Jongmin Mun / jongminm@usc.edu / jong-min.org / (+1) 984 317 2288

Education

Ph.D. in Data Sciences and Operations, **University of Southern California**, Los Angeles, CA 2023 – Present
M.A. and B.A. in Statistics, Yonsei University, Seoul, South Korea 2023

Employment

Artificial Intelligence Researcher (as mandatory military service) 2020 – 2021
Center for Army Analysis and Simulation, Republic of Korea Army, Gyeryong, South Korea

Publications

Mun J., Kwak, S., and Kim, I. (2024) "Minimax optimal two-sample testing under **local differential privacy**." Under revision, *Journal of Machine Learning Research.* [preprint] [Python package]

- Quantified privacy-utility trade-off in testing whether two histograms from non-interactive ϵ-local-DP noisy counts originated from different multinomial distributions
  - Developed private permutation test based on U-statistics, providing exact cutoff for hypothesis rejection with guaranteed Type I error control
  - Established minimax optimality of privacy mechanisms that add Laplace noise or flip bits in one-hot vectors, combined with proposed permutation U-statistic test
  - Proved that for sample size n, two k-category multinomial distributions must be separated by at least $k^{1/4}(n\epsilon^2)^{-1/2}$ in ℓ_2 distance to achieve meaningful power, demonstrating degradation from non-private rate $n^{-1/2}$
- Quantified privacy-utility trade-off in testing whether two non-interactive ϵ-DP samples originated from different continuous distributions
  - Demonstrated that adding Laplace noise or flipping bits in one-hot vectors of binned data, combined with proposed permutation U-statistic test, achieve minimax optimality
  - Proved that for sample size n, two d-dimensional s-Hölder or s-Besov smooth distributions must be separated by at least $(n\epsilon^2)^{-2s/(4s+3d)}$ in ℓ_2 distance to achieve meaningful power, demonstrating degradation from non-private rate $n^{-2s/(4s+d)}$
- Implemented proposed algorithms in object-oriented design, using PyTorch for GPU-accelerated computation
- Numerically validated proposed algorithms, handling sample sizes of $\approx$ 400,000 and categories $\approx$ 1,000 via GPU cluster

Mun J., Dubey, P., and Fan, Y. (2025) "High-Dimensional Sparse Clustering via Iterative Semidefinite Programming Relaxed K-Means" Under review, *Advances in Neural Information Processing.* [preprint]

- Designed an semidefinite programming-based iterative algorithm for clustering where the number of features exceeds the sample size, but only few unknown variables are relevant to clustering

Mun J.†, Bang, S., and Kim, J. (2025) "Weighted support vector machine for **extremely imbalanced data**." *Computational Statistics and Data Analysis* [published version] [preprint]

- Derived exact formula for class-weighted SVM hinge loss when minority class with subgroup structure is augmented with synthetic samples generated via Gaussian mixture generative model, derived from asymptotically optimal oracle formula

Nam J.*, **Mun J.***†, Jo S., and Kim J. (2024) "Prediction of forest fire risk for artillery military training using weighted support vector machine for **imbalanced data**." *Journal of Classification* [published version] [preprint]

- Achieved 99% improved forest fire prediction on highly imbalanced dataset by integrating class-weighted loss SVM and subgroup-aware data augmentation using Gaussian mixture generative model

Namgung, J.*, **Mun J.***, Park, Y., Kim, J., and Park, B. (2024) "Sex differences in autism spectrum disorder using **class imbalance** adjusted functional connectivity." *Neuroimage* [published version]

- Developed minority class data augmentation method for brain connectivity dataset with highly imbalanced sex distribution, leveraging Gaussian mixture generative model and dimension reduction via diffusion map embedding
- Identified three new regions of sex differences in autism brain connectivity using ANOVA with the proposed method

Park, Y., Kwon, Y., Lee, D., Kim, S., **Mun J.**, Kim, J., Jung, H., Cheon, J., Chang, J., and Park, J. (2024) "In-vivo integration of soft neural probes through high-resolution printing of liquid electronics on the cranium." *Nature Communications.* [published version]

- Verified that signals acquired from novel probes are similar to those from conventional probes, in cluster structure and phase locking, by applying PCA, k-means clustering, and goodness-of-fit test

* : Co-first authors; †: Work done as Artificial Intelligence Researcher at *Center for Army Analysis and Simulation*