

COMPUTATIONAL STATISTICS: TIME SERIES AND DATA MINING
(Spine title: Plib)
(Thesis format: Monograph)

by

Jiaqi Bi

Graduate Program in Epidemiology and Biostatistics

A thesis submitted in partial fulfillment
of the requirements for the degree of
Masters of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

© Jiaqi Bi 2024

THE UNIVERSITY OF WESTERN ONTARIO
School of Graduate and Postdoctoral Studies

CERTIFICATE OF EXAMINATION

Supervisor:

.....
Dr. Yun-Hee Choi

Joint Supervisor:

.....
Dr. Osvaldo Espin-Garcia

Supervisory Committee:

.....
Dr.

.....
Dr.

Examiners:

.....
Dr.

.....
Dr.

.....
Dr.

The thesis by

Jiaqi Bi

entitled:

Computational Statistics: Time Series and Data Mining

is accepted in partial fulfillment of the
requirements for the degree of
Masters of Science

.....
Date

.....
Chair of the Thesis Examination Board

Abstract

This is a really silly abstract.

Keywords: Time series analysis, data mining

Contents

Certificate of Examination	ii
Abstract	iii
List of Figures	v
List of Tables	vi
List of Appendices	vii
1 Introduction	1
1.1 Background	1
1.2 Motivation	1
2 Theorems	2
2.1 Basic Theorems	2
Bibliography	3
A Proofs of Theorems	4
Curriculum Vitae	5

List of Figures

List of Tables

List of Appendices

Appendix A Proofs of Theorems	4
---	---

Chapter 1

Introduction

1.1 Background

The Breast Cancer type 1/2, usually referred as BRCA1/2, are proteins that consists of genes that code for BRCA1 in humans. BRCA1/2 are human tumor suppressor genes, that are responsible for repairing the DNA [1]. When the mutation exists on these genes may cause the impairments of proper functions, which can lead to the possibility of capturing the breast, ovarian, or other specific cancers [2, 3, 4]. Inheriting one of these mutations does not guarantee developing cancer disease, but the mutation can increase the risk of getting those cancers [5]. In medicine, these cancer types are classified as Hereditary Breast and Ovarian Cancer Syndrome (HBOC) [6]. There are many medical and statistical development in modelling the competing risk of breast cancer and ovarian cancer, such as applying parametric survival analysis based on two cancer outcomes developed by Choi et al. [7], and some clinical trials via risk-reducing approach of treatments on breast cancer patients [8]. Importantly, since the study is based on the specific disease, there may be selection bias from the sampling process. To minimize the sampling bias, an ascertainment correction should be applied to the likelihood calculation. Thus, with different choices of baseline hazard function, there were several frailty distributions being introduced such as Gamma frailty, and log-normal frailty to incorporate the family structure in the survival analysis.

1.2 Motivation

One important factor of making the correct inference and prediction is to address the missing data problem within the dataset. In the observations of BRCA1/2 joining the Polygenic Risk Score (PRS), there are many individuals contain the missing data in the PRS.

Chapter 2

Theorems

2.1 Basic Theorems

Theorem 2.1.1 $e^{i\pi} = -1$

Bibliography

- [1] JA Duncan, JR Reeves, and TG Cooke. BRCA1 and BRCA2 proteins: roles in health and disease. *Molecular pathology*, 51(5):237, 1998.
- [2] Julia B Greer and David C Whitcomb. Role of BRCA1/2 mutations in pancreatic cancer. *Gut*, 2006.
- [3] Bruce G Haffty, Elizabeth Harrold, Atif J Khan, Pradip Pathare, Tanya E Smith, Bruce C Turner, Peter M Glazer, Barbara Ward, Daryl Carter, Ellen Matloff, et al. Outcome of conservatively managed early-onset breast cancer by BRCA1/2 status. *The Lancet*, 359(9316):1471–1477, 2002.
- [4] Yong-Wen Huang. Association of BRCA1/2 mutations with ovarian cancer prognosis: an updated meta-analysis. *Medicine*, 97(2), 2018.
- [5] Bernard Friedenson. The BRCA1/2 pathway prevents hematologic cancers in addition to breast and ovarian cancers. *BMC cancer*, 7(1):1–11, 2007.
- [6] Michael P Lux, Peter A Fasching, and Matthias W Beckmann. Hereditary breast and ovarian cancer: review and future perspectives. *Journal of molecular medicine*, 84:16–28, 2006.
- [7] Yun-Hee Choi, Hae Jung, Sandra Buys, Mary Daly, Esther M John, John Hopper, Irene Andrulis, Mary Beth Terry, and Laurent Briollais. A competing risks model with binary time varying covariates for estimation of breast cancer risks in brca1 families. *Statistical Methods in Medical Research*, 30(9):2165–2183, 2021.
- [8] Yun-Hee Choi, Mary Beth Terry, Mary B Daly, Robert J MacInnis, John L Hopper, Sarah Colonna, Sandra S Buys, Irene L Andrulis, Esther M John, Allison W Kurian, et al. Association of risk-reducing salpingo-oophorectomy with breast cancer risk in women with brca1 and brca2 pathogenic variants. *JAMA oncology*, 7(4):585–592, 2021.

Appendix A

Proofs of Theorems

Proof of Theorem 2.1.1

$$e^{i\pi} = \cos(\pi) + i \sin(\pi) \tag{A.1}$$

$$= -1 \tag{A.2}$$

■

Curriculum Vitae

Name: Jiaqi Bi

**Post-Secondary
Education and
Degrees:** La La School
La La Land
1996 - 2000 M.A.

University of Western Ontario
London, ON
2008 - 2012 Ph.D.

**Honours and
Awards:** NSERC PGS M
2006-2007

**Related Work
Experience:** Teaching Assistant
The University of Western Ontario
2008 - 2012

Publications:

La La