Empirical Bayes False Discovery Rate controlling methodology

Daniel Yekutieli

Statistics and OR
Tel Aviv University

October 13, 2025

Plan of my talk

1. FDR controlling methodology of Benjamini and Hochberg ’95
2. Review Bayesian FDR for the two group model (Efron et al ’01, Storey ’01 and ’02, Sun and Cai ’07, Efron ’12)
3. Compare eBayes FDR controlling procedures to the BH procedure
4. Review notion of invariant decision problems (widely studied in 1950’s - 60’s, Berger ‘06) and present examples of application of eBayes in invariant problems
5. Introduce joint research project with Prof. Zhang and Hannig

Yosef Hochberg 1945-2013

Yosef Hochberg 1945-2013

PhD at UNC in 1974, advised by Pranab K. Sen

J. R. Statist. Soc. B (1995)
57, *No.* 1, *pp.* 289–300

Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing

By YOAV BENJAMINI† and YOSEF HOCHBERG

Tel Aviv University, Israel

[Received January 1993. Revised March 1994]

SUMMARY

The common approach to the multiplicity problem calls for controlling the familywise error rate (FWER). This approach, though, has faults, and we point out a few. A different approach to problems of multiple significance testing is presented. It calls for controlling the expected proportion of falsely rejected hypotheses—the false discovery rate. This error rate is equivalent to the FWER when all hypotheses are true but is smaller otherwise. Therefore, in problems where the control of the false discovery rate rather than that of the FWER is desired, there is potential for a gain in power. A simple sequential Bonferroni-type procedure is proved to control the false discovery rate for independent test statistics, and a simulation study shows that the gain in power is substantial. The use of the new procedure and the appropriateness of the criterion are illustrated with examples.

Keywords: BONFERRONI-TYPE PROCEDURES; FAMILYWISE ERROR RATE; MULTIPLE-COMPARISON PROCEDURES; *p*-VALUES

Multiple hypotheses testing framework

- m tested null hypotheses $H_1 \cdots H_m$
- m_0 null hypotheses $(P_i \sim U[0,1])$,
  $m_1 = m - m_0$ false null hypotheses $(P_i \leq U[0,1])$
- Rejecting a null hypothesis is a discovery, a false discovery is erroneously rejecting a true null hypothesis
- R is the number of discoveries and V is the number of false discoveries

$$FWE := \Pr(V > 0)$$

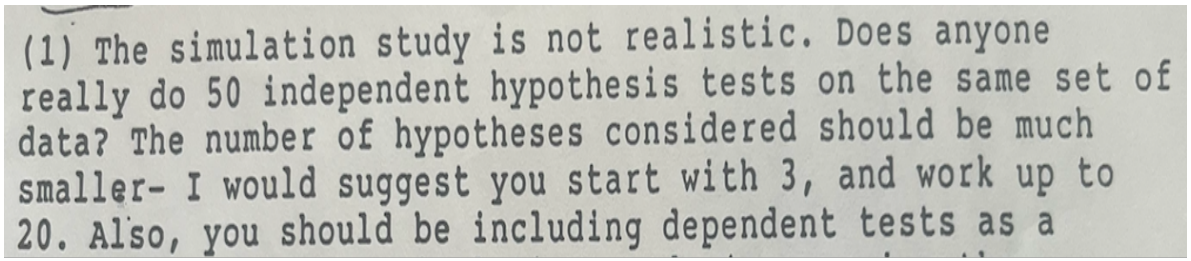
$$FDR := \mathbb{E}\, FDP, \quad FDP = \begin{cases} 0 & \text{if } R = 0 \\ V/R & \text{if } R > 0 \end{cases}$$

Level q BH procedure

1. Sort the p-values $P_{(1)} \leq P_{(2)} \leq \cdots \leq P_{(m)}$
2. Compare $P_{(i)}$ with $i \cdot q/m$
3. Let $r = \max\{i : P_{(i)} \leq i \cdot q/m\}$
4. Discoveries are $H_{(1)} \cdots H_{(r)}$

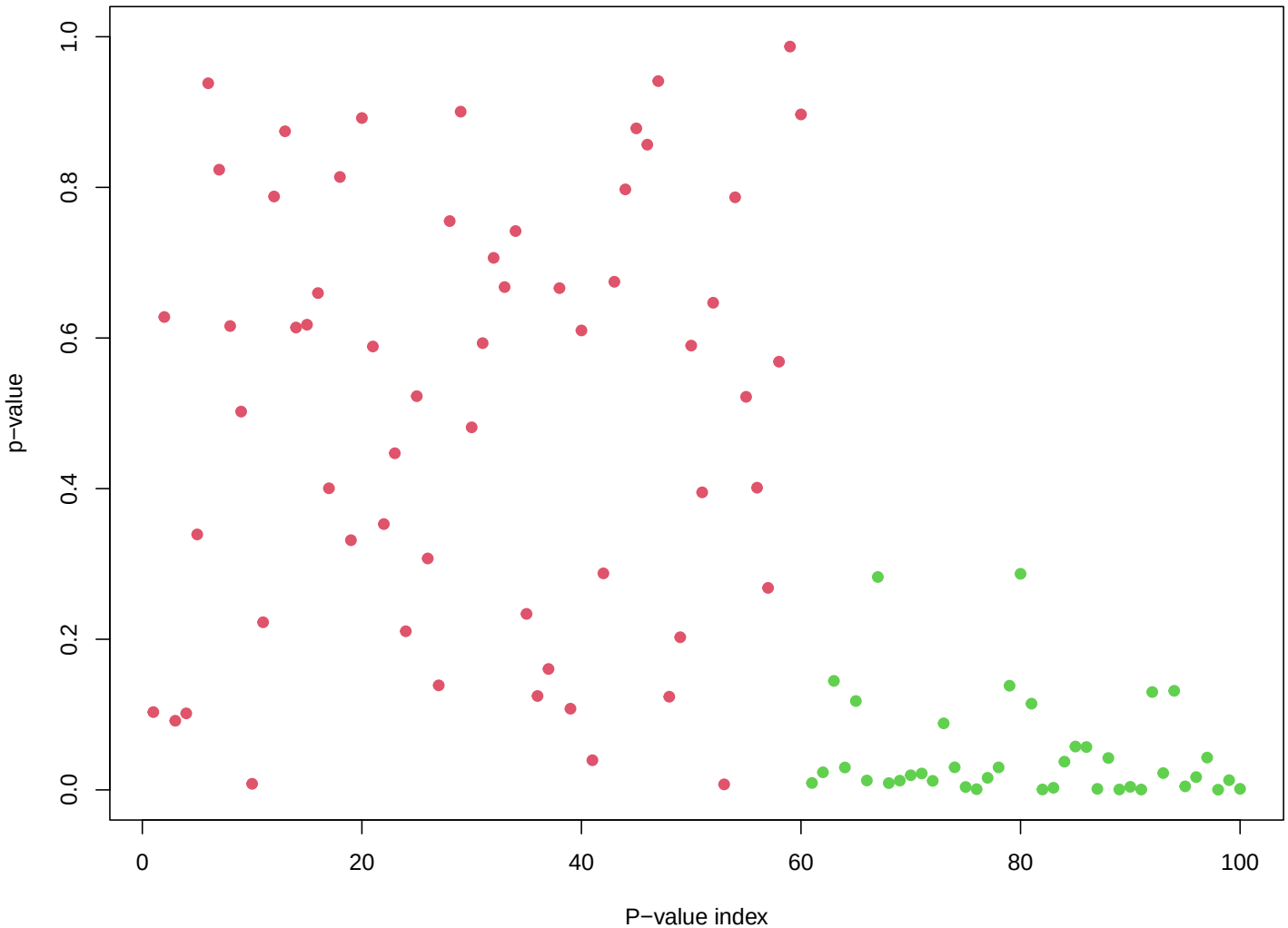
Context

- Consider small testing problems

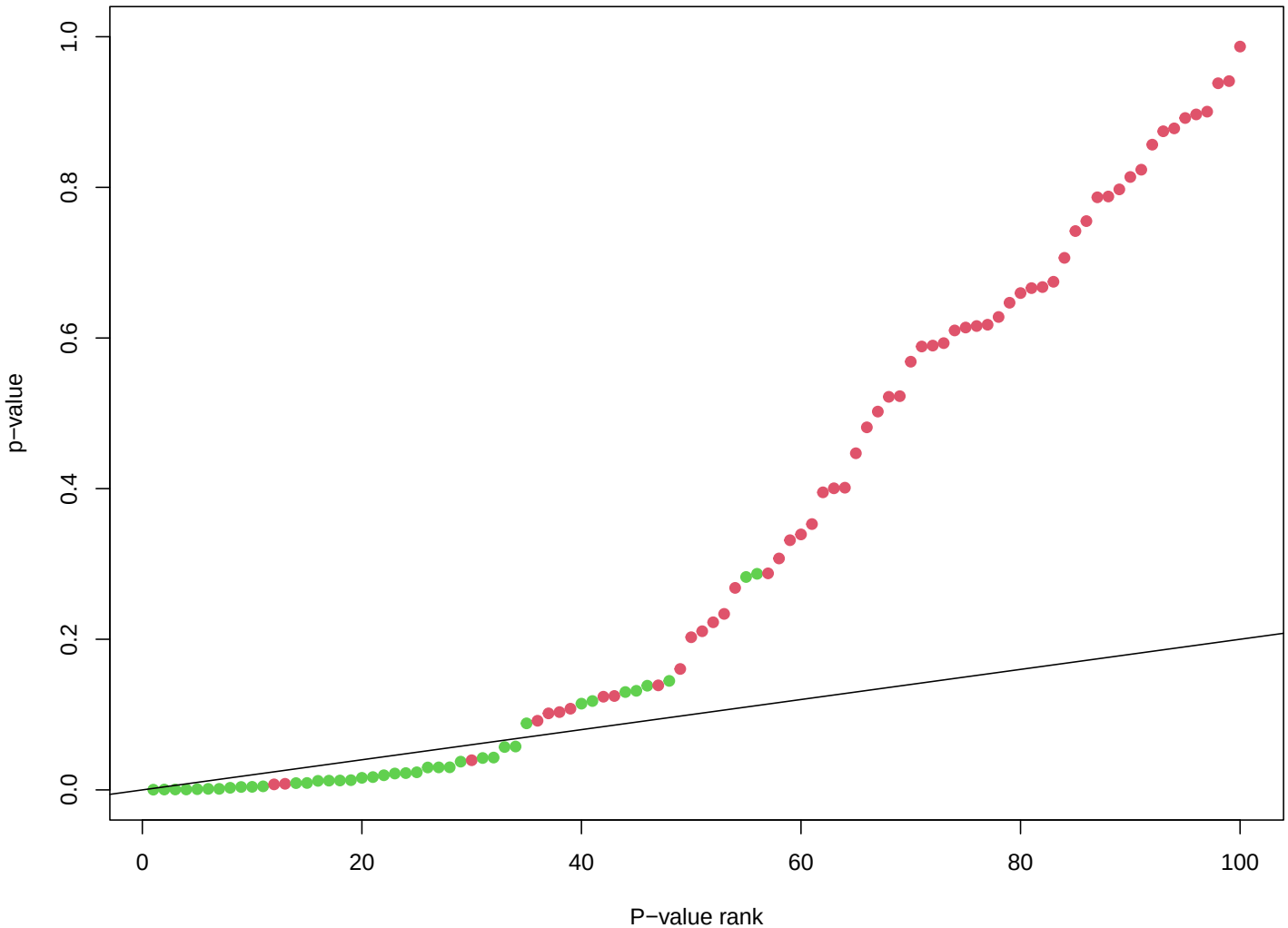

(1) The simulation study is not realistic. Does anyone really do 50 independent hypothesis tests on the same set of data? The number of hypotheses considered should be much smaller- I would suggest you start with 3, and work up to 20. Also, you should be including dependent tests as a

- Compare with the Bonferroni FWE controlling procedure:
  - Discoveries are H_i with $P_i \leq q/m$

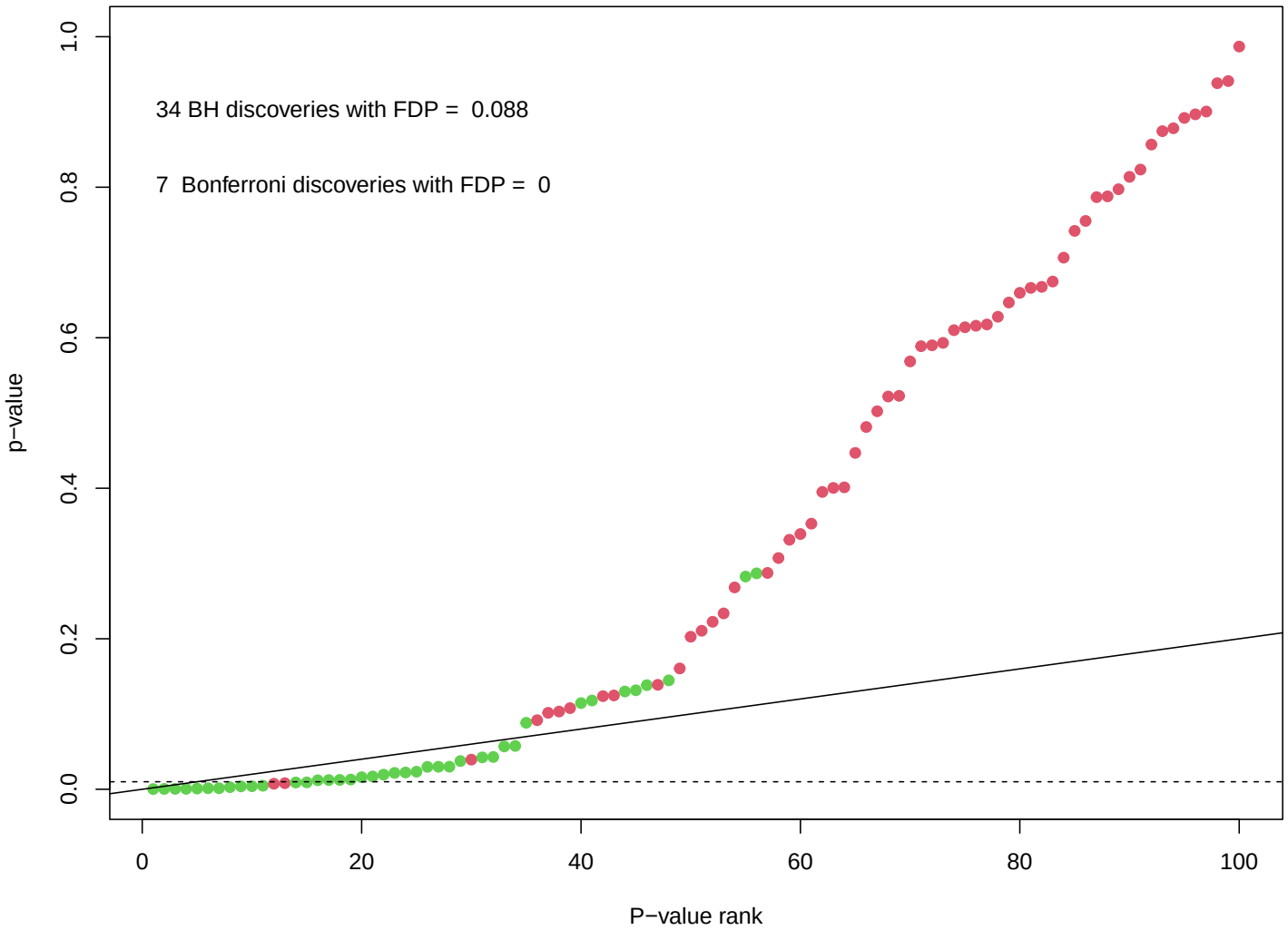
$P_1 \cdots P_{60}$ true null p-values, $P_{61} \cdots P_{100}$ false null p-values



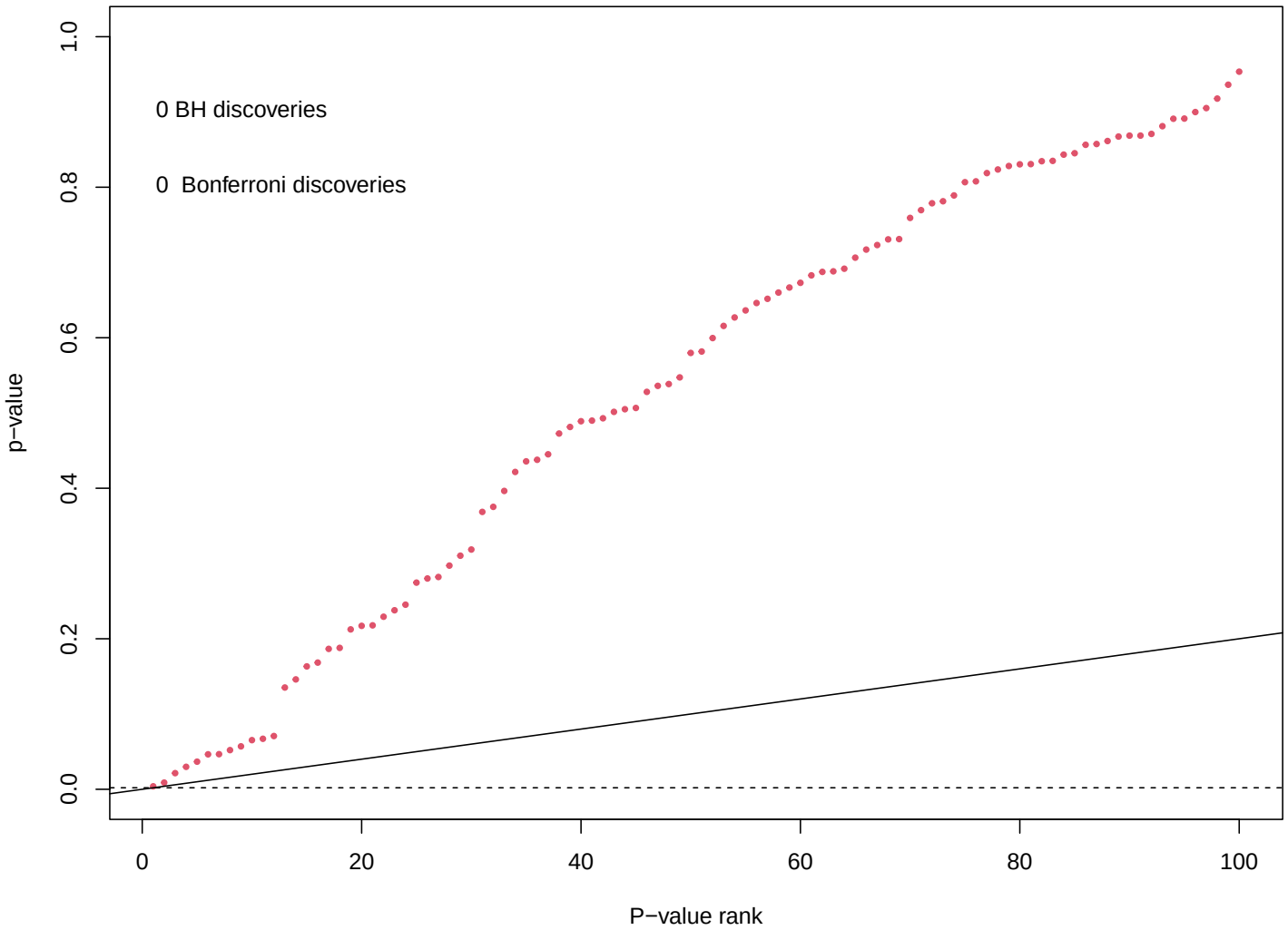
Level $q = 0.20(!?)$ BH procedure



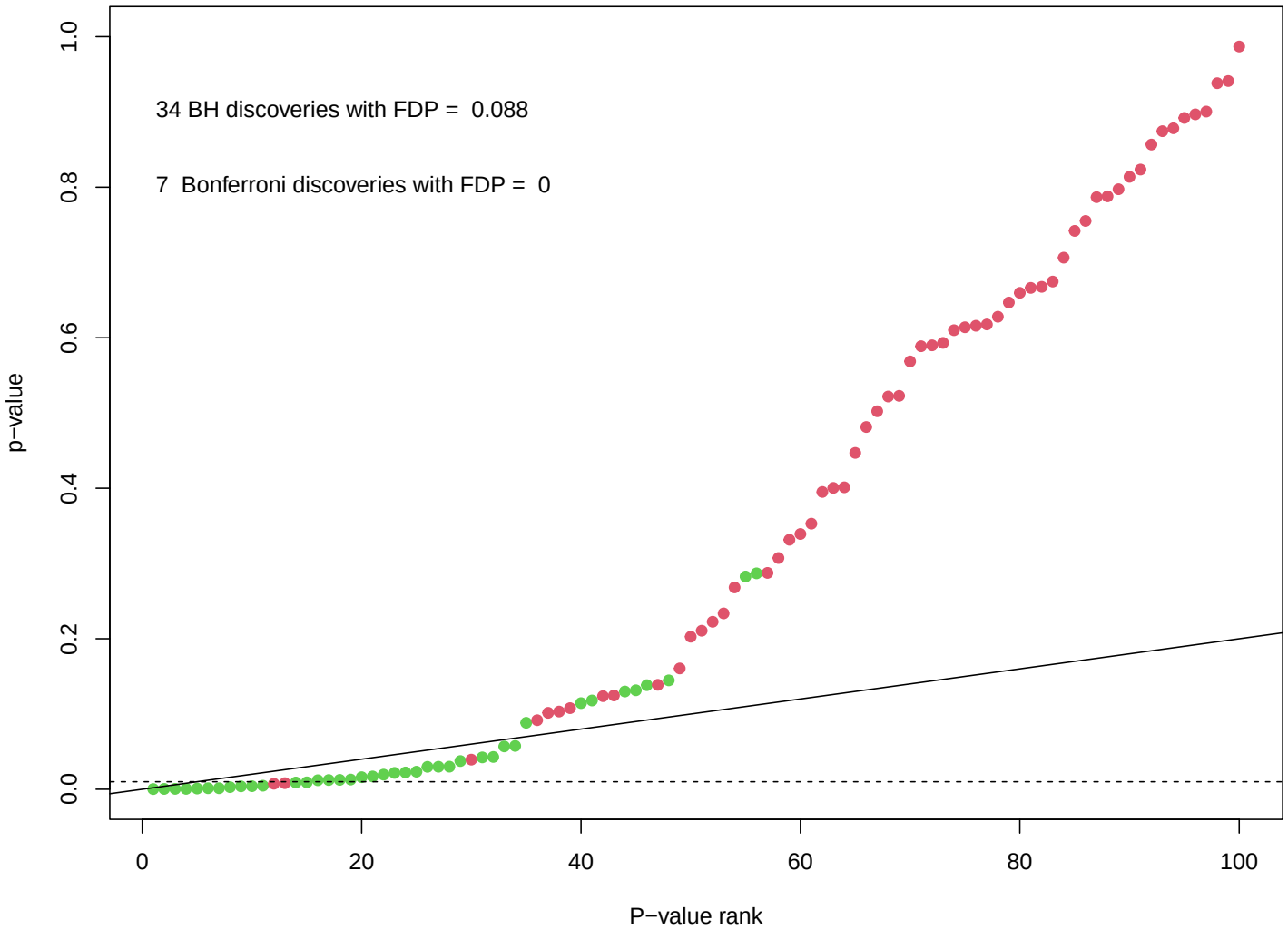
Compare with $\alpha = 0.20$ Bonferroni procedure



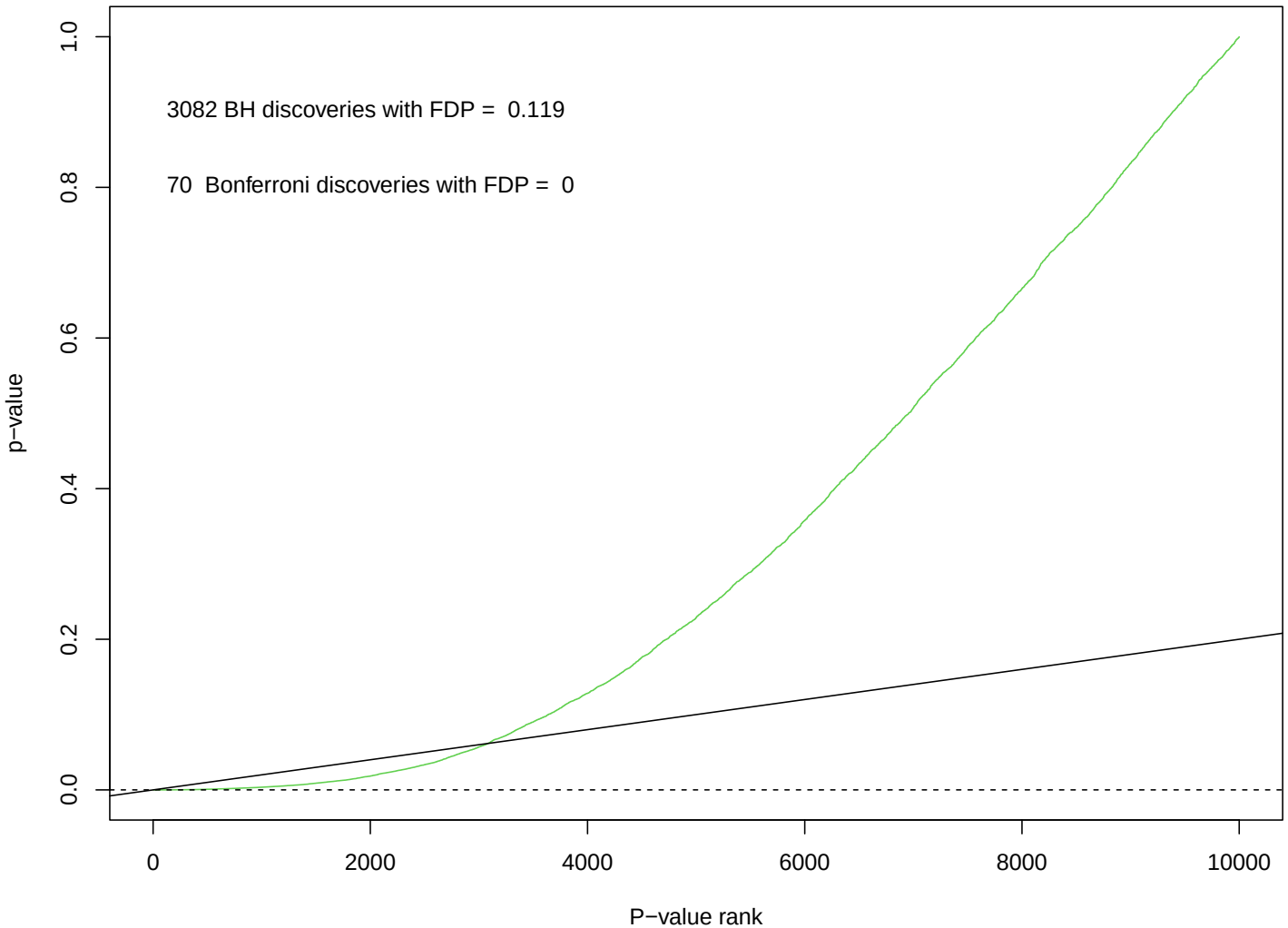
$P_1 \cdots P_{100}$ true null p-values



Return to 60% true null p-values simulation



Same simulation but now $m = 10000$



Are the FDR and the BH procedure new?

– Soric, JASA ‘89

“ . . . It is mainly the discoveries that are reported and included into science . . . unless the proportion of false discoveries is kept small there is danger that a large part of science is untrue”

– Simes, Biometrika ‘86

1. Sort the p-values $P_{(1)} \leq P_{2)} \leq \cdots \leq P_{(m)}$
2. Compare $P_{(i)}$ with $i \cdot q/m$
3. If $\exists i : P_{(i)} \leq i \cdot q/m$ then reject the global null that $H_1 \cdots H_m$ are true null hypotheses.

BH procedure $\rightarrow$ FDR control

BH '95

- Independence $FDR \leq m_0 \cdot q/m$

Benjamini and DY '01

- Independence $FDR = m_0 \cdot q/m$
- Positive dependence $FDR \leq m_0 \cdot q/m$
- Geneal dependence $FDR \leq (1 + 1/2 + \cdots + 1/m) \cdot m_0 \cdot q/m$

Simulations

- Robustness to dependence

Interpretation of level 0.05 BH procedure results

Test m null hypotheses out of which m_0 are true null hypotheses.
e.g. Genome-Wide Association Study trying to find genetic variants associated with a trait or a disease.

- All noise regime ($m_0 = m$)
  – Any discovery is false $FDR \equiv FWE$
  – Make 0 discoveries with prob. 95%
- Signal and noise regime ($m_0 < m$)
  – Many discoveries $FWE \approx 1$ but $FDR \approx 0.05$
  – A randomly selected discovery is true with prob. 95%

Bayesian FDR for the two group mixture model

Two group model:

- Parameter $\vec{H} = (H_1, \cdots, H_m)$: $H_i \overset{iid}{\sim} Bernouli(1-\pi_0)$, where $H_i = 0, 1$ corresponds to a true or false null hypothesis.
- Statistic vector $\vec{Z} = (Z_1, \cdots, Z_m)$: $Z_i | H_i = 0 \sim f_0 = N(0,1)$, $Z_i | H_i = 1 \sim f_1$
- Performing a multiple testing procedure = applying classifier:

$$\mathcal{R}(\vec{Z}) = (R_1(Z_1), \cdots, R_m(Z_m)) \in \{0,1\}^m$$

where R_i is indicator for making a discovery and V_i is indicator for making a false discovery, i.e. $V_i = 1 \quad \Leftrightarrow \quad R_i = 1 \wedge H_i = 0$

- For $R = R_1 + \cdots + R_m$ and $V = V_1 + \cdots + V_m$.

$$FDP = V / \max(R, 1)$$

Relation between Bayesian FDR and the BH FDR

- Efron defines the Bayes FDR (= Storey's pFDR):

$$Fdr = Pr(H_i = 0|R_i = 1) = \mathbb{E}_{\vec{H},\vec{Z}}(V/R|R > 0)$$

- For the two group model the BH FDR is given by

$$FDR = \mathbb{E}_{\vec{Z}|\vec{H}}\{V/\max(R, 1)\}$$

- Thus we may express

$$Fdr = \mathbb{E}_{\vec{H}}\{FDR \,/\, \Pr_{\vec{Z}|\vec{H}}(R > 0)\}$$

For $\pi_0 < 1$ (otherwise $FDR \equiv 0$) with $\Pr_{\vec{Z}|\vec{H}}(R > 0) \approx 1$ we have $Fdr \approx FDR$

Bayes rules for classification problems

The classification loss is $L(\vec{H}, \vec{Z}) = \sum_{i=1}^{m} L(H_i, Z_i)$ for

$$L(H_i, Z_i) = \lambda_1 \cdot I(H_i = 0, R_i(Z_i) = 1) + \lambda_2 \cdot I(H_i = 1, R_i(Z_i) = 0)$$

with type-I and type-II error penalties λ_1 and λ_2.

- The Bayes rule minimizing $\mathbb{E}_{\vec{H},\vec{Z}} L(\vec{H}, \vec{Z})$:

$$R_i^{Bayes}(Z_i) = 1 \quad \Leftrightarrow \quad \Pr(H_i = 0|Z_i) \ \leq \ \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

  where this conditional probability the local FDR

$$fdr(Z_i) = \Pr(H_i = 0|Z_i) = \frac{\pi_0 \cdot f_0(Z_i)}{\pi_0 \cdot f_0(Z_i) + \pi_1 \cdot f_1(Z_i)}$$

- The Bayes FDR for any classification rule is

$$Fdr(R_i) = E_{Z_i|R_i=1} fdr(Z_i)$$

Optimality of Bayes classifier

- The $Fdr = q$ Bayes classifier is

$$R_i^{Bayes}(Z_i; q) = I(fdr(Z_i) \leq \lambda(q))$$

  for $\lambda(q)$ yielding $Fdr(R_i^{Bayes}) = q$

- $R_i^{Bayes}(Z_i; q)$ has more power than any other R_i with $Fdr(R_i) = q$
  - Larger discovery probability:

$$\Pr_{\vec{H},\vec{Z}}(R_i = 1) \leq \Pr_{\vec{H},\vec{Z}}(R_i^{Bayes} = 1)$$

  - Smaller type II error, $Fnr(R_i) = \Pr_{\vec{H},\vec{Z}}(H_i = 1 | R_i = 0)$:

$$Fnr(R_i^{Bayes}) \leq Fnr(R_i)$$

Comparing the eBayes classifier to the BH procedure

To implement the Bayes classifier we need to know the distribution of $H_i \rightarrow$ eBayes estimate for π_0

- For the two group mixture model this eBayes classifier implemented in *q-value* package is almost identical to the KBY (Krieger, Benjamini, DY '06) procedure that incorporates an estimate for π_0 in the BH procedure
- In the continuous parameter mixture model, $\theta_i \sim f(\theta)$ and $Z_i|\theta_i \sim N(\theta_i, 1)$, DY '12 shows that for discovering that $0 < \theta_i$ the Bayes classifier is much more powerful than a KBY procedure applied to $P_i = 1 - \Phi(Z_i)$. This is because under the composite null hypothesis that $\theta_i < 0$ the distribution of P_i is $\geq U[0, 1]$

Comparing the Bayes classifier to the BH procedure (cont.)

Replicability analysis, Heller and DY '14:

- Replicability analysis aims to discover SNP associations that are present in more than one study when combining information from several Genome-wide Association Studies.
- Phrase Replicability analysis as a multiple group mixture model and show that the eBayes classifier is much more powerful than the BH procedure applied to replicability p-values.
- The BH procedure has very little power for discovering replicability because the no-replicability null is a composite null hypothesis and also because the replicability p-value is a suboptimal summary for the multivariate statistics.

Invariance under group of transformations G

Invariant decision problem, Berger '06:

1. Invariant family of distribution: parameter θ and data X, with likelihood
$$X \sim f(x|\theta).$$
$g \in G$ specifies a transformation on the parameters, $g(\theta)$, and a transformation on the data, $Y = \tilde{g}(X)$, such that
$$Y \sim f(x|g(\theta)).$$

2. Invariant estimator / decision rule / action:
$$\hat{\theta}(Y) = \hat{\theta}(\tilde{g}(X)) = g(\hat{\theta}(X)).$$

3. Invariant loss function:
$$L(\theta, \hat{\theta}(X)) = L(g(\theta), g(\hat{\theta}(X))).$$

Invariant decision problems (cont.)

- G partitions the parameter space Θ into orbits:

$$\Theta(\theta_0) = \{g(\theta_0) : g \in G\}$$

- The Risk is constant on $\Theta(\theta_0)$,

$$R(\theta_0, \hat{\theta}) = R(g(\theta_0), \hat{\theta}),$$

  thus equals the average risk for the Haar measure on G.
- This implies that 0.95 credible intervals based on this prior cover the fixed unknown θ with 0.95 probability.
- And for any fixed θ the corresponding Bayes rule is the invariant estimator that minimizes $R(\theta, \hat{\theta})$.

Problem: to implement this approach need an oracle to disclose $\Theta(\theta)$, the orbit of the unknown parameter value $\rightarrow$ eBayes...

Example 1: Two group model $\rightarrow$ invariant decision problem

Fixed $\vec{H}$. G is the group of permutations, $h(g)$ is Uniform on G.
$\Theta(\vec{H})$ is the set of m-vector with $m_0 \times 0's$ and $m_1 \times 1's$.

- As before, $Z_i|H_i = 0 \sim f_0$ and $Z_i|H_i = 1 \sim f_1$
- Note that $(\vec{H}, \vec{Z})$, multiple testing procedures, and the classifying loss are invariant to permutations.
- Denoting $\Theta(H_i = 0)$ the subset of m-vectors with $H_i = 0$ yields

$$\Pr(H_i = 0) = \frac{\sum_{\vec{h} \in \Theta(H_i=0)} \Pr(\vec{H} = \vec{h})}{\sum_{\vec{h} \in \Theta(\vec{H})} \Pr(\vec{H} = \vec{h})} = \frac{\binom{m-1}{m_0-1}}{\binom{m}{m_0}} = \frac{m_0}{m}$$

the Bayes classifier is a little different

$$\Pr(H_i = 0|\vec{Z}) = \frac{\sum_{\vec{h} \in \Theta(H_i=0)} \Pi_{j=1}^{m} f_{h_j}(Z_j)}{\sum_{\vec{h} \in \Theta(\vec{H})} \Pi_{j=1}^{m} f_{h_j}(Z_j)} \neq \frac{m_0 \cdot f_0(Z_i)}{m_0 \cdot f_0(Z_i) + m_1 \cdot f_1(Z_i)}$$

and NOW we get optimal classifying for every fixed $\vec{H}$

Example 2: Normal covariance matrix estimation

Joint work with Malgorzata Bogdan, Jonas Wallin and Daniel Xiang

- Data $X_{n\times m} = (X^1, \cdots, X^n)^T$ with $X^i \overset{iid}{\sim} N(0, \Sigma_{m\times m})$.
- Parameter $\Sigma = \Gamma^T \Delta \Gamma$, with $\Delta_{m\times m} = \mathrm{diag}(D)$ for vector of ordered eigenvalues D and eigenvector matrix Γ.
- G is $\mathrm{O}(m)$, the group of $m \times m$ orthogonal matrices.
- Thus $\Theta(\Sigma)$ is the set of all covariance matrices with the same ordered eigenvalues vector D (= same eigenvalues vector empirical distribution).
- We are working on a hierarchical Bayes modeling approach that uses finite Polya Tree models to generate the eigenvalue distribution and $\mathrm{O}(m)$ prior distribution for the eigenvectors matrix.

Example 3: Linear Regression

Joint work with Malgorzata Bogdan, Jonas Wallin and Asaf Weinstein

- Normal linear regression model with known σ

$$\vec{Y} = \mathbf{X}_{n\times p}\vec{\beta} + \vec{\epsilon},\ \vec{\epsilon} \sim N(0, \sigma^2 I_{n\times n})$$

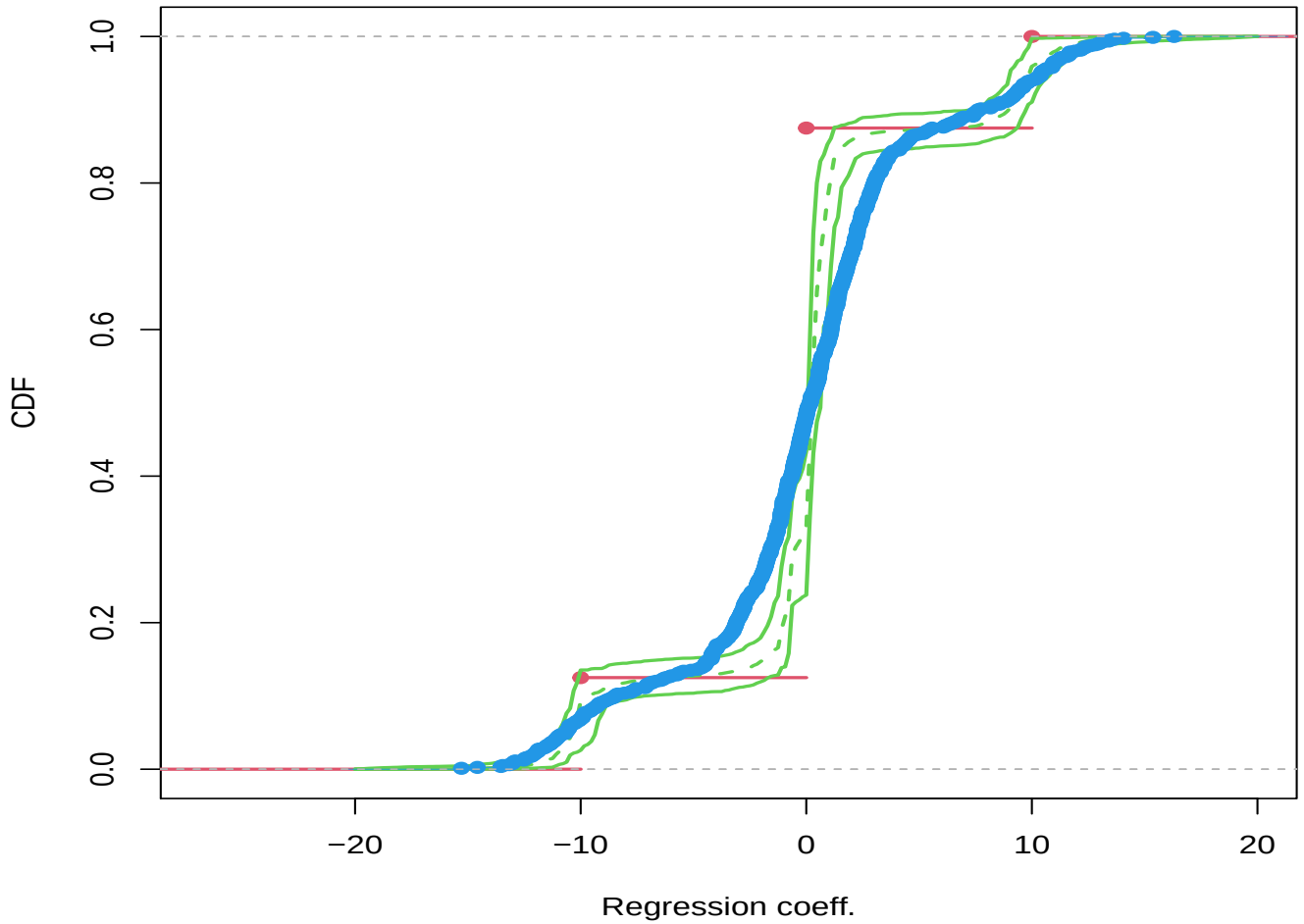
  X-matrix columns such that $\vec{X}_i^T \vec{X}_i = \rho_1$ and $\vec{X}_i^T \vec{X}_j = \rho_2$ for $i \neq j$
- G is the set of all permutation of the model coefficient vector. Thus the orbit for the model consists of all models with the same coefficient vector empirical distribution.
- We present a sampler for the posterior parameter distribution for hierarchical Bayes modeling that uses finite Polya Trees (FPT) to generate the marginal distribution of the model coefficients and and Uniform prior on model coefficient permutations.

Simulated example

- Simulate $\vec{Y} = \mathbf{X}\vec{\beta} + \vec{\epsilon}$
  a. $\mathbf{X}_{4000\times 800}$ with $X_{i,j} \sim N(0, 1/4000)$
  b. $\vec{\beta}$ consists of $100 \times -10's$, $600 \times 0's$ and $100 \times 10's$
  c. $\vec{\epsilon}_{4000\times 1}$ with $\epsilon_j \sim N(0, 2^2)$
- We compare $\sqrt{\sum_{i=1}^{800}(\beta_i - \hat{\beta}_i)^2/800}$ for five estimates for $\vec{\beta}$:
  - Maximum Likelihood Estimate
  - LASSO and Ridge estimates (R GLMNET)
  - hierarchical Bayes posterior mean (for MCMC that inputs $\vec{Y}, \mathbf{X}$ and samples coefficients order statistics and coefficient permutations)
  - Oracle Bayes posterior mean (for MCMC that inputs coefficient order statistics, $\vec{Y}, \mathbf{X}$ and samples coefficient permutations)

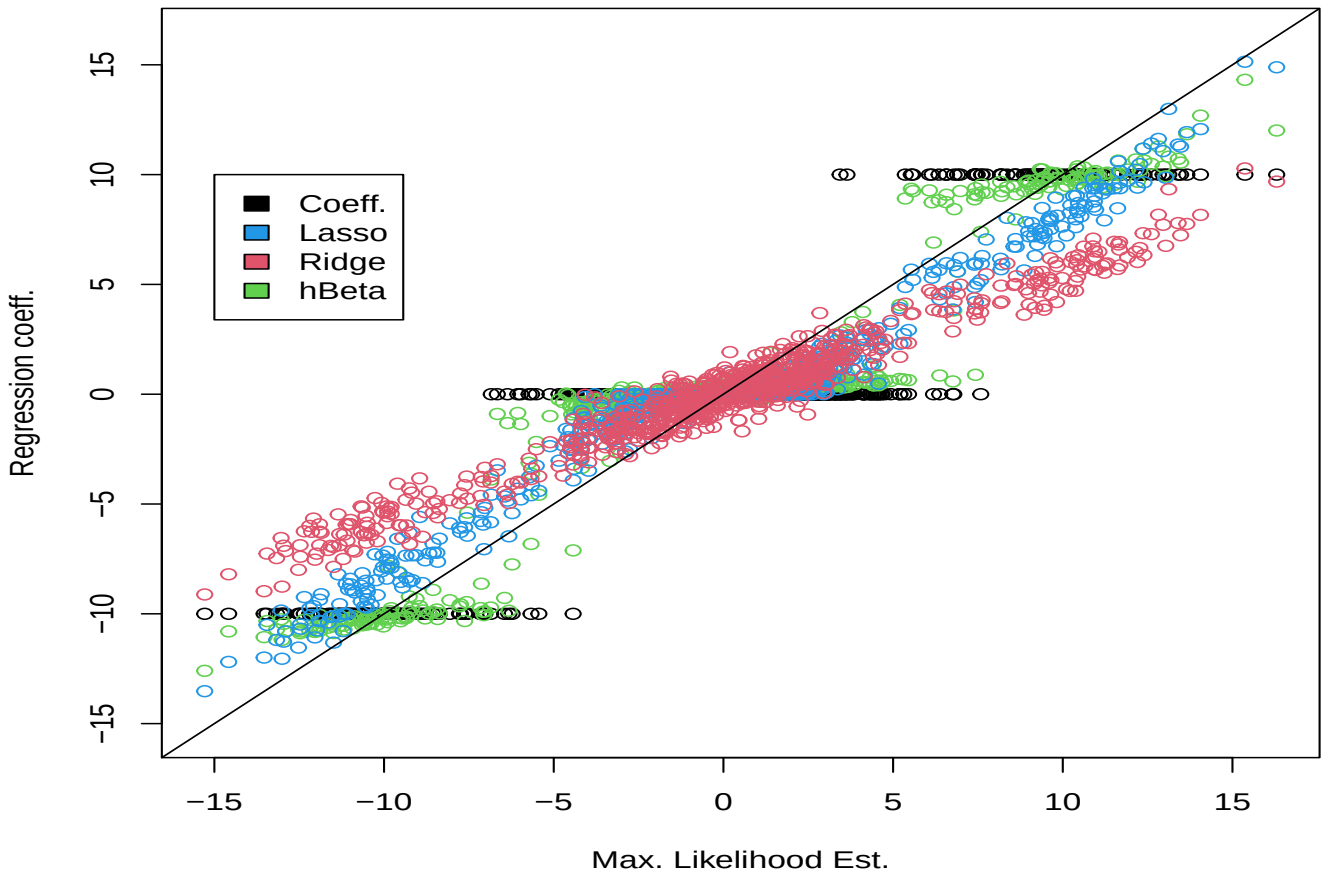
Regression coefficient eCDF's

Oracle , hBayes FPT median and 95% CI, MLE



Shrinkage plot – MLE on X axis

sqrt-MSE's: MLE 2.34, Ridge 2.56, Lasso 1.69, Gibbs 0.999, Oracle 0.996 (not shown)



Discussion

- The idea of using Bayes rules for minimizing frequentist risk for compound decision problems was suggested by Herbert Robbins in the early 1950's.
- Weinstein '21 generalizes this approach to FDR control and selective inference for permutation invariant problems.
- Our modeling approach aims at providing optimal results for fixed unknown $\vec{H}$ in large statistical problems: GLM, Graph community detection, NN ....
- Application of the hBayes approach produces posterior samples which specify $\Theta(\vec{H})$ that may be expressed in terms of a distribution – therefore we are essentially performing eBayes. However, the distribution we are considering is a summary of the fixed unknown $\vec{H}$, not a posited population distribution.

Research project funded by the BSF

"Hierarchical Bayes framework for Binary Expansion Statistic"

- Explore binary expansion statistics and inference methods to assess complex dependencies in large-scale datasets with a hierarchical Bayes modeling approach, whose basis is a mixture of finite Polya trees corresponding to multiple dyadic partitions of the p-dimensional unit cube.
- Challenges: (1) Develop a comprehensive binary expansion testing framework based on the hBayes approach. (2) Addressing estimation challenges within the binary expansion. (3) Reexamining the foundations of modeling and inference through Bayesian, fiducial, and frequentist perspectives.

– THANKS –