Poster Title

Ronald A. Fisher
(*xxx@stat.ucla.edu)

UCLA Department of Statistics

Objectives & Motivations

- Develop novel and modern techniques in high-dimensional statistics and sparse regularization for structure learning of Bayesian networks (BNs) from big data.
- Inspired by applications in computational biology, e.g., construction of gene networks from high-throughput genomic data. Hence, focused on methods that scale to thousands of variables for both continuous and discrete data.

Penalized Likelihood and High-D Theory

- n i.i.d. observations of $X = (X_1, \ldots, X_p)$;
  adding interventions for causal learning is possible.
- Bayesian network (BN) for X, parameterized by coefficients $B := (\beta_{ij})$, interpreted as a weighted adjacency matrix, *always a directed acyclic graph (DAG)*.
- Item ...

Theory: High-D regime, $p \gg n \to \infty$, degree-growth $d \log p/n = o(1)$,

1. **Deviation bounds:**
2. **Sparsity bounds:**
3. **Model selection consistency:**
4. **Uniform control of SEM coeffs.:** All estimated DAGs $\widehat{B}(\pi)$ are close to their true DAG $\widetilde{B}(\pi)$, for all orderings π. Leveraged lattice property of neighborhood regression and led to study of abstract neighborhood regression and connections with PCGs and their algebraic properties.

References

Paper 1
Paper 2
Preprint 1, arXiv:xxxx.

PCGs and neighborhood regression

- General framework for studying neighborhood regression. Regressing X_j on a subset S of the rest of variables $X_S = \{X_i, i \in S\}$:
- Partial correlation graphs (PCGs): Natural setting for studying these relations. General setup in a Hilbert space. Express everything in terms of