Politecnico di Torino

ScuDo
Scuola di Dottorato ~ Doctoral School
WHAT YOU ARE, TAKES YOU FAR

Doctoral Dissertation
Doctoral Program in Pure & Applied Mathematics (36^{th}cycle)

Statistical models for understanding biomedical data

By

Hiu Ching Yip

******

Supervisor(s):
Prof. Gianluca Mastrantonio, Supervisor
Prof. Enrico Bibbona, Co-Supervisor

Doctoral Examination Committee:

Politecnico di Torino
Tuesday 8th October, 2024

Declaration

I hereby declare that, the contents and organization of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

Hiu Ching Yip
Tuesday 8th October, 2024

* This dissertation is presented in partial fulfillment of the requirements for **Ph.D. degree** in the Graduate School of Politecnico di Torino (ScuDo).

Statistical models for understanding biomedical data

Hiu Ching Yip

This work is a collection of statistical models that can be applied in biomedical research. It focuses on the development of statistical methods in the following research areas: cross-species comparison in bio-acoustic analysis, meta-analysis of randomized clinical trials and preferential attachment of an evolving network.

A statistical model is presented for each of the corresponding areas. In particular, these models and their implementation integrated novel techniques and methodologies in order to tackle challenges that are induced by the characteristics of real-world data sets. The first is a spatio-temporal model for bio-acoustic data that consists of periodic artifacts and is temporally non-stationary. The second is a modified proportional hazards model that can account for the impacts of the follow-up time of the trials in mixed treatment comparison. The third is a correction of a specific preferential attachment model that exhibit the power-law degree sequence which is empirically observed in real-world network data.

The models and the methodologies involved in the implementation have solid theoretical basis. More importantly, these models are useful quantitative tools for biomedical researchers to analyze and learn from data sets that possess the problematic characteristics encountered in this work. Extensive simulation studies and testings on real data sets were conducted.