Department of Biostatistics and Medical Informatics Seminar

Didong Li, Ph.D.

Assistant Professor

Biostatistics

University of North Carolina at Chapel Hill

https://sites.google.com/view/didongli/

Friday, Feb 20, 2026
12:00-1:00 pm
Morgridge Hall Seminar Room 7560
Zoom Meeting: 966 3372 9112
https://uwmadison.zoom.us/j/96633729112?pwd=tHFc9i1dAAgmXe05uWtw8wXBIQZxGB.1
Passcode: 621125

Statistics in the Age of AI: Theory, Methods, and Data

Abstract: Artificial Intelligence (AI) has surged in popularity, creating both opportunities and challenges for statistics. In this talk, I will present three recent directions from my lab that reflect our efforts to engage with the age of AI. First, I will discuss theoretical results for generative models, including statistical foundations that connect latent dimension, approximation error, and model complexity as well as how generative models survive data contamination during recursive training. Second, I will discuss a method to use embeddings from large language models to enhance high-dimensional hypothesis testing, a widely used statistical tool in scientific domains, motivated by problems in cancer genomics where traditional methods are underpowered. I will also discuss extensions to genetic studies, where we curated annotations for 8.9 billion genetic variants from the human genome, and obtained embeddings of these 8.9 billion variants for downstream tasks . Finally, I will switch to an infrastructural view, introducing STimage-1K4M, one of the first and largest publicly available spatial transcriptomics datasets curated by my group, consisting of 1,149 slides and more than 4 million pathology image–gene expression pairs across 50 tissue types. This resource has been downloaded over 230,000 times on HuggingFace and has facilitated the training of multiple foundation models.

Bio: Dr. Didong Li is an Assistant Professor of Biostatistics at the University of North Carolina at Chapel Hill, with secondary affiliations in the Department of Statistics and Operations Research, the Carolina Center for Interdisciplinary Applied Mathematics, the Lineberger Comprehensive Cancer Center, and the Gillings Center for Artificial Intelligence and Public Health. He received a PhD in Mathematics from Duke, completed postdoctoral training at Princeton Computer Science and UCLA Biostatistics, and was a visiting scholar at the Gladstone Institute. His research focuses on theory and methods development for robust inference with complex and high-dimensional data, specifically in generative AI, manifold learning, nonparametric Bayes, information geometry, and spatial statistics. He has applied these methods to electronic health record data, large-scale genetic data, and spatial transcriptomics.

School of Medicine and Public Health
UNIVERSITY OF WISCONSIN–MADISON