Beomjo Park, PhD

beomjop@gmail.com | beomjop | beomjopark | https://beomjopark.github.io | Mountain View, CA

PROFESSIONAL SUMMARY

Senior Research Data Scientist specializing in statistical (causal) inference, machine learning, and optimization for large-scale decision systems. Develops rigorous, award-winning methodologies that bridge theory and practice, from experimentation design and measurement to decision-making under uncertainty. Proven record of architecting scalable systems and optimization frameworks with measurable impact across generative ML, advertising, and scientific applications.

EXPERIENCE

Research Data Scientist, Google, Mountain View, CA — Aug 2023 – Present

- Architected a US patent-pending ML optimization framework for generative video ad portfolios. Built a scalable inference and constrained optimization algorithm to dynamically generate diverse, ROI maximizing video collections from component assets, yielding 13% effective revenue and 2% engagement uplift.
- Developed a decision-theoretic ad quality metric to disentangle systemic biases from user intents using maximum distributional divergence on censored data. Established cross-functional adoption of the metric across multiple teams, steering strategy roadmaps through quantified headroom and prioritization.
- Developed a statistical ablation framework to improve resource efficiency in ad video generation. Unlocked $130M in projected ARR and reduced generation overhead by replacing over 2M under-performing videos.

Data Scientist Intern, Google, Mountain View, CA — May 2022 – Aug 2022

- Engineered Bayesian causal mediation models to quantify the impact of YouTube ad quality on user engagement.
- Improved quality metric by 5% using a latent assignment model to resolve structural missingness in human-eval.

Graduate Researcher, Carnegie Mellon University, Pittsburgh, PA — Jan 2019 – Aug 2023

- Pioneered a novel framework for Robust Universal Inference, developing methods that guarantee uniformly valid confidence sets even under model misspecification, lack of regularity, and data contamination [1][2][8]. Proved the real-world utility of the methods in causal discovery using causal structural equation model [1].
- Modeled spatio-temporal heat transport field of global oceans from large-scale autonomous profiling float observations that are partially missing, heterogeneous, and sparsely distributed. Engineered latent Gaussian process models that uncover climatological insights in collaboration with domain scientists [3].

Graduate Researcher, Korea University, Seoul, Korea — Sep 2016 – Jul 2019

- Developed hierarchical Bayesian model with nonparametric mixture processes, applied to a meta-analysis in medical studies [4] and functional data analysis.
- Designed and evaluated model selection criteria for scalable Variational inference [5][7].
- Contributed to and currently maintain the `bsamGP` open-source `R` package on CRAN. Enhanced the end-user application and built discipline-specific worked examples to facilitate real-world adoption [6].

Research Assistant, NCSoft (NLP lab), Korea — Jul 2016 – Dec 2016

- Designed a hierarchical Bayesian latent model to extract key features and their importance affecting individual baseball players' seasonal performance in Korea Baseball Championship.

HONORS AND AWARDS

Student paper award by Statistical Learning and Data Science Section, *American Statistical Association*. — Aug 2023
Outstanding intern presentation by YouTube Ads QUADS team, *Google*. — Aug 2022
PhD TA of the year (2021-2022) by Dept. of Statistics, *Carnegie Mellon University*. — May 2022
SG graduate student paper presentation award (3rd place) by the *Korean Statistical Society*. — Nov 2017
National Science Scholarship by *Korea Student Aid Foundation*. — Fall 2015
High Honors (with scholarship) by *Korea University*. — Mar 2011 – Aug 2016

EDUCATION

Ph.D. in Statistics & Data Science, *Carnegie Mellon University*, Pittsburgh, PA — Aug 2018 – Aug 2023
M.Sc. in Statistics, *Korea University*, Seoul, Korea — Sep 2016 – Aug 2018
B.Sc. in Industrial Engineering & B.Ec. in Statistics, *Korea University*, Seoul, Korea — Mar 2010 – Aug 2016

PUBLICATIONS

[1] **Park, B.**, Balakrishnan, S. & Wasserman, L. (2025) Robust Universal Inference For Misspecified Models. *Biometrika* ▶ Winner of ASA Statistical Learning and Data Science Student Paper Award.

[2] **Park, B.** (2023) Robust Inference: A Price of Misspecification and How to be Resilient. *Carnegie Mellon Thesis*

[3] **Park, B.**, Kuusela, M., Giglio, D. & Gray, A. (2022) Spatio-Temporal Local Interpolation of Global Ocean Heat Transport using Argo Floats: A Debiased Latent Gaussian Process Approach. *Annals of Applied Statistics*

[4] Jo, S., **Park, B.**, Chung, Y., Kim, J., Lee, E. & Choi, T. (2021) Bayesian semiparametric mixed effects models for meta-analysis of literature data: An application to cadmium toxicity studies. *Statistics In Medicine*.

[5] Lim, D., **Park, B.**, Nott, D. J., Choi, T., & Xueue, W. (2020) Sparse signal shrinkage and outlier detection in high-dimensional quantile regression with variational Bayes. *Statistics and Its Interface*.

[6] Jo, S., Choi, T., **Park, B.**, & Lenk, P.J. (2019) `bsamGP`: An `R` Package for Bayesian Spectral Analysis Models using Gaussian Process Priors. *Journal of Statistical Software*.

[7] Ong, V. M., Mensah, K. M., Nott, D. J., Jo, S., **Park, B.**, & Choi, T. (2017) A variational Bayes approach to a semiparametric regression using Gaussian process priors. *Electric Journal of Statistics*.

PREPRINTS

[8] **Park, B.**, Balakrishnan, S. & Wasserman, L. (2023) Nonparametric Functional Estimation under Contamination.

CONFERENCE PRESENTATIONS

Park, B., Balakrishnan, S. & Wasserman, L. (Aug. 2023) Robust Universal Inference. *JSM*, Toronto, Canada.

Park, B., & Kuusela, M. (Aug. 2020) Spatio-Temporal Local Interpolation for Quantifying Global Ocean Heat Transport from Autonomous Observations. (Contributed Talk) *JSM*, virtual.

Park, B., & Choi, T. (Jul. 2018) Bayesian Hierarchical Varying-coefficient Mixed Model. (Poster session) *The third East Asia Chapter of ISBA Conference*, Seoul, Korea.

Park, B., & Choi, T. (Nov. 2017) Bayesian Multivariate Hierarchical Semiparametric Mixed Model with Gaussian Process Priors. *The Korean Statistical Society Autumn Conference*, Seoul, Korea.
▶ 3rd place on SG Graduate Student Paper Presentation Award.

TEACHING EXPERIENCES

	Teaching Assistant, Carnegie Mellon University
Aug 2018 – July 2023	Carnegie Mellon Undergrad Research Experience program (Sports Analytics & Optum Camp), Introduction to Statistical Inference (head TA), Advanced Methods for Data Analysis (head TA), Probability and Mathematical Statistics (head TA), Statistical Graphics and Visualization, Statistical Computing.
	Teaching Assistant, Korea University
Sep 2016 – Aug 2017	Mathematical Statistics, Research Methods II, Statistical Computing Methods, Elementary Computational Statistics.

LANGUAGES

Languages:	English (Proficient), Korean (Native)
Programming:	R†, Python, MATLAB, C++, SQL

† Current maintainer of `bsamGP` package on CRAN.