Series of lectures on Bayesian selective inference
Lecture 2: Bayesian FDR controlling testing procedure

Daniel Yekutieli

Statistics and OR
Tel Aviv University

Spring School of Research Unit "Structural Inference in Statistics"
17-21 March 2014, Konigs Wusterhausen

Quick review

Our goal is to provide selective inference: (a) making correct statistical discoveries (b) providing valid inference for our discoveries

Frequentist perspective:

1. BH procedure correctly discovers non-null effects and classifies sign of effects
2. FCR control a frequentist mechanism for constructing valid marginal CI's for selected parameters

Bayesian perspective (i.e. two group model):

1. Derived the Bayes classifier (test statistic = local FDR)
2. Two group model applies for a randomly selected selected component
3. Bayesian FDR is controlled by eBayes
4. BH can be expressed as eBayes classifier whose statistic is the p-value

Replicability in multiple GWAS – work with Ruth Heller

Genome-wide Association Studies try to identify genetic variants that are associated with a given phenotype.

- Replicability analysis aims to discover associations between SNP and phenotype that are present in more than one of the studies (i.e. for each SNP, test null hypothesis that the SNP is associated with the phenotype in 1 or less studies)
- Meta-analysis combines several GWAS for increased power to discover genetic variants that are associated in at least one study (i.e. for each SNP, test null hypothesis that the SNP is associated with the phenotype in 0 studies)

Kraft, Zeggini and Ioannidis ’09 effects in GWAS may be as small as population genetic biases, important to see associations in several studies conducted using a similar, but not identical, study base.

Analyses of Type 2 diabetes GWAS

Data from 6 GWAS testing association with T2D, same 2.5×10^6 SNPs in each study.

- Frequentist FDR analysis (Benjamini, Heller and Yekutieli '09)
  1. Compute p-value for each SNP to test (1) no association (2) no-replication
  2. Apply BH procedure at level 0.05 to each set of 2.5M p-values
  3. Results: 466 associated SNP, replicated associations for 113 SNP in 5 genomic regions
- Bayesian FDR analysis (Heller and Yekutieli '13)
  1. eBayes level 0.05 FDR controlling approach for testing (1) no association (2) no-replication
  2. Results: 803 associated SNP, replicated associations for 219 SNP in 17 genomic regions

Surprise: Bayesian FDR procedure usually don't offer considerably more power than the BH procedure!

Is it real?

Extensive simulation:

- Bayesian FDR procedure has more power than BH procedure for discovering associations, and considerably more (7-15 fold) power for discovering replicated associations!
- Bayesian FDR procedure controls the FDR at nominal level (simulation mean FDP = 0.05) for large studies, slightly under-conservative (simulation mean FDP = 0.07) for smaller studies.
- BH procedure slightly over-conservative (simulation mean FDP = 0.04) for testing no association, highly over-conservative (simulation mean FDP < 0.001) for testing no replication.

Plan

1. eBayes replicability analysis
2. GWAS analysis results
3. Why is the eBayes proc much more power than BH?
4. Illustrate on simulated data

Bayesian FDR replicability analysis

Notations

- SNP's are indexed by $j = 1 \cdots M$ $(= 2.5 \times 10^6)$
- Studies are indexed by $i = 1 \cdots n$ $(= 6)$
- The Parameter for SNP j is the association status $\vec{H}_j = (H_{1j} \cdots H_{nj})$ with $H_{ij} \in \{-1, 0, 1\}$
- The observation vector for SNP j is $\vec{Z}_j = (Z_{1j} \cdots Z_{nj})$ where Z_{ij} is log-OR z-score for testing no association between SNP j and T2D in Study i.

Hypotheses of interest for n studies

- $\mathcal{H} = \{\vec{h} = (h_1, \ldots, h_n) : h_i \in \{-1, 0, 1\}\}$
- The null hypotheses we test correspond to $\mathcal{H}^0 \subseteq \mathcal{H}$:

1. H^0_{NA} is the no association null hypothesis that the SNP is not associated with the phenotype in any of the studies that corresponds to

$$\mathcal{H}^0_{NA} = \{(0, 0, \cdots, 0)\}$$

2. H^0_{NR} is the no replicability null hypothesis that the SNP is positively and negatively associated with the phenotype in at most one study that corresponds to

$$\mathcal{H}^0_{NR} = \{\vec{h} : \ \#(h_i = -1) \leq 1 \ \cap \ \ \#(h_i = 1) \leq 1\}$$

Generalization of the two-group model

- $\Pr(\vec{H}_j = \vec{h}) = \pi(\vec{h})$ for $\vec{h} \in \mathcal{H}$.
- Conditional on the association status $\vec{H}_j = \vec{h}$,

$$f(\vec{z}_j|\vec{H}_j = \vec{h}) = \prod_{i=1}^{n} f_{i,h_i}(z_{ij})$$

  with $f_{i,-1}(z), f_{i,-1}(z)$ and $f_{i,-1}(z)$ the marginal z-score density in study i for SNP's that are negatively dependent, independent and positively dependent with T2D
- The marginal (mixture) density is

$$f(\vec{z}_j) = \sum_{\vec{h}\in\mathcal{H}} \pi(\vec{h}) \cdot f(\vec{z}_j|\vec{H} = \vec{h})$$

The Bayes FDR for n studies

- For $\mathcal{H}^0 \subset \mathcal{H}$, the *local Bayes FDR* for $\vec{z}_j$ is

$$
\begin{aligned}
fdr_{\mathcal{H}^0}(\vec{z}_j) &= Pr(\vec{H}_j \in \mathcal{H}^0 | \vec{z}_j) = \sum_{\vec{h} \in \mathcal{H}^0} Pr(\vec{H}_j = \vec{h} | \vec{z}_j) \\
&= \sum_{\vec{h} \in \mathcal{H}^0} \frac{\pi(\vec{h}) \cdot f(\vec{z}_j | \vec{H} = \vec{h})}{f(\vec{z}_j)}
\end{aligned}
$$

- The *Bayes FDR* for subset $\mathcal{Z} \subseteq R^n$ is

$$
Fdr_{\mathcal{H}^0}(\mathcal{Z}) = Pr(\vec{H}_j \in \mathcal{H}^0 | \vec{z}_j \in \mathcal{Z}) = E_f(fdr_{\mathcal{H}^0}(\vec{z}_j) | \vec{z}_j \in \mathcal{Z}).
$$

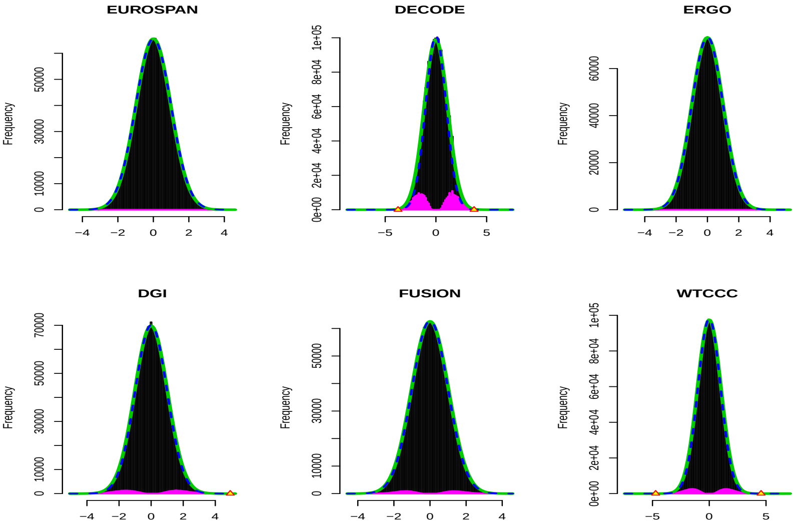
- The optimal rejection region among all possible rejection regions that are constrained to have a Bayes FDR of at most level q, is

$$
\mathcal{Z}_{OR,\mathcal{H}^0} = \{\vec{z} : fdr_{\mathcal{H}^0}(\vec{z}) \leq \delta(q)\}
$$

Empirical Bayes approach

1. For each study use locfdr to estimate the z-score densities
2. Use EM algorithm to find MLE for π
3. Compute local fdr's for each SNP
4. Use local fdr's to construct tests no-association and no-replicability

locfdr plots



The composite likelihood

- Given the marginal z-score densities we can compute the likelihood for SNP j

$$L(\vec{\pi};\vec{z}_j,f) = \Pr(\vec{z}_j|\ \vec{\pi}) = \sum_{\vec{h}\in\mathcal{H}} \pi(\vec{h})\cdot f(\vec{z}_j|\vec{H}=\vec{h})$$

- Note that to compute the complete likelihood we need to know the joint distribution of $(\vec{H}_1\cdots\vec{H}_M)$ and the joint distribution of $(\vec{Z}_1\cdots\vec{Z}_M)$ given $(\vec{H}_1\cdots\vec{H}_M)$
- Instead we consider the composite likelihood that have similar MLE in large problems with local dependencies

$$L^{CL}(\vec{\pi};\vec{z},f) = \Pr(\vec{z}_1\cdot\vec{z}_M|\ \vec{\pi}) = \prod_{j=1}^{M} L(\vec{\pi};\vec{z}_j,f)$$

- We use EM the find MLE for $\vec{\pi}$

eBayes testing procedure

- The local FDR is

$$\widehat{fdr}_{\mathcal{H}^0}(\vec{z}_j) = \sum_{\vec{h} \in \mathcal{H}^0} \hat{\pi}(\vec{h}) \prod_{i=1}^{n} \hat{f}_{i,h_i}(z_{ij}) / \hat{f}(\vec{z}_j)$$

- The Bayes FDR for rejection region Γ is

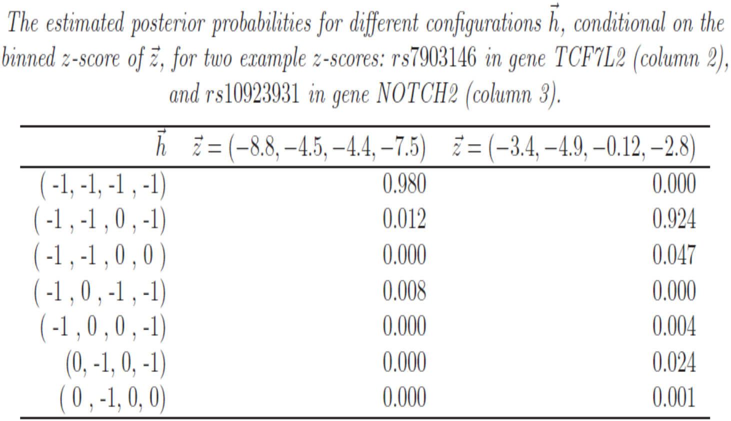
$$\widehat{Fdr}_{\mathcal{H}^0}(\Gamma) = \frac{\sum_{k:\vec{z}_k \in \Gamma} \widehat{fdr}_{\mathcal{H}^0}(\vec{z}_k)}{\#\{k : \vec{z}_k \in \mathcal{Z}\}}$$

- The eBayes optimal rejection region is

$$\Gamma_q = \{\vec{z}_j : \widehat{fdr}_{\mathcal{H}_0}(\vec{z}_j) \leq \hat{\delta}(q)\ \}$$

where $\hat{\delta}(q)$ is the largest threshold for which $\widehat{Fdr}_{\mathcal{H}^0}(\Gamma) \leq q$

Posterior configuration probabilities for two SNPs



The estimated posterior probabilities for different configurations $\vec{h}$, conditional on the binned z-score of $\vec{z}$, for two example z-scores: rs7903146 in gene TCF7L2 (column 2), and rs10923931 in gene NOTCH2 (column 3).

$\vec{h}$	$\vec{z} = (-8.8, -4.5, -4.4, -7.5)$	$\vec{z} = (-3.4, -4.9, -0.12, -2.8)$
(-1, -1, -1 , -1)	0.980	0.000
(-1 , -1 , 0 , -1)	0.012	0.924
(-1 , -1 , 0 , 0)	0.000	0.047
(-1 , 0 , -1 , -1)	0.008	0.000
(-1 , 0 , 0 , -1)	0.000	0.004
(0, -1, 0, -1)	0.000	0.024
(0 , -1, 0, 0)	0.000	0.001

Analysis results

For the SNPs with strongest evidence towards replicability in 17 distinct regions discovered by the empirical Bayes replicability analysis: the estimated Bayes FDR for replicability and for association (column 5-6); the adjusted p-values from the analysis of BHY09 for replicability and for association (column 7-8).

				Empirical Bayes Fdr		BHY09 adjusted p-values	
	chr	pos	gene	Replicability	Association	Replicability	Association
rs7903146	10	114758349	TCF7L2	2.40e-11	4.61e-22	0.00e+00	0.00e+00
rs10440833	6	20688121	CDKAL1	1.60e-05	8.06e-08	9.06e-09	0.00e+00
rs5015480	10	94465559	non-coding	1.10e-03	7.74e-05	8.78e-04	1.12e-07
rs4402960	3	185511687	IGF2BP2	3.14e-03	6.87e-04	0.0205	3.51e-05
rs5215	11	17408630	KCNJ11	8.91e-03	4.50e-03	1.00e+00	0.0236
rs757110	11	17418477	ABCC8	9.98e-03	6.16e-03	1.00e+00	0.0267
rs4933734	10	94414567	KIF11	0.0111	2.96e-04	1.00e+00	1.55e-05
rs10923931	1	120517959	NOTCH2	0.0134	2.70e-03	1.00e+00	3.45e-04
rs11187033	10	94262359	IDE	0.0189	2.07e-03	0.0186	7.07e-06
rs319602	5	134222164	TXNDC15	0.0202	7.07e-03	1.00e+00	0.0364
rs849134	7	28196222	JAZF1	0.0210	7.80e-03	9.84e-01	1.16e-03
rs6883047	5	134272055	PCBD2	0.0235	8.55e-03	1.00e+00	0.0471
rs10832778	11	17394073	B7H6	0.0282	0.0164	1.00e+00	1.53e-01
rs13070993	3	12217797	SYN2	0.0370	0.0235	1.00e+00	0.0369
rs10433537	3	12198485	TIMP4	0.0360	0.0233	1.00e+00	0.0386
rs10113282	8	96038252	C8orf38	0.0387	0.0102	1.00e+00	0.0408
rs1554522	17	25913172	KSR1	0.0436	0.0145	1.00e+00	2.13e-01

Why does BH have less power than Bayes classifier?

We define the $Fdr = q$ p-value based classifier: $R_i = I\{P_i \leq p(q)\}$ with $p(q)$ such that $Fdr(P_i \leq p(q)) \ = \ q$.

1. <u>The $Fdr = q$ p-value based classifier is suboptimal</u>

$$\Pr\{P_i \leq p(q)\} \ < \ \Pr\{fdr(Z_i) \leq \delta(q)\}$$

2. The BH procedure is $I\{P_i \leq \hat{p}(q)\}$, since $\hat{p}(q)$ is derived based on an overly conservative estimate of Fdr

$$\widehat{Fdr}(P_i \leq p) = \frac{p}{\#\{p_j \leq p\}/m} \ > \ Fdr(P_i \leq p)$$

therefore $\hat{p}(q) < p(q)$ and thus $\underline{\Pr(P_i \leq \hat{p}(q)) < \Pr(P_i \leq p(q))}$

Return to continuous parameter-value simulation

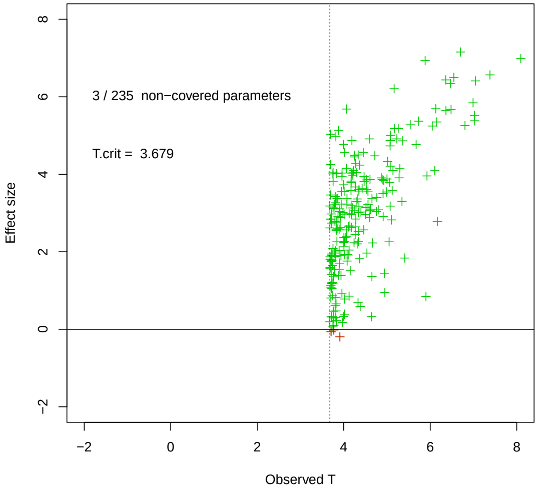
Generate $m = 10,000$ iid (θ_i, Y_i):

- Parameter $\theta_i \sim \pi(\theta_i)$ with

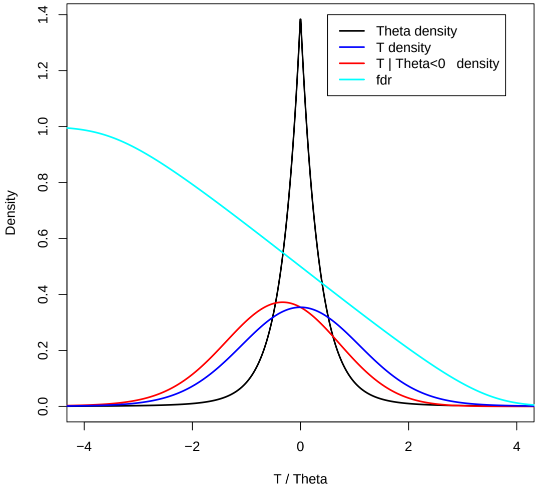
$$\pi(\theta_i) \ = \ 0.9 \cdot \frac{3 \cdot e^{-3 \cdot |\theta_i|}}{2} + 0.1 \cdot \frac{1 \cdot e^{-1 \cdot |\theta_i|}}{2} \tag{1}$$

- Observations $T_i \sim N(\theta_i, 1)$
- P-values $P_i = 1 - \Phi(|T_i|)$

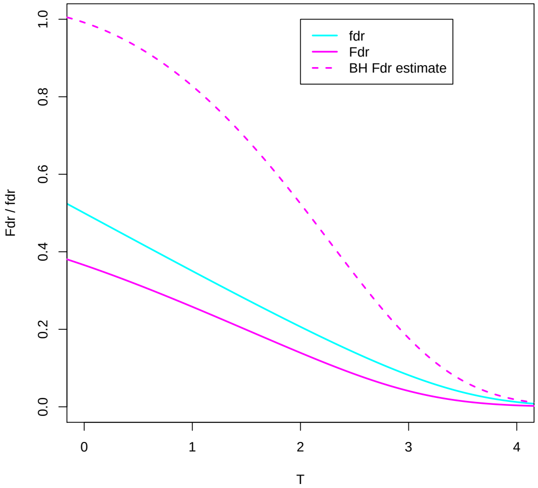
BH $q = 0.05$ results



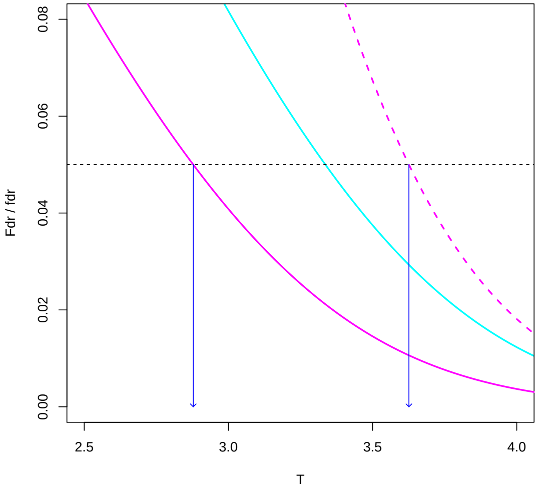
Theta and T densities and the local fdr



The Bayesian FDR and the BH eBayes estimate



Fdr = 0.05 testing procedure



Simplified 2 GWAS analysis simulation

- $\mathcal{H} = \{\ (0,0),(3,0),(0,3),(3,3)\}$
- $\pi(0,0) = 0.85, \pi(3,0) = 0.05, \pi(0,3) = 0.05, \pi(3,3) = 0.05$
- $Z_i = (Z_{i1}, Z_{i2})$ with $Z_{i1} \overset{iid}{\sim} N(h_1, 1)$ and $Z_{i2} \overset{iid}{\sim} N(h_2, 1)$

We consider two type of null sets :

1. No association
$$\mathcal{H}_0^{NA} = \{(0,0)\}$$
2. No replication
$$\mathcal{H}_0^{NR} = \{(0,0),(3,0),(0,3)\}$$

Computations

$$\begin{aligned} f(z_i) \;=\; & \phi(z_{i1}) \cdot \phi(z_{i2}) \cdot \pi(0,0) + \phi(z_{i1}-3) \cdot \phi(z_{i2}) \cdot \pi(3,0) \\ + \; & \phi(z_{i1}) \cdot \phi(z_{i2}-3) \cdot \pi(0,3) + \; \phi(z_{i1}-3) \cdot \phi(z_{i2}-3) \cdot \pi(3,3) \end{aligned}$$

- No association local fdr

$$fdr_{NA}(z_i) = \frac{\phi(z_{i1}) \cdot \phi(z_{i2}) \cdot \pi(0,0)}{f(z_i)}$$

- No replication local fdr

$$fdr_{NR}(z_i) = \frac{\sum_{h \in \mathcal{H}_0^{NR}} \; \phi(z_{i1}-h_1) \cdot \phi(z_{i2}-h_2) \cdot \pi(h_1,h_2)}{f(z_i)}$$

Computations (cont.)

- Marginal p-values $P_{i1} = 1 - \Phi(z_{i1}),\ P_{i2} = 1 - \Phi(z_{i2})$
- No association p-value

$$P_i^{NA} = 1 - F_{\chi^2_2}(-2 \cdot log(P_1) - 2 \cdot log(P_2))$$

- No replication p-value $P_i^{NR} = max(P_1, P_2)$

Over conservativeness of BH *Fdr* estimates

Recall, the BH procedure is based on

$$\widehat{Fdr}(P_i \leq p) = \frac{p}{\#(P_i \leq p)/m}$$

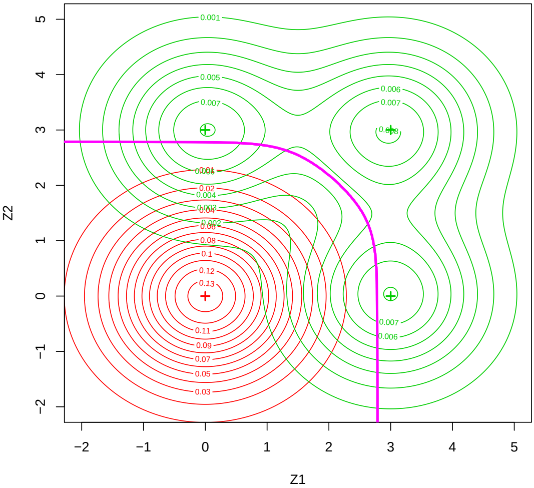
1. Actual *Fdr* value for testing no association

$$\begin{aligned}
\Pr(H_i \in \mathcal{H}_0^{NA} |\, P_i^{NA} \leq p) &= \frac{\Pr(P_i^{NA} \leq p | H_i = (0,0)) \cdot \Pr(H_i = (0,0))}{\Pr(P_i^{NA} \leq p)} \\
&\approx \frac{p}{\#(P_i^{NA} \leq p)/m} \cdot \pi(0,0)
\end{aligned}$$

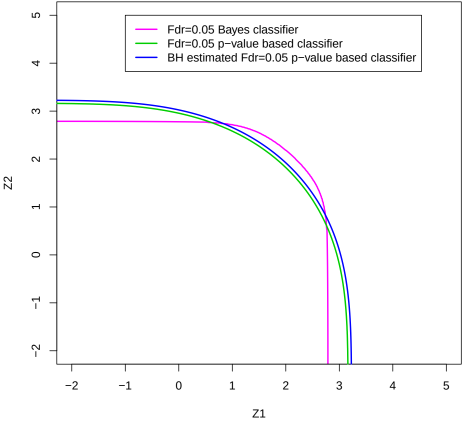
2. Actual *Fdr* value for testing no replication

$$\Pr(H_i \in \mathcal{H}_0^{NR} |\, P_i \leq p) = \frac{\sum_{h \in \mathcal{H}_0^{NR}} \Pr(P_i^{NR} \leq p | H_i = h) \cdot \pi(h)}{\Pr(P_i^{NR} \leq p)}$$

No association $Fdr = 0.05$ Bayes classifier



$Fdr = 0.05$ Bayes classifier and p-value based classifiers

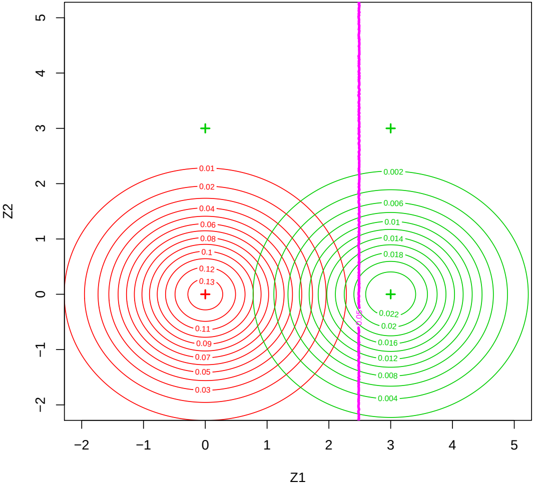


Power: 0.112, 0.108, 0.104

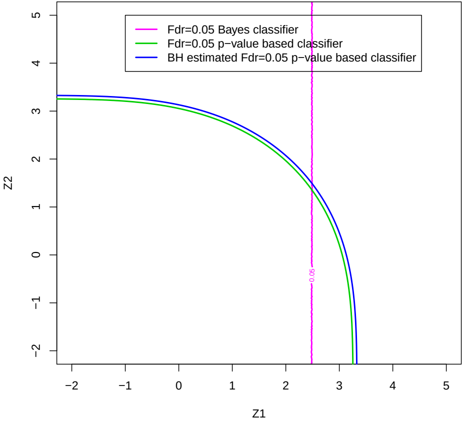
Change in hyper-parameter values

- $\mathcal{H} = \{\ (0,0), (3,0), (0,3), (3,3)\}$
- $\pi(0,0) = 0.85$
  $\underline{\pi(3,0) = 0, \pi(0,3) = 0.15}$
  $\underline{\ \pi(3,3) = 0}$
- $Z_i = (Z_{i1}, Z_{i2}),\ Z_{i1} \overset{iid}{\sim} N(h_1, 1)\ and\ \ Z_{i2} \overset{iid}{\sim} N(h_2, 1)$

No association $Fdr = 0.05$ Bayes classifier

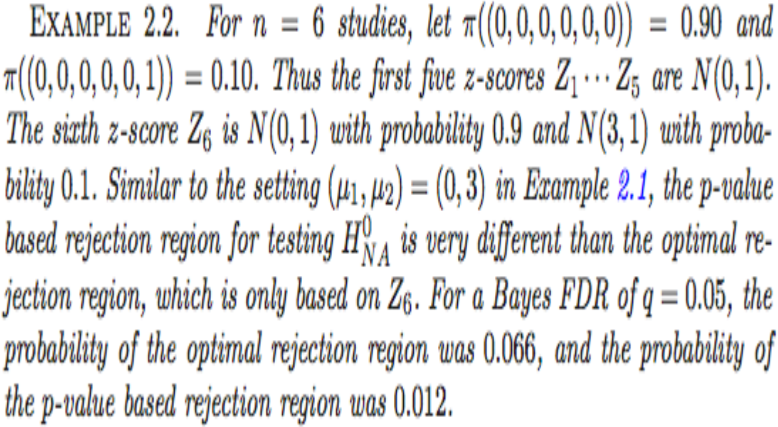


$Fdr = 0.05$ Bayes classifier and p-value based classifiers



Power: 0.110, 0.081, 0.076

Difference in power between for 6 studies



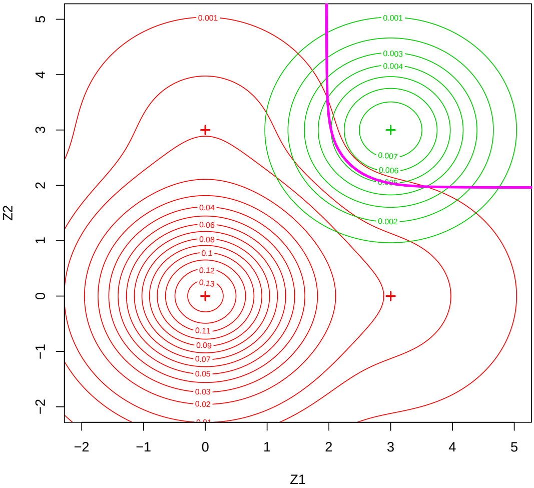
EXAMPLE 2.2. *For $n = 6$ studies, let $\pi((0,0,0,0,0,0)) = 0.90$ and $\pi((0,0,0,0,0,1)) = 0.10$. Thus the first five z-scores $Z_1 \cdots Z_5$ are $N(0,1)$. The sixth z-score Z_6 is $N(0,1)$ with probability 0.9 and $N(3,1)$ with probability 0.1. Similar to the setting $(\mu_1, \mu_2) = (0,3)$ in Example 2.1, the p-value based rejection region for testing H^0_{NA} is very different than the optimal rejection region, which is only based on Z_6. For a Bayes FDR of $q = 0.05$, the probability of the optimal rejection region was 0.066, and the probability of the p-value based rejection region was 0.012.*

Return to original hyper-parameter values

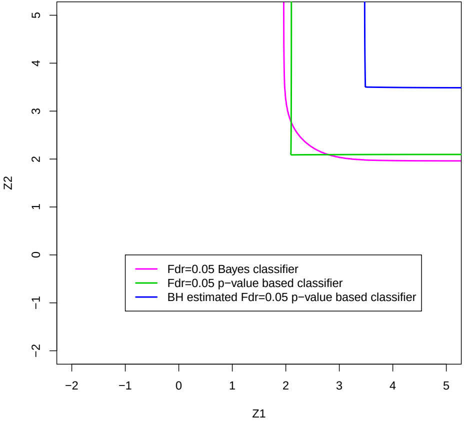
- $\mathcal{H} = \{\ (0,0), (3,0), (0,3), (3,3)\}$
- $\pi(0,0) = 0.85$
  $\underline{\pi(3,0) = 0.05, \pi(0,3) = 0.05}$,
  $\underline{\pi(3,3) = 0.05}$
- $Z_i = (Z_{i1}, Z_{i2}),\ Z_{i1} \overset{iid}{\sim} N(h_1, 1)\ \textit{and}\ \ Z_{i2} \overset{iid}{\sim} N(h_2, 1)$

However now we classify <u>no replication</u>

No replication $Fdr = 0.05$ Bayes classifier



Bayes classifier and p-value based classifiers



Power: 0.0359, 0.0351, 0.0049

Final comments

- For scalar Z_i and if $\mathcal{H}_0$ is a single point in the parameter space (i.e. simple null hypothesis) use BH procedure
- For high dimensional Z_i or non-simple $\mathcal{H}_0$ try deriving a Bayesian classifier Prior distribution $\pi(\boldsymbol{h})$ is the marginal distribution of H_i in data population
- R package: *repFDR*

A few references

- Benjamini Y., Heller R., Yekutieli D., (2009) “Selective Inference in Complex Research.” *JRSS A*, **267**, 1–17.
- Efron, B. (2010) “Large-Scale Inference.” Cambridge, United Kingdom.
- Heller R., Yekutiel D., (2013) “Replicability analysis for genome-wide association studies.” *arXiv* 1209.2829
- Kraft, P., Zeggini, E. and Ioannidis, J. (2009) “Replication in Genome-wide Association Studies.” *Statistical science*, **24** (4), 561 – 573.