

# Publications List

Hien Duy Nguyen

May 18, 2021

## Peer-Reviewed Publications

### Journal Articles

- [1] D V Fryer, I Strümke, and H Nguyen. Explaining the data or explaining a model? Shapley values that uncover non-linear dependencies. *PeerJ Computer Science*, to appear, 2021.
- [2] M Vladimirova, S Girard, H Nguyen, and J Arbel. Sub-Weibull distributions: generalizing sub-Gaussian and sub-Exponential properties to heavier-tailed distributions. *Stat*, 9:e318, 2020.
- [3] E Redivo, H Nguyen, and M Gupta. Bayesian clustering of skewed and multimodal data using geometric skew normal distributions. *Computational Statistics and Data Analysis*, 152:107044, 2020.
- [4] T T Nguyen, H D Nguyen, F Chamroukhi, and G J McLachlan. Approximation by finite mixtures of continuous density functions that vanish at infinity. *Cogent Mathematics and Statistics*, 7:1750861, 2020.
- [5] H D Nguyen, F Forbes, and G J McLachlan. Mini-batch learning of exponential family finite mixture models. *Statistics and Computing*, 30:731–748, 2020.
- [6] H D Nguyen, J Arbel, H Lü, and F Forbes. Approximate Bayesian computation via the energy statistic. *IEEE Access*, 8:131683–131698, 2020.

- [7] D V Fryer, I Strümke, and H D Nguyen. Shapley value confidence intervals for attributing variance explained. *Frontiers in Applied Mathematics and Statistics*, 6:587199, 2020.
- [8] J Bagnall, A Jones, N Karavarsamis, and H Nguyen. The fully-visible Boltzmann machine and the Senate of the 45th Australian Parliament in 2016. *Journal of Computational Social Science*, 3:55–81, 2020.
- [9] J Arbel, O Marchal, and H D Nguyen. On strict sub-Gaussianity, optimal proxy variance and symmetry for bounded random variables. *ESAIM: Probability and Statistics*, 24:39–55, 2020.
- [10] L Truong, H Nguyen, H Nguyen, and H Vu. Pedestrian overpass use and its relationship with digital and social distractions, and overpass characteristics. *Accident Analysis and Prevention*, 131:234–238, 2019.
- [11] H D Nguyen, Y Yee, G J McLachlan, and J P Lerch. False discovery rate control for grouped or discretely supported p-values with application to a neuroimaging study. *SORT*, 43:1–22, 2019.
- [12] H D Nguyen and G J McLachlan. On approximation via convolution-defined mixture models. *Communications in Statistics - Theory and Methods*, 48:3945–3955, 2019.
- [13] H D Nguyen, F Chamroukhi, and F Forbes. Approximation results regarding the multiple-output mixture of linear experts model. *Neurocomputing*, 366:208–214, 2019.
- [14] H D Nguyen. Asymptotic normality of the time-domain generalized least squares estimator for linear regression models. *Stat*, 8(e248), 2019.
- [15] A T Jones, J J Bagnall, and H D Nguyen. BoltzMM: an R package for maximum pseudo-likelihood estimation of fully-visible Boltzmann machines. *Journal of Open Source Software*, 4:1193, 2019. <https://doi.org/10.21105/joss.01193>.
- [16] D Fryer, H Nguyen, and P Orban. studentlife: tidy handling and navigation of a valuable mobile-health dataset. *Journal of Open Source Software*, 4(40), 2019.

- [17] F Chamroukhi and H D Nguyen. Model-based clustering and classification of functional data. *WIREs Data Mining and Knowledge Discovery*, e1298, 2019.
- [18] P Orban, C Dansereau, L Desbois, V Mongeau-Perusse, C-E Giguere, H Nguyen, A Mendrek, E Stip, and P Bellec. Multisite generalizability of schizophrenia diagnosis classification based on functional brain connectivity. *Schizophrenia Research*, 192:167–171, 2018.
- [19] H D Nguyen, D H Wang, and G J McLachlan. Randomized mixture models for probability density approximation and estimation. *Information Sciences*, 467:135–148, 2018.
- [20] H D Nguyen, G J McLachlan, J F P Ullmann, V Voleti, W Li, E M C Hillman, D C Reutens, and A L Janke. Whole-volume clustering of time series data from zebrafish brain calcium images via mixture modeling. *Statistical Analysis and Data Mining*, 11:5–16, 2018.
- [21] H D Nguyen and G J McLachlan. Some theoretical results regarding the polygonal distribution. *Communications in Statistics - Theory and Methods*, 47:5083–5095, 2018.
- [22] H D Nguyen and G J McLachlan. Chunked-and-averaged estimators for vector parameters. *Statistics and Probability Letters*, 137:336–342, 2018.
- [23] H D Nguyen, A T Jones, and G J McLachlan. Stream-suitable optimization algorithms for some soft-margin support vector machine variants. *Japanese Journal of Statistics and Data Science*, 1:81–108, 2018.
- [24] H D Nguyen and F Chamroukhi. An introduction to the practical and theoretical aspects of mixture-of-experts modeling. *WIREs Data Mining and Knowledge Discovery*, e1246, 2018.
- [25] H D Nguyen. Near universal consistency of the maximum pseudolikelihood estimator for discrete models. *Journal of the Korean Statistical Society*, 47:90–98, 2018.
- [26] L R Lloyd-Jones, H D Nguyen, and G J McLachlan. A globally convergent algorithm for lasso-penalized mixture of linear regression models. *Computational Statistics and Data Analysis*, 119:19–38, 2018.

- [27] A T Jones, H D Nguyen, and G J McLachlan. logKDE: log-transformed kernel density estimation. *Journal of Open Source Software*, 3:870, 2018.
- [28] C Oyarzun, A Sanjurjo, and H Nguyen. Response functions. *European Economic Review*, 98:1–31, 2017.
- [29] H D Nguyen, G J McLachlan, P Orban, P Bellec, and A L Janke. Maximum pseudolikelihood estimation for a model-based clustering of time series data. *Neural Computation*, 29:990–1020, 2017.
- [30] H D Nguyen and G J McLachlan. Progress on a conjecture regarding the triangular distribution. *Communications in Statistics - Theory and Methods*, 46:11261–11271, 2017.
- [31] H D Nguyen. An introduction to MM algorithms for machine learning and statistical estimation. *WIREs Data Mining and Knowledge Discovery*, 7(e1198), 2017.
- [32] H D Nguyen and I A Wood. Asymptotic normality of the maximum pseudolikelihood estimator for fully visible Boltzmann machines. *IEEE Transactions on Neural Networks and Learning Systems*, 27:897–902, 2016.
- [33] H D Nguyen and I A Wood. A block successive lower-bound maximization algorithm for the maximum pseudolikelihood estimation of fully visible Boltzmann machines. *Neural Computation*, 28:485–492, 2016.
- [34] H D Nguyen, G J McLachlan, and I A Wood. Mixtures of spatial spline regressions for clustering and classification. *Computational Statistics and Data Analysis*, 93:76–85, 2016.
- [35] H D Nguyen, G J McLachlan, J F P Ullmann, and A L Janke. Spatial clustering of time-series via mixture of autoregressions models and Markov Random Fields. *Statistica Neerlandica*, 70:414–439, 2016.
- [36] H D Nguyen, G J McLachlan, J F P Ullmann, and A L Janke. Laplace mixture autoregressive models. *Statistics and Probability Letters*, 110:18–24, 2016.

- [37] H D Nguyen and G J McLachlan. Maximum likelihood estimation of triangular and polygonal distributions. *Computational Statistics and Data Analysis*, 106:23–36, 2016.
- [38] H D Nguyen and G J McLachlan. Linear mixed models with marginally symmetric nonparametric random-effects. *Computational Statistics and Data Analysis*, 106:151–169, 2016.
- [39] H D Nguyen and G J McLachlan. Laplace mixture of linear experts. *Computational Statistics and Data Analysis*, 93:177–191, 2016.
- [40] H D Nguyen, L R Lloyd-Jones, and G J McLachlan. A universal approximation theorem for mixture-of-experts models. *Neural Computation*, 28:2585–2593, 2016.
- [41] H D Nguyen, L R Lloyd-Jones, and G J McLachlan. A block minorization-maximization algorithm for heteroscedastic regression. *IEEE Signal Processing Letters*, 23:1031–1135, 2016.
- [42] L R Lloyd-Jones, H D Nguyen, G J McLachlan, W Sumpton, and Y-G Wang. Mixture of time dependent growth models with an application to blue swimmer crab length-frequency data. *Biometrics*, 72:1255–1265, 2016.
- [43] H D Nguyen and G J McLachlan. Maximum likelihood estimation of Gaussian mixture models without matrix operations. *Advances in Data Analysis and Classification*, 9:371–394, 2015.
- [44] H D Nguyen, G J McLachlan, N Cherbuin, and A L Janke. False discovery rate control in magnetic resonance imaging studies via Markov random fields. *IEEE Transactions on Medical Imaging*, 33:1735–1748, 2014.
- [45] L R Lloyd-Jones, H D Nguyen, Y-G Wang, and M F O’Neill. Improved estimation of size-transition matrices using tag-recapture data. *Canadian Journal of Fisheries and Aquatic Sciences*, 71:1385–1394, 2014.
- [46] D Chen, A Shah, H Nguyen, D Loo, K Inder, and M Hill. Online quantitative proteomics p-value calculator for permutation-based statistical testing of peptide ratios. *Journal of Proteomics Research*, 13:4184–4191, 2014.

- [47] H D Nguyen, M M Hill, and I A Wood. A robust permutation test for quantitative SILAC proteomics experiments. *Journal of Integrated OMICS*, 2(80-93), 2012.
- [48] K L Inder, Y Z Zheng, M J Davis, H Moon, D Loo, H Nguyen, J A Clements, R G Parton, L J Foster, and M M Hill. Expression of PRTF in PC-3 cells modulated cholesterol dynamics and actin cytoskeleton impacting secretion pathways. *Molecular and Cellular Proteomics*, 11(M111.012245), 2012.

### Conference Papers

- [49] D Fryer, H Nguyen, and P Castellazzi.  $k$ -means on positive definite matrices, and an application to clustering in radar image sequences. In *Proceedings of the IEEE Symposium Series on Computational Intelligence*, 2020.
- [50] H D Nguyen. An introduction to approximate Bayesian computation. In *Proceedings of the Research School on Statistics and Data Science (RSSDS)*, 2019.
- [51] F Chamroukhi, F Lecocq, and H D Nguyen. Regularized estimation and feature selection in mixtures of Gaussian-gated experts models. In *Proceedings of the Research School on Statistics and Data Science (RSSDS)*, 2019.
- [52] H D Nguyen, A T Jones, and G J McLachlan. Positive data kernel density estimation via the logKDE package for R. In *Proceedings of the Sixteenth Australasian Data Mining Conference*, 2018.
- [53] H D Nguyen and G J McLachlan. Iteratively-reweighted least-squares fitting of support vector machines: a majorization-minimization algorithm approach. In *Proceedings of the 2017 Future Technologies Conference (FTC)*, 2017.
- [54] H D Nguyen. A two-sample Kolmogorov-Smirnov-like test for Big Data. In *Proceedings of the Fifteenth Australasian Data Mining Conference*, 2017.
- [55] H D Nguyen and G J McLachlan. Asymptotic inference for hidden process regression models. In *Proceedings of the IEEE Statistical Signal Processing Workshop*, 2014.

- [56] H D Nguyen, A L Janke, N Cherbuin, G J McLachlan, P Sachdev, and K J Anstey. Spatial false discovery rate control for magnetic resonance imaging studies. In *Proceedings of the 2013 Digital Imaging: Techniques and Applications (DICTA) Conference*, 2013.
- [57] H D Nguyen and I A Wood. Variable selection in statistical models using population-based incremental learning with applications to genome-wide association studies. In *Proceedings of the 2012 IEEE Congress on Evolutionary Computation (CEC)*, 2012.

## Book Chapters

- [58] G J McLachlan, S K Ng, and H D Nguyen. EM Algorithm. In *Wiley StatsRef: Statistics Reference Online*. Wiley, Chichester, 2021.
- [59] H D Nguyen and A T Jones. Big data-appropriate clustering via stochastic approximation and Gaussian mixture models. In *Data Analytics: Concepts, Techniques, and Applications*. CRC Press, 2018.
- [60] H D Nguyen, G J McLachlan, and M M Hill. Permutation tests with false discovery corrections for comparative-profiling proteomics experiments. In *Methods in Molecular Biology: Proteomics Bioinformatics*. Springer, 2017.

## Other Publications

- [61] H Nguyen, editor. *Statistics and Data Science: Proceedings of the 2019 Research School on Statistics and Data Science (RSSDS)*, Singapore, 2019. Springer.
- [62] A T Jones, H D Nguyen, and J J Bagnall. BoltzMM: Boltzmann Machines with MM Algorithms. Software published in the Comprehensive R Archive Network, 2019. <https://CRAN.R-project.org/package=BoltzMM>.
- [63] H D Nguyen, A T Jones, and G J McLachlan. logKDE: Computing log-transformed kernel density estimates for positive data. Software published in the Comprehensive R Archive Network, 2018. <https://CRAN.R-project.org/package=logKDE>.

- [64] J Bagnall, A T Jones, and H Nguyen. Analysing the voting patterns of the Senate of the 45th Australian Parliament via fully-visible Boltzmann machines. Poster presented at UseR! 2018, 2018. <https://hal.archives-ouvertes.fr/hal-01838443v1>.
- [65] H D Nguyen. A novel algorithm for clustering of data on the unit sphere via mixture models. In *JSM Proceedings: Statistical Computing Section*, 2017.
- [66] G J McLachlan and H D Nguyen. Contribution to the discussion of paper by M. Drton and M. Plummer. *Journal of the Royal Statistical Society B*, 79:365, 2017.
- [67] A T Jones and H D Nguyen. lowmemtkmeans: Low memory use trimmed k-means. Software published in the Comprehensive R Archive Network, 2016. <https://CRAN.R-project.org/package=lowmemtkmeans>.
- [68] H D Nguyen. NostalgIR: Advanced text-based plots. Software published in the Comprehensive R Archive Network, 2015. <https://CRAN.R-project.org/package=NostalgIR>.
- [69] H D Nguyen. *Finite mixture models for regression problems*. PhD thesis, University of Queensland, 2015. <https://doi.org/10.14264/uql.2015.584>.

## Submitted Manuscripts, Working Papers, and Preprints

- [70] D C Phan, L T Truong, H D Nguyen, and R Tay. Modelling the safety effects of train commuters' access modes. 2021.
- [71] T T Nguyen, H D Nguyen, F Chamroukhi, and F Forbes. A non-asymptotic penalization criterion for model selection in mixture of experts models. 2021.
- [72] T T Nguyen, F Chamroukhi, H D Nguyen, and F Forbes. A non-asymptotic model selection in block-diagonal mixture of polynomial experts models. 2021.
- [73] H D Nguyen, D Fryer, and G McLachlan. Order selection with confidence for finite mixture models. 2021.



- [74] H D Nguyen and F Forbes. Global implicit function theorems and the online expectation–maximisation algorithm. 2021.
- [75] D Fryer, I Strümke, and H Nguyen. Shapley values for feature selection: the good, the bad, and the axioms. 2021.
- [76] F Forbes, H D Nguyen, T T Nguyen, and J Arbel. Approximate Bayesian computation with surrogate posteriors. 2021.
- [77] J-B Durand, F Forbes, C D Phan, L Truong, H D Nguyen, and F Dama. Bayesian non-parametric spatial prior for traffic crash risk mapping: a case study of Victoria, Australia. 2021.
- [78] F Almasi, H Nguyen, D Heydarian, S Nikbin, C J Jenvey, E Halliwell, R Sohi, E N Ponnampalam, A Desai, M Jois, and M J Stear. Quantification of the variation in grazing behaviour among sheep. 2021.
- [79] S Urchs, A Tam, P Orban, C Moreau, Y Benhajali, H D Nguyen, A C Evans, and P Bellec. Subtypes of functional connectivity associate robustly with asd diagnosis. 2020.
- [80] S Urchs, H D Nguyen, C Moreau, C Dansereau, A Tam, A C Evans, and P Bellec. Reproducible functional connectivity endophenotype confers high risk of ASD diagnosis in subset of individuals. 2020.
- [81] M Terrett, D Fryer, T Doody, H Nguyen, and P Castellazzi. SARGDV: Efficient identification of groundwater-dependent vegetation using synthetic aperture radar. 2020.
- [82] R Sohi, A Carroll, H Nguyen, Z Almasi, J Miller, J Trompf, A Bervan, B I Godoy, M Stear, A Desai, and M Jois. Determination of ewe behaviour around lambing time and prediction of parturition seven days prior to lambing by smart technology in an extensive farm system. 2020.
- [83] T T Nguyen, H D Nguyen, F Chamroukhi, and G J McLachlan. An  $l_1$ -oracle inequality for the lasso in mixture-of-experts regression models. 2020.

- [84] T T Nguyen, F Chamroukhi, H D Nguyen, and G J McLachlan. Approximation of probability density functions via location-scale finite mixtures in Lebesgue spaces. 2020.
- [85] H D Nguyen, T T Nguyen, F Chamroukhi, and G McLachlan. Approximations of conditional probability density functions in Lebesgue spaces via mixture of experts models. 2020.
- [86] H D Nguyen and D V Fryer. A binary-response regression model based on support vector machines. 2020.
- [87] H D Nguyen. Universal inference with composite likelihoods. 2020.
- [88] H D Nguyen. Finite sample inference for generic autoregressive models. 2020.
- [89] D V Fryer, I Strümke, and H Nguyen. Explaining the data or explaining a model? Shapley values that uncover non-linear dependencies. 2020.
- [90] H D Nguyen. A note on the convergence of the Gaussian mean shift algorithm. <https://arxiv.org/abs/1703.02337>, 2017.
- [91] H D Nguyen, G J McLachlan, J F P Ullmann, and A L Janke. Faster functional clustering via Gaussian mixture models. <https://arxiv.org/abs/1608.05481>, 2016.