

## THÔNG TIN VỀ LUẬN ÁN TIẾN SĨ

1. Họ và tên nghiên cứu sinh: Đào Thanh Tùng
2. Giới tính: Nam
3. Ngày sinh: 05/11/1975
4. Nơi sinh: Hà Tĩnh
5. Quyết định công nhận nghiên cứu sinh: Số 4438/QĐ-ĐHKHTN ngày 26/11/2015 của Hiệu trưởng Trường Đại học Khoa học Tự nhiên - ĐHQGHN.
6. Các thay đổi trong quá trình đào tạo: Quyết định gia hạn số 2402/QĐ-ĐHKHTN ngày 02/8/2019 của Hiệu trưởng Trường Đại học Khoa học Tự nhiên - ĐHQGHN.
7. Tên đề tài luận án: **Lựa chọn biến, số thành phần và ước lượng tham số bằng phương pháp VB cho các mô hình GLMM và MRDE-MN.**
8. Chuyên ngành: Lý thuyết xác suất và thống kê toán học
9. Mã số: 9460112.02
10. Cán bộ hướng dẫn khoa học:
  - Hướng dẫn chính: PGS.TS. Trần Minh Ngọc
  - Hướng dẫn phụ: TS. Trần Mạnh Cường
11. Tóm tắt các kết quả mới của luận án:
  - Luận án đề xuất thuật toán mới, nhanh nhằm thực hiện đồng thời việc lựa chọn biến và ước lượng tham số cho mô hình GLMM trong trường hợp nhiều biến.
  - Luận án là công trình đầu tiên nghiên cứu mô hình MRDE-MN và đề xuất thuật toán thực hiện đồng thời ba mục tiêu là lựa chọn biến, xác định số thành phần và ước lượng tham số cho mô hình MRDE-MN.Các thuật toán đề xuất trong luận án đã được đánh giá là có hiệu suất tốt thông qua thử nghiệm trên dữ liệu mô phỏng và dữ liệu thật.
12. Khả năng ứng dụng thực tiễn:
  - Luận án đã xây dựng được hai thuật toán nhanh, có hiệu suất tốt, có thể áp dụng trong thực tiễn trên dữ liệu thật với mô hình GLMM và mô hình MRDE-MN.

- Các mô hình GLMM và MRDE-MN trong trường hợp nhiều biến đều rất khó giải quyết nhưng hai thuật toán đã xây dựng được trong luận án đều có khả năng thực hiện tốt, hiệu suất cao.

13. Các hướng nghiên cứu tiếp theo:

- Lựa chọn nhóm biến trong GLMM bằng cách sử dụng Lasso phạt nhóm.
- Lựa chọn biến được sắp xếp trong GLMM bằng cách sử dụng phạt tuyệt đối tổng hợp.
- Mở rộng phương pháp cho lựa chọn biến và lựa chọn số thành phần trong các mô hình trộn với mỗi thành phần là một mạng nơron.

14. Các công trình công bố liên quan đến luận án:

- [1] Dao Thanh Tung, Minh-Ngoc Tran, Tran Manh Cuong, Bayesian Adaptive Lasso with Variational Bayes for Variable Selection in High-dimensional Generalized Linear Mixed Models, *Communications in Statistics – Simulations and Computations*, **0**, 1 – 14, 2018.  
DOI: 10.1080/03610918.2017.1387663.
- [2] Dao Thanh Tung, Minh-Ngoc Tran, Flexible Multivariate Regression Density Estimation, *Communications in Statistics – Theory and Methods*, 2020, DOI: 10.1080/03610926.2020.1723633.

Ngày 20 tháng 8 năm 2020

**Người hướng dẫn luận án**

**Nghiên cứu sinh**

**TS. Trần Mạnh Cường**

**Đào Thanh Tùng**

## INFORMATION ON DOCTORAL THESIS

1. Full name: Dao Thanh Tung
2. Sex: Male
3. Date of birth: November 5<sup>th</sup>, 1975
4. Place of birth: Ha Tinh
5. Admission decision number: No 4438/QĐ-ĐHKHTN, dated on November 26<sup>th</sup>, 2015 by Rector of VNU University of Science.
6. Changes in academic process:  
Extension decision number 2402/QĐ-ĐHKHTN, dated on August 02<sup>nd</sup>, 2019 by Rector of VNU University of Science.
7. Official thesis title: **Selection of variables, component numbers and parameter estimates by VB method for GLMM and MRDE-MN models.**
8. Major: Theory of Probability and Mathematical Statistics.
9. Code: 9460112.02
10. Supervisors:  
Assoc.Prof.Dr. Tran Minh Ngoc  
Dr. Tran Manh Cuong
11. Summary of the new findings of the thesis
  - The thesis proposes a new, fast algorithm to perform simultaneous selection of variables and parameter estimation for GLMM model in case of multi-variables.
  - The thesis is the first work to study the MRDE-MN model and propose the algorithm to perform three objectives simultaneously including variable selection, componential number determination and parameter estimation for MRDE-MN model.The algorithms proposed in the thesis have been evaluated as having good performance through testing on simulation data and real data.
12. Practical applicability, if any:

- The thesis has built two fast algorithms with good performance, which can be applied in practice on real data with the GLMM model and the MRDE-MN model.
- GLMM and MRDE-MN models in High-dimensional are difficult to solve, but the two algorithms built in the thesis have good performance and high efficiency.

13. Further research directions, if any

- Extend the method for grouped variable selection in GLMMs by using the group lasso penalty.
- Extend the method for ordered variable selection in GLMMs by the composite absolute penalty.
- Extend the method for variable selection and component selection in the mixture of experts model where each component is a neural network.

14. Thesis-related publications:

- [1] Dao Thanh Tung, Minh-Ngoc Tran, Tran Manh Cuong, Bayesian Adaptive Lasso with Variational Bayes for Variable Selection in High-dimensional Generalized Linear Mixed Models, *Communications in Statistics – Simulations and Computations*, **0**, 1 – 14, 2018.  
DOI: 10.1080/03610918.2017.1387663.
- [2] Dao Thanh Tung, Minh-Ngoc Tran, Flexible Multivariate Regression Density Estimation, *Communications in Statistics – Theory and Methods*, 2020, DOI: 10.1080/03610926.2020.1723633.

*Date: August 20<sup>th</sup>, 2020*

**Supervisor**

**PhD Student**

**Dr. Tran Manh Cuong**

**Dao Thanh Tung**