Please join us for the Division of Epidemiology, Biostatistics, and Environmental Health’s second colloquium of the fall semester on Wednesday, October 9th from 11:00 -12:00 pm in the University Center Senate Chamber (Room 261).

Please RSVP by Friday, October 4th to Dr. Meredith Ray (maray@memphis.edu).

Saunak Sen, PhD, MStat
Professor and Chief of Biostatistics, Department of Preventative Medicine
University of Tennessee Health Science Center
11:00 am-12:00 pm

**Topic: Sparse matrix linear models for structured high-throughput data**

High-throughput technologies have radically altered our ability to answer complex, large-scale scientific questions. Examples include data from high-throughput chemical genetic screens, mass spectrometers, and sequencing technologies. Data from such platforms are often structured, but such structure is commonly ignored in the statistical analysis, for computational simplicity. Further, due to large data volume, one expects few associations in the data, i.e. we expect signals to be sparse. We present flexible and computationally efficient methods for fitting sparse matrix linear models to structured high-throughput data. We induce model sparsity using an L1 penalty and consider the case when the response matrix and the covariate matrices are large. Since standard estimation methods for these penalized regression models fail, we devise fast, memory-efficient estimation algorithms that use the matrix structure of the model. We evaluate our method’s performance on simulated data and three high-throughput datasets: a large-scale chemical genetic screen in E coli, an environmental exposure assessment using mass spectrometry, and a genetic analysis of a plant transcriptome. For future work, matrix linear models can potentially incorporate different loss and penalty functions or be extended to multi-dimensional (tensor-valued) responses. Our algorithms implemented in the Julia programming language and are publicly available.