Matteo Sesia

Bridge Hall 401H, Marshall School of Business, University of Southern California, Los Angeles, CA 90089
sesia@marshall.usc.edu — https://msesia.github.io

Professional Experience

Assistant Professor, Thomas Lord Department of **Computer Science** (by courtesy)
University of Southern California, Viterbi School of Engineering. (01/2023–present)

Assistant Professor, Department of **Data Sciences and Operations** (tenure-track)
University of Southern California, Marshall School of Business. (06/2020–present)

Other Professional Activities

Visiting professor, Collegio Carlo Alberto (Torino, Italy). (05/2023)

Research intern, Adobe Inc. (San Jose, California). (06/2017–08/2017)

Education

Ph.D. in Statistics, Stanford University (2020). Advisor: Emmanuel Candès.
Thesis title: New methods for variable importance testing with applications to genetic studies.

M.S. in Physics of Complex Systems, Politecnico di Torino & Université Paris-Sud (2015).
Graduated *cum laude* (highest honors).

M.A. in Statistics and Applied Mathematics, Collegio Carlo Alberto (2015).
Graduated *with distinction* (highest honors).

B.S. in Engineering Physics, Politecnico di Torino (2013).
Graduated *cum laude* (highest honors).

Funding and awards

GenAI research grant (co-PI), University of Southern California, 2023–2024. ($100,000)

Dr. Douglas Basil Award for Junior Business Faculty (USC Marshall), 2023.

Amazon Research Award (sole PI), 2022–2023. ($120,000)

NSF research grant DMS 2210637 (sole PI), 2022–2025. ($160,000)

Jerome H. Friedman Applied Statistics Dissertation Award (Stanford University), 2020.

International Master's Scholarship, Université Paris-Saclay, 2014–2015.

Allievi Honors Program, Collegio Carlo Alberto, 2011–2015.

Teaching

University of Southern California:

GSBA 506b: Applied Managerial Statistics (graduate), Spring 2023.

BUAD 310: Applied Business Statistics (undergraduate), Fall 2020, Fall 2021.

Università di Torino and Collegio Carlo Alberto:

Predictive uncertainty in ML with conformal inference (graduate), May 2023.

Stanford University:

Stats 390: Consulting Workshop (graduate), Summer 2018.

Stats 195: Introduction to R (undergraduate), Spring 2018, Spring 2020.

Publications and Preprints

Publications in primary research areas *

[1] Z. Liang,* M. Sesia,* W. Sun.* Integrative conformal p-values for powerful out-of-distribution testing with labeled outliers. *J. R. Stat. Soc. B* (2023, recently accepted) https://arxiv.org/abs/2208.11111

[2] M. Bashari, A. Epstein, Y. Romano, M. Sesia. Derandomized novelty detection with FDR control via conformal e-values. *NeurIPS* (2023, recently accepted) https://arxiv.org/abs/2302.07294

[3] Z. Liang, Y. Zhou, M. Sesia. Conformal inference is (almost) free for neural networks trained with early stopping. *ICML* (2023). https://proceedings.mlr.press/v202/liang23i.html

[4] S. Bates,* E. Candès,* L. Lei,* Y. Romano,* M. Sesia*. Testing for outliers with conformal p-values. *Annals of Statistics* (2023). https://doi.org/10.1214/22-aos2244

[5] B. Einbinder,* Y. Romano,* M. Sesia,* Y. Zhou.* Training uncertainty-aware classifiers with conformalized deep learning. *NeurIPS* (2022) https://arxiv.org/abs/2205.05878

[6] M. Sesia, S. Favaro. Conformal frequency estimation with sketched data. *NeurIPS* (2022) https://arxiv.org/abs/2204.04270

[7] S. Li,* Z. Ren,* C. Sabatti,* M. Sesia.* Transfer learning in genome-wide association studies with knockoffs. *Sankhya B* (2022) https://doi.org/10.1007/s13571-022-00297-y

[8] N. Fingerhut, M. Sesia, Y. Romano. Coordinated double machine learning. *ICML* (2022). https://proceedings.mlr.press/v162/fingerhut22a.html

[9] S. Li,* M. Sesia,* Y. Romano, E. Candès, C. Sabatti. Searching for robust associations with a multi-environment knockoff filter. *Biometrika* (2021). https://doi.org/10.1093/biomet/asab055

[10] M. Sesia, Y. Romano. Conformal regression with conditional histograms. *NeurIPS* (spotlight) (2021). https://arxiv.org/abs/2105.08747

[11] M. Sesia, S. Bates, E. Candès, J. Marchini, C. Sabatti. False discovery rate control in genome-wide association studies with population structure. *Proc. Natl. Acad. Sci. U.S.A.*, 118 (40) (2021). https://doi.org/10.1073/pnas.2105841118

[12] C. Chia,* M. Sesia,* C.-S. Ho, S. Jeffrey, J. Dionne, E. Candès, R. Howe. Interpretable classification of bacterial Raman spectra with knockoff wavelets. *IEEE J. Biomed. Health. Inform.* (2021). https://doi.org/10.1109/JBHI.2021.3094873

[13] Y. Romano,* M. Sesia,* E. Candès. Classification with valid and adaptive coverage. *NeurIPS* (spotlight) (2020). https://arxiv.org/abs/2006.02544

[14] S. Bates, M. Sesia, C. Sabatti, E. Candès. Causal inference in genetic trio studies. *Proc. Natl. Acad. Sci. U.S.A.*, 117 (39) 24117-24126 (2020). https://doi.org/10.1073/pnas.2007743117

[15] M. Sesia, E. Katsevich, S. Bates, E. Candès, C. Sabatti. Multi-resolution localization of causal variants across the genome. *Nature Commun.*, 11, 1093 (2020). https://doi.org/10.1038/s41467-020-14791-2

[16] M. Sesia, E. Candès. A comparison of some conformal quantile regression methods. *Stat*, 9:e261 (2020). http://dx.doi.org/10.1002/sta4.261

[17] Y. Romano,* M. Sesia,* E. Candès. Deep knockoffs. *J. Am. Stat. Assoc.* (2019). https://doi.org/10.1080/01621459.2019.1660174

[18] M. Sesia, C. Sabatti, E. Candès. Rejoinder: "Gene hunting with hidden Markov model knockoffs". *Biometrika*, 106, 35–45 (2019). https://doi.org/10.1093/biomet/asy075

*Asterisks indicate equal contributions.

[19] M. Sesia, C. Sabatti, E. Candès. Gene hunting with hidden Markov model knockoffs. *Biometrika*, 106, 1–18 (2019). https://doi.org/10.1093/biomet/asy033

Publications from consulting and collaborations

[20] M. A. Juratli, D. Roy, E. Oppermann, M. Sesia, A. Schnitzbauer, J. Hoelzen, S. Katou, H. Morgul, B. Strucker, A. Pascher, M. Heikenwalder, W. O. Bechstein. Long-term effect of liver resection on circulating stem cells expressing PD-L1 in patients with hepatocellular carcinoma – Pilot study. *Zeitschrift für Gastroenterologie* (2022) https://www.thieme-connect.com/products/ejournals/abstract/10.1055/s-0042-1754779

[21] M. A. Juratli, D. Roy, E. Oppermann, M. Sesia, A. Schnitzbauer, J. Hoelzen, S. Katou, H. Morgul, B. Strucker, A. Pascher, W. O. Bechstein. Long-term dynamics of circulating tumor cells and their prognostic relevance after liver resection in patients with hepatocellular carcinoma. *Zeitschrift für Gastroenterologie* (2022) https://www.thieme-connect.com/products/ejournals/html/10.1055/s-0042-1754884

[22] J. Hoelzen, K. Sander, M. Sesia, D. Roy, E. Rijcken, A. Schnabel, B. Strücker, M. Juratli, A. Pascher. Robotic-assisted esophagectomy leads to significant reduction in postoperative acute pain: A retrospective clinical trial. *Ann. Surg. Oncol.* (2022) https://doi.org/10.1245/s10434-022-12200-0.

[23] A. Fayazi, M. Sesia, K. S. Anand. Hyperoxemia among pediatric intensive care unit patients receiving oxygen therapy. *J. Pediatr. Intensive Care* (2021). https://doi.org/10.1055/s-0041-1740586

Preprints

[24] M. Beraha, S. Favaro, and M. Sesia. Frequency and cardinality recovery from sketched data: a novel approach bridging Bayesian and frequentist views. (2023) https://arxiv.org/abs/2309.15408

[25] M. Sesia, Y. X. Wang, Xin Tong. Adaptive conformal classification with noisy labels. (2023) https://arxiv.org/abs/2309.05092

[26] M. Sesia, S. Favaro, E. Dobriban. Conformal frequency estimation using discrete sketched data with coverage for distinct queries. (2022) https://arxiv.org/abs/2211.04612

[27] S. Favaro, M. Sesia. Bayesian nonparametric estimation of coverage probabilities and distinct counts from sketched data. (2022) https://arxiv.org/pdf/2209.02135

[28] M. Sesia, T. Sun. Searching for subgroup-specific associations while controlling the false discovery rate. (2022) https://arxiv.org/abs/2205.08653

Patents

M. Sesia and Y. Abbasi-Yadkori (Adobe Inc). "Recommendation system using linear stochastic bandits and confidence interval generation". US 11,100,559. August 24, 2021.

Professional Service

Journal referee:[†] Annals of Statistics (7), Bayesian Analysis (1), Biometrika (5), Briefings in Bioinformatics (1), Electronic Journal of Statistics (1), Human Genetics (1), Information Sciences (1), Journal of Machine Learning Research (3), Journal of the American Statistical Association (5), Journal of the Royal Statistical Society B (4), Nature Communications (1), Operations Research (1), SIAM Journal on Mathematics of Data Science (1), Statistics and Computing (2), Statistics and Probability Letters (1), Statistics in Medicine (1), Statistical Science (2).

Conference referee: COLT (1), ISIT (1), NeurIPS (20), ICML (12).

Ad-hoc grant reviewer: Israel Science Foundation 2022.

[†]Brackets contain the number of manuscripts reviewed.

Grant review panelist: National Science Foundation 2023.

Department Service

Statistics seminar organizer (2021–2022, 2022–2023); faculty hiring committee (2021–2022, 2022–2023); PhD admission committee (2020–2021, 2021–2022)

Students

Current USC students: Cora Liang (Math, PhD, 5th year), Tianmin Xie (Data Sciences and Operations, PhD, 2nd year), Yanfei Zhou (Data Sciences and Operations, PhD, 3rd year).

Past USC students: Yibin Xiong (Applied Mathematics, Undergraduate, Summer–Fall 2022), Xuanqi Zhang (Financial Engineering, Master, Fall 2021).

Dissertation committees: Gregory Faletto (Data Sciences and Operations).

Qualifying exam committees: Yiqiu Shen (Data Sciences and Operations).

Presentations

Invited presentations

University of Washington, Department of Biostatistics. Nov. 9, 2023, in Seattle, WA.
University of Pittsburgh, Department of Statistics. Oct. 23, 2023, in Pittsburgh, PA.
Polytechnic University of Milan, Department of Mathematics. Sept. 21, 2023, in Milan, Italy.
Joint Statistical Meeting, Invited Poster Session. Aug. 6, 2023, in Toronto, Canada.
Computational Genomics Summer Institute, July 15, 2023, in Big Bear Lake, CA.
Boston University, New England Statistics Symposium (short course). June 4, 2023, in Boston, MA.
Università Cattolica, Department of Statistical Sciences. May 18, 2023, in Milan, Italy.
Collegio Carlo Alberto, Statistics Seminar. May 11, 2023, in Torino, Italy.
National Institute of Statistical Sciences. April 19, 2023, in Washington, D.C. (remote).
University of California - Irvine, Department of Mathematics. March 6, 2023, in Irvine, CA.
International Conference on Statistics and Data Science, Dec. 13, 2022, in Florence, Italy.
University of Southern California, Department of Computer Science, Oct. 27, 2022, in Los Angeles, CA.
University of California - Santa Cruz, Department of Statistics. October 17, 2022, in Santa Cruz, CA.
SIAM Symposium on Mathematics of Interpretable ML, Sept. 30, 2022, in San Diego, CA.
Computational Genomics Summer Institute, July 7, 2022, in Big Bear Lake, CA.
International Seminar on Selective Inference, June 2, 2022, (remote).
University of Southern California, Department of Economics, Feb. 25, 2022, in Los Angeles, CA.
Merck & Co., Inc. Feb. 10, 2022, in Morristown, NJ (remote).
Yale University, School of Public Health. Feb. 16, 2021, in New Haven, CT (remote).
University of Nottingham, School of Mathematical Sciences. Dec. 17, 2020, in Nottingham, UK (remote).
University of Milan - Bicocca, DEMS Department. Dec. 2, 2020, in Milan, Italy (remote).
Johns Hopkins University, Mathematical Institute for Data Science. Feb. 18, 2020, in Baltimore, MD.
University of Southern California, DSO Department. Jan. 27, 2020, in Los Angeles, CA.
University of California - Davis, Department of Statistics. Jan. 6, 2020, in Davis, CA.
Regeneron Pharmaceuticals, Inc. Sept. 17, 2019, in Eastview, NY.
23andMe, Inc. May 21, 2019, in Mountain View, CA.
Stanford University, Statistics Department. July 16, 2019, in Stanford, CA.
Stanford University, Statistics Industrial Affiliates Meeting. Feb. 22, 2019, in Stanford, CA.
Collegio Carlo Alberto, Statistics Seminar. Dec. 19, 2018, in Torino, Italy.
Université Grenoble Alpes, Bayes in Grenoble Seminar. July 10, 2018, in Grenoble, France.

Invited discussions

International Seminar on Selective Inference, Nov. 15, 2023, (remote).
International Seminar on Selective Inference, Nov. 17, 2022, (remote).
International Seminar on Selective Inference, Oct. 22, 2020, (remote).

Contributed presentations

Joint Statistical Meeting, Topic-Contributed Paper Session. Aug. 9, 2023, in Toronto, Canada.
RMDS Lab, IM DATA Conference 2022, Aug. 14, 2022, in Los Angeles, CA.
ICML, spotlight presentation. July 19, 2022, in Baltimore, MD.
NeurIPS, spotlight presentation. Dec. 9, 2021 (remote).
RMDS Lab, IM DATA Conference 2021, Oct. 28, 2021, in Pasadena, CA (remote).
RMDS Lab, IM DATA Conference 2020, Nov. 2, 2020, in Pasadena, CA (remote).
NeurIPS, spotlight presentation. Dec. 8, 2020 (remote).
Royal Statistics Society Conference, Sept. 3–6, 2018, in Cardiff, United Kingdom.
Workshop on Model Selection, Regularization and Inference, July. 12–14, 2018, in Vienna, Austria.
Computational and Methodological Statistics Conference, Dec. 16–18, 2017, in London, United Kingdom.

Contributed poster presentations

ICML. July 19, 2022, in Baltimore, MD.
American Society for Human Genetics Annual Meeting, Oct. 18–22, 2021, virtual meeting.
ICML Workshop on distribution-free uncertainty quantification, July 24, 2021, virtual meeting.
American Society for Human Genetics Annual Meeting, Oct. 27–30, 2020, virtual meeting.
American Society for Human Genetics Annual Meeting, Oct. 15–19, 2019, in Houston, TX.
Higher-Order Asymptotics and Post-Selection Inference Workshop, Aug. 17–19, 2019, in St. Louis, MO.
American Society for Human Genetics Annual Meeting, Oct. 16–20, 2018, in San Diego, CA.