Peng Yang

6100 Main St, Houston, TX, 77005

☐ +1(626)-474-7651 • ✉ py11@rice.edu • 🌐 pengyang0411.github.io

 https://github.com/pengyang0411

EDUCATION

Rice University   **Jan. 2020 - Jun. 2024**
Ph.D. in Statistics, GPA: 3.98/4.00   *Houston, TX*
- Advisor: Dr. Ying Yuan, Dr. Liang Li, Dr. Ziyi Li.
- ASA best student paper award on Statistics in Genomics and Genetics (2023).

Rice University   **Aug. 2018 - Dec. 2019**
Master in Statistics   *Houston, TX*

Ningbo University   **Sep. 2014 - Jul. 2018**
Bachelor of Mathematics and Applied Mathematics   *Ningbo, China*
- Ningbo University Scholarship (2015).
- National First Award of The Mathematical Contest in Modeling (2016).
- Team Honorable Mention in 2016 Mathematical Contest in Modeling (2016).
- Excellent Paper Award in Shenzhen Mathematical Modeling Challenge (2016).
- Zhengjiang State Scholarship (2017).

TECHNICAL SKILLS

- **Languages:** Python, SQL, R, SAS, Matlab, C, Maple
- **Software:** Linux, GitHub, AWS, LaTeX
- **Research:** Bayesian Statistics, Machine Learning, Optimization, A/B Testing, Mixture Models
- **Packages:**
  - **ICeITH** model developer on GitHub (`https://github.come/pengyang0411/ICeITH`).
  - **MERIT** model developer on GitHub (`https://github.com/pengyang0411/MDR`).

PUBLICATIONS

[1]. **Yang, P.**, Yuan, Y. (2022). Design and Sample Size Determination for Multiple-dose Randomized Trials for Dose Optimization. Review in **Biometrics**.

[2]. **Yang, P.**, Yuan, Y., Li, Z. (2022). A Novel Bayesian Model for Assessing Intratumor Heterogeneity of Tumor Infiltrating Leukocytes with Multi-region Gene Expression Sequencing. Review in **JASA (ACS)**.

[3]. **Yang, P.**, Yuan, Y. (2022). A Bayesian Design to Adaptively Borrow Information by Incorporating Longitudinal Measurements. In preparation.

[4]. Liu, S., Park, Y., **Yang, P.**, Mandrekar S., Yuan, Y. (2022). A Bayesian Adaptive Design for Avatar Driven Cancer Clinical Trials. In preparation.

[5]. Cao, S., Wang, J. R., Ji, S.,**Yang, P.**, Dai, Y., Guo, S., ... & Wang, W. (2022). Estimation of tumor cell total mRNA expression in 15 cancer types predicts disease progression. **Nature biotechnology**, 1-10.

[6]. Bahrambeigi, V., Lee, J.J., Branchi, V., Henry, J.T., Rajapakshe, K.I., Stephens, B.M., Dhebat, S., Hurd, M.W., Sun, R., **Yang, P.**. and Wang, W., 2022. Transcriptomic Profiling of Plasma Extracellular Vesicles Enables Reliable Annotation of the Cancer-specific Transcriptome and Molecular Subtype. bioRxiv.

EXPERIENCE

MD Anderson Cancer Center   **Aug. 2019 - Present**
Research Assistant   *Houston, TX*

Experimental Design: *Determine the sample size of a multi-arm randomized clinical trial*
- Proposed a generalized power function to identify the correct dose.
- Developed a maximum type I error function to control the overall type I error.
- Determined sample size and decision rules by balancing generalized power and type I error.
- Created user-friendly package MERIT in R to assist trial enrollment and decision-making

A/B Testing: *Draw treatment effect by information borrowing from historical data*
- Drew causal inference for experiment arms suffer from limited sample size by borrowing historical data.
- Achieved adaptive information borrowing by incorporating longitudinal biomarker data for similarity measure.
- Developed a Bayesian hierarchical model and optimized it by Metropolis-Hastings Random-Walk Sampling Algorithm.

Mixture Model and Optimization: *Evaluating the ITH with multi-region gene expression design*
- Proposed a Bayesian hierarchical model to leverage the cell type profiles as prior knowledge to decompose the mixture gene expression data while accounting for within-subject correlation.
- Developed a variational inference algorithm for optimization and significantly reduced computing time.
- Discovered the relation of proposed intratumor heterogeneity to survival outcomes that potentially serve for diagnosis.
- Achieved the best accuracy (correlation: 0.97; AUC: 0.85) and robustness benchmarked with other state-of-the-art methods.

Clustering: *Spatial clustering for cancer cell subclone identification on spatial transcriptomic data*
- Proposed a kernel-based dissimilarity measure that integrates high-dimensional spatial and transcriptomic profiles to identify cancer cell subclones by hierarchical and k-means clustering using R.
- Evaluated the sensitivity of weights between two kernels and came up with a better data normalization strategy.

Variable Selection: *Profile likelihood-based gene selection method*
- Developed a profile likelihood-based approach to rank the most informative genes and select the top of them to increase the identifiability of DeMixT likelihood using R and C.
- Conducted extensive simulations to assess the robustness of gene ranking performance over 4,000 patient samples in 16 different cancer types.

Rice University   **Aug. 2018 - Present**
Research Assistant   *Houston, TX*

Casual Inference: *Estimating causal effect by de-confounding mediators for Uber eats*
- Developed a simulation scheme for the causation of customers experiencing delay and their future engagement by incorporating potential mediators from a three-sided market.
- Implemented propensity score matching, inverse probability weighting, doubly robust regression, and instrumental variable approach to draw the correct causal effect.
- Proposed a strategy to benefit the company by distributing promo codes to neutralize the negative causal effect of customers experiencing a delay.

Deep Learning: *Satellite image segmentation*
- Performed data augmentation to enlarge the dataset and dimension reduction for outlier detection in Python.
- Adopted U-Net architecture for image segmentation and tuning hyper-parameters to achieve better convergence.
- Implemented bootstrap and dropout layers to reduce overfitting, and ensemble approach for better accuracy.

EXTRACURRICULAR ACTIVITIES

○ Soccer, Billiards, Bouldering, Yo-Yo