JOHNNY MYUNGWON LEE

Edinburgh, United Kingdom · johnny.myungwon.lee@ed.ac.uk · in/johnnymwlee/

Education

Doctor of Philosophy in Statistics *Edinburgh, United Kingdom*
University of Edinburgh, School of Mathematics *Sep 2022 - Present*

- **Supervisor:** Dr. Daniel Paulin, Dr. Miguel de Carvalho
- **Research Interest:** Bayesian Statistics, Machine Learning, Computational Statistics, Meta Learning

Bachelor of Science (Honours) in Mathematics and Statistics *Edinburgh, United Kingdom*
University of Edinburgh, School of Mathematics *Sep 2016 - May 2022*

- 1st Class Honours Degree
- **Thesis:** *"Semismooth Newton Augmented Lagrangian Method for solving Lasso Problems with Implementation in R"* under the supervision of Dr. Daniel Paulin
- **Core Courses:** Bayesian Theory, Bayesian Data Analysis, Incomplete Data Analysis, Multivariate Data Analysis, Biomedical Data Science, Fundamental of Optimisation, Introductory Applied Machine Learning, Generalised Regression Model, Statistical Research Skills, Differential Equation, Python Programming

Bachelor of Science in Statistics *Seoul, Republic of Korea*
Seoul National University, Department of Statistics *Sep 2018 - Jun 2019*

- 3rd Year Exchange Student / Study Abroad Programme
- **Core Courses:** Time Series Analysis, Mathematical Statistics, Computational Statistics, Data Mining Methods and Lab, Understanding Big Data, Computer Application for Scientific Computation

Publications

Published

Lee, M.W., de Carvalho, M., Paulin, D., Pereira, S., Trigo, R., & Da Camara, C. 2023. Bayesian Regularisation for Tail Index Regression. 37th IWSM Proceedings, 37(1): 493---496.

In Review

Lee, M.W., de Carvalho, M., Paulin, D., Pereira, S., Trigo, R., & Da Camara, C. 2023. Bayesian Regularisation with Smoothing for Tail Index Regression. 37th IWSM Proceedings, 37(1): 493---496.

In Prep

Lee, M.W., Renfrew, M., Paulin, D. 2022. Semismooth Newton Augmented Lagrangian Method for Solving Lasso Problems with Implementation in R.

Presentations

Contributed Presentations

Lee, M.W.. (*Upcoming*). 2024 International Society for Bayesian Analysis World Meeting,Venice, Italy. - "Bayesian Regularisation with Smoothing for Tail Index Regression"

Lee, M.W.. 2024. PG Colloqium, School of Mathematics, University of Edinburgh, Edinburgh, United Kingdom. - "Bayesian Regularisation with Smoothing for Tail Index Regression"

Lee, M.W.. 2024. Student Conference, Maxwell Institute for Mathematical Sciences, Edinburgh, United Kingdom. - "Bayesian Regularisation with Smoothing for Tail Index Regression"

Lee, M.W.. 2023. Bayesian Young Statistician Meeting, junior-International Society for Bayesian Analysis, Online - "Bayesian Regularisation with Smoothing for Tail Index Regression"

Lee, M.W.. 2023. 37th International Workshop for Statistical Modelling, Dortmund, Germany - "Bayesian Regularisation for Tail Index Regression"

Lee, M.W.. 2021. Department of Industrial Engineering, Seoul National University, Seoul, Republic of Korea - "Natural Language Processing on Solid-state Battery Cell Patent Documents."

Teaching Experience

2023/2024	**Bayesian Data Analysis (MATH11175),** Tutor
	Bayesian Theory (MATH11177), Tutor
	Statistical Research Skills (MATH11188), Tutor
	Incomplete Data Analysis (MATH11185), Tutor
	Generalised Regression Model (MATH11187), Tutor
	Introduction to Data Science (MATH08077), Tutor
2022/2023	**Biomedical Data Science (MATH11174),** Super Tutor (Course Instructor) / Examination Setter

Experiences

LG Chemical Patent Analysis & Samsung Securities News Curation Industry Research *Seoul, Republic of Korea*
Data Mining Laboratory Intern, Seoul National University *Jun 2021 - Aug 2021*

- Supervised by Prof. Sungzoon Cho, Dr. Hoonsik Shin and Dr. Hyejin Lee.
  Analysed 1,198 solid-state battery cell patent documents to classify sentences into either problems or solutions by building unsupervised learning models such as **BERT, char-CNN and RNN** with NLTK, Scikit-learn and TensorFlow
- **Increased 20% of model accuracy (F1 score)** by performing text pre-processing / post-processing using **RE**
- **Decreased parsing error rate from 35% to 3%** by suggesting alternative parsing method for SEC Edgar 10-K reports using BeautifulSoup, RE, Gensim in Samsung Securities News Curation Project team

Financial Value Insight Data Research: ETFs & Glassdoor Research *Seoul, Republic of Korea*
Data Mining Laboratory Intern, Seoul National University *Jun 2019 - Aug 2019*

- Supervised by Prof. Sungzoon Cho, Dr. Hyunyong Kim and Dr. Elaine Pak.
  Integrated large-scale web-crawling on *morningstar.com, finance.naver.com, comp.fnguide.com* | *glassdoor.com & jobplanet.com* using **Python** packages such as Selenium Webdriver, BeautifulSoup, RE and Pandas
- **Identified correlation between company's annual profit and anonymous company's reviews** through natural language processing such as doc2vec and NTF-IDF
- Nominated by team leader as a presenter in the weekly lab seminar on behalf of the Glassdoor research team.
- **Gathered and recalculated the missing financial fundamentals** for 435 Exchange Traded Fund from Korea Exchange. Involved in building database schema to store collected data using SQL and Django Framework

DataAssembly *Edinburgh, United Kingdom*
Co-founder & Consultant, Buchanan Institute, University of Edinburgh *Feb 2018 - Aug 2019*

- A student-led data consultancy **DataAssembly**, providing pro-bono data anlaytics to small local businesses and third sector industries through student's engagement in the real-world problem using **R** and **Python**
- Represented **DataAssembly** at DataFest Event 2018 organised by the Scottish Government and the Data Lab
- Acknowledged, fulfilled and identified the needs of team members and clients, by approaching clients from door to door and guiding students from non-statistical backgrounds with structured curriculum and opportunities.

Projects

Semismooth Newton Augmented Lagrangian Method for Solving Lasso Problems with Implementation in R *Edinburgh, United Kingdom*
Supervised by Dr. Daniel Paulin *Sep 2021 - Apr 2022*

- With the advantage of the second order sparsity of the problem, we aimed to reduce the expensive computational cost and result in a fast and highly efficient algorithm to solve Lasso regression
- Implemented the algorithm in **R** with **C++** on UCI/Statlib and Methylation Profiling data with cross-validation and showed having lower Mean Squared Error than glmnet
- R package available on Github, `https://www.github.com/johnnymdoubleu/lassoSSNAL`

News Letter Mall Project *Seoul, Republic of Korea*
Supervised by Dr. Hyejin Lee *Mar 2021 - Aug 2021*

- Identified the need of rapid supply of accurate news in the Korean stock market to mitigate investment risk
- With Kiwoom Securities trading's **REST API**, Naver's **OpenAPI**, and telegram **API**, developed a real-time bot sending out newsletters. Developed an auto trading algorithm based on the information crawled.
- Utilised **Mecab** tokeniser from **KoNLPy** package to pre-process and censor unwanted AI generated news and advertisements.

Pattern Analysis of New York Airplanes Data Research *Seoul, Republic of Korea*
Sep 2018 - Dec 2018

- Analysed 120 million flights' arrival and departure data of commercial flights within the USA using **R** & **SQL**
- Provided graphical and visual summary of the best **time of day / day of week / time of year** to fly to optimise delay hours with supervised learning such as linear regression, logistic regression, ridge regression, lasso regression, linear discriminant analysis, quadratic discriminant analysis, decision trees, random forest and boosting.

Skills

Machine Learning:Python / R / MATLAB | **Front End**:HTML / CSS / JS | LaTeX
Microsoft Office Specialist–Masters (*Jun 2017*) — Word Expert | Excel Expert | PowerPoint Core | Access Core

Military Service *Daegwallyeong, Republic of Korea*
Honourably Discharged from *36th Infantry Division, Department of Intelligence* *Aug 2019 - Mar 2021*