TrungTin Nguyen

Curriculum Vitae

Inria Grenoble Rhone-Alpes
Member of STATIFY Team
Inovallée
655 Av. de l'Europe
38334 Montbonnot-Saint-Martin, France
✉ *trung-tin.nguyen@inria.fr*
🖫 **Homepage**: *trung-tinnguyen.github.io*
January 31, 2022

"Essentially, all models are wrong, but some are useful" - George E.P. Box (1987).

Academic Appointments

2022–2023 **Postdoctoral Fellow**, *Inria Grenoble-Rhône-Alpes*, France.
Mentors: Senior Researcher Florence Forbes, Senior Lecturer Hien Duy Nguyen, and Associate Researcher Julyan Arbel.
Topic: Simulation-based inference for Bayesian inverse problems.

Tertiary Education

2018–2021 **Ph.D. in Statistics and Data Science**, *Université de Caen Normandie*, Caen, France.
Advisor: Professor Faïcel Chamroukhi.
Thesis title: Model selection and approximation in high-dimensional mixtures of experts models: from theory to practice.
Manuscript: pdf.

2017–2018 **Master of Science, Technology and Health**, *Université d'Orléans*, Orléans, France, GPA: 18/20.
Master 2 Program of Mathematics. Major in Applied Mathematics. Mention *"Très Bien"*.
Advisor: Professor Le Thi Hoai An.
Thesis title: Reinforcement learning for resource allocation problems using a partially observable Markov decision process.

2013–2017 **Bachelor of Science**, *Vietnam National University-Ho Chi Minh Univeristy of Science*, Ho Chi Minh City, Vietnam, GPA: 9.17/10.
Honors Program in Mathematics and Computer Science. Major in Probability and Statistics, minor in Numerical Analysis. Rank: 2/1557: Summa Cum Laude.
Advisor: Professor Dang Duc Trong.
Thesis title: Multiplicative censoring model.

Research Interests

Statistical learning
Supervised, unsupervised and visualization of *high-dimensional data*, *model selection in clustering and regression for functional and heterogeneous data*, *statistical convergence for deep hierarchical mixtures of experts (MoE)*, *approximate Bayesian computation*.

Machine learning
Deep generative models (*variational autoencoders*, generative adversarial networks), *reinforcement learning*, optimal transport (*Wasserstein distance*).

Optimization
Robust and effective optimization algorithms for deep neural network (stochastic gradient descent, Adam,. . .), deep hierarchical MoE (*generalized expectation–maximization (EM) algorithm*, EM algorithm, MM algorithm,. . .), DC algorithm.

Biostatistics
Statistical learning and machine learning for large biological data sets (omics data), e.g., *genomics*, *transcriptomics* and proteinomics.

Publications

Journal Publications	3 papers
Journal Submissions	4 papers
Conference	1 paper

Ph.D. Thesis

2021   TrungTin Nguyen. *Model Selection and Approximation in High-dimensional Mixtures of Experts Models: From Theory to Practice*. PhD Thesis, Normandie Université, December 2021.

Model selection in mixture of experts models

2021   TrungTin Nguyen, Hien Duy Nguyen, Faicel Chamroukhi, and Florence Forbes. A non-asymptotic penalization criterion for model selection in mixture of experts models. *arXiv preprint arXiv:2104.02640. Under revision, Electronic Journal of Statistics*, 2021.

2021   TrungTin Nguyen, Faicel Chamroukhi, Hien Duy Nguyen, and Florence Forbes. Non-asymptotic model selection in block-diagonal mixture of polynomial experts models. *arXiv preprint arXiv:2104.08959. Under revision, Journal of Multivariate Analysis*, 2021.

2020   TrungTin Nguyen, Hien D Nguyen, Faicel Chamroukhi, and Geoffrey J McLachlan. An l_1-oracle inequality for the Lasso in mixture-of-experts regression models. *arXiv preprint arXiv:2009.10622. Under revision, ESAIM: Probability and Statistics*, 2020.

Approximate Bayesian computation

2021   Florence Forbes, Hien Duy Nguyen, TrungTin Nguyen, and Julyan Arbel. Supporting Information Approximate Bayesian computation with surrogate posteriors. *Preprint. Under revision, Statistics and Computing*, August 2021.

2021   Florence Forbes, Hien Duy Nguyen, TrungTin Nguyen, and Julyan Arbel. Approximate Bayesian computation with surrogate posteriors. *Preprint. Under revision, Statistics and Computing*, February 2021.

2021   Julyan Arbel, Florence Forbes, Hien Duy Nguyen, and TrungTin Nguyen. Approximate Bayesian computation with surrogate posteriors. In *ISBA 2021 - World Meeting of the International Society for Bayesian Analysis*, Marseille, France, June 2021.

Approximation capabilities of the mixture of experts models

2021   TrungTin Nguyen, Faicel Chamroukhi, Hien D Nguyen, and Geoffrey J McLachlan. Approximation of probability density functions via location-scale finite mixtures in Lebesgue spaces. *arXiv preprint arXiv:2008.09787. To appear. Communications in Statistics - Theory and Methods*, 2021.

2021   Hien Duy Nguyen, TrungTin Nguyen, Faicel Chamroukhi, and Geoffrey John McLachlan. Approximations of conditional probability density functions in Lebesgue spaces via mixture of experts models. *Journal of Statistical Distributions and Applications*, volume 8, page 13, 2021.

2020   TrungTin Nguyen, Hien D Nguyen, Faicel Chamroukhi, and Geoffrey J McLachlan. Approximation by finite mixtures of continuous density functions that vanish at infinity. *Cogent Mathematics & Statistics*, volume 7, page 1750861. Cogent OA, 2020.

Conference, Seminar, Workshop Presentations

10/2021   Model Selection and Approximation in High-dimensional Mixtures of Experts Models: From Theory to Practice at Jed 2021: Journée scientifique de l'École Doctorale 2021, Le Havre, France (Oral presentation: pdf).

09/2021   Approximation and non-asymptotic model selection in mixture of experts models at Journée Thématique: "Intelligence Artificielle - Applications et défis mathématiques", INSA Rouen Normandie, Rouen, France (Poster session: pdf).

06/2021   Non-asymptotic model selection in mixture of polynomial experts models at MHC2021 Mixtures Hidden Markov model Clustering, Institut de Mathématique d'Orsay, Paris, France (Poster session: pdf).

04/2021   Non-asymptotic model selection for the Gaussian-gated localized mixture of experts regression models at MiMo 2021: Workshop on Mixture Models, Laboratoire de Mathématiques Raphaël Salem, Université de Rouen Normandie, France (Invited speaker: pdf).

Projects

2019–2021   **Member of the LANDER (Latent Analysis, Adversarial Networks, and DimEnsionality Reduction)**.
- Principal investigator:
  - Florence Forbes (Mistis Inria Grenoble Rhone-Alpes, France),
  - Hien Duy Nguyen (School of Mathematics and Physics, University of Queensland, Australia).
- Other participants:
  - Queensland University of Technology, Brisbane, Australia,
  - University of Queensland, Brisbane, Australia,
  - Swinburne University of Technology, Melbourne, Australia
  - Université de Caen Normandie, France.
- Website: `https://team.inria.fr/statify/projects/lander/`
- My submitted publications in this project: *Model selection criteria with non-asymptotic approaches for Gaussian-gated Localized Mixture of Experts* and *Non-asymptotic model selection in Block-diagonal Localized Mixture of Experts*.

Professional Memberships

08/2021- 12/2023   Institute of Mathematical Statistics (IMS).

01/2020- 12/2021   International Society for Bayesian Analysis (ISBA).

Professional Services

Journal Reviewing

Computational Statistics and Data Analysis (Elsevier): 2 papers. See certificate.

Biometrical Journal (Wiley): 2 papers. See certificate.

Neurocomputing (Elsevier): 1 papers.

Communications in Statistics - Theory and Methods (Taylor Francis): 2 papers. See certificate.

Conference Reviewing/Program Committee

Proceedings of the Research School on Statistics and Data Science (RSSDS 2019) (Springer): 2 papers

Fellowships and Awards

2018–2021   Ph.D. scholarship in Statistics and Data Science granted by Ministère de l'Enseignement Supérieur et de la Recherche, France.

2017   Highest Distinction Graduation Award, Vietnam National University–Ho Chi Minh University of Science.

2014–2017   Scholarship of the National Program for the Development of Mathematics 2010–2020 of Vietnam Institute for Advanced Study in Mathematics (VIASM).

2013–2017   Outstanding Student Award, Deparament of Mathematics and Computer Science, Vietnam National University–Ho Chi Minh University of Science.

Research Experiences

Inria Grenoble-Rhône-Alpes Research Centre.

09/2020–01/2021   ***Visiting Doctoral Fellowship: 4 months***.
- Collaboration: Florence Forbes, Senior Researcher (Director of Research) at INRIA Grenoble Rhone-Alpes, head of team Statify.
- Abstract: Model selection criteria with non-asymptotic approaches for Gaussian-gated Localized Mixture of Experts

Lab of Mathematics Nicolas Oresme, Université de Caen Normandie, France.

2018–2021   ***Ph.D. Thesis: Model selection and approximation in high-dimensional mixtures of experts models: from theory to practice***.
- Advisor: Professor Faïcel Chamroukhi, Normandie Univ, UNICAEN, CNRS, LMNO, Caen, France.
- Abstract: This Ph.D. thesis aims to study the approximation capabilities, model estimation and selection properties, of a rich family of mixtures of experts (MoE) models in a high-dimensional setting, including MoE with Gaussian experts and softmax (SGaME) or Gaussian gating functions (GLoME).
  - Firstly, we improve upon universal approximation results in the context of unconditional mixture distributions, and study such capabilities for MoE models in a variety of contexts, including conditional probability density functions (PDF) approximation and approximate Bayesian computation. More precisely, we prove that to an arbitrary degree of accuracy, location-scale mixtures of a continuous PDF can approximate any continuous PDF, uniformly, on a compact set; location-scale mixtures of an essentially bounded PDF, respectively of conditional PDF, can approximate any PDF, respectively any continuous conditional PDF whenever the input and output variables are both compactly supported, in Lebesgue spaces.
  - Next, we establish non-asymptotic risk bounds that take the form of weak oracle inequalities, provided that lower bounds on the penalties hold true, in high-dimensional regression scenarios for a variety of MoE regression models, including GLoME and SGaME, based on an inverse regression strategy or a Lasso penalization, respectively. We show that the performance in Jensen–Kullback–Leibler type loss of our penalized maximum likelihood estimator is roughly comparable to that of oracle model, given large enough the constant in front of the penalty. This penalty is only known up to a multiplicative constant, proportional to the dimension of model and is calibrated by slope heuristic criterion.
  - Finally, to support our theoretical results and the statistical study of non-asymptotic model selection in a variety of MoE models, we perform numerical studies by considering simulated and real data, which highlight the performance of our finite-sample oracle inequality results.

Theoretical and Applied Computer Science Laboratory LITA, Université de Lorraine, France.

2017–2018   ***Master Thesis: Reinforcement learning for resource allocation problems using a partially observable Markov decision process (POMDP)***.
- Advisor: Professor Le Thi Hoai An, director of Theoretical and Applied Computer Science Laboratory LITA, Université de Lorraine, France.
- Abstract: I have generalized successfully Reinforcement Learning algorithms, which are created by my Advisor for Markov Decision Process, to the more general POMDP framework by using the DC (Difference of Convex functions) programming and DCA (DC Algorithms) for a realistic application, fleet management.
- Grade: 17/20.

Vietnam National University–Ho Chi Minh Univeristy of Science, Vietnam.

2016–2017   ***Bachelor Thesis: Multiplicative censoring model***.
- Advisor: Professor Dang Duc Trong, head of faculty Mathematics and Computer Science, Vietnam National University-Ho Chi Minh Univeristy of Science, Vietnam.
- Abstract: I studied classical kernels to estimate density f from the multiplicative censoring model, given by $Y_i = X_i U_i, 1 \leq i \leq n$, where $(X_i)_{1 \leq i \leq n}$ are i.i.d. with unknown density f in $\mathbf{R}$, $(U_i)_{1 \leq i \leq n}$ are i.i.d. with uniform distribution $\mathcal{U}([0,1])$. Only the sample $(Y_i)_{1 \leq i \leq n}$ is observed and $(U_i)_{1 \leq i \leq n}$ and $(X_i)_{1 \leq i \leq n}$ are independent sequences. Then simulation experiments are implemented by using software $\mathbf{R}$ with different bandwidth selection procedures.
- Grade: 10/10.

Teaching Experiences

Fall 2018: **Teaching Assistant, Mathematical and numerical foundations of modeling and simulation using partial differential equations instructed by Professor Jing-Rebecca Li (Ecole Polytechnique, France)**, *French-Vietnam Master 2 in Applied Mathematics*, *Vietnam National University–Ho Chi Minh Univeristy of Science*, Vietnam.

Fall 2017: **Teaching Assistant, Principles of Mathematical Analysis**, *Department of Mathematics and Computer Science, Vietnam National University–Ho Chi Minh Univeristy of Science*, Vietnam.

Selected Academic Experiences

07/2019 **Participated in 3rd International Summer School on Deep Learning (39 hours)**, *Warsaw, Poland*.
Including some featured courses:
- *Deep Generative Models* by Aaron Courville (University of Montréal, Canada).
- *Dive into Deep Learning* by Alex Smola (Amazon, USA).
- *Mathematics of Deep Learning* by Rene Vidal (Johns Hopkins University, USA).

12/2018 **Attended the course "Optimization theory for Statistics and Machine Learning"**, *Caen, France*, by Dr. Hien Duy Nguyen.

11/2018 **Participated to training program "Apprentissage statistique et science des données (18 hours)**, *Caen, France*, by Professor Faicel Chamroukhi.

06-09/2018 **Accomplished an online course Deep Learning Specialization**, *Stanford University, USA*, instructed by Professor Andrew Ng.
Including 5 courses:
- Neural Networks and Deep Learning. Grade: 100%.
- Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization. Grade: 100%.
- Structuring Machine Learning Projects. Grade: 98.3%.
- Convolutional Neural Networks. Grade: 98.9%.
- Sequence Models. Grade: 100%.
Course Certificates:
`https://www.coursera.org/account/accomplishments/specialization/4HZQ4ET4HCSQ`

Languages

Vietnamese	**Mother tongue**	
English	**IELTS 7.0/9.0**	*Excellent in reading, listening, good at writing and speaking.*
French	**Elementary: A2**	*Can understand isolated phrases and common expressions.*

Computer Skills

Programming Languages	Advanced R, Advanced Python, MATLAB, C++, SAS.
Operating Systems	Linux, macOS, Microsoft Windows.

Referees

Faïcel Chamroukhi
Professor of Statistics and Data Science
Université de Caen Normandie, France
✉ faicel.chamroukhi@unicaen.fr

Hien Duy Nguyen
Senior Lecturer
School of Mathematics and Physics
University of Queensland, Brisbane, Australia
✉ h.nguyen7@uq.edu.au

Florence Forbes
Senior Researcher (Director of Research)
INRIA Grenoble Rhone-Alpes, France
✉ florence.forbes@inria.fr

Associate Researcher Julyan Arbel
INRIA Grenoble Rhone-Alpes, France
✉ julyan.arbel@inria.fr

Geoffrey McLachlan
Professor of Statistics
University of Queensland, Brisbane, Australia
✉ g.mclachlan@uq.edu.au

Le Thi Hoai An
Full Professor of Exceptional Class
University of Lorraine, France
✉ hoai-an.le-thi@univ-lorraine.fr