

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**
Houston, TX
 - *Ph.D. in Statistics, GPA: 3.98/4.00*
 - 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**
Houston, TX
 - *Master in Statistics*
- **Ningbo University** **Sep. 2014 - Jul. 2018**
Ningbo, China
 - *Bachelor of Mathematics and Applied Mathematics*
 - 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:** Clinical Trial Design, Bayesian Statistics, Machine Learning, Optimization, Mixture Models
- **Packages:**
 - **MERIT** shiny app developer on Trial Design (www.trialdesign.org).
 - **ICeITH** model developer on GitHub (<https://github.com/pengyang0411/ICeITH>).
 - **SAM** prior model developer (R package and shiny app under developing)

PUBLICATIONS

- [1]. **Yang, P.**, Li, D., Lin, R., Huang, B. and Yuan, Y. (2023). Design and Sample Size Determination for Multiple-dose Randomized Trials for Dose Optimization. Review in **Biometrics**.
- [2]. **Yang, P.**, Zhao, Y., Nie, L., Vallejo, J. and Yuan, Y. (2023). SAM: Self-adapting Mixture Prior to Dynamically Borrow Information from Historical Data in Clinical Trials. Revision in **Biometrics**.
- [3]. **Yang, P.**, Yuan, Y., Li, Z. (2023). A Novel Bayesian Model for Assessing Intratumor Heterogeneity of Tumor Infiltrating Leukocytes with Multi-region Gene Expression Sequencing. Revision in **AOAS**.
- [4]. **Yang, P.**, Yuan, Y. (2022). A Bayesian Design to Adaptively Borrow Information by Incorporating Longitudinal Measurements. In preparation.
- [5]. Liu, S., Park, Y., **Yang, P.**, Mandrekar S., Yuan, Y. (2022). A Bayesian Adaptive Design for Avatar Driven Cancer Clinical Trials. In preparation.
- [6]. 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.
- [7]. 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

Research Assistant

Aug. 2019 - Present

Houston, TX

Dose Optimization: *Sample size determination for multi-arm randomized clinical trial*

- Proposed a multi-dose phase II clinical trial for sample size determination and dose optimization that maximizes the benefit-risk tradeoff.
- Generalized standard type I error and power to accommodate the unique characteristics of dose optimization and derived a decision rule.
- Developed Shiny app on www.trialdesign.org to help researchers implement the design and improve the development of treatment.

Information Borrowing: *Dynamic information borrowing*

- Developed a self-adapting mixture (SAM) prior to leveraging the historical information to the current trial design
- Proposed a data-driven way to determine the mixture weight to achieve dynamic information borrowing with different degrees of prior-data conflicts

Information Borrowing: *Joint modelling*

- Proposed a joint Bayesian model that leverages the biomarker measurements from both pediatric and adult data to adaptively measure the congruence between two groups.

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.

Rice University

Research Assistant

Aug. 2018 - Present

Houston, TX

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.

POSTER PRESENTATION

- A Novel Bayesian Model for Assessing Intratumor Heterogeneity of Tumor Infiltrating Leukocytes with Multi-region Gene Expression Sequencing, Advances in Statistical and Computational Methods for Analysis of Biomedical, Genetic, and Omics Data, 2023.
- A Novel Bayesian Model for Assessing Intratumor Heterogeneity of Tumor Infiltrating Leukocytes with Multi-region Gene Expression Sequencing, Eastern North American Region (ENAR) 2023.