{2.49}$$

interchanging the parameter and dimension, we have

$$\sqrt{m}(\hat{\mu}-\mu_0) = \frac{-\mu_0}{\sqrt{m}(1+\mu_0)}\times\sum_{j=2}^{m}\left[\sqrt{2\mu_0}\sum_{k=2}^{n}\zeta_k{}^{1/2}w_{j,k}\eta_{j,k-1}+\eta_{j-1,1}^2-1\right]+o_p\left(n^{1/2}\right). \tag{2.50}$$

2.2.7 Asymptotic normality for $\hat{\lambda}$ and $\hat{\mu}$

In the previous section, we have expressed the scaled difference of the parameter estimators $\hat{\lambda}, \hat{\mu}$ and the true values λ_0, μ_0 as a linear combination of $w_{j,k}\gamma_{j-1,k}$ and $w_{j,k}\eta_{j,k-1}$ respectively. Our goal in here is to show that

$$\begin{bmatrix}\sqrt{n}(\hat{\lambda}-\lambda_0)\\ \sqrt{m}(\hat{\mu}-\mu_0)\end{bmatrix} \to_{\mathcal{D}} N\left(\mathbf{0},\quad \mathbf{\Sigma_1}\right) \tag{2.51}$$

where

$$\mathbf{\Sigma_1} = \begin{bmatrix} 0 & \frac{2\lambda_0{}^2}{1+\lambda_0} \\ \frac{2\mu_0{}^2}{1+\mu_0} & 0 \end{bmatrix}.$$

To proceed with the proof of (2.51), notice that a random vector $(Y_1, \ldots, Y_d)$ is normally distributed if and only if any linear combination $\sum_{i=1}^{d} a_i Y_i$ is a normal random variable. Thus, for (2.51) to hold, it is sufficient to show that for $t \in \mathbb{R}$,

$$\sqrt{n}(\hat{\lambda} - \lambda_0) + t\sqrt{m}(\hat{\mu} - \mu_0) \to N\left(0, 2\left(\frac{{\lambda_0}^2}{1+\lambda_0} + t\frac{{\mu_0}^2}{1+\mu_0}\right)\right). \tag{2.52}$$

Notice that from (2.49) and (2.50) we have

$$\begin{aligned} &- n\frac{(1+\lambda_0)}{\lambda_0}\left(\hat{\lambda} - \lambda_0\right) - tm\frac{(1+\mu_0)}{\mu_0}\left(\hat{\mu} - \mu_0\right) \\ &= \sum_{k=2}^{n}\left[\sqrt{\frac{2\lambda_0}{n}}\sum_{j=2}^{m}{\xi_j}^{1/2}\gamma_{j-1,k}w_{j,k} + \left(\gamma_{1,k}^2 - 1\right) + t\sqrt{\frac{2\mu_0}{m}}\sum_{j=2}^{m}{\zeta_k}^{1/2}\eta_{j,k-1}w_{j,k}\right] \\ &+ \frac{t}{\sqrt{m}}\sum_{j=2}^{m}(\eta_{j,1}^2 - 1). \end{aligned} \tag{2.53}$$

Since we are dealing with functions of dependent random variables, following the strategies used by Ying in [24], we show the asymptotic normality of (2.52) by first viewing it as a combination of a martingale difference sequence and the sum of a sequence of *i.i.d.* chi-squared random variables. To this end, let us verify that we indeed have a martingale difference sequence in (2.52).

Definition 2.2.9. *Let $\lambda_0, \mu_0, \xi_j, \zeta_k, w_{j,k}, \eta_{j,k}$ and $\gamma_{j,k}$ be the same as previously defined, and let*

- $\mathcal{E}_k(m,n) = \sqrt{\frac{2\lambda_0}{n}}\sum_{j=2}^{m}{\xi_j}^{1/2}\gamma_{j-1,k}w_{j,k} + (\gamma_{1,k}^2 - 1) + t\sqrt{\frac{2\mu_0}{m}}\sum_{j=2}^{m}{\zeta_k}^{1/2}\eta_{j,k-1}w_{j,k}$
- $\mathcal{F}_k = \sigma(x_{1,l}, x_{2,l}, \ldots, x_{m,l}), l \leq k$

Claim 2.2.10. *For a fixed pair of m and n, $\mathcal{E}_k(m,n))$ is a martingale-difference array with respect to the σ- filtration $\mathcal{F}_k$.*

Proof of claim (2.2.10). We need to show the following

(1) $\mathcal{E}_k(m,n))$ is $\mathcal{F}_k$-measurable.

(2) $\mathbb{E}\left[\sum_{l=2}^{k}\mathcal{E}_l(m,n)) - \sum_{l=2}^{k-1}\mathcal{E}_l(m,n))|\mathcal{F}_{k-1}\right] = 0.$

Since, $\gamma_{j-1,k}, w_{j,k}$ and $\eta_{j,k}$ are functions of $x_{j,k}$ and $x_{j,k-1}$, so $\mathcal{E}_k(m,n))$ is $\mathcal{F}_k$-measurable by definition. Moreover, this implies that

$$\mathbb{E}\left[\gamma_{j-1,k}|\mathcal{F}_{k-1}\right] = \mathbb{E}[\gamma_{j-1,k}] = \mathbb{E}\left[w_{j,k}\right] = \mathbb{E}\left[w_{j,k}|\mathcal{F}_{k-1}\right] = 0$$

by independence, and

$$\mathbb{E}\left[\eta_{j,k-1}|\mathcal{F}_{k-1}\right] = \eta_{j,k-1},$$

since $\eta_{j,k-1} \in \mathcal{F}_{k-1}$. Therefore, $\mathcal{E}_k(m,n))$ is adapted to the filtration $\mathcal{F}_{k-1}$. Next, to show 2.2.10(2), notice that again by indenpendence and measurability, we have

$$\mathbb{E}\left[\sum_{l=2}^{k}\mathcal{E}_l(m,n))|\mathcal{F}_{k-1}\right]$$

$$= \sqrt{\frac{2\lambda_0}{n}}\sum_{j=2}^{m}\sum_{l=2}^{k}\xi_j{}^{1/2}\mathbb{E}\left[\gamma_{j-1,l}w_{j,l}|\mathcal{F}_{k-1}\right] + \frac{1}{\sqrt{n}}\sum_{l=2}^{k}\left(\mathbb{E}\left[\gamma_{1,l}^2|\mathcal{F}_{k-1}\right] - 1\right)$$

$$+ t\sqrt{\frac{2\mu_0}{m}}\zeta_l^{1/2}\sum_{j=2}^{m}\sum_{l=2}^{k}\mathbb{E}\left[\eta_{j,l-1}w_{j,l}|\mathcal{F}_{k-1}\right]$$

$$= \sqrt{\frac{2\lambda_0}{n}}\sum_{j=2}^{m}\sum_{l=2}^{k}\xi_j{}^{1/2}\mathbb{E}\left[\gamma_{j-1,l}|\mathcal{F}_{k-1}\right]\mathbb{E}\left[w_{j,l}|\mathcal{F}_{k-1}\right] + \frac{1}{\sqrt{n}}\sum_{l=2}^{k}\left(\mathbb{E}\left[\gamma_{1,l}^2|\mathcal{F}_{k-1}\right] - 1\right)$$

$$+ t\sqrt{\frac{2\mu_0}{m}}\zeta_l^{1/2}\sum_{j=2}^{m}\sum_{l=2}^{k}\mathbb{E}\left[\eta_{j,l-1}|\mathcal{F}_{k-1}\right]\mathbb{E}\left[w_{j,l}|\mathcal{F}_{k-1}\right]$$

$$= \sqrt{\frac{2\lambda_0}{n}}\sum_{j=2}^{m}\sum_{l=2}^{k-1}\xi_j{}^{1/2}\mathbb{E}\left[\gamma_{j-1,l}w_{j,l}|\mathcal{F}_{k-1}\right] + \frac{1}{\sqrt{n}}\sum_{l=2}^{k-1}\left(\mathbb{E}\left[\gamma_{1,l}^2|\mathcal{F}_{k-1}\right] - 1\right)$$

$$+ t\sqrt{\frac{2\mu_0}{m}}\zeta_l^{1/2}\sum_{j=2}^{m}\sum_{l=2}^{k-1}\mathbb{E}\left[\eta_{j,l-1}w_{j,l}|\mathcal{F}_{k-1}\right]$$

$$= \mathbb{E}\left[\sum_{l=2}^{k-1}\mathcal{E}_l(m,n))|\mathcal{F}_{k-1}\right]$$

$$= \sum_{l=2}^{k-1}\mathcal{E}_l(m,n).$$

□

Since (2.53) can be expressed as the row sum of a martingale-difference array, we can then utilize the martingale central limit theorem to show its asymptotic normality. Notice that, even though $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ are estimated based on the assumption that some of the observations are missing, asymptotically they can be viewed as the same as those estimated in the complete-data case, as long as the magnitude restriction of the missing rows and columns are satisfied. Also, unlike the central limit theorem for sums of *i.i.d.* random variables, we use a weaker version of the Lindeberg condition for the row sums of martingle-difference arrays.

Theorem 2.2.11 (Pollard 1984)**.** *Let $\{\mathcal{E}_k(m,n)\}$ be a martingale-difference array, and let*

$$\nu_{m,k} = \mathbb{E}\left[\mathcal{E}_k^2(m,n)|\mathcal{F}_{k-1}\right]$$

be a sequence of conditional variances for $k = 2, \dots, n$. If, as $m \to \infty$,

(1) $\sum_{k=2}^{n} \nu_{m,k} \to_p \nu,,$ *where* $\nu > 0$

(2) for every $\delta > 0, \sum_{k=2}^{n} \mathbb{P}\left(\mathcal{E}_k^2(m,n)\left\{|\mathcal{E}_k(m,n)| > \delta\right\}|\mathcal{F}_{k-1}\right) \to_p 0$

then $\sum_{k=2}^{n} \mathcal{E}_k(m,n) \to_D N(0,\nu)$.

Remark 2.2.12. *Using Chebychev's inequality, theorem 2.2.11* (2) *is equivalent to*

$$\sum_{k=2}^{n} \mathbb{E}\left[\mathcal{E}_k^4(m,n)\right] \to_p 0.$$

In order to utilize Pollard's central limit theorem, we need to show that

$$\sum_{k=2}^{n} \mathbb{E}\left[\mathcal{E}_k^2(m,n)|\mathcal{F}_{k-1}\right] \to_p 2\left(\frac{{\lambda_0}^2}{1+\lambda_0} + t\frac{{\mu_0}^2}{1+\mu_0}\right), \tag{2.54}$$

and

$$\sum_{k=2}^{n} \mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] \to_p 0. \tag{2.55}$$

To show (2.54), we look at each of the cross terms and squared terms from $\mathcal{E}_k^2(m,n)$. Notice that $w_{j,k}$ is independent of $\gamma_{j',k}$, for $j > j'$, thus

$$\mathbb{E}\left[(\gamma_{j-1,k}w_{j,k})(\gamma_{1,k}^2-1)|\mathcal{F}_{k-1}\right] = \mathbb{E}\left[w_{j,k}|\mathcal{F}_{k-1}]\mathbb{E}[\gamma_{j-1,k}(\gamma_{1,k}^2-1)|\mathcal{F}_{k-1}\right] = 0.$$

Similarly, since $\eta_{j,1}$ is independent of $\gamma_{j-1,k}$, we have

$$\mathbb{E}\left[\gamma_{j_1-1,k}w_{j_1,k}(\eta_{j_2,1}^2-1)|\mathcal{F}_{k-1}\right] = \mathbb{E}\left[(\eta_{j_2,1}^2-1)|\mathcal{F}_{k-1}\right]\mathbb{E}\left[\gamma_{j_1-1,k}w_{j,k}|\mathcal{F}_{k-1}\right] = 0.$$

Next, for $j_1 \neq j_2$,

$$\mathbb{E}\left[\gamma_{j_1-1,k}w_{j_1,k}\eta_{j_2-1,k}w_{j_2,k}|\mathcal{F}_{k-1}\right] = \mathbb{E}\left[w_{j_1,k}|\mathcal{F}_{k-1}\right]\mathbb{E}\left[\gamma_{j_1-1,k}\eta_{j_2-1,k}w_{j_2,k}|\mathcal{F}_{k-1}\right] = 0,$$

and for $j_1 = j_2$,

$$\mathbb{E}\left[\gamma_{j-1,k}w_{j,k}\eta_{j-1,k-1}w_{j,k}\right] = \mathbb{E}\left[\gamma_{j-1,k}w_{j,k}|\mathcal{F}_{k-1}]\mathbb{E}[\eta_{j-1,k-1}w_{j,k}|\mathcal{F}_{k-1}\right] = 0.$$

Thus all the cross terms have expectation zero conditioning on $\mathcal{F}_{k-1}$. Now, for the squared terms, we have

$$\mathbb{E}\left[\gamma_{j-1,k}^2w_{j,k}^2|\mathcal{F}_{k-1}\right] = \mathbb{E}\left[\gamma_{j-1,k}^2|\mathcal{F}_{k-1}]\mathbb{E}[w_{j,k}^2|\mathcal{F}_{k-1}\right] = 1$$

and

$$\mathbb{E}\left[\eta_{j,k-1}^2w_{j,k}^2\right] = \eta_{j,k-1}^2\mathbb{E}\left[w_{j,k}^2\right] = \eta_{j,k-1}^2.$$

As a result, we can write

$$\sum_{k=2}^{n} \mathbb{E}\left[\mathcal{E}_k^2(m,n)|\mathcal{F}_{k-1}\right] = \frac{2\lambda_0}{n}\sum_{k=2}^{n}\sum_{j=2}^{m}\xi_j \mathbb{E}\left[\gamma_{j-1,l}^2\right]\mathbb{E}\left[w_{j,k}^2\right] + \frac{1}{n}\sum_{k=2}^{n}\mathbb{E}\left[(\gamma_{1,k}^2-1)^2\right]$$

$$+ \frac{2t^2\mu_0}{m}\sum_{k=2}^{n}\sum_{j=2}^{m}\zeta_k\eta_{j,k-1}^2\mathbb{E}\left[w_{j,k}^2\right]$$

$$= \frac{2}{n}\sum_{k=2}^{n}(\lambda_0+1) + \frac{2t^2\mu_0}{m}\sum_{j=2}^{m}\eta_{j,k}^2$$

$$= 2\left(\lambda_0 + 1 + t^2\mu_0(1 + \frac{1}{m}\sum_{k=2}^{n}\sum_{j=2}^{m}\zeta_k(\eta_{j,k-1}^2 - 1))\right)$$

$$= 2\left(\lambda_0 + 1 + t^2\mu_0\left(1 + o_p(n^{-1}\right)\right),$$

where the last equality is due to the fact that

$$\mathbb{P}(\sum_{k=2}^{n}\zeta_k(\eta_{j,k-1}^2 - 1) > n^{\epsilon}) \le \frac{\sum_{k=2}^{n}{\zeta_k}^2\mathbb{E}[(\eta_{j,k-1}^2-1)^2]}{n^{2\epsilon}}$$

$$\le \frac{2n}{n^{2\epsilon+1}}$$

$$= o_p(n^{-1}).$$

Therefore, we have

$$\sum_{k=2}^{n}\mathbb{E}[\mathcal{E}_k^2(m,n)|\mathcal{F}_{k-1}] \to_p 2(\lambda_0 + 1 + t^2\mu_0).$$

Now, to show (2.55), first notice that by the multinomial theorem, we have

$$\sum_{k=2}^{n}\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] = \sum_{k=2}^{n}\sum_{|s|=4}\binom{4}{s_1, s_2, s_3}\left(\sqrt{\frac{2\lambda_0}{n}}\sum_{j=2}^{m}\xi_j^{1/2}\gamma_{j-1,k}w_{j,k}\right)^{s_1}$$

$$\times\left(\frac{\gamma_{1,k}^2-1}{\sqrt{n}}\right)^{s_2}\left(t\sqrt{\frac{2\mu_0}{m}}\sum_{j=2}^{m}\zeta_k^{1/2}\eta_{j,k-1}w_{j,k}\right)^{s_3},$$

where $|s| = s_1 + s_2 + s_3$.

Combination type	$(\mathbf{s_1}, \mathbf{s_2}, \mathbf{s_3})$
Type A	$(0,1,3)$,$(0,3,1)$ $(1,0,3)$,$(1,1,2)$,$(1,2,1)$,$(1,3,0)$ $(2,1,1)$,$(3,0,1)$,$(3,1,0)$
Type B	$(4,0,0)$, $(0,0,4)$, $(0,4,0)$
Type C	$(2,2,0)$, $(0,2,2)$, $(2,0,2)$

Table 2.2: An list of possible exponent combinations for $\mathbb{E}\left[\epsilon_k^4(m,n)|\mathcal{F}_{k-1}\right]$.

There are 15 possible combinations for (s_1, s_2, s_3), as organized in Table 2.2 above. To show that each of them converge to zero in probability, we will utilize the assumptions that $\zeta_k, \xi_j \leq o(n^{-1/2})$, $\sum_{k=2}^{n}\zeta_k, \sum_{j=2}^{m}\xi_j \leq 1$, and vairous independence properties from lemma (2.2.4). Now, looking at each combination type, we see that the index combination from Type A all contain a term with an exponent of 1. In particular, when $s_2 = 1$, we can easily show that $\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] = 0$ by factoring out $\mathbb{E}\left[(\gamma_{1,k}^2 - 1)|\mathcal{F}_{k-1}\right]$, which is equal to zero, from the rest of the term. On the other hand, when $s_2 = 3$, we can factor out $\mathbb{E}\left[\gamma_{j-1,k} w_{j,k} | \mathcal{F}_{k-1}\right]$ or $\mathbb{E}\left[\eta_{j,k-1} w_{j,k} | \mathcal{F}_{k-1}\right]$, in either case will again give us $\mathbb{E}\left[\mathcal{E}_k^4(m,n) | \mathcal{F}_{k-1}\right] = 0$. Now, when $s_2 = 2$, we have

$$
\begin{aligned}
&\sum_{k=2}^{n} \mathbb{E}\left[\mathcal{E}_k^4(m,n) | \mathcal{F}_{k-1}\right] \\
&= \frac{2t\sqrt{\lambda_0 \mu_0}}{n\sqrt{mn}} \sum_{k=2}^{n} \mathbb{E}\left[\left(\gamma_{1,k}^2 - 1\right)^2 \sum_{j_1=2}^{m} \sum_{j_2=2}^{m} \xi_j{}^{1/2} \gamma_{j_1-1,k} w_{j_1,k} \zeta_k{}^{1/2} \eta_{j_2,k-1} w_{j_2,k} | \mathcal{F}_{k-1}\right] \\
&= \frac{O(1)}{n^2} \sum_{k=2}^{n} \zeta_k{}^{1/2} \mathbb{E}\left[\left(\gamma_{1,k}^2 - 1\right)^2 | \mathcal{F}_{k-1}\right] \\
&\times \sum_{j_1=2}^{m} \sum_{j_2=2}^{m} \xi_j{}^{1/2} \eta_{j_2,k-1} \mathbb{E}\left[\gamma_{j_1-1,k} | \mathcal{F}_{k-1}\right] \mathbb{E}\left[w_{j_2,k}^2 | \mathcal{F}_{k-1}\right] \\
&= 0
\end{aligned}
$$

The rest of the cases in the Type A category can similarly expressed due to the independence between $\gamma_{j-1,k}, \eta_{j,k-1}$, and $w_{j,k}$, so they are omitted here. Now, to investigate the cases in Type B, first, consider the case when $s_1 = 4$ (which is similar to the case when $s_3 = 4$) :

$$
\sum_{k=2}^{n} \mathbb{E}\left[\mathcal{E}_k^4(m,n) | \mathcal{F}_{k-1}\right] = \frac{4\lambda_0{}^2}{n^2} \sum_{k=2}^{n} \sum_{j_1=2}^{4} \sum_{j_2=2}^{4} \sum_{j_3=2}^{4} \sum_{j_4=2}^{4} \sqrt{\xi_{j_1} \xi_{j_2} \xi_{j_3} \xi_{j_4}}
$$

$$
\times \mathbb{E}\left[\gamma_{j_1-1,k} w_{j_1,k} \gamma_{j_2-1,k} w_{j_2,k} \gamma_{j_3-1,k} w_{j_3,k} \gamma_{j_4-1,k} w_{j_4,k} | \mathcal{F}_{k-1}\right].
$$

Notice that if all the indices are distinct, then by independence we can factor out any one of $\mathbb{E}[w_{j_l,k}|\mathcal{F}_{k-1}]$, which would give us $\sum_{k=2}^{n}\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] = 0$. Therefore, we look at the two other cases: Case 1, when all indices are the same and Case 2, when there are two distinct pairs.

Case 1, $j = j_1 \dots j_4$:

$$\begin{aligned}
\sum_{k=2}^{n}\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] &= \frac{4{\lambda_0}^2}{n^2}\sum_{j=2}^{m}{\xi_j}^2\sum_{k=2}^{n}\mathbb{E}\left[\gamma_{j-1,k}^4 w_{j,k}^4|\mathcal{F}_{k-1}\right] \\
&= \frac{4{\lambda_0}^2}{n^2}\sum_{j=2}^{m}{\xi_j}^2\sum_{k=2}^{n}\mathbb{E}\left[\gamma_{j-1,k}^4|\mathcal{F}_{k-1}\right]\mathbb{E}\left[w_{j,k}^4|\mathcal{F}_{k-1}\right] \\
&\leq \frac{O(1)}{n^3}(m-1)(n-1) \to 0.
\end{aligned}$$

Case 2, $j_1 = j_3$, $j_2 = j_4$: without loss of generality, assume $j_1 > j_2$, then

$$\sum_{k=2}^{n}\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] = \frac{4{\lambda_0}^2}{n^2}\sum_{k=2}^{n}\sum_{j_1=2}^{m}\xi_{j_1}\sum_{j_2=2}^{m}\xi_{j_2}\mathbb{E}\left[(\gamma_{j_1-1,k}w_{j_1,k})^2(\gamma_{j_2-1,k}w_{j_2,k})^2|\mathcal{F}_{k-1}\right]$$

$$= \frac{4{\lambda_0}^2}{n^2}\sum_{k=2}^{n}\sum_{j_1=2}^{m}\xi_{j_1}\sum_{j_2=2}^{m}\xi_{j_2}\mathbb{E}\left[(\gamma_{j_1-1,k}\gamma_{j_2-1,k})^2|\mathcal{F}_{k-1}\right]$$

$$\times\,\mathbb{E}\left[w_{j_1,k}^2\mathcal{F}_{k-1}\right]\mathbb{E}\left[w_{j_2,k}^2|\mathcal{F}_{k-1}\right]$$

$$\frac{4{\lambda_0}^2}{n^2}\sum_{k=2}^{n}\sum_{j_1=2}^{m}\xi_{j_1}\sum_{j_2=2}^{m}\xi_{j_2}\mathbb{E}\left[(\gamma_{j_1-1,k}\gamma_{j_2-1,k})^2|\mathcal{F}_{k-1}\right]$$

$$\leq \frac{O(1)}{n^2}(n-1) \to 0.$$

The last combination in Type B is when $s_2 = 4$, which gives us

$$\sum_{k=2}^{n}\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] = \sum_{k=2}^{n}\mathbb{E}\left[\left(\frac{\gamma_{1,k}^2-1}{\sqrt{n}}\right)^4|\mathcal{F}_{k-1}\right]$$

$$= \frac{1}{n^2}\sum_{k=2}^{n}\mathbb{E}\left[(\gamma_{1,k}^2-1)^4|\mathcal{F}_{k-1}\right]$$

$$= O(1)\frac{(n-1)}{n^2} \to 0,$$

thus $\sum_{k=2}^{n}\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] = 0$ for all combinations in Type B. Now, for the combinations in Type C, notice that when $(s_1, s_2, s_3) = (2, 2, 0)$ (and similarly for

$(0,2,2))$, we have by independence,

$$\sum_{k=2}^{n}\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] = \frac{2\lambda_0}{n^2}\sum_{k=2}^{n}\mathbb{E}\left[\left(\sum_{j=2}^{m}\xi_j{}^{1/2}\gamma_{j-1,k}w_{j,k}\right)^2\left(\gamma_{1,k}^2-1\right)^2|\mathcal{F}_{k-1}\right]$$

$$= \frac{2\lambda_0}{n^2}\sum_{k=2}^{n}\sum_{j=2}^{m}\xi_j\mathbb{E}\left[\left(\gamma_{1,k}^2-1\right)^2|\mathcal{F}_{k-1}\right]\mathbb{E}\left[\gamma_{j-1,k}^2|\mathcal{F}_{k-1}\right]\mathbb{E}\left[w_{j,k}^2|\mathcal{F}_{k-1}\right]$$

$$= \frac{4\lambda_0}{n^2}\sum_{k=2}^{n}\sum_{j=2}^{m}\xi_j\mathbb{E}\left[\gamma_{j-1,k}^2|\mathcal{F}_{k-1}\right]$$

$$= \frac{O(1)}{n} \to 0.$$

Finally, for the case $(s_1,s_2,s_3)=(2,0,2)$, we have

$$\sum_{k=2}^{n}\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right]$$
$$= \frac{2\lambda_0\mu_0}{mn}$$
$$\times \sum_{k=2}^{n}\mathbb{E}\left[\sum_{j_1=2}^{n}\sum_{j_2=2}^{n}\sum_{j_3=2}^{n}\sum_{j_4=2}^{n}\xi_{j_1}^{1/2}\xi_{j_2}^{1/2}\zeta_k\gamma_{j_1-1,k}w_{j_1,k}\gamma_{j_2-1,k}w_{j_2,k}\eta_{j_3,k-1}w_{j_3,k}\eta_{j_4,k-1}w_{j_4,k}|\mathcal{F}_{k-1}\right].$$

Similarly to previous arguments, if all j_l's distinct, then $\sum_{k=2}^{n}\mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right]=0$ by independence, since we can factor out any one of the random variables whose expectation is zero. On the other hand, if we have $j_1 = j_3$ and $j_2 = j_4$, then

$$
\begin{aligned}
\sum_{k=2}^{n} \mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] &= \frac{4\mu_0\lambda_0}{mn}\sum_{k=2}^{n}\zeta_k \sum_{j_1=2}^{m}\sum_{j_2=2}^{m} \mathbb{E}\left[\left(\gamma_{j_1-1,k} w_{j_1,k}\right)^2 \left(\eta_{j_2,k-1} w_{j_2,k}\right)^2 |\mathcal{F}_{k-1}\right] \\
&= \frac{4\mu_0\lambda_0}{mn}\sum_{k=2}^{n}\zeta_k \sum_{j_1=2}^{m} \xi_{j_1} \\
&\times \sum_{j_2=2}^{m} \mathbb{E}\left[\gamma_{j_1-1,k}^2|\mathcal{F}_{k-1}\right] \mathbb{E}\left[w_{j_1,k}^2\right] \mathbb{E}\left[\eta_{j_2,k-1}^2 \mathcal{F}_{k-1}\right] \mathbb{E}\left[w_{j_2,k}^2|\mathcal{F}_{k-1}\right] \\
&= O(1)\frac{m-1}{mn} \to 0.
\end{aligned}
$$

Since in each case we have $\sum_{k=2}^{n} \mathbb{E}\left[\mathcal{E}_k^4(m,n)|\mathcal{F}_{k-1}\right] \to 0$, this implies that $\mathcal{E}_k(m,n)$ satisfies the weak Lindeberg's condition. Therefore, we can use Theorem (2.2.11) to conclude that

$$
\sum_{k=2}^{n} \mathcal{E}_k(m,n) \to_D N\left(0, 2\left(\lambda_0 + 1 + t^2\mu_0\right)\right).
$$

Furthermore, since $\eta_{j,1}^2 - 1$ is a sequence of *i.i.d.* centered chi-squared random variables with mean 0 and variance $2t^2$, it follows that

$$
\sum_{k=2}^{n} \mathcal{E}_k(m,n) + \frac{t}{\sqrt{m}}\sum_{j=2}^{m}(\eta_{j-1,1}^2 - 1) \to_D N\left(0, 2\left(\lambda_0 + 1 + t^2(\mu_0+1)\right)\right) \tag{2.56}
$$

As a result, we can write

$$\begin{bmatrix} -n\frac{(1+\lambda_0)}{\lambda_0}(\hat{\lambda}-\lambda_0) \\ \\ -m\frac{(1+\mu_0)}{\mu_0}(\hat{\mu}-\mu_0) \end{bmatrix} \to_D N\left(\mathbf{0}, \begin{bmatrix} 2\,(1+\lambda_0) & 0 \\ \\ 0 & 2\,(1+\mu_0) \end{bmatrix}\right) \tag{2.57}$$

which implies that

$$\begin{bmatrix} \sqrt{n}\left(\hat{\lambda}-\lambda_0\right) \\ \\ \sqrt{m}\,(\hat{\mu}-\mu_0) \end{bmatrix} \to_D N\left(\mathbf{0}, \begin{bmatrix} \frac{2{\lambda_0}^2}{(1+\lambda_0)} & 0 \\ \\ 0 & \frac{2{\mu_0}^2}{(1+\mu_0)} \end{bmatrix}\right). \tag{2.58}$$

Finally, to show the asymptotic normality for

$$\begin{bmatrix} \sqrt{n}\left(\hat{\lambda}-\lambda_0\right) & \sqrt{n}\,(\hat{\mu}-\mu_0) & \sqrt{n}\,(\hat{\sigma}^2-\sigma_0^2) \end{bmatrix}',$$

first notice that from the consistency result, we have

$$\begin{aligned} &\sqrt{mn-m^*n^*}\left(\hat{\lambda}\hat{\mu}\hat{\sigma}^2-\lambda_0\mu_0\sigma_0^2\right) \\ &= (1+o(1))\sqrt{mn}\left(\hat{\lambda}\hat{\mu}\hat{\sigma}^2-\lambda_0\hat{\mu}\hat{\sigma}^2+\lambda_0\hat{\mu}\hat{\sigma}^2-\lambda_0\mu_0\hat{\sigma}^2+\lambda_0\mu_0\hat{\sigma}^2-\lambda_0\mu_0\sigma_0^2\right) \\ &= (1+o(1))\left(\sqrt{m}\hat{\mu}\hat{\sigma}^2\sqrt{n}\left(\hat{\lambda}-\lambda_0\right)+\sqrt{n}\lambda_0\hat{\sigma}^2\sqrt{m}\,(\hat{\mu}-\mu_0)+\sqrt{m}\lambda_0\mu_0\sqrt{n}\left(\hat{\sigma}^2-\sigma_0^2\right)\right). \end{aligned} \tag{2.59}$$

Assuming $\frac{n}{m} \to \rho$, where ρ is a positive constant, we can express (2.59) as

$$-\sqrt{n}\left(\hat{\sigma}^2 - \sigma_0^2\right) = \frac{\sigma_0^2}{\lambda_0}\sqrt{n}\left(\hat{\lambda} - \lambda_0\right) + \sqrt{\rho}\frac{\sigma_0^2}{\mu_0}\sqrt{m}\left(\hat{\mu} - \mu_0\right). \tag{2.60}$$

Since $\sqrt{n}\left(\hat{\sigma}^2 - \sigma_0^2\right)$ is a linear combination of two asymptotically normal random variables, we have that $\sqrt{n}\left(\hat{\sigma}^2 - \sigma_0^2\right)$ is asymptotically normal as well. Therefore, we only need to find its covariance structure. Notice that

$$\begin{aligned}
&\operatorname{Cov}\left(\sqrt{n}\left(\hat{\sigma}^2 - \sigma_0^2\right), \sqrt{n}\left(\hat{\mu} - \mu_0\right)\right) \\
&= \mathbb{E}\left[\left(-\frac{\sigma_0^2}{\lambda_0}\sqrt{n}\left(\hat{\lambda} - \lambda_0\right) - \rho^{1/2}\frac{\sigma_0^2}{\mu_0}\sqrt{m}\left(\hat{\mu} - \mu_0\right)\right)\left(\rho^{1/2}\sqrt{m}\left(\hat{\mu} - \mu_0\right)\right)\right] \\
&= -\rho\frac{\sigma_0^2}{\mu_0}\mathbb{E}\left[\sqrt{m}\left(\hat{\mu} - \mu_0\right)^2\right] = \frac{-2\rho\sigma_0^2\mu_0}{1+\mu_0}.
\end{aligned} \tag{2.61}$$

Similarly,

$$\operatorname{Cov}\left(\sqrt{n}\left(\hat{\sigma}^2 - \sigma_0^2\right), \sqrt{n}\left(\hat{\lambda} - \lambda_0\right)\right) = \frac{-2\sigma_0^2\mu_0}{1+\mu_0}, \tag{2.62}$$

and

$$\begin{aligned}
\mathbb{E}\left[\left(\sqrt{n}\left(\hat{\sigma}^2 - \sigma_0^2\right)\right)^2\right] &= \mathbb{E}\left[\left(\frac{\sigma_0^2\sqrt{n}}{\lambda_0}\left(\hat{\lambda} - \lambda_0\right) - \frac{\rho^{1/2}\sigma_0^2\sqrt{m}}{\mu_0}\left(\hat{\mu} - \mu_0\right)\right)^2\right] \\
&= \frac{\sigma_0^4}{{\lambda_0}^2}\mathbb{E}\left[\sqrt{n}\left(\hat{\lambda} - \lambda_0\right)^2\right] + \rho\frac{\sigma_0^2}{{\mu_0}^2}\mathbb{E}\left[\sqrt{m}\left(\hat{\mu} - \mu_0\right)^2\right] \\
&\to 2\sigma_0^2\left[\frac{1}{1+\lambda_0} + \frac{\rho}{1+\mu_0}\right].
\end{aligned} \tag{2.63}$$

Thus, (2.61) – (2.63) implies that

$$\begin{bmatrix} \sqrt{n}\left(\hat{\lambda}-\lambda_0\right) \\ \sqrt{n}\left(\hat{\mu}-\mu_0\right) \\ \sqrt{n}\left(\hat{\sigma}^2-\sigma_0^2\right) \end{bmatrix} \rightarrow_D N\left(\mathbf{0}, \begin{bmatrix} 0 & \frac{{\lambda_0}^2}{1+\lambda_0} & \frac{-2\lambda_0{\sigma_0^2}^2}{1+\lambda_0} \\ \frac{{\mu_0}^2}{1+\mu_0} & 0 & \frac{-2\mu_0\sigma_0^2}{1+\mu_0} \\ \frac{-2\mu_0\sigma_0^2}{1+\mu_0} & \frac{-2\lambda_0\sigma_0^2}{1+\lambda_0} & 2\sigma_0^4\left[\frac{1}{1+\lambda_0}+\frac{\rho}{1+\mu_0}\right] \end{bmatrix}\right). \tag{2.64}$$

2.3 Approximate likelihood estimation for randomly missing data

So far, we have introduced and analyzed the approximate likelihood estimator for the case when observations are missing in a block satisfying certain regularity conditons. In here, we propose an alternative method to estimate λ, μ and σ^2 when the sampling sites have randomly missing observations. As before, let X be a realization of the O-U process defined throughout this paper, and $X \in \mathbb{R}^{n\times m}$. Suppose each sampling site (u_i, v_j) has a probability of p, $0 \le p < 1$ of being missing. Let $X^{(o)} = \{X_j^{(o)}\}_{j=1}^m$ denote the set of data that is available, where each $X_j^{(o)}$ is the j^{th} column of $X^{(o)}$. In this section, we define the following notations:

- $n_j :=$ number of available observations for each column $X_j^{(o)}$
- $K_j :=$ set of indices for each available observation in $X_j^{(o)}$
- $\Sigma_{j,j'}^{(o)} := \mathbb{E}[X_j^{(o)}, X_{j'}^{(o)}]$, an $n_j \times n_{k'}$ covariance matrix between column $X_j^{(o)}$ and column $X_{j'}^{(o)}$
- $B_{j,j'}^{(o)} := \{e^{-\mu|v_k - v_{k'}|}\}_{(k,k')\in K_j \times K_{j'}}$, such that $\Sigma_{j,j'}^{(o)} = \sigma^2 e^{-\lambda|u_j - u_{j'}|} B_{j,j'}^{(o)}$

We propose to approximate the likelihood function of (λ, μ, σ^2) given $X^{(o)}$ based on the Markov property of the full-observation case:

$$f\left(\lambda, \mu, \sigma^2|\theta\right) = f\left(X_1^{(o)}\right) \Pi_{j=2}^m f\left(X_j^{(o)}|X_{j-1}^{(o)}\right). \tag{2.65}$$

where each of the conditional variable is a multivariate normal with mean $m_{j|j-1}^{(o)}$

and covariance matrix $\sigma^2 B^{(o)}_{j|j-1}$. Let $l\left(\lambda, \mu, \sigma^2 | X^{(o)}\right) = -2\log\left(f(\lambda, \mu, \sigma^2|X^{(o)}\right)$, then

$$\begin{aligned} l\left(\lambda, \mu, \sigma^2 | X^{(o)}\right) = & \left(\sum_{j=1}^{m} n_j\right) \log\left(2\pi\sigma^2\right) + \sum_{j=1}^{m} \ln\left|B^{(o)}_{j|j-1}\right| \\ & + \frac{1}{\sigma^2}\left(X^{(o)}_1\right)'\left(B^{(o)}_{11}\right)^{-1} X^{(o)}_1 \\ & + \frac{1}{\sigma^2}\sum_{j=2}^{m}\left[\left(X^{(o)}_j - m^{(o)}_{j|j-1}\right)'\left(B^{(o)}_{j|j-1}\right)^{-1}\left(X^{(o)}_j - m^{(o)}_{j|j-1}\right)\right] \end{aligned} \quad (2.66)$$

where

$$m^{(o)}_{j|j-1} = \begin{cases} 0 & j = 1 \\ e^{-\lambda|u_j - u_{j-1}|} B^{(o)}_{j,j-1}\left(B^{(o)}_{j-1,j-1}\right)^{-1} X^{(o)}_{j-1} & j = 2\dots m \end{cases} \quad (2.67)$$

$$B^{(o)}_{j|j-1} = \begin{cases} \sigma^2 B^{(o)}_{11} & j = 1 \\ \sigma^2\left[B^{(o)}_{jj} - e^{-2\lambda|u_j - u_{j-1}|} B^{(o)}_{j,j-1}\left(B^{(o)}_{j-1,j-1}\right)^{-1} B^{(o)}_{j-1,j}\right] & j = 2,\dots m. \end{cases} \quad (2.68)$$

Comparing (2.66) to the likelihood functions defined in (1.8) and (2.6), a main difference is that the covariance matrix $B^{(o)}_{j|j-1}$ no longer has an explicit tridiagonal inverse. This is not necessarily infeasible in the sense of computation, especially since today's computers have become much more efficient in handling large matrices. However, due to the missing observations, we cannot find an explicit way to express the inverse of $B^{(o)}_{j|j-1}$, which means that the approximation technique used in [24] for the quadratic form may not be a good tool for the asymptotic analyses here. One way to tackle this problem is that, instead of trying to find an explicit form for the inverse of $B^{(o)}_{j|j-1}$, we look at ways to approximate the matrix using a version of power series expansion with finitely many terms. To this end, consider the following definition

Definition 2.3.1. *Let* $G := \{g_{kl}\}_{k,l=1}^{N} \in \mathbb{R}^{N\times N}$ *be a positive-definite tridiagonal matrix. Define the terms of* $L := \{l_{kl}\}_{k,l=1}^{N}$ *by the following:*

- $l_{kk} = \sqrt{g_{kk}}, k = 1$ *or* N
- $l_{k,k-1} = \dfrac{g_{k-1,k}}{l_{k-1,k-1}}$ *and* $l_{k,k} = \sqrt{g_{k,k} - l_{k,k-1}^2}, k = 2, \ldots, N-1$

Then L *is a lower bidiagonal matrix and* $G = LL'$.

To see how the bidiagonal matrices in definition (2.3.1) relate to our problem, notice that

$$B_{j|j-1}^{(o)} = B_{j,j}^{(o)} - e^{-2\lambda|u_j - u_{j-1}|} B_{j,j-1}^{(o)} (B_{j-1,j-1}^{(o)})^{-1} B_{j-1,j}^{(o)}, \tag{2.69}$$

the fact that $\left(B_{j-1,j-1}^{(o)}\right)^{-1}$ is a tridiagonal matrix allows us to choose a lower bidiagonal matrix L_j, as defined in (2.3.1), such that $\left(B_{j-1,j-1}^{(o)}\right)^{-1} = L_{j-1}L_{j-1}'$. Therefore, we have

$$\begin{aligned} B_{j|j-1}^{(o)} &= B_{j,j}^{(o)} - e^{-2\lambda|u_j - u_{j-1}|} B_{j,j-1}^{(o)} L_{j-1} L_{j-1}' B_{j-1,j}^{(o)} \\ &= (L_j^{-1})' L_j^{-1} - e^{-2\lambda|u_j - u_{j-1}|} (B_{j,j-1}^{(o)} L_{j-1})(B_{j,j-1}^{(o)} L_{j-1})' \\ &= (L^{-1})_j' (I - C_{j|j-1} C_{j|j-1}') L_j^{-1}, \end{aligned} \tag{2.70}$$

where

$$C_{j|j-1} = e^{-\lambda|u_j - u_{j-1}|} L_j' B_{j,j-1}^{(o)} L_{j-1}.$$

As a result,

$$\left(B_{j|j-1}^{(o)}\right)^{-1} = L_j \left(I - C_{j|j-1} C_{j|j-1}'\right)^{-1} L_j'. \tag{2.71}$$

Notice that, using spectral decomposition, a $d \times d$ positive semi-definite matrix, say M can be represented as $\sum_{k=1}^{d} \lambda e_k e_k'$, where $\{e_k, k = 1, 2, \ldots, d\}$ is an orthonormal basis of the eigenspace of M. This implies that, we can represent $(I - C_{j|j-1} C_{j|j-1}')^{-1}$ as a convergent power series if all the eigenvalues of $C_{j|j-1} C_{j|j-1}'$ are less than 1. To this end, consider two random variables $Y_{j,j-1}$ and $Z_{j,j-1}$, defined as follow:

- $Y_{j,j-1} := L_j' \left(X_j^{(o)} - B_{j,j-1}^{(o)} \left(B_{j-1,j-1}^{(o)}\right)^{-1} X_{j-1}^{(o)}\right) L_j$
- $Z_{j,j-1} := B_{j,j-1}^{(o)} \left(B_{j-1,j-1}^{(o)}\right)^{-1} X_{j-1}^{(o)}$

Then $Y_{j,j-1}$ and $Z_{j,j-1}$ are two multivariate normal random variables with mean 0 and covariance matrices $\sigma^2(I - C_{j|j-1} C_{j|j-1}')$ and $\sigma^2 C_{j|j-1} C_{j|j-1}'$ respectively. Assuming both random variables have a non-degenerate distribution, we have that their covariance matrices are positive-definite, which implies that they each have a spectral decomposition with positive eigenvalues . Let δ be an eigenvalue of $I - C_{j|j-1} C_{j|j-1}'$, then

$$\det(I - C_{j|j-1} C_{j|j-1}' - \delta I) = 0 \tag{2.72}$$

which implies that $1 - \delta$ is an eigenvalue of $C_{j|j-1} C_{j|j-1}'$. By definition of positive-definite matrices, we have that $1-\delta>0$, which implies $0<\delta<1$. Thus, we can approximate $\left(I-C_{j|j-1}C'_{j|j-1}\right)^{-1}$ using a finite number of terms in the power series. Let

$$G^*_{j|j-1}=\sum_{k=0}^{K}\left(C_{j|j-1}C'_{j|j-1}\right)^k, \tag{2.73}$$

we propose to approximate (2.66) by:

$$\begin{aligned} l^*\left(\lambda,\mu,\sigma^2|X^{(o)}\right) &= \left(\sum_{j=1}^{m}n_j\right)\log\left(2\pi\sigma^2\right)+\sum_{j=1}^{m}\ln|B^{(o)}_{j|j-1}| \\ &+\frac{1}{\sigma^2}\left(X^{(o)}_1\right)'\left(B^{(o)}_{11}\right)^{-1}X^{(o)}_1 \\ &+\frac{1}{\sigma^2}\sum_{j=2}^{m}\left[\left(X^{(o)}_j-m^{(o)}_{j|j-1}\right)'\left(L_jG^*_{j|j-1}L'_j\right)\left(X^{(o)}_j-m^{(o)}_{j|j-1}\right)\right] \end{aligned} \tag{2.74}$$

In this definition, we have replaced the inverse of $\Sigma^{(o)}_{j|j-1}$ with a finite sum of terms that does not involve inversing any matrices, which could potentially ease the process of analysing the asymptotic properties of the estimates for $\hat{\lambda},\hat{\mu}$, and $\hat{\sigma^2}$, which are solutions to

$$\left\{\frac{\partial}{\partial\lambda}l^*=0,\frac{\partial}{\partial\mu}l^*=0,\frac{\partial}{\partial\sigma^2}l^*=0\right\}. \tag{2.75}$$

Intuitively, the number of terms to use in the approximate likelihood function would depend on the true parameter values as well as the grid size. We investigate the effect of grid size and parameter values have on the accuracy of the power series approximation using simulated data. In Figure (2.3) below, we see an example of how accuracy of the power series approximation changes as the dimension of the sampling grid increases. Although investigating the theoretical properties

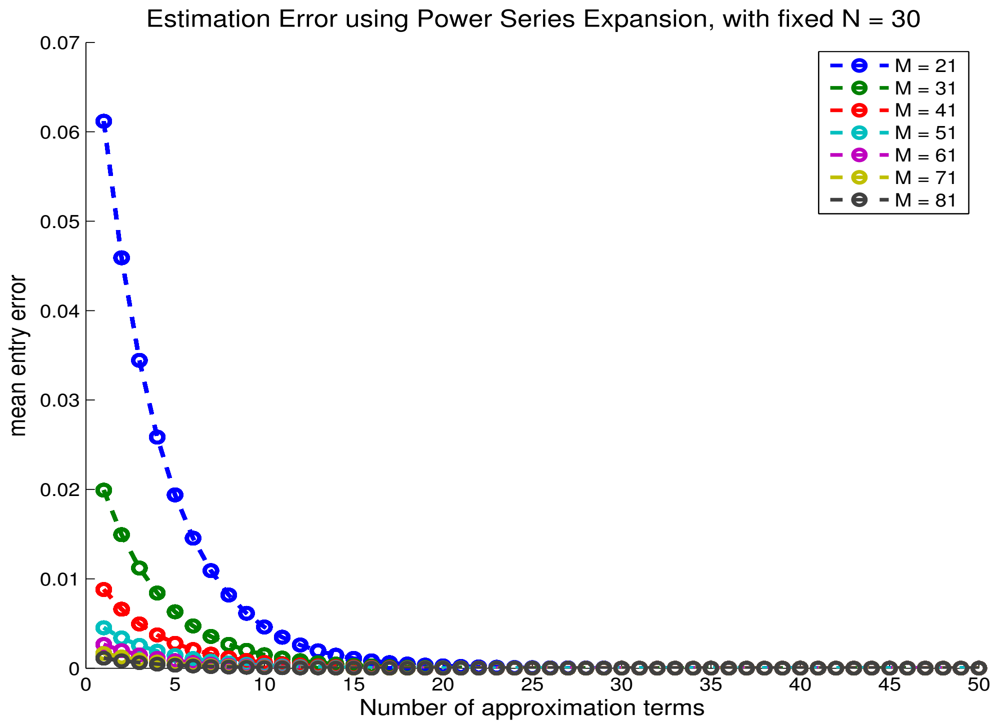


Figure 2.3: Comparing accuracy of the power series estimation of $\Sigma^{(o)}_{j|j-1}$ with varying grid sizes (N fixed, M varies). Shown in figure is the mean entry difference between $\left(\Sigma^{(o)}_{j|j-1}\right)^{-1}$ and $\sum_{j=2}^{M} G^{*}_{j|j-1}$.

of estimators from (2.66) would be of great value, in this thesis we devote our attention to inspect the numerical aspect of this estimator. In particular, our goal is to show, through numerical experiements, that estimators from (2.66) and (2.6) behave similarly, which can hopefully be used as a basis for developing theoretical properties in future investigations.

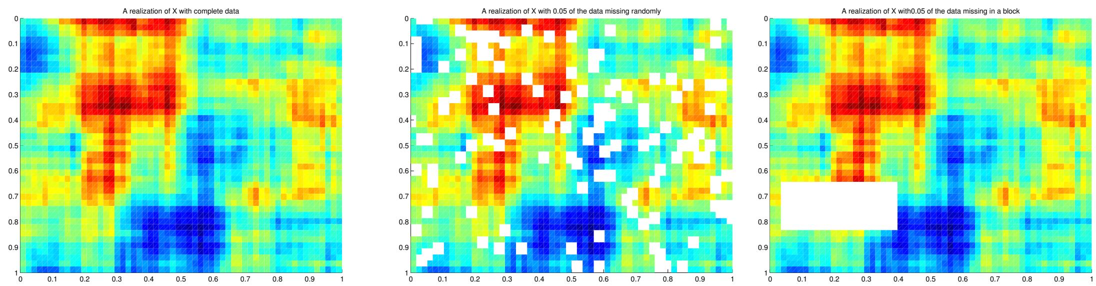
Chapter 3

Implementation

In this chapter we investigate the approximated likelihood estimators through a series of numerical experiements. The main steps involved in the implementation includes the following: data simulation, missingness simulation, and numerical experiments using large number of realizations. When evaluating the approximated likliehood functions, the maxima are obtained using Newton's method. To this end, let X be an $M \times N$ O-U field with parameter values λ_0, μ_0 and σ_0^2. In the following numerical experiments, we simulate X using SZ, where

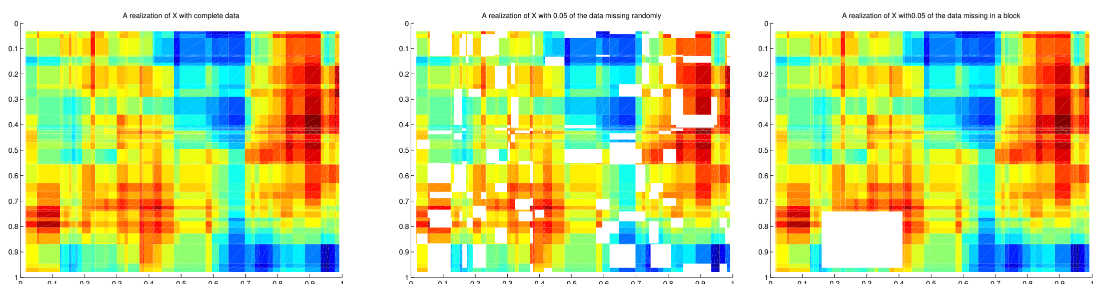
- $Z \sim N_{MN}(0, I)$ is an $MN \times 1$ vector of standard normal random variables
- S is the $MN \times MN$ cholesky decomposition (i.e. $SS' = \text{Cov}(X)$) of the covariance matrix of the O-U field with parameters λ_0, μ_0 and μ_0.

As an example, consider X with parameters $M = 59, N = 43,\ \lambda = 3.6,\ \mu = 2.1$ and $\sigma^2 = 5.9$ and the following two cases: X_{eq} (equally-spaced sampling sites) and X_{arb} (arbitrarily-spaced sampling sites). Figures (3.1) - (3.2) illustrate two different realizations of X with missing observations either following a block or a randomly distributed pattern. Figures (3.7) and (3.4) then illustrates the corresponding likelihood functions in each spacing arrangement and missing data scienarios. As we will see from this example, the spacing of the sampling grid does not appear to have much effect on the behavior of the likelihood functions, as they both have very similar shapes with a minimum close to the true parameter value.



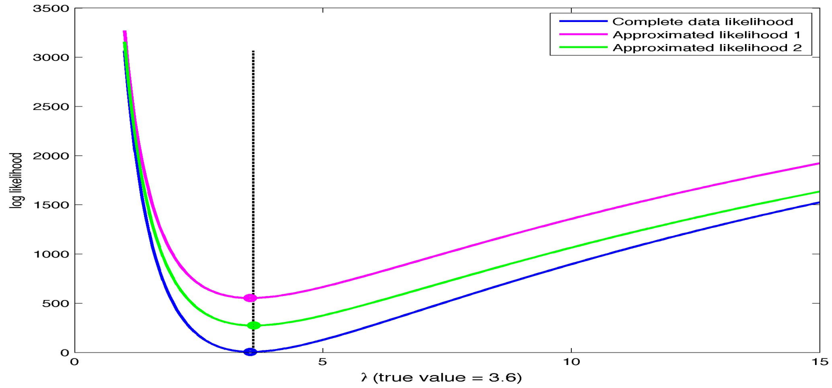
(a) Complete data (b) Randomly missing data (c) Missing block

Figure 3.1: A realization of X_{eq} with two types of data missingness

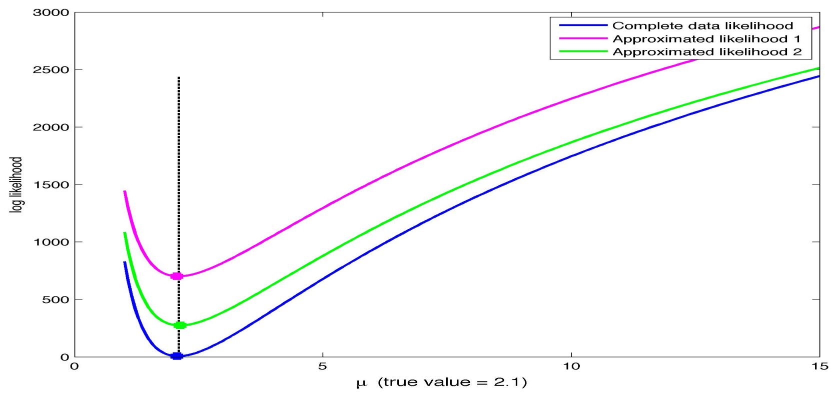


(a) Complete data (b) Randomly missing data (c) Missing block

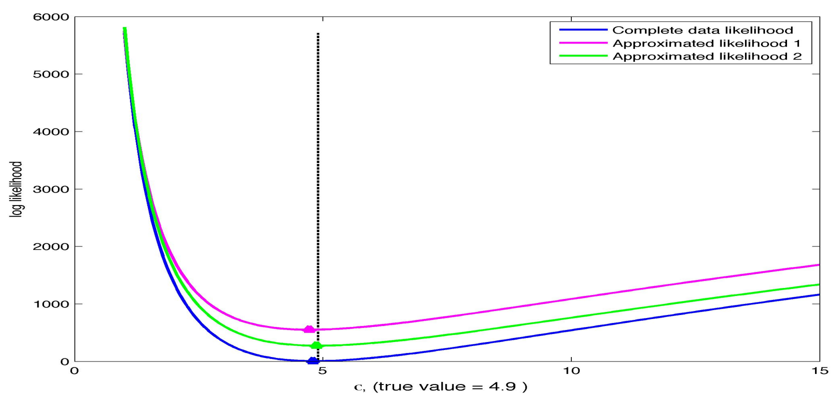
Figure 3.2: A realization of X_{arb} with two types of data missingness



(a) Estimation of λ

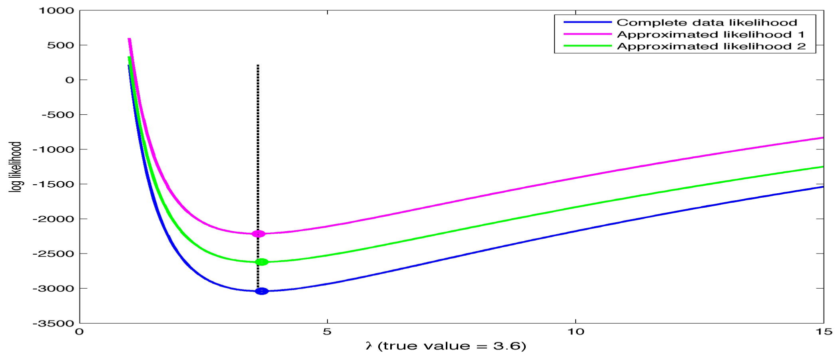


(b) Estimation of μ

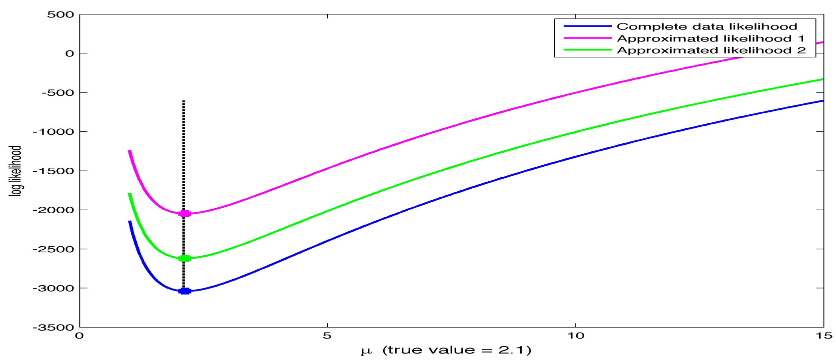


(c) Estimation of σ^2

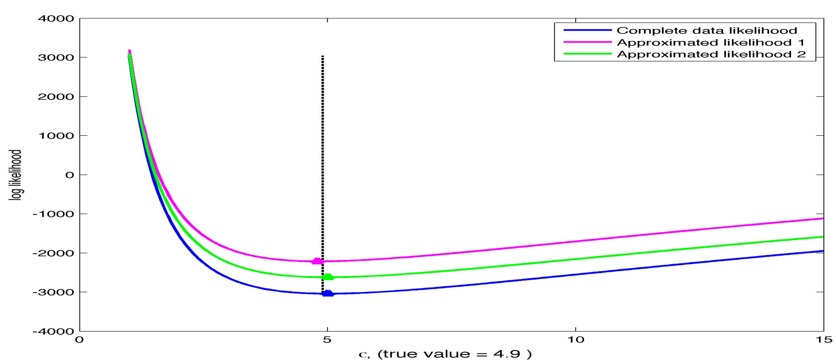
Figure 3.3: A comparison of the approximated likelihood functions to the complete-data likelihood for X_{eq}. The blue plot is the complete data likelihood, while the red plot indicates the likelihood function for the randomly missing data case, and the green plot is the approximated likelihood function for the missing block case.



(a) Estimation of λ



(b) Estimation of μ



(c) Estimation of σ^2

Figure 3.4: A comparison of the approximated likelihood functions to the complete-data likelihood for X_{arb}. The blue plot is the complete data likelihood, while the red plot indicates the likelihood function for the randomly missing data case, and the green plot is the approximated likelihood function for the missing block case.

3.1 Illustrative example with simulated data

The purpose of this example is to investigate the numerical properties of the approximated likelihood estimators with respect to a particular set of parameter values, where

- $M = 39, N = 33$
- $\lambda_0 = 3.2, \mu_0 = 5.1$ and $\sigma_0^2 = 4.18$

The experiment is set up according to the following : we generate 300 equally-spaced realizations using the parameter values above, and with each realization we look at four different levels of missing observations : $[0.01, 0.06, 0.11, 0.16]$. With each missing level and each realization, we then estimate the values of λ, μ and σ^2 using the approximated likelihood estimation proposed for each of the missing pattern. In particular, for the case when the observations are randomly missing, we used a finite series of eight terms to approximate the matrix inverse. We then compare the summary statistics from the resulting estimates to those obtained using the EM algorithm, where in each iteration the missing values are replaced with a conditional expectation drawn from the distribution based on the current parameter estimate (See Appdendix B for a more detail description of the implementation steps).

Notice that, in this experiment, instead of attempting to compute the asymptotic distributions of $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$, we intend to assess the properties of the parameter estimaters under our proposed method in a smaller dimensional setting. By doing so, we hope to provide a realistic snapshot of how our estimation scheme will perform in practical scenarios, where the potential applications could be modeling data collected from agricultural experiments, weather monitoring stations and public health studies.

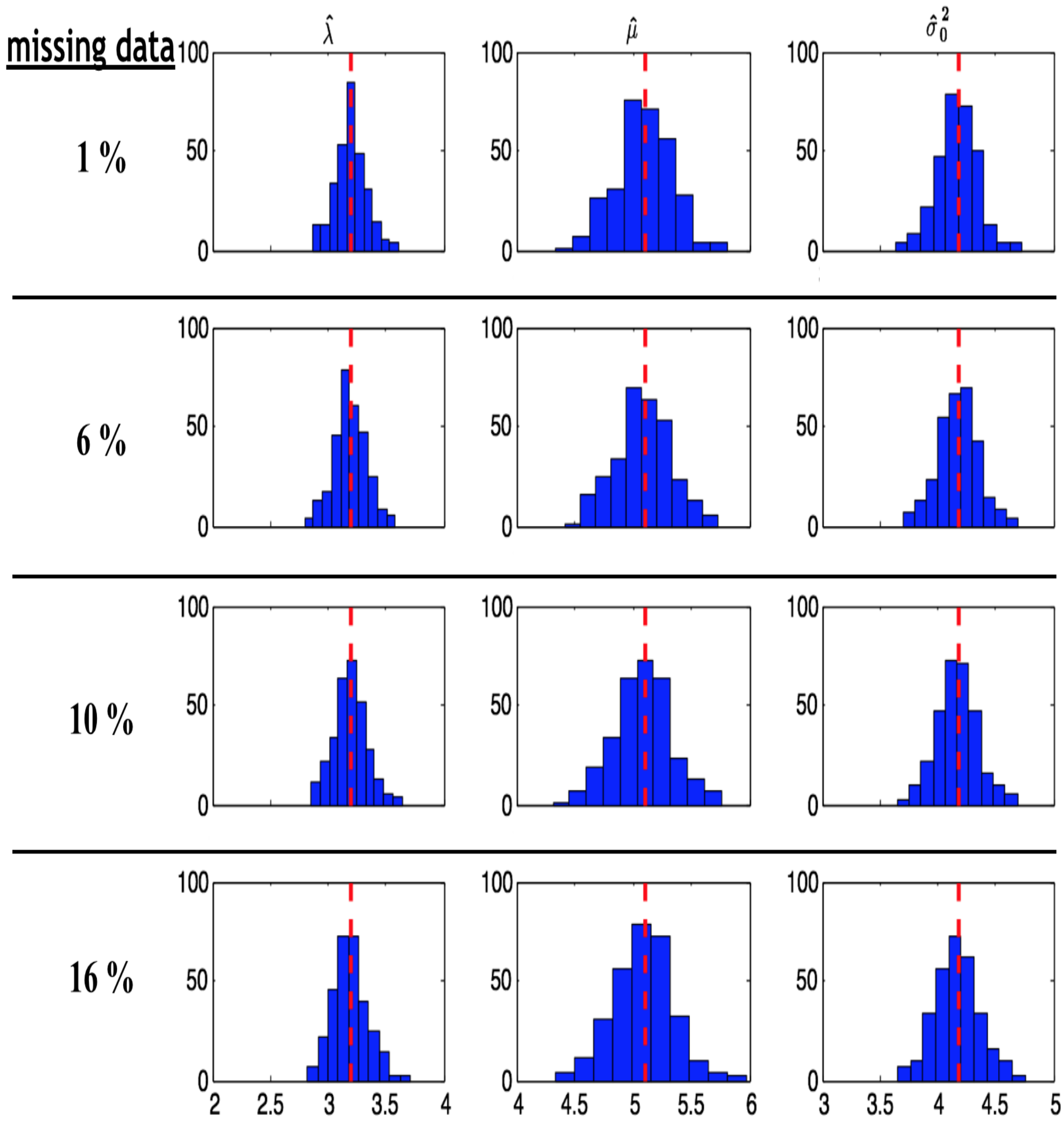


Figure 3.5: Histograms of $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ obtained from the approximated likelihood estimators, where data is missing in a single block. Notice that the spread of the distribution is proportional to the value of λ_0, μ_0 and σ_0^2

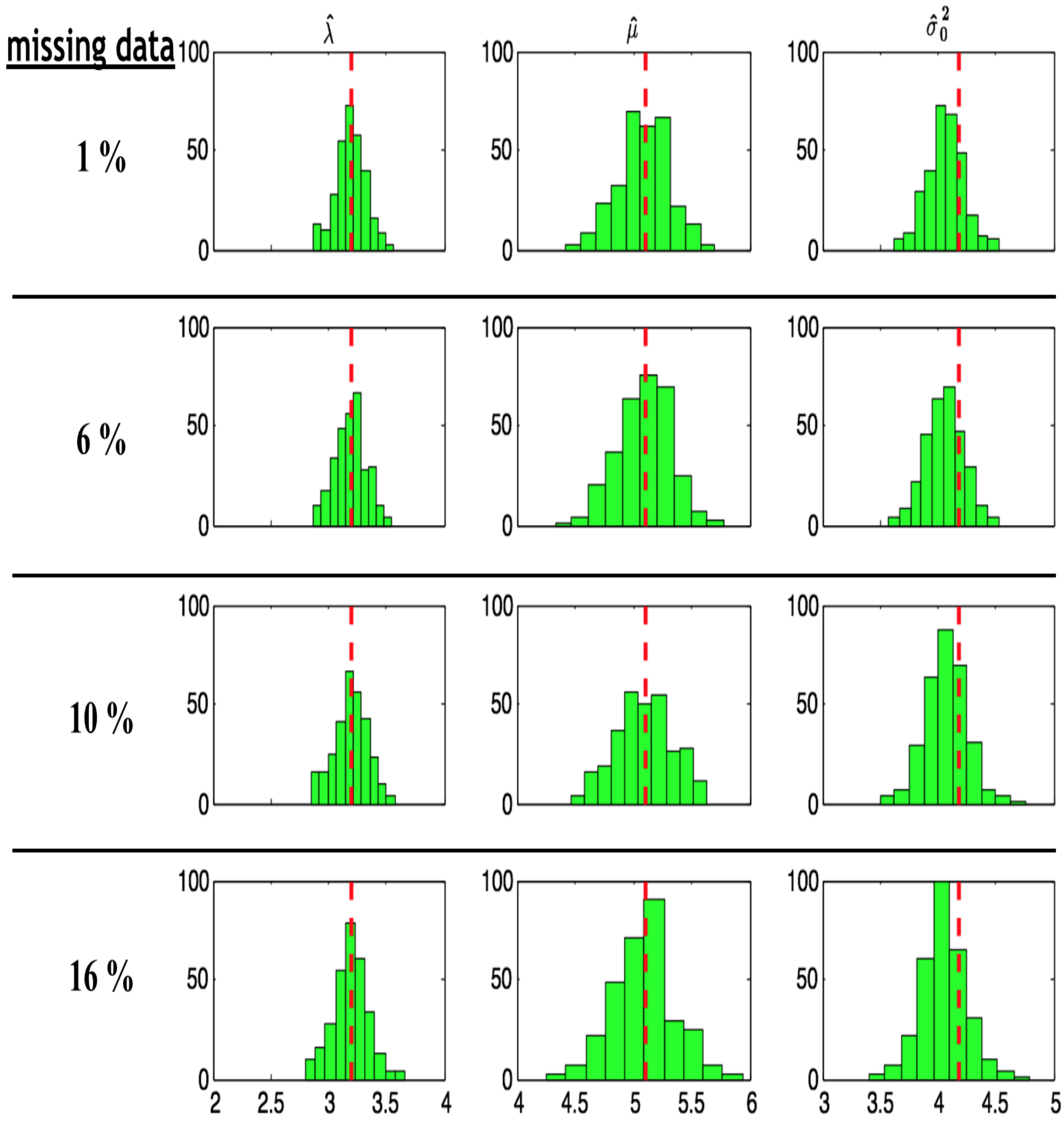


Figure 3.6: Histograms of $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ obtained from the approximated likelihood estimators, where data is missing randomly throughout the field. Similarly, in here the spread of the distribution also appears to be proportional to the value of λ_0, μ_0 and σ_0^2

Bias						
Missingness Pattern	Estimation Method	Parameter	% of missing observations			
			1	6	11	16
Block	Approximated Likelihood	$\hat{\lambda}$	0.0083	0.0084	0.0111	0.0123
		$\hat{\mu}$	0.0086	0.0085	0.0110	0.0139
		$\hat{\sigma}^2$	0.0091	0.0096	0.0107	0.01584
	Expectation Maximization	$\hat{\lambda}$	-0.0377	0.1299	0.2418	0.4196
		$\hat{\mu}$	-0.0193	0.2149	0.3113	0.4399
		$\hat{\sigma}^2$	0.0252	-0.0724	-0.1026	-0.1337
Random	Approximated Likelihood	$\hat{\lambda}$	0.00934	0.00838	0.0074	0.0091
		$\hat{\mu}$	0.0108	0.0049	0.0097	0.0106
		$\hat{\sigma}^2$	0.1204	0.1245	0.11	0.1296
	Expectation Maximization	$\hat{\lambda}$	-0.0349	0.1692	0.3187	0.4786
		$\hat{\mu}$	-0.0201	0.2876	0.4945	0.7504
		$\hat{\sigma}^2$	0.0213	-0.1791	-0.3259	-0.5227

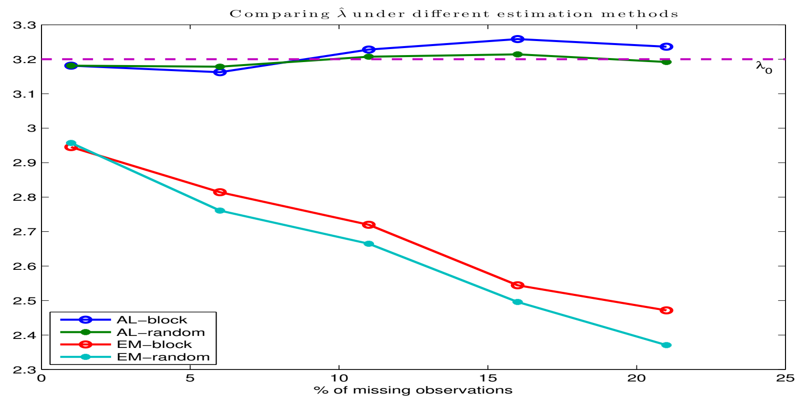
Table 3.1: Summary statistics comparing the bias of $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ estimated using the approximated likelihood function versus the EM algorithm. Two scenarios (missing block, randomly missing) are simulated with varying proportion of the observations missing.

Root Mean Squared Error						
Missingness Pattern	Estimation Method	Parameter	% of missing observations			
			1	6	11	16
Block	Approximated Likelihood	$\hat{\lambda}$	0.1370	00.1370	0.1418	0.1477
		$\hat{\mu}$	0.2311	0.2352	0.2377	0.2529
		$\hat{\sigma}^2$	0.1696	0.1696	0.1753	0.1858
	Expectation Maximization	$\hat{\lambda}$	0.4148	0.4197	0.4609	0.5607
		$\hat{\mu}$	0.5343	0.5630	0.6116	0.6994
		$\hat{\sigma}^2$	0.5853	0.6174	0.6439	0.6745
Random	Approximated Likelihood	$\hat{\lambda}$	0.1270	0.1313	0.1434	0.1492
		$\hat{\mu}$	0.2158	0.2203	0.2395	0.2366
		$\hat{\sigma}^2$	0.3413	0.3310	0.3510	0.3623
	Expectation Maximization	$\hat{\lambda}$	0.4111	0.4214	0.4882	0.5908
		$\hat{\mu}$	0.5373	0.5859	0.7015	0.8810
		$\hat{\sigma}^2$	0.5815	0.6422	0.7231	0.8531

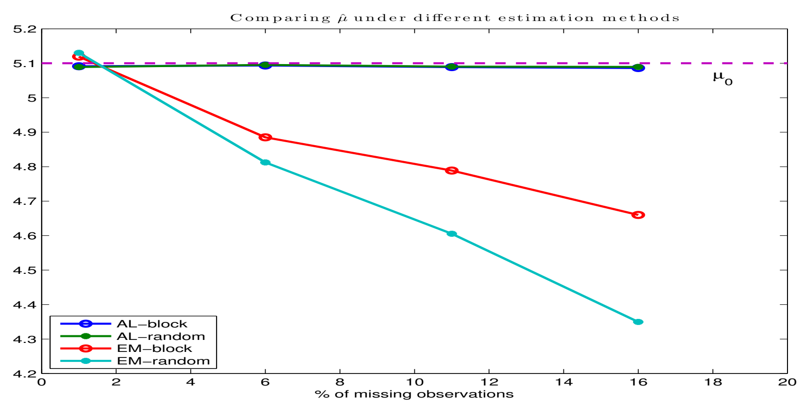
Table 3.2: Summary statistics comparing the root mean squared error of $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ estimated using the approximated likelihood function versus the EM algorithm. Two scenarios (missing block, randomly missing) are simulated with varying proportion of the observations missing.

Standard Deviation						
Missingness Pattern	Estimation Method	Parameter	% of missing observations			
			1	6	11	16
Block	Approximated Likelihood	$\hat{\lambda}$	0.1350	0.1369	0.1418	0.1474
		$\hat{\mu}$	0.2314	0.2354	0.2378	0.2529
		$\hat{\sigma}^2$	0.1695	0.1696	0.1753	0.1854
	Expectation Maximization	$\hat{\lambda}$	0.4138	0.3998	0.3931	0.3726
		$\hat{\mu}$	0.5328	0.5212	0.5272	0.5446
		$\hat{\sigma}^2$	0.5857	0.6142	0.6367	0.6623
Random	Approximated Likelihood	$\hat{\lambda}$	0.1269	0.1313	0.1434	0.1492
		$\hat{\mu}$	0.2159	0.2206	0.2398	0.2568
		$\hat{\sigma}^2$	0.3301	0.3409	0.3495	0.3614
	Expectation Maximization	$\hat{\lambda}$	0.3469	0.3705	0.3866	0.4103
		$\hat{\mu}$	0.4623	0.4983	0.5114	0.5373
		$\hat{\sigma}^2$	0.5821	0.6177	0.6465	0.6754

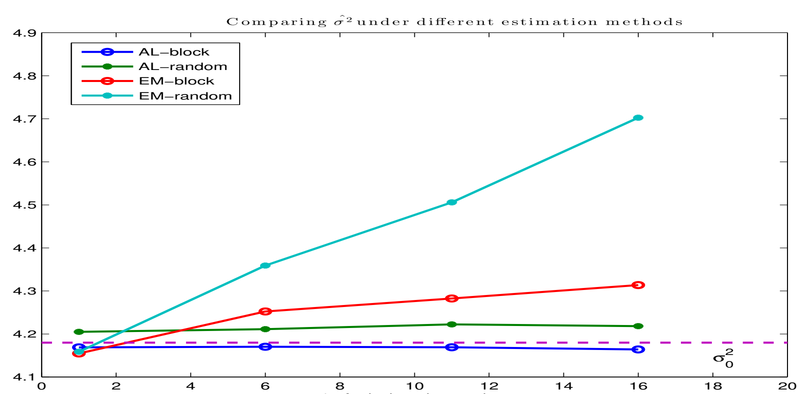
Table 3.3: Summary statistics comparing the standard deviation of $\hat{\lambda}, \hat{\mu}$ and $\hat{\sigma}^2$ estimated using the approximated likelihood function versus the EM algorithm. Two scenarios (missing block, randomly missing) are simulated with varying proportion of the observations missing.



(a) Empirical distribution of $\hat{\lambda}$



(b) Empirical distribution of $\hat{\mu}$



(c) Empirical distribution of $\hat{\sigma}^2$

Figure 3.7: A comparison of the approximated likelihood functions to the complete-data likelihood for X_{eq}. The blue plot is the complete data likelihood, while the red plot indicates the likelihood function for the randomly missing data case, and the green plot is the approximated likelihood function for the missing block case.

Estimation method	Approximate Likelihood		Expectation Maximation	
Missing pattern	block	random	block	random
Computation time (seconds per realization)	4.93	3.95	9.54	16.68

Table 3.4: A comparison of computational time for λ, μ and σ^2 using each estimation method

3.2 Remark

From the numerical experiments we noticed that, unlike the EM algorithm, both approximated likelihood methods rely only on information given by the available data and do not involve any iterative steps and imputations in the algorithm. These features result in a speedier estimation process, as they eliminate the need to repetitvely compute matrix inverses (see Table 3.4). Although the estimates show an increasing level of bias as number of missing observations increase, this is not unexpected and we see that the overall accuracy is still within a reasonable range. This suggests that the estimators will likely provide good results even in the presence of missing observations.

Chapter 4

Conclusion

In this work, we proposed an approximate likelihood estimation method for the O-U process, defined on a two-dimensional lattice with missing sampling sites under two scenarios of data missingness. By imposing the Markov property from the full observaion case on the approximated likelihood function, we elimated the computational burden of computing high dimensional inverse for the missing block case. Moreover, the asymptotics for the approximate likelihood estimate in the missing block case show simlar result compared to the MLE, as long as the size of the missing grid is under control. While for the case with randomly missing observations, we replace the inverse of the covariance matrix with a finite matrix series approximation and show that numerically, they yield similar results compared to the missing block case.

Based on these preliminary results, it is reasonable to set up a conjecture assuming similar asymptotics for the approximated likelihood estimators between these two missing data scenarios. The main challenge in proceeding with analyzing the randomly missing observation case is that, due to the pattern of the missing sampling sites, we can no longer analyze the likelihood funtion by utilizing the tridiagonal inverse matrices that result in the full data and missing block cases. However, we could potentially develop theoretical analysis by approximating the inverse of the covariance matrix using a finite number of binomial expression terms described in section 2.6, this would be an interesting direction for future work.

Bibliography

[1] Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *J. Roy. Statist. Soc.* Ser. B **36**, 192-236.

[2] Cressie,N. (1993). *Statistics for Spatial Data.* Wiley, New York, 1993.

[3] Dempster, A.P., Laird, N., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc* Ser. B, **39** 1–38.

[4] Doob, J.L. (1942). The Brownian Movement and Stochastic Equations *Ann. Math. 43 (2): 351-369*

[5] Roderick J.A. Little and Donald B. Rubin (2002). *Statistical Analysis with Missing Data. – second edition.* Wiley series in probability and statistics. ISBN 0-471-18386-5

[6] Mardia, K.V. and Marshall, R.J. (1984) Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135-146.

[7] McLeish, D.L. (1974). Dependent Central Limit Theorems and Invariance Principles *Ann. Prob.* **2** (4) 620–628

[8] Orchard, T. and Woodbury, M.A. (1972). A Missing Information Principle: Theory and Applications. *Proc. of the 6th Symp. on Math. Stat. and Prob.* **1** 697 –715

[9] Pollard, D (1984) *Convergence of Stochastic Processes.* Springer, New York.

[10] Rubin, D. (1976). Inference and missing data *Biometrika* **63** 3 581–92.

[11] Rubin, D. (1987). *Multiple imputation for nonresponse in surveys.* New York" John Wiley and Sons.

[12] Roderick J.A. Little and Donald B. Rubin (2002). *Statistical Analysis with Missing Data. – second edition.* Wiley series in probability and statistics. ISBN 0-471-18386-5

[13] Sacks, J., Schiller, S.B. and Welch W. J. (1989). Designs for computer experiments. *Technometrics* **31** 41-47.

[14] Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1989). Design and analysis of computer experiments (with discussion) *Statist. Sci.* **4** 409-435.

[15] Schafer, J.L.(1997). *Analysis of incomplete multivariate data.* Monographs on Statistics and Applied Probability 72, Chapman and Hall/CRC.

[16] Sprott, J. C. (2004). A method for approximating missing data in spatial patterns. Computers and Graphics, 28, 113-117.

[17] Stein, M.L., Chi,Z. and Welty, L.J. (2004) Approximaing likelihoods for large spatial data sets *J.R. Statist. Soc. B* **66**, Part 2, 275-296

[18] Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography* **46**, Supplement: Proceedings. International Geographical Union. Commission on Quantitative Methods 234–240.

[19] Varin, C., Reid, N. and Firth, D. (2011). An Overview of Composite Likelihood Methods *Statistica Sinica* **21**, 5-42.

[20] Vecchia A.V. (1988). Estimation and model identification for continuous spatial processes. *J.R. Statist. Soc. B* **50** No. 2 297–312.

[21] Waller, L.A., and Gotway, C.A. (2004). *Spatial data in applied spatial statistics for public health data.* John Wiley & Sons, Inc., Hoboken, NJ, USA. dii: 10.1002/0471662682.ch3

[22] Wu, C.F. Jeff (1983) On the Convergence Properties of the EM Algorithm *Ann. Stat. 11(1) 95-103*

[23] Yates, F. (1933). The analysis of replicated experiments when the field results are incomplete. *Empire Journal of Experimental Agriculture* **1** (2) , 129–42.

[24] Ying, Z (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme *Ann. Stat.* **21**(3) 1567-1590

Appendix A: Derivatives involved in obtaining the MLE

First and second derivatives of $B^{-1}(\mu)$

Let $B^{-1}(\mu) \in \mathbb{R}^{N \times N}$ be the tridiagonal inverse matrix from the OU process. Let $\bar{b}_{kl}$ be the kl^{th} entry of B^{-1}. For $k = 1, \ldots, N$, define the following:

- $b_k = e^{-\mu\zeta_k}$
- $b_k^* = \frac{\zeta_k b_k^2}{(1-b_k^2)^2}$
- $b_k^{**} = \frac{-\zeta_k b_k^2(1+b_2^2)}{(1-b_k^2)^3}$

Then we have:

- $\frac{\partial}{\partial\mu} b_{11} = -2b_2^*, \frac{\partial^2}{\partial\mu} b_{11} = -4b_2^{**}$
- $\frac{\partial}{\partial\mu} b_{NN} = -2b_N^*, \frac{\partial^2}{\partial\mu} b_{NN} = -4b_N^{**}$
- $\frac{\partial}{\partial\mu} b_{kk} = -2(b_k^* + b_{k+1}^*), \frac{\partial^2}{\partial\mu} b_{kk} = -4(b_k^{**} + b_{k+1}^{**}), k = 2, \ldots, N-1$
- $\frac{\partial}{\partial\mu} b_{k,k-1} = \frac{\zeta_k b_k(1+b_k^2)}{(1-b_k^2)^2}, \frac{\partial^2}{\partial\mu} b_{k,k-1} = \frac{\zeta_k^2 b_k}{(1-b_k^2)^3}[1 + 6b_k^2 + b_k^4]$

Partial derivatives of $l(\theta|X)$

The MLE $\hat{\theta}$ is the root of the gradient of $(\theta|X)$, using the Newton's, the update mechanism is as follow:

$$\begin{bmatrix} \lambda_{p+1} \\ \\ \mu_{p+1} \end{bmatrix} = \begin{bmatrix} \lambda_p \\ \\ \mu_p \end{bmatrix} - \begin{bmatrix} h_{\lambda\lambda} & h_{\lambda\mu} \\ \\ h_{\mu\lambda} & h_{\mu\mu} \end{bmatrix}^{-1} \begin{bmatrix} dl_\lambda \\ \\ dl_\mu \end{bmatrix}_{(\lambda,\mu)=(\lambda_i,\mu_i)} \tag{1}$$

where

$$\begin{aligned} dl_\lambda = {} & N\sum_{j=2}^{M} \frac{2\xi_j e^{-2\lambda\xi_i}}{1-e^{-2\lambda\xi_j}} + \frac{2}{\sigma^2}\sum_{j=2}^{M} \frac{\xi_j e^{-\lambda\xi_j} X_{\cdot,j-1}^T B^{-1}(X_{\cdot,j} - e^{-\lambda\xi_j}X_{\cdot,j-1})}{1-e^{-2\lambda\xi_j}} \\ & - \frac{2}{\sigma^2}\sum_{j=2}^{M} \frac{\xi_j e^{-2\lambda\xi_j}}{(1-e^{-2\lambda\xi_j})^2}(X_{\cdot,j} - e^{-\lambda\xi_j}X_{\cdot,j-1})^T B^{-1}(X_{\cdot,j} - e^{-\lambda\xi_j}X_{\cdot,j-1}) \end{aligned} \tag{2}$$

$$\begin{aligned} dl_\mu = {} & M\sum_{k=2}^{N} \frac{2\zeta_k e^{-2\mu\zeta_i}}{1-e^{-2\mu\zeta_k}} + \frac{1}{\sigma^2}\Bigg[X_{\cdot,1}^T D_\mu B^{-1} X_{\cdot,1} \\ & + \sum_{j=2}^{M} \frac{\xi_j e^{-2\lambda\xi_j}}{(1-e^{-2\lambda\xi_j})^2}(X_{\cdot,j} - e^{-\lambda\xi_j}X_{\cdot,j-1})^T D_\mu B^{-1}(X_{\cdot,j} - e^{-\lambda\xi_j}X_{\cdot,j-1})\Bigg] \end{aligned} \tag{3}$$

$$\begin{aligned} h_{\lambda\lambda} = {} & N\sum_{j=2}^{M} \frac{-4e^{-2\lambda\xi_j}}{(1-e^{-2\lambda\xi_j})^2} - \frac{2}{\sigma^2}\sum_{j=2}^{M} \xi^2 e^{-\lambda\xi_j}\Bigg[\frac{(1+e^{-2\lambda\xi_j})X_{\cdot,j-1}^T B^-1(X_{\cdot,j} - e^{-\lambda\eta_j}X_{\cdot,j-1})}{(1-e^{-2\lambda\xi_j})^2} \\ & - \frac{e^{-\lambda\xi_j}X_{\cdot,j-1}^T B^{-1}X_{\cdot,j-1}}{1-e^{-2\lambda\xi_j}} + \frac{2e^{-2\lambda\eta_j}X_{\cdot,j-1}B^{-1}(X_{\cdot,j} - e^{-\lambda\eta_j}X_{\cdot,j-1})}{(1-e^{-2\lambda\eta_j})^2} \\ & - \frac{2e^{-\lambda\eta_j}(1+e^{-2\lambda\eta_j})(X_{\cdot,j} - e^{-\lambda\eta_j}X_{\cdot,j-1})^T B^{-1}(X_{\cdot,j} - e^{-\lambda\eta_j}X_{\cdot,j-1})}{(1-e^{-2\lambda\eta_j})^3}\Bigg] \end{aligned} \tag{4}$$

$$h_{\lambda\mu} = \frac{2}{\sigma^2} \sum_{j=2}^{M} \Bigg[\frac{\xi_j e^{-\lambda\eta_j} X_{\cdot,j-1}^T D_\mu B^{-1} (\mathcal{X} 1)}{1 - e^{-2\lambda\eta_j}} - \frac{\xi_j e^{-2\lambda\eta_j} (X_{\cdot,j} - e^{-\lambda\eta_j} X_{\cdot,j-1})^T D_\mu B^{-1} (X_{\cdot,j} - e^{-\lambda\eta_j} X_{\cdot,j-1})}{(1 - e^{-2\lambda\eta_j})^2} \Bigg] \tag{5}$$

$$h_{\mu\lambda} = \frac{1}{\sigma^2} \sum_{j=2}^{M} 2\xi_j e^{-2\lambda\eta_j} \Bigg[\frac{X_{\cdot,j-1}^T D_\mu B^{-1} (X_{\cdot,j} - e^{-\lambda\eta_j} X_{\cdot,j-1})}{1 - e^{-2\lambda\eta_j}} - (X_{\cdot,j} - e^{-\lambda\eta_j} X_{\cdot,j-1})^T D_\mu B^{-1} (X_{\cdot,j} - e^{-\lambda\eta_j} X_{\cdot,j-1}) \Bigg] \tag{6}$$

$$h_{\mu\mu} = M \sum_{k=2}^{N} \frac{-4\zeta_k^2 e^{-2\mu\zeta_j}}{(1 - e^{-2\mu\zeta_j})^2} + \frac{1}{\sigma^2} \Bigg[X_{\cdot,1}^T D_\mu^2 B^{-1} X_{\cdot,1} + \frac{k=j}{M} \frac{(X_{\cdot,j} - e^{-\lambda\eta_j} X_{\cdot,j-1})^T D_\mu^2 B^{-1} (X_{\cdot,j} - e^{-\lambda\eta_j} X_{\cdot,j-1})}{1 - e^{-2\lambda\eta_j}} \Bigg] \tag{7}$$

Appendix B: Implementation steps for the EM algorithm

The EM algorithm is one of the most widely-used methods in model-based inferences when there are unobserved or missing data, or when the likelihood function cannot be found explicitly. Based on the idea of 'guess, update and repeat', it iteratively computes the MLE of parameters using available information and updates its corresponding expected likelihood function. Many examples of its application has been chronicled in multiple papers since before its formal introduction in the 1970s. While some theoretical foundation was laid by Orchard and Woodbury [8], it was Dempster, Lair and Rubin [3] that first gave the EM algorithm its name and wide popularity via their classic paper in 1977, where they defined a generalized framework for the two-step method and presented theoretical details on asymptotic convergence. Later, Wu [22] furthered investigated the convergence properties on the sequence of EM estimates. To formally describe the algorithm, let $\mathcal{Y}$ be a sample space and Θ be a parameter space. Consider a set of n observations $Y \in \mathcal{Y}$, we can write $Y := (Y_{\text{obs}}, Y_{\text{mis}})$, where Y_{jobs} is the set of observed data and Y_{mis} is the set of unobserved or missing data. Notice that by unobserved we mean data from hidden variables, while missing data refers to the case when the data could have been observed directly. We assume Y has pdf or pmf $f(Y|\theta)$, where $\theta \in \Theta$ unknown. One way to estimate θ is by maximizing the incomplete data likelihood given by Y_{obs}, defined as :

$$L(\theta|Y_{obs}) := \int f(Y_{obs}, Y_{mis}|\theta) dY_{mis}.$$

Then, the goal is to find θ^* s.t.

$$\theta^* = \text{argmax}_{\theta \in \Theta} L(\theta|Y_{obs}),$$

and if $L(\cdot|Y_{obs})$ is differentiable and unimodal, we can find the MLE by solving for $\frac{\partial}{\partial\theta} \ln L(\theta|Y_{obs}) = 0$. The iterative step comes in when explicit solutions are not available, in which case, instead of solving for the maximum likelihood, we look at the expected likelihood function given Y_{jobs}. We define the following: for $p = 0, 1, 2, \ldots$

E-step : Let $l(\theta|Y) = \ln L(\theta|Y)$. Compute

$$Q(\theta|\theta^{(p)}) := \mathbb{E}_{\theta^{(p)}}(l(\theta|Y)|Y_{obs}) = \int l(\theta|Y) f(Y_{mis}|Y_{obs}, \theta = \theta^{(p)}).$$

M-step : Find $\theta^{(p+1)}$ s.t.

$$Q(\theta^{(p+1)}|\theta^{(p)}) \geq Q(\theta|\theta^{(p)}) \text{ for all } \theta \in \Theta.$$

Repeat until $|\theta^{(p+1)} - \theta^{(p)}| < \epsilon$

In our case, suppose there are a total of $n^{(u)}$ observations missing at random for the realization X, and let $n^{(o)}$ be the number of remaining observations, so that $MN = n^{(o)} + n^{(u)}$, we can partition X into 'observed' and 'unobserved' compartments:

$$X = \begin{bmatrix} X^{(o)} \\ \\ X^{(u)} \end{bmatrix} \sim N_{MN}\left(\begin{bmatrix} \mathbf{0}_{n^{(o)}\times 1} \\ \\ \mathbf{0}_{n^{(u)}\times 1} \end{bmatrix}, \begin{bmatrix} \Sigma_{oo} & \Sigma_{ou} \\ \\ \Sigma_{uo} & \Sigma_{uu} \end{bmatrix} \right), \tag{8}$$

where

- $\Sigma_{oo} \in \mathbb{R}^{n^{(o)}\times n^{(o)}}$ is the covariance matrix for the observed data,
- $\Sigma_{uu} \in \mathbb{R}^{n^{(u)}\times n^{(u)}}$ is the covariance matrix for the unobserved data, and
- $\Sigma_{uo} \in \mathbb{R}^{n^{(u)}\times n^{(o)}}$ is the covariance matrix for the unobserved and the observed data, and $\Sigma_{ou} = \Sigma_{uo}^T$.

The E-step of the EM algorithm in our case requires the conditional distribution of $X^{(u)}|X^{(o)}, \theta^{(p)}$, where $\theta^{(p)}$ is the set of current estimates for the parameters λ, μ and σ^2. Notice that from (8), we have

$$f(X^{(u)}|X^{(o)}, \theta^{(p)}) = (2\pi)^{-\frac{n^{(u)}}{2}} |S^{(u)}|^{-1/2} e^{(X^{(u)} - \mathbf{m}^{(u)})'(S^{(u)})^{-1}(X^{(u)} - \mathbf{m}^{(u)})}, \tag{9}$$

where $S^{(u)} = \Sigma_{uu} - \Sigma_{uo}(\Sigma_{oo})^{-1}\Sigma_{ou}$, and $\mathbf{m}^{(u)} = \Sigma_{uo}(\Sigma_{oo})^{-1}X^{(o)}$. This means that

$$Q(\theta|\theta^{(p)}) := \mathbb{E}_{\theta}[l(\theta, \mathcal{X})|\mathcal{X}^{(o)}, \theta^{(p)}] = \int l(\theta|\mathcal{X}) f(X^{(u)}|X^{(o)}, \theta^{(p)}) dX^{(u)}, \quad (10)$$

$$= MN \ln|\Sigma| + \mathbb{E}[\mathcal{X}'\Sigma^{-1}\mathcal{X}|X^{(o)}, \theta^{(p)}], \quad (11)$$

$$= MN \ln|\Sigma| + \sum_{i,j=1}^{MN} \sum_{i'j'=1}^{MN} \phi_{ij} \mathbb{E}[x_{ij} x_{i'j'}|X^{(o)}, \theta^{(p)}], \quad (12)$$

where ϕ_{ij} is the ij-th entry of Σ^{-1}. In particular, the value of $\mathbb{E}[x_{ij,x_{i'j'}}|X^{(o)}, \theta^{(p)}]$ depends on the (un)availability of the sampling site. Thus, the E-step essentially uses a conditional random field evaluated using $S^{(u)}$ and $\mathbf{m}^{(u)}$ at the current parameter update $\theta^{(p)}$, which implies that the likelihood function can be expressed using equation (1.8) with respect to the conditional data. Schematically, the algorithm can be summarized in the following steps.

1. For $p = 0, 1, 2 \ldots$, generate a random field $\mathcal{X}^{(p)}$ with 'complete' observation under parameter estimates $\sigma^{2(p)}$, $\lambda^{(p)}$, and $\mu^{(p)}$.

2. Define a mapping $\mathcal{M} : \Theta \to \Theta$, and choose $\theta^{(p+1)} \in \mathcal{M}(\theta^{(p)})$. In our case, the mapping can be expressed as

$$\{\theta \in \Theta : \mathbb{E}[l(\theta)|\mathcal{X}^{(p)}, \theta^{(p)}] \leq \mathbb{E}[l(\theta)|\mathcal{X}^{(p)}, \theta^{(p)}]\}.$$

In particular, each element of $\theta^{(p+1)}$ is updated according to the following:

$$\lambda^{(p+1)} = \{\lambda : \frac{\partial}{\partial\lambda}\mathbb{E}[l(\theta)|\mathcal{X}^{(p)}, \theta^{(p)}] = 0\}$$
$$\mu^{(p+1)} = \{\mu : \frac{\partial}{\partial\mu}\mathbb{E}[l(\theta)|\mathcal{X}^{(p)}, \theta^{(p)}] = 0\}$$
$$\sigma^{2(p+1)} = \frac{(\mathcal{X}^{(p)})_1^T B^{-1}(\mu^{(p+1)})\mathcal{X}_1^{(p)}}{MN} + \frac{\sum_{i=2}^{M} \frac{(\mathcal{X}_i^{(p)} - e^{-\lambda^{(p+1)}}\mathcal{X}_{i-1}^{(p)})^{-1}B^{-1}(\mu^{(p+1)})(\mathcal{X}_i^{(p)} - e^{-\lambda^{(p+1)}}\mathcal{X}_{i-1}^{(p)})}{1 - e^{-2\lambda^{(p+1)}\xi_i}}}{MN}$$

3. Repeat until convergence or maximum step number is achieved.

In other words, this is partly an imputation-based method using the neighboring available observations of the missing sites. The motivation for this approach is based on the assumption that observations that are closer together have higher correlations, thus it is reasonable to estimate parameter values from the incomplete random field, where the missing sites are replaced with conditional means based on information near them. The appendix lists the explicit derivatives of the likelihood function, as well as the expressions implemented in the Newton's method for find the MLE.

Properties of the EM estimates

In general, the $Q\left(\cdot|\theta^{(p)}\right)$ is a monotone function of p, and $\theta^{(p)}$ is guaranteed convergence to at least a stationary point of $l\left(\theta|Y\right)$.

Theorem B.1 (Dempster, Laird and Rubin,1976). *Every GEM algorithm increases $l(\theta|Y_{ons})$ at each iteration, that is*

$$l\left(\theta^{(p+1)}|Y_{obs}\right) \geq l\left(\theta^{(p)}|Y_{obs}\right).$$

Proof.

$$\begin{aligned} f(Y|\theta) &= f(Y_{obs}, Y_{mis}|\theta) \\ &= f(Y_{obs}|\theta) \cdot f(Y_{mis}|Y_{obs}, \theta) \end{aligned}$$

thus the corresponding log-likelihood decomposition is

$$l(\theta|Y) = l(\theta|Y_{obs}) + \ln f(Y_{mis}|Y_{obs}, \theta).$$

From the same decomposition, we also have

$$l(\theta|Y_{obs}) = l(\theta|Y) - \ln f(Y_{mis}|Y_{obs}, \theta) \tag{13}$$

Now, at iteration p in the E-step, (13) is expressed as :

$$l(\theta^{(p)}|Y_{obs}) = Q(\theta|\theta^{(p)}) - H(\theta|\theta^{(p)}), \tag{14}$$

where

$$Q(\theta|\theta^{(p)}) = \int l(\theta|Y) f(Y_{mis}|Y_{obs}, \theta = \theta^{(p)}),$$

and

$$H(\theta|\theta^{(p)}) = \int [\ln f(Y_{mis}|Y_{obs},\theta)] f(Y_{mis}|Y_{obs},\theta=\theta^{(p)}) dY_{mis}.$$

Note that since $ln(\cdot)$ is concave, by Jensen's inequality

$$H(\theta|\theta^{(p)}) \leq H(\theta^{(p)}|\theta^{(p)})$$

Thus, the difference in two consecutive iterations in (14) is :

$$\begin{aligned} & l(\theta^{(p+1)}|Y_{obs}) - l(\theta^{(p)}|Y_{obs}) \\ &= \underbrace{[Q(\theta^{(p+1)}|\theta^{(p)}) - Q(\theta^{(p)}|\theta^{(p)})]}_{>0 \text{ by definition of M-step}} - \underbrace{[H(\theta^{(p+1)}|\theta^{(p)}) - H(\theta^{(p)}|\theta^{(p)})]}_{<0 \text{ by Jensen's inequality}}. \end{aligned}$$

This implies,

$$l(\theta^{(p+1)}|Y_{obs}) \geq l(\theta^{(p)}|Y_{obs}) \text{ for all } p,$$

with equality if and only if both

$$Q(\theta^{(p+1)}|\theta^{(p)}) = Q(\theta^{(p)}|\theta^{(p)}) \text{ and } H(\theta^{(p+1)}|\theta^{(p)}) = H(\theta^{(p)}|\theta^{(p)}).$$

□

Thus, we can see from Theorem (B.1) that, if the likelihood function is bounded, then $l(\theta^{(p)}|Y_{obs})$ must converge to some value l^*, and if $l(\theta^{(p)}|Y_{obs})$ is continuous, then that implies $\theta^{(p)} \rightarrow \theta^*$ as well. Of course, the first question that arises is whether l^* and $\theta*$ correspond to the MLE of the problem. To answer this question, we look at the following theorem from Wu [22]:

Theorem B2 (Wu,1983)**.** *Let $\{\theta_p\}$ be a general EM (GEM) sequence generated by $\theta_{p+1} \in M(\theta_p)$ and suppose that*

1. *M is a closed point-to-set mapping over the compliment of the solution set $\mathcal{S}$, and*
2. *$L(\theta_{p+1}) > L(\theta_p)$ for all $\theta_p \notin \mathcal{S}$ (and vice versa if using negative likelihood function).*

Then all the limit points of $\{\theta_p\}$ are stationary points of Lm and $L(\theta_p)$ converges monotonically to $L^* = L(\theta^*)$ for some $\theta^* \in \mathcal{S}$.

Furthermore, it is stated that a sufficient condition for the closeness of M is that:

$$Q(\theta|\theta^{(p)}) \text{ is continuous in both } \theta \text{ and } \theta^{(p)}. \tag{15}$$

To show that the Q function in our problem is continuous, it is enough to show that it is continuous in both λ and μ, since σ^2 can be completely derived from them. Notice that, since

$$MN|\Sigma| = MN[\log(2\pi) + \log(\sigma^2)] + N\sum_{i=2}^{M}\log(1 - e^{-2\lambda\xi_i}) + M\sum_{k=2}^{N}\log(1 - e^{-2\mu\xi_k}) \tag{16}$$

is clearly a continuous function for λ, μ and σ^2. On the other hand,

$$\Sigma^{-1} = \sigma^2 A(\lambda)^{-1} \otimes B(\mu)^{-1}$$

with the terms of $A(\lambda)^{-1}$ being in the form of:

- $\bar{a}_{11} = \frac{1}{1-a_2^2}$
- $\bar{a}_{MM} = \frac{1}{1-a_M^2}$
- $\bar{a_{ii}} = \frac{1}{1-a_i^2} + \frac{1}{1-a_{i+1}^2} - 1, i = 2, \ldots, M-1$
- $\bar{a}_{i,i-1} = \frac{-a_i}{1-a_i^2} = \bar{a}_{i-1,i}, i = 2, \ldots, M$

with $\bar{a}_{ij} = 0$ if $|i-j| > 1$, and $a_i = e^{-\lambda|u_i - u_{i-1}|}$. The terms of $B^{-1}(\mu)$ are expressed similarly. Since each of a_i^2 is strictly between 0 and 1, we have that the terms for $A^{-1}(\lambda)$ are continuous for λ. This implies that the entries of Σ^{-1} are continuous for both λ and μ. Since $Q(\theta|\theta^{(p)})$ depends on the term $|\Sigma|$ and Σ^{-1}, we have that $Q(\theta|\theta^{(p)})$ is continuous in λ and μ as well. To show continuity in $\lambda^{(p)}$ and $\mu^{(p)}$, note that

$$\mathbb{E}[x_{ij}x_{i'j'}|\mathcal{X}^{(o)}, \theta^{(p)}] = \begin{cases} x_{ij}x_{i'j'} & \text{if both } x_{ij}x_{i'j'} \text{ are observed} \\ x_{ij}\mathbb{E}[x_{i'j'}|\mathcal{X}^{(o)}, \theta^{(p)}] & \text{if only } x_{ij} \text{ is observed, and vice versa.} \end{cases} \tag{17}$$

Now, the terms $\mathbb{E}[x_{i'j'}|\mathcal{X}^{(o)}, \theta^{(p)}]$ and $\mathbb{E}[x_{ij}x_{i'j'}|\mathcal{X}^{(o)}, \theta^{(p)}]$ are elements of $m^{(u)}|_{\theta=\theta^{(p)}}$ and $S^{(u)}|_{\theta=\theta^{(p)}}$ respectively. Thus, we need to show that $m^{(u)}|_{\theta=\theta^{(p)}}$ and $S^{(u)}|_{\theta=\theta^{(p)}}$ are continuous in $\theta^{(p)}$. Note that

$$S^{(u)}|_{\theta=\theta^{(p)}} = (\Sigma_{uu} - (\Sigma_{oo})^{-1}\Sigma_{ou})|_{\theta=\theta^{(p)}} \tag{18}$$

since Σ_{uu} is a sub matrix of Σ formed by deleting the rows and columns corresponding to the observed data , each element of $\Sigma_{uu}|_{\theta=\theta^{(p)}}$ can be expressed as either $\sigma^{2(p)}$ or

$$\text{Cov}(x, x')|_{\theta=\theta^{(p)}} = \sigma^{2(p)} e^{-\lambda^{(p)} u* - \mu^{(p)} v*} \tag{19}$$

where $u*$ and $v*$ are some constant strictly between 0 and 1. Thus, $S_{uu}|\theta_{\theta^{(p)}}$ is continuous for $\lambda^{(p)}, \mu^{(p)}$ and $\sigma^{2(p)}$. Similarly, the elements of $\Sigma_{ou}|_{\theta=\theta^{(p)}}$ are of the same form since it is also a sub matrix of $\Sigma_{..}$ It remains to show continuity of $\Sigma_{oo}^{-1}|_{\theta=\theta^{(p)}}$. To this end, consider an $n \times n$ matrix A, then its determinant can be expressed as

$$\det(A) = \Sigma_\sigma (a_{1\alpha} a_{2\beta} \cdots a_{n\nu}) \det P_\sigma \tag{20}$$

where

- each σ corresponds to a distinct permutation of $(1, 2, \ldots, n)$,
- $(\alpha, \beta, \ldots, \nu)$ is the set of indices with respect to that permutation, and
- P_σ is a permutation matrix whose determinant is either 1 or -1.

Using Cramer's rule, A^{-1} can be expressed as

$$A^{-1} = \frac{C^T}{det(A)}, \tag{21}$$

where C is a matrix of co-factors for A, which again is a linear combination of the products of the elements of A, but with order $n-1$ instead of n. Now, the terms of Σ_{oo} follow the form in (19), thus the elements of Σ_{oo}^{-1} are therefore continuous. Since products of continuous functions are still continuous, continuity of $S^{(u)}|_{\theta=\theta^{(p)}}$

is therefore satisfied. It follows that $m^{(u)}|_{\theta=\theta^{(p)}} = \Sigma_{uo}(\Sigma_{oo})^{-1}$ is continuous as $\Sigma_{uo} = \Sigma_{ou}^T$.

Appendix C: MATLAB code for numerical experiments

This section lists the MATLAB source code that is used for all the simulation studies in this paper

Simulation of randomly missing observations in the random field (OU_MISS.m)

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% OU_MISS.m
%% Function to generate missing observations for OU_SIM
%% Author  : Sami Cheong
%% Date    : 7/29/14
%% Version : 1
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% function X_miss= OU_MISS(X,miss_level)
```

```
% INPUT:
% X          = complete-observation Gaussian random field.
% miss_level = % of the observations that is missing
% OUTPUT:
% X_miss     = X with randomly missing values according to miss_level
function [X_miss,miss_ind]= OU_MISS(X,miss_level)
N=size(X,1);
M=size(X,2);
X=reshape(X,[N*M,1]);
% Create missing observations using Bernoulli distribution with
% probability defined by 'miss_level':
miss_ind=binornd(1,miss_level,N*M,1);
% Or randomly permute the sampling sites:
% perm_ind = randperm(N*M);
% Choose the % of observations missing as represented by the index:
%miss=perm_ind(1:floor(miss_level*(N*M)));
X_miss=X;
X_miss(miss_ind==1)=NaN;
X_miss=reshape(X_miss,[N,M]);
miss_ind=reshape(miss_ind,[N,M]);
```

Simulation of missing blocks in the random field (OU_BLOCKMISS.m)

```
%% OU_BLOCKMISS.m
%% Function to simulate a block of missing observations in a random
%% field
```

```
%% function [Y,K1,J1] = OU_BLOCKMISS1(X,missing_prop)
%%
%% INPUT :
%% X              = matrix of realization of a random field
%%                  with complete observations
%% missing_prop   = proportion of the observations that are missing
function [Y,K1,J1] = OU_BLOCKMISS(X,missing_prop)
if (missing_prop <= 0 || missing_prop >=0.45)
    error('missing level must be strictly between 0 and 0.45')
else
end
% get dimension:
N = size(X,1);
M = size(X,2);
% Total number of observations
% generate the block dimension according to level of missingness:
% C is the area of the missing block,
% we need to generate dimensions of the missing block
Y=X;
C=round(1/missing_prop);
factor = [9 8 7 6 5 4 3 2];
i=1;
while i < length(factor)
    if mod(C,factor(i))==0
            C1=C/factor(i);
            C2=C/C1;
            i=i+1;
    else
        C1=floor(C/3);
```

```
        C2=floor(C/C1);
        i=i+1;
    end
end
N1=round(N/C1);
M1=round(M/C2);
% generate an index to start the missing block:
k1=randi([3 floor(N/4)],1);
j1=randi([3 floor(M/4)],1);
K1=k1:1:min((k1+(N1-1)),N);
J1=j1:1:min(j1+(M1-1),M);
% Assign NaN to the resulting block
Y(K1,J1)=NaN;
```

Conditional random field (OU_COND.m)

```
%% Evaluate the conditonal mean and variance given incomplete X
% OU_COND.m
% function [X_cond,S_oo,S_uu,S_ou,S_uo]=OU_COND(X,Gamma)
% INPUT:
% X        = N-by-M centered Gaussian random (OU) field
% (with missing observations).
% Gamma    = Covariance structure of the OU field,
% evaluated at the current parameter value.
%
% OUTPUT:
% X_cond   = Random field where missing observations are replaced with
% conditonal mean.
```

```
% Cov_cond = conditional covariance matrix for the unobserved samples
function [X_cond,S_oo,S_uu,S_ou,S_uo]=OU_COND(X,Gamma)
% reshape X into a MN-by-1 vector:
N=size(X,1);
M=size(X,2);
X= reshape(X,[N*M,1]);
% get unobserved and observed indices:
[Unobs_ind] = find(isnan(X)==1);
[Obs_ind]=find(isnan(X)==0);
% Partition the covariance matrix S= [S_oo | S_ou; S_uo | S_uu]:
S_oo = Gamma(Obs_ind,Obs_ind);
S_uu = Gamma(Unobs_ind, Unobs_ind);
S_ou = Gamma(Obs_ind,Unobs_ind);
S_uo=S_ou';
X_obs=X(Obs_ind);

Xstar=S_oo\X_obs;
X_unobs = S_uo*Xstar ;
X_cond=X;
X_cond(Unobs_ind)=X_unobs;
X_cond=reshape(X_cond,[N,M]);
```

Inverse of the complete-data covariance matrix, matrix square root and related derivatives

Explicit form of B^{-1}

```
%% This function computes the exact inverse of the
%%covariance matrices used in the OU process
function B_inv = OU_COVINV(b)
% INPUT:
% b = vector of elements that form the sq matrx.
% e.g. b(1:N)=exp(-mu.*nu(1:N));
% Assign values to border elements:
N = length(b);
B_inv(1,1)=1/(1-(b(2))^2);
B_inv(N,N)=1/(1-(b(N))^2);
for k=2:N-1
   B_inv(k,k)  = 1/(1-(b(k+1))^2)+1/(1-(b(k))^2)-1 ;
end
for k=2:N
   B_inv(k,k-1)=-b(k)/(1-(b(k))^2);
   B_inv(k-1,k)=-b(k)/(1-(b(k))^2);
end
```

Bidiagonal matrix square-root of B^{-1}

```
%% This function finds the lower-bidiagonal
%% matrix square-root for a symmatric tridiagonal matrix
% INPUT:
% G = symmetric tridiagonal matric
% OUTPUT:
% L = lower bidiagonal matrix such that G=L*L'
function [L]=OU_SQRTM(G)
if norm(G-G') > 0
```

```
    error('G must be symmetric and tridiagonal')
else
end
K=size(G,1);
L=zeros(K,K);
L(1,1) = sqrt(G(1,1));
L(K,K) = sqrt(G(K,K));
for j=2:K
   L(j,j-1)=G(j-1,j)/L(j-1,j-1);
   L(j,j)  =sqrt(G(j,j)-L(j,j-1)^2);
end
```

Computing $D_\mu B^{-1}$ and $D^2_\mu B^{-1}$

```
%% This function computes the derivative of
%% the inverse of the covariance matrix B^-1(\mu):
% function DB_inv = OU_DBINV(mu,v)
function DB_inv = OU_DBINV(mu,v)
% INPUT:
% mu = parameter value for B(mu)
% v  = Vector for the vertical coordinate of the random field
% OUTPUT:
% DB_inv = A matrix of derivatives for B^-1(\mu)
% Assign values to border elements:
N = length(v);
DB_inv=zeros(N,N);
% standaridze vectors into column vectors
if size(v,2)~=1
```

```
   v=v;
else
end
nu =[0; abs(v(2:N)−v(1:N−1))];
b = exp(−mu*nu);
b_sq=b.^2;
DB_inv(1,1) = −2*nu(2)*b_sq(2)/((1−b_sq(2))^2);
DB_inv(N,N) = −2*nu(N)*b_sq(N)/((1−b_sq(N))^2);
for j=2:N−1
    DB_inv(j,j)  = −2*(nu(j)*b_sq(j)/((1−b_sq(j)))^2 ...
        +nu(j+1)*b_sq(j+1)/((1−b_sq(j+1)))^2);
end
for j=2:N

    DB_inv(j−1,j) = nu(j)*b(j)*(1+b_sq(j))/((1−b_sq(j))^2);
    DB_inv(j,j−1) =DB_inv(j−1,j);
end



%% This function computes the 2nd derivtive
%% of the inverse of the covariance matrix B^−1(\mu):
% function DB2_inv = OU_DBINV(mu,v)
function DB2_inv = OU_D2BINV(mu,v)
% INPUT:
% mu = parameter value for B(mu)
% v  = Vector for the vertical coordinate of the random field
% OUTPUT:
% DB_inv = A matrix of derivatives for B^−1(\mu)
% Assign values to border elements:
```

```
N = length(v);
DB2_inv=zeros(N,N);
% standaridze vectors into column vectors
if size(v,2)~=1
   v=v;
else
end
nu =[0; abs(v(2:N)-v(1:N-1))];
b = exp(-mu.*nu);
b_sq=b.^2;
b_star=(nu.*b_sq.*(ones(N,1)+b_sq))/((ones(N,1)-b_sq).^3);
DB2_inv(1,1) = 4*b_star(2);
DB2_inv(N,N) = 4*b_star(N);
for j=2:N-1
    DB2_inv(j,j)  = 4*(b_star(j)+b_star(j+1));
end
for j=2:N
    DB2_inv(j-1,j) = ...
        ((nu(j))^2*b(j)*(1+6*(b(j))^2+(b(j))^4))/((1-b_sq(j))^2);
    DB2_inv(j,j-1) =DB2_inv(j-1,j);
end


%% This function finds the lower-bidiagonal
%% matrix square-root for a symmatric tridiagonal matrix
% INPUT:
% G = symmetric tridiagonal matric
% OUTPUT:
% L = lower bidiagonal matrix such that G=L*L'
```

```
function [L]=OU_SQRTM(G)
if norm(G-G') > 0
    error('G must be symmetric and tridiagonal')
else
end
K=size(G,1);
L=zeros(K,K);
L(1,1) = sqrt(G(1,1));
L(K,K) = sqrt(G(K,K));
for j=2:K
   L(j,j-1)=G(j-1,j)/L(j-1,j-1);
   L(j,j)  =sqrt(G(j,j)-L(j,j-1)^2);
end
```

Estimation schemes, likelihood functions and related derivatives

Complete-data likelihood function

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% OU_LIKE.m
%% Function to evaluate the -2 log likelihood function
%% of the parameter values given an observation of
%% the OU process
%% Author  : Sami Cheong
%% Date    : 7/29/14
%% Version : 1
```

```
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% function likelihood = OU_LIKE(lambda,mu,sigma2,u,v,X)
function likelihood = OU_LIKE(lambda,mu,sigma2,u,v,X)
N=size(X,1);
M=size(X,2);
% eta=zeros(M,1);
% nu=zeros(N,1);
% if u(1) == 0
%     eta(1) = u(1);
% else
%     eta(1) = 0;
% end
% if v(1) == 0
%     nu(1)  = v(1);
% else
%     nu(1) = 0;
% end
% work with column vectors:
if size(u,2)~=1
    u=u';
end
if size(v,2)~=1
    v=v';
end
eta = [0;abs(u(2:M)-u(1:M-1))];
nu  = [0;abs(v(2:N)-v(1:N-1))];
b = exp(-mu.*nu);
% define the different terms in the likelihood function
```

```
term_pi = M*N*log(2*pi);
term_sigma  = M*N*log(sigma2);
term_lambda = N*sum(log(1-exp(-2*lambda.*eta(2:M))));
term_mu = M*sum(log(1-exp(-2*mu.*nu(2:N))));
% initialize the terms for the quadratic form n
long_term = zeros(M,1);
Xa = zeros(size(X));
Binv = OU_COVINV(b);
Xa(:,1)= X(:,1);
for i=2:M
   Xa(:,i)= X(:,i)-(exp(-lambda.*eta(i))*X(:,i-1));
   long_term(i)=(Xa(:,i)'*Binv*Xa(:,i))/(1-exp(-2*lambda.*eta(i)));
end
likelihood = term_pi+term_sigma+term_lambda+term_mu +...
      (X(:,1)'*Binv*X(:,1)+ sum(long_term(2:M)))/sigma2;
```

Parameter updates based on the Hessian matrix of the complete-data likelihood function

```
%% function hess = OU_LIKE_HESS_UPDATE(lambda,mu,sigma2,u,v,X)
%% function to evaluate the hessian matrix
%% of the complete data likelihood
%% generate structure :H=[h_lmblmb, h_lmu; h_mul, h_mumu]
%% Uses other functinos : OU_DBINV(mu,v) , OU_D2BINV(mu,v)
function [update] = OU_LIKE_HESS_UPDATE(lambda,mu,sigma2,u,v,X)
% H=[H_11, H_12; H_21, H_22]
% define the components of the Hessian matrix:
M=size(X,2);
```

```
N=size(X,1);
% standardize things to be column vectors :
if size(u,2)~=1
u=u';
else
end
if size(v,2)~=1
v=v';
end
eta=[0;abs(u(2:end)-u(1:end-1))];
zeta=[0;abs(v(2:end)-v(1:end-1))];
zeta_expmu  = exp(-mu.*zeta);
zeta_expmu2 = ones(N,1)-(zeta_expmu.^2);
eta_explmb  = exp(-lambda.*eta);
eta_explmb2 = ones(M,1)-(eta_explmb.^2);
% compute B^-1:
B_inv=OU_COVINV(zeta_expmu);
% initialize stuff
%X_q=NaN(size(X));
for j=2:M
% define terms in H_11 = h_lambdalambda
X_q(:,j)=X(:,j)-eta_explmb(j)*X(:,j-1);
h_lmblmb_1(j)  =  -4*(eta(j)*eta_explmb(j))^2/((eta_explmb2(j))^2);
h_lmblmb_2(j)  = ((eta(j))^2*eta_explmb(j)*(1+(eta_explmb(j))^2)*...
    (X(:,j-1)')*B_inv*X_q(:,j))/((eta_explmb2(j))^2);
h_lmblmb_3(j)  = (eta(j)*eta_explmb(j))^2*(X(:,j-1)')*...
    B_inv*X(:,j-1)/(eta_explmb2(j));
h_lmblmb_4(j)  = 2*(eta(j)*eta_explmb(j))^2*(1+(eta_explmb(j))^2)*...
    X_q(:,j)'*B_inv*X_q(:,j)/((eta_explmb2(j))^3);
```

```
h_lmblmb_5(j) = 2*(eta(j)*eta_explmb(j))^2*eta_explmb(j)*...
    (X(:,j-1)')*B_inv*X_q(:,j)/((eta_explmb2(j))^2);
% split the H_12 = h_lambdamu:
h_lmbmu_1(j) = (eta(j)*eta_explmb(j)*X(:,j-1)'*...
    OU_DBINV(mu,v)*X_q(:,j))/(eta_explmb2(j));
h_lmbmu_2(j) = ((eta(j)*(eta_explmb(j))^2)*X_q(:,j)'*...
    OU_DBINV(mu,v)*X_q(:,j))/((eta_explmb2(j))^2);
h_mulmb_1(j) = (eta(j)*((eta_explmb(j))^2)*X(:,j-1)'*...
    OU_DBINV(mu,v)*X_q(:,j))/(eta_explmb2(j));
h_mulmb_2(j) = (eta(j)*((eta_explmb(j))^2)*X_q(:,j)'*...
    OU_DBINV(mu,v)*X_q(:,j));
% define the lambda term in h_mumu:
h_mumu_star(j) = ...
   (X_q(:,j)'*OU_D2BINV(mu,v)*X_q(:,j))/(eta_explmb2(j));
dl_mu_star(j) =...
   (X_q(:,j)'*OU_DBINV(mu,v)*X_q(:,j))/(eta_explmb2(j));
% define the terms in the first derivative of l(theta):
dl_lmb_1(j)=(2*eta(j)*(eta_explmb(j))^2)/(eta_explmb2(j));
dl_lmb_2(j)=...
(eta(j)*eta_explmb(j)*X(:,j-1)'*B_inv*X_q(:,j))/(eta_explmb2(j));
dl_lmb_3(j)=...
eta(j)*(eta_explmb(j)/(eta_explmb2(j)))^2*((X_q(:,j)')*B_inv*X_q(:,j));
end
for k=2:N
h_mumu_1(k)=-4*(zeta(k)*zeta_expmu(k))^2/((zeta_expmu2(k))^2);
dl_mu_1(k)= 2*zeta(k)*(zeta_expmu(k))^2/(zeta_expmu2(k));
end
%% Put all the terms together:
h_lmblmb = sum(N*h_lmblmb_1(2:M))-...
```

```
            (2/sigma2)*(sum(h_lmblmb_2(2:M)-h_lmblmb_3(2:M)-...
            h_lmblmb_4(2:M)+h_lmblmb_5(2:M)));
h_lmbmu=(2/sigma2)*sum(h_lmbmu_1(2:M)-h_lmbmu_2(2:M));
h_mulmb = (2/sigma2)*(sum(h_mulmb_1(2:M)-h_mulmb_2(2:M)));
h_mumu = sum(M*h_mumu_1(2:N)) +...
        (1/sigma2)*(X(:,1)'*OU_D2BINV(mu,v)*X(:,1) +...
        sum(h_mumu_star(2:M)));
% Define the Hessian matrix
H=[h_lmblmb,h_lmbmu; h_mulmb h_mumu];
% Partial derivatives:
dl_lmb=sum(N*dl_lmb_1(2:M)+(2/sigma2)*...
    dl_lmb_2(2:M)-(2/sigma2)*dl_lmb_3(2:M));
dl_mu = sum(M*dl_mu_1(2:N))+(1/sigma2)*(X(:,1)'*...
    OU_DBINV(mu,v)*X(:,1)+sum(dl_mu_star(2:M)));
update = H\[dl_lmb;dl_mu];
```

Estimation schemes

Obtaining MLE of σ^2 using existing $\hat{\lambda}$ and $\hat{\mu}$

```
%% OU_SIG_LIKE.m
%% function sigma2_hat = OU_SIG_LIKE(lambda_hat,mu_hat,u,v,X)
%% This function returns the likelihood estimate of sigma2
%% evaluated with mu_hat and lambda_hat:
%% function sigma2_hat = OU_SIG_LIKE(lambda_hat,mu_hat,u,v,X);
%% INPUT :
%% lambda_hat, mu_hat : current estimate of lambda and mu.
%% u , v : horizontal and vertical strip respectively.
```

```
%% X : random field with complete data or missing data imputed.
%% OUTPUT:
%% sigma2_hat = MLE of sigma based on input of lambda and mu
%% Date : 6/25/14 by Sami Cheong
function sigma2_hat = OU_SIG_LIKE(lambda_hat,mu_hat,u,v,X)
N=size(X,1);
M=size(X,2);
eta=zeros(M,1);
nu=zeros(N,1);
eta(1)=u(1);
nu(1) =v(1);
eta(2:M)=abs(u(2:M)-u(1:M-1));
nu(2:N)=abs(v(2:N)-v(1:N-1));
b(1)=0;
b(2:N)=exp(-mu_hat.*nu(2:N));
% Define the term used for the estimate:
Xa=zeros(size(X));
long_term=zeros(M,1);
Xa(:,1)=X(:,1);
Binv=OU_COVINV(b);
for i=2:M
   Xa(:,i)=X(:,i)-(exp(-lambda_hat*eta(i))*X(:,i-1));
   long_term(i)=(Xa(:,i)'*Binv*Xa(:,i))/(1-exp(-2*lambda_hat*eta(i)));
end
sigma2_hat=((X(:,1)'*Binv*X(:,1)+ sum(long_term(2:M))))/(M*N);
```

Obtaining MLE of λ, μ and σ^2 using the EM algorithm

```
%% OU_EM.m
%% Function to implement the EM algorithm.
%% Author  : Sami Cheong
%% Date    : 7/29/14
%% Version : 1
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% function likelihood = OU_EM(lambda,mu,sigma2,u,v,X)
% INPUT :
% X_miss
% = OU field with randomly missing observations.
% u
% = horizontal coordinate of input grid between [0,1]
% v
% = vertical coordinate of input grid between [0,1]
% lambdap, mup, sigma2p
% = initial parameter values for the algorithm.
% OUTPUT:
% theta_new
% = EM estimation of the parameter lambda, mu, sigma2
% lmb_vec,sig_vec,mu_vec
% = Vector of the EM estimates at each iteration.
% iter_vec
% = vector of indices that keeps track of steps.
%
function [theta_new,lmb_vec,sig_vec,mu_vec,iter_vec,lmn_vec] =...
    OU_EM1(X_miss,u,v,lambdap,mup,sigma2p)
% set tolerance and max number of iterations
tol=0.0001;
```

```
max_step=100;
iter=1;
lambda_int=lambdap;
mu_int=mup;
%sigma_int=sigma2p;
% Initial parameter values:
%theta_p=[lambdap;mup;sigma2p];
theta_new=zeros(3,1);
err=1000;
% Initialize:
lmb_vec = NaN(max_step,1);
sig_vec = NaN(max_step,1);
mu_vec  = NaN(max_step,1);
lmn_vec=NaN(max_step,1);
err_new=2000;
while abs(err_new-err) > tol && iter < max_step
    err_new=err;
% Evaluate covariance matrix wrt current parameter value:
    [~,~,~,Gamma]=OU_SIM(u,v,lambdap,mup,sigma2p);
% Generate conditional random field:
    [X_cond,~,~,~,~]=OU_COND1(X_miss,Gamma);
    theta_new(1)= fminsearch(@(lambda) OU_LIKE(lambda,mup,...
        sigma2p,u,v,X_cond),lambda_int);
    theta_new(2)= fminsearch(@(mu) OU_LIKE(lambdap,mu,...
        sigma2p,u,v,X_cond),mu_int);
    theta_new(3)= OU_SIG_LIKE(theta_new(1),theta_new(2),u,v,X_cond);
% Keep track of the estimates at each iteration
    lmb_vec(iter)=lambdap;
    mu_vec(iter)=mup;
```

```
    sig_vec(iter)=sigma2p;
    lambdap = theta_new(1);
    mup = theta_new(2);
    sigma2p = theta_new(3);
% Keep track of the value of the likelihood function:
    lmn_vec(iter)=OU_LIKE(theta_new(1),...
        theta_new(2),theta_new(3),u,v,X_cond);
    theta_p=[lambdap;sigma2p;mup];
% Keep track of estimation error:
    err = sum(abs(theta_new-theta_p));
% Update iteration
     iter=iter+1;
end
[~,~,~,Gamma]=OU_SIM(u,v,theta_new(2),theta_new(3),theta_new(3));
% Generate conditional random field:
[X_cond,~,~,~,~]=OU_COND(X_miss,Gamma);
theta_new(3)= OU_SIG_LIKE(theta_new(1),theta_new(2),u,v,X_cond);

iter_vec=1:iter-1;
lmb_vec=lmb_vec(iter_vec);
mu_vec=mu_vec(iter_vec);
sig_vec=sig_vec(iter_vec);
lmn_vec=lmn_vec(iter_vec,:);
```

Obtaining MLE of λ, μ and σ^2 using Newton's method

```
%% OU_LIKE_NEWTON.m
% function [theta] = OU_LIKE_NEWTON(theta0, delta)
```

```
%
% Function to implement Newton's method on l(X) to obtain MLE
% where X is an OU process with complete observation
% INPUT:
% lambda0,mu0,sigma20 = initial guess for the three parameter
%                       values
% u,v                 = input grid for the random field
% X                   = set of observations from which we wish to
%                       approximate the parameter
% delta               = accuracy we set for the estimate,
%                       delta = ||theta-theta0||_2
function [theta] = OU_LIKE_NEWTON(lambda0,mu0,sigma20,u,v,X, delta)
%format long e
lambda0=OU_SUB_reset_bound(lambda0);
mu0=OU_SUB_reset_bound(mu0);
sigma20=OU_SUB_reset_bound(sigma20);
% Evaluate the initial value wrt the complete data likelihood
l_0 = OU_LIKE(lambda0,mu0,sigma20,u,v,X);
if abs(l_0) <= delta
           %% check to see if initial guess satisfies
  return;                        %% convergence criterion.
end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%
%% MAIN ROUTINE
%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
max_iter=2000;
iter=0;
```

```
error = 1000;
while (error > delta && iter < max_iter)
  l_0 = OU_LIKE(lambda0,mu0,sigma20,u,v,X);
% update parameters lambda and mu using Newton's method
  theta_update = ...
      [lambda0; mu0]-OU_LIKE_HESS_UPDATE(lambda0,mu0,sigma20,u,v,X);
  lambda0 = OU_SUB_reset_bound(theta_update(1));
  mu0= OU_SUB_reset_bound(theta_update(2));
% update sigma2
  sigma20 =  OU_SIG_LIKE(lambda0,mu0,u,v,X);
  sigma2_update = sigma20;
% measure error in the likelihood function
  error= abs(l_0 - OU_LIKE(lambda0,mu0,sigma20,u,v,X));
% update iteration
  iter = iter +1;
% print stuff
%fprintf('\n Newton iteration  = %d, delta = %d,\n', iter,error)
%fprintf('\n lambda = %d, mu = %d, sigma2 = %d, \n', ...
% lambda0, mu0,sigma20)
%theta = [theta_update(1);theta_update(2);sigma2_update];
end;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%% Sub function to make sure the estimates don't go nuts.
function theta_reset = OU_SUB_reset_bound(x)
    if (x <= 1 || x > 100)
    x = 2+randi(5);
fprintf('\n Parameter values reset to default,value = %d, \n',x)
else
end
```

```
theta_reset = x;
return;
```

Obtaining ALE of λ, μ and σ^2 based on Markov property assumption

- With block-missing observations

```
%% OU_APPROXLIKE.m
%% This code implements the approximate likelihood function
%% based on th Markov property assumption in partitioned data
% Author: Sami Cheong
% Version : 0
% Date  : 10/27/2014
% function likelihood=OU_APPROXLIKE(X_miss,lambda,mu,sigmas)
% The goal is to approximate l_mn with l_1 + l_2,
% where l_1 is the likelihood for complete
% column observations, and l_2 is the likelihood for incomplete
% observations
function likelihood=OU_APPROXLIKE(lambda,mu,sigmas,u,v,X)
N=size(X,1);
M=size(X,2);
% Set of indices for the random field:
J=1:M;
K=1:N;
% identify columns with missing data (NaN):
Miss_mat=isnan(X);
Miss_col=find(any(Miss_mat));
```

```
Miss_row=find(isnan(X(:,Miss_col(1)))==1);
Nprime=length(Miss_row);
Mprime=length(Miss_col);
% J is the horizontal axis of the field
Jprime=Miss_col;
J2prime=Jprime;
% Kprime is the indices that for the usable information in the
% missing block columns
if Miss_row(1) > ceil((N)/2)
    Kprime= 1:1:(Miss_row(1)-1);
    K2prime=Miss_row(end)+1:1:N;
else
    Kprime=(1+Miss_row(end)):1:N;
    K2prime=1:1:Miss_row(1)-1;
end
eta=[];
nu=[];
if u(1) == 0
    eta(1) = u(1);
else
    eta(1) = 0;
end
if v(1) == 0
    nu(1)  = v(1);
else
    nu(1) = 0;
end
eta(2:M) = abs(u(2:M)-u(1:M-1))
nu(2:N)  = abs(v(2:N)-v(1:N-1))
```

```
% covariance structure for the complete data:
b  = exp(-mu.*nu)
Binv = OU_COVINV(b)
% covariance structure for the incomplete data:
b0 = exp(-mu.*nu(Kprime))
B0inv = OU_COVINV(b0);
b1=exp(-mu.*nu(K2prime));
B1inv=OU_COVINV(b1);
% Define indices and dimensions for l_1 and l_2
J1=setdiff(J,Jprime);
K1=setdiff(K,Kprime);
J2=Jprime;
K2=Kprime;
J3=J2prime;
K3=K2prime;
M1=length(J1);
N1=N;
M2=length(J2);
N2=length(K2);
M3=length(J3);
N3=length(K3);
% Define quadratic term for l_1:
J11=setdiff(J1,1);
K11=setdiff(K1,1);
Q1=0;
for j = setdiff(J11,Jprime(end)+1)
    Xj_star=X(:,j)-exp(-lambda*eta(j))*X(:,j-1);
    s1=(Xj_star'*Binv*Xj_star)/(1-exp(-2*lambda*eta(j)));
    Q1=s1+Q1;
```

```
end
l_1= M1*N1*log(2*pi*sigmas)...
    +M1*sum(log(1-exp(-2*mu.*nu(2:N))))+...
    N1*sum(log(1-exp(-2*lambda.*eta(J11))))...
    +(X(:,1)'*Binv*X(:,1)+Q1)./sigmas;
% Define the usable observations:
X0 = X(K2,J2);
%Define quadratic term for l_2:
Q2=0;
for j = 2:M2;
    X0j_star=X0(:,j)-exp(-lambda*eta(J2(j)))*X0(:,j-1);
    Q2=Q2+...
    ((X0j_star'*B0inv*X0j_star)/(1-exp(-2*lambda*eta(J2(j)))));
end
K22=setdiff(K2,1);
J22=setdiff(J2,1);
l_2= M2*N2*log(2*pi*sigmas)...
    +M2*sum(log(1-exp(-2*mu*nu(K22))))+...
    N2*sum(log(1-exp(-2*lambda*eta(J22))))...
    +(X0(:,1)'*B0inv*X0(:,1)+Q2)./sigmas;
% Define the usable observations:
X1=X(K3,J3);
Q3=0;
for j=2:M3
    X1j_star=X1(:,j)-exp(-lambda*eta(J3(j)))*X1(:,j-1);
    Q3=Q3+...
    ((X1j_star'*B1inv*X1j_star)/(1-exp(-2*lambda*eta(J2(j)))));
end
K33=setdiff(K3,1);
```

```
J33=setdiff(J3,1);
l_3= M3*N3*log(2*pi*sigmas)...
    +M3*sum(log(1-exp(-2*mu*nu(K33))))+...
    N3*sum(log(1-exp(-2*lambda*eta(J33))))...
    +(X1(:,1)'*B1inv*X1(:,1)+Q3)./sigmas;
% Sum up the likelihood functions from partitioned data
likelihood = l_1+l_2+l_3;
```

- With randomly-missing observations

```
%% OU_BINAPPROX.m
%% function [approx_like,approx_err] =
%% OU_BINAPPROX_LAMBDA(X_miss,u,v,A,B,mu,lambda...
% ,sigma2,num_term_approx)
%
% Implements the approximation likelihood function for a
% realization of an OU field with missing observations
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% INPUT:
% X_miss            :
% set of available observations in the
% OU field arranged in a 2D matrix
% u, v              :
% horizontal and vertical input grids
% A, B              :
% Covariance matrices for the horizontal
% and vertical components
```

```
% mu,lambda,sigmas  :
% parameters for the model
% num_term_approx   :
% number of terms to use in the power series
% calc_error        :
% indicates whether to return the error of
% approximating the inverse of the conditional
% covariance matrix
% OUTPUT:
% approx_like       :
% value of the approximated likelihood based on the
% description above
% approx_error      :
% difference between the direct inverse and the
% approximation
% Version : 1
% Date    : 7/23/15
% Author  : Sami Cheong
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [approx_log_like,approx_err] ...
    = OU_BINAPPROX(X_miss,u,v,A,B,...
      mu,lambda,sigma2,...
      num_term_approx,calc_error)
% initialize components of the log likelihood
approx_err_temp=0   ;
M = size(X_miss,2);
N = size(X_miss,1);
log_like_vec=NaN(M,1);
```

```
quadratic_term = NaN(M,1);
num_of_obs = NaN(M,1);
% get set of indices with available observations:
v1_avail_ind = find(~isnan(X_miss(:,1)));
Xo_1     = X_miss(v1_avail_ind,1);
B11      = B(v1_avail_ind,v1_avail_ind);
v1_avail     = v(v1_avail_ind);
% standardize vectors to be column vectors
if size(v1_avail,2)~=1
    v1_avail = v1_avail';
    else
end
% The first column of X_miss is treated as a ...
% special case since it does not have conditional density
zeta_11 = [0;abs(v1_avail(2:end)-v1_avail(1:end-1))];
B11_inv = OU_COVINV(exp(-mu.*zeta_11));
num_of_obs(1)       = length(Xo_1);
quadratic_term(1)   = Xo_1'*B11_inv*Xo_1;
% assign values to the negative log-likelihood function
log_like_vec(1)     = num_of_obs(1)*log(2*pi*sigma2) + ...
                      det(log(B11)) + quadratic_term(1);
%% For the rest of the columns:
for j=2:M
% find indices of available sites:
vj_avail_ind  = find(~isnan(X_miss(:,j)));
vj1_avail_ind = find(~isnan(X_miss(:,j-1)));
% get the corresponding data:
Xo_j          = X_miss(vj_avail_ind,j);
Xo_j1         = X_miss(vj1_avail_ind,j-1);
```

```
% covariance matrices:
B_jj    = B(vj_avail_ind,vj_avail_ind);
B_jj1   = B(vj_avail_ind,vj1_avail_ind);
B_j1j   = B_jj1';
% number of available observations for column j:
num_of_obs(j) = length(Xo_j);

% get distance between each sampling sites:
vj_avail      = v(vj_avail_ind);
vj1_avail      = v(vj1_avail_ind);
% standardize the vectors to be column vectors:
if size(vj_avail,2)~=1
    vj_avail = vj_avail';
    else
end
if size(vj1_avail,2)~=1
    vj1_avail = vj1_avail';
    else
end
% components for the covariance function:
zeta_jj = [0;abs(vj_avail(2:end)-vj_avail(1:end-1))];
zeta_j1j1 = [0;abs(vj1_avail(2:end)-vj1_avail(1:end-1))];

% inverse of B_jj and B_j1j1:
B_jj_inv = OU_COVINV(exp(-mu.*zeta_jj));
B_j1j1_inv = OU_COVINV(exp(-mu.*zeta_j1j1));

% LU decomposition of B_j1j1_inv
```

```
L_j = OU_SQRTM(B_jj_inv);
L_j1= OU_SQRTM(B_j1j1_inv);
% Define the distance between each column
eta_j  =abs(u(j)-u(j-1));
% Define conditional mean:
mo_j   = exp(-lambda*eta_j)*B_jj1*B_j1j1_inv*Xo_j1;
% Define conditional covariance matrix So_j:
So_j = B_jj - exp(-2*lambda*eta_j)*B_jj1*B_j1j1_inv*B_j1j;
% Define inverse of So_j
So_j_inv =inv(So_j);
% Define the terms used in the power series expansion:
T_j = L_j'*B_jj1*L_j1;
Tstar_j=exp(-2*lambda*eta_j)*T_j*(T_j');
% initialize the power sum
Tstar_j_terms=...
    NaN(size(Tstar_j,1),size(Tstar_j,2),num_term_approx);
% The power sum depends on user input num_term_approx:
for k=1:num_term_approx
Tstar_j_terms(:,:,k)=Tstar_j^k;
end
Tstar_j_sum=sum(Tstar_j_terms,3);
% Identify matrix used for the power series expansion
I_j=eye(num_of_obs(j),num_of_obs(j));
% So_j expressed in a different form:
Ao_j = ((L_j')\(I_j-Tstar_j))/(L_j);
% Approximate the inverse of the conditional variance:
Ao_j_inv = L_j*(I_j+Tstar_j_sum)*(L_j');
% Approximated quadratic form:
quadratic_term(j)=(Xo_j-mo_j)'*So_j_inv*(Xo_j-mo_j);
```

```
% log likelihood for the jthe column:
log_like_vec(j)=...
    num_of_obs(j)*log(2*pi*sigma2)+log(det(So_j))+...
    (1/sigma2)*quadratic_term(j);
% keep track of error between true inverse and power series approx.
if calc_error == 1
% Inverse of So_j:
So_j_inv =inv(So_j);
approx_err_temp=approx_err_temp+norm(So_j_inv-Ao_j_inv);
else
end
end
approx_err=approx_err_temp/M;
approx_log_like=sum(log_like_vec);
```

CURRICULUM VITAE

SAMI CHEONG

EDUCATION

2016 Ph.D., Mathematics, University of Wisconsin - Milwaukee.
Dissertation : Parameter Estimation for the Spatial Ornstein-Uhlenbeck process with Missing Observations.
Advisor: Professor Jugal Ghorai.

2012 M.S., Mathematics, University of Wisconsin - Milwaukee.
Thesis: A Study of the Mathematical Modeling of Translation Initiaion in Protein Synthesis
Co-advisors: Professor Jugal Ghorai and Professor Gabriella Pinter

2009 B.S., Mathematics with Music minor, University of Wisconsin-Whitewater.

WORK EXPERIENCE

2015 Data Scientist Co-Op, Monsanto Company, Saint Louis, MO.

2014 Summer Research Assistant, University of Wisconsin-Milwaukee, Milwaukee, WI.

2012-2014 Research Intern, Humand and Molecular Genetics Center, Medical College of Wisconsin, Milwaukee, WI.

2011-2014 Graduate Mentor, Undergradate BioMath (UBM) program, University of Wisconsin-Milwaukee, WI.

2009-2015 Graduate Teaching Assistant, University of Wisconsin-Milwaukee, WI.

PAPERS and PRESENTATIONS

2016 Joint Mathematics Meetings, Seattle, WA,
Talk: "Estimating parameters for the spatial Ornstein-Uhlenbeck process with missing observations".

2014 SIAM Annual Meeting, Chicago,IL
Poster Presentation: "Gene Selection from Microarray Data : An Exploratory Approach".

2014 Chen, Y., Cabrera, S., Jia, S. Kaldunski, M., Kramer, J. **Cheong, S.**, Geoffrey, R., Roethle, M., Woodliff, J., Greenbaum, C., Wang, X. and Hessner, M., *Molecular signatures differentiate immune states in Type 1 Diabetes families*, Diabetes. 2014 Apr 23.

2013 Bradley, W.T., Chatterjee, S., **Cheong, S.**, Huang, S., Lois, B., Poddar, A., *Network Analytics and Visualization in Healthcare*, Technical Report, CRSC-TR13-09.

2012 MAA Mathfest,Madison,WI,
Talk:"Parameter Estimation: the Basics!"

2012 MAA WI Chapter Meeting,Milwaukee School of Engineering,
Talk:"A Study of the Mathematical Model of Protein Synthesis Initiation".

2011 MAA WI Chapter MeetingUniversity of Wisconsin-Stout,
Talk: "A Study of the Mathematical Modeling of Protein Synthesis".

2009 National Conference on Undergraduate Research,University of Wisconsin-La Crosse,
Talk:"Appearance of Surface Color during Simulated Twilight".

WORKSHOPS ATTENDED

2015 Advanced Python Workshop, GIS day, University of Wisconsin-Milwaukee, Milwaukee, WI.

2014 SAS Visual Analytics Workshop, University of Wisconsin-Milwaukee, Milwaukee, WI.

2013 Mathematical and Statistical Modeling Workshop for Graduate Students, Statistical and Applied Mathematical Institute (SAMSI)/ North Carolina State University, Rayleigh, NC.

2012 Joint MBI-NIMBioS-CAMBAM Summer Graduate Workshop on Stochastics Applied to Biological Systems, The Ohio State University, Columbus, OH.

TECHNICAL SKILLS

- Knowledge in MATLAB, R, Python, HTML/CSS
- Passed Actuarial Exam P (2009)