A practical evaluation of recent methods in high-dimensional inference

Charles Zheng

Stanford University

May 5, 2015

Problem and motivation

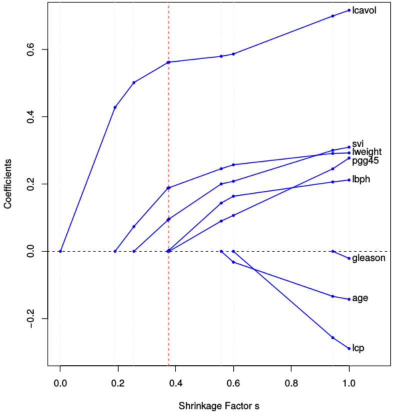
- $x \in \mathbb{R}^p, y \in \mathbb{R}$ have a joint distribution P where $y|x \sim N(x^T\beta, \sigma^2)$
- Observe $X = (x_1, \ldots, x_n)^T$, $Y = (y_1, \ldots, y_n)$ iid
- Problem: test $H_i : \beta_0 = i$ for $i = 1, \ldots, p$
- Motivation: x are SNPs (mutations), y is phenotype

Methods

	Control	$p > n$	
Classical inference (Pearson 1930)	Marginal	No	
Covariance test (Lockhart et al. 2014)	FWER?	Yes	
Debiased lasso (Javanmard et al. 2014)	Marginal	Yes	
Knockoffs (Barber et al. 2014)	FDR	?	

The LASSO path

$$\hat{\beta}_\lambda = \text{argmin}_\beta \frac{1}{2}||X\beta - Y||^2 + \lambda||\beta||_1$$



(Image credit: ??)

Covariance test

- (2014) Lockhart, Taylor, Tibshirani (x 2)
- Standard assumptions $Y \sim N(X\beta, \sigma^2 I)$ + large p asymptotics
- *See also* non-asymptotic exact test (Lee, Sun x 2, Taylor 2015)

Step	Predictor entered	Forward stepwise	Lasso
1	lcavol	0.000	0.000
2	lweight	0.000	0.052
3	svi	0.041	0.174
4	lbph	0.045	0.929
5	pgg45	0.226	0.353
6	age	0.191	0.650
7	lcp	0.065	0.051
8	gleason	0.883	0.978

Debiased regularized M-estimators

- (2014) Javanmard and Montanari
- Standard assumptions + sparsity condition on β + large *n* and *p* asymptotics