New Procedures for Controlling FDR

– and some "interesting" issues

Jiayang Sun and Zhongfa Zhang
Department of Statistics
Case Western Reserve University
Cleveland, OH 44106

zxz8@cwru.edu
jiayang.sun@cwru.edu

Seventh Purdue International Symposium on Statistics,
19-24 June 2003

Outline

1. Introduction and Motivation
2. Mean and Variance of Sample FDR
3. Relationship between FDR and FWER
4. New Procedure - comparisons with BH and STS procedures, and microarray data application
5. Connection to Change Point Problem and Mixture Estimation

1. Introduction and Motivation

One of challenging problems in analyzing large data sets is multiple or simultaneous testing of a large number of hypotheses. Powerful multiple testing procedures are useful in

- finding disease genes or detecting diffferential gene expressions from microarray data;
- locating significant source activity in a brain from EEG or fMRI data; and
- analyzing effectiveness of a new biomedical engineering treatment in preventing pressure sores.

Multiple Testing Problem

Goal: Test m null hypotheses $H_1, \ldots H_m$.

Reality: m_0 of them are true, $m_1 = m - m_0$ are not.

Data: realization of test statistics $T_1, \ldots, T_m$, or corresponding p-values $p_1, \ldots, p_m$.

Outcome of a Test Procedure

counts	accept	reject	total
null true	U	V	m_0
alt true	T	S	m_1
total	m-R	R	m

where R is observable (so is $m - R$), but not V and T.

Errors and Power

Family-Wise Error Rate:

$$\begin{aligned} \text{FWER} &= P_{H_i' \text{s true}}(\text{ reject at least one of them}) \\ &= P(V > 0) \end{aligned}$$

counts	accept	reject	total
null true	U	V	m_0
alt true	T	S	m_1
total	m-R	R	m

False Discovery Rate	False Nondiscovery Rate
$\text{FDR} = E(\frac{V}{R} I_{\{R>0\}})$	$\text{FNR} = E(\frac{T}{m-R} I_{\{m-R>0\}})$
Benjamini and Hochberg (95)	Genovese and Wasserman (01)
	Sakar

The BH Procedure

Benjamini and Hochberg (95) proposed the BH procedure:

reject all H_i's with $p_i \leq p_{(k)}$

where $p_{(1)}, \ldots, p_{(m)}$ are sorted p-values and

$$k = \max\{i : \ \ 1 \leq i \leq m, \ p_{(i)} \leq \frac{i}{m}\alpha\}.$$

They showed that this procedure in the independent pvlaues case controls FDR at α:

$$FDR \leq \frac{m_0}{m}\alpha.$$

Generalizations: The bound holds for positively

dependent p-value case (Benjamini and Yekutieli, 01); equality holds for continuous case (Storey, Taylor and Siegmund, 02) using a beautiful martingale argument.

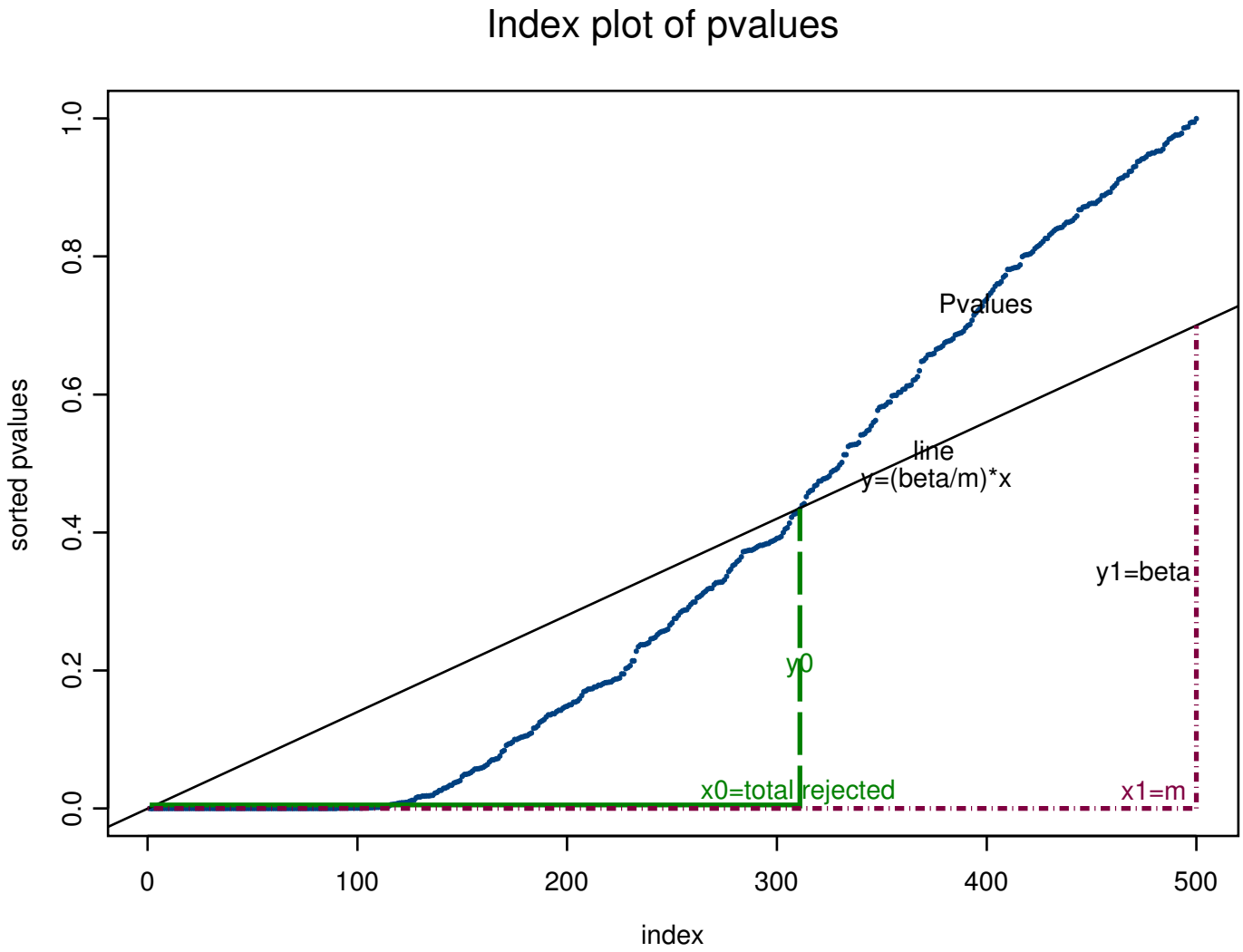
2. Mean and Variance of V/R

We have a different proof for the expected FDR:

$$E(\frac{V}{R}1_{\{R\geq 1\}}|p'_1,\ldots,p'_{m_1}) = \frac{m_0}{m}\alpha$$

and also the variance of sample FDR

$$\frac{m_0}{m(m-m_0+1)}\alpha(1-\alpha) \leq \mathrm{Var}(\frac{V}{R}1_{\{R\geq 1\}}) \leq \frac{m_0}{m}\alpha(1-\alpha)$$



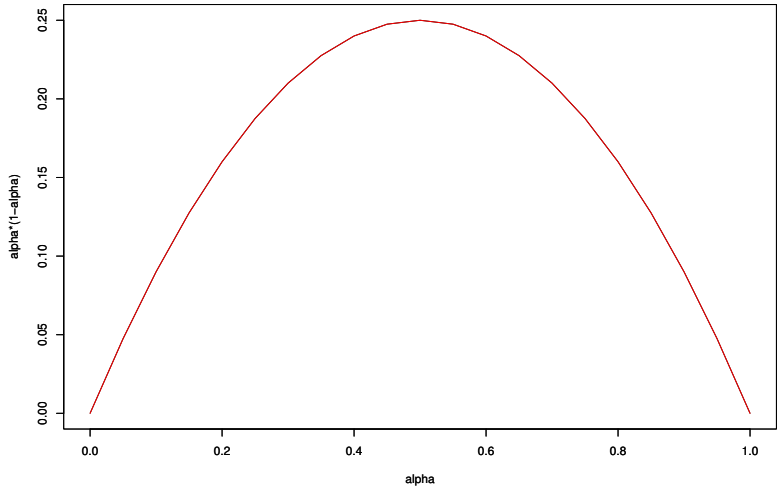
FDR:

$$\frac{Y_0}{X_0} = \frac{y_1}{x_1} \Rightarrow \frac{V}{R} \approx_9 \frac{Y_0 * m_0}{X_0} = \frac{\alpha * m_0}{m}$$

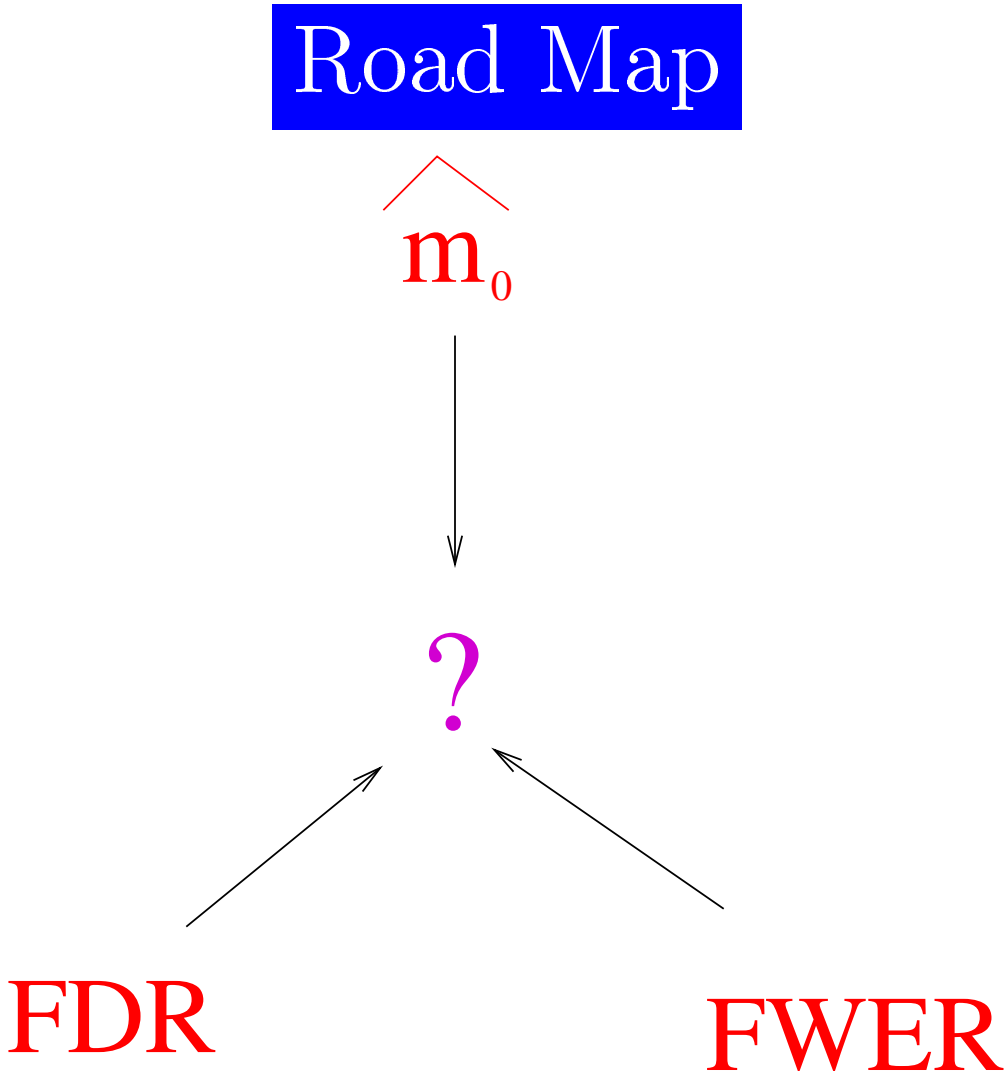
Variance of sample FDR:

$$\frac{m_0}{m(m-m_0+1)}\alpha(1-\alpha) \leq \mathrm{Var}\left(\frac{V}{R}1_{\{R\geq 1\}}\right) \leq \frac{m_0}{m}\alpha(1-\alpha)$$

Left: $p_i' = 0$, Right: p_i' is uniform.



(implication on page 24 later)


Road Map
$\hat{m}_0$
?
FDR
FWER

3. Relationship between FDR and FWER

$$\boxed{FDR = E(\frac{V}{R} I_{\{R>0\}}) \leq P(V \geq 1) = FWER}$$

where the equality holds if $m_0 = m$.

If T controls FDR at α and S controls FWER at α, then T controls FWER at $\alpha' \geq \alpha$; thus, generally, T is more powerful than S.

Q: Can we find a functional relationship between FDR and FWER:

$$FDR = g(FWER, f, m_0), \quad \text{where } f \text{ is alt. dis.?}$$

If yes, can borrow our expertise in building simultaneous confidence bands.

Generic Multiple Testing Procedures

Consider *generic* multiple tests of the form: reject all these H_i' that satisfies "$p_i \leq c$" where c is a critical value that can be fixed or estimated.

The BH (Benjamini and Hochberg, 95) procedure is a generic procedure:

$$c = p_{(k)}.$$

Other examples of generic test procedures include:

	pw	Bonferroni	fixed	first r	reg classifier
c	α	α/m	t	$p_{(r)}$	$\sup\{t : \hat{P}\{H_1 = 1 \mid p_1 = t\} > 0.5\}$

Back to the Relationship

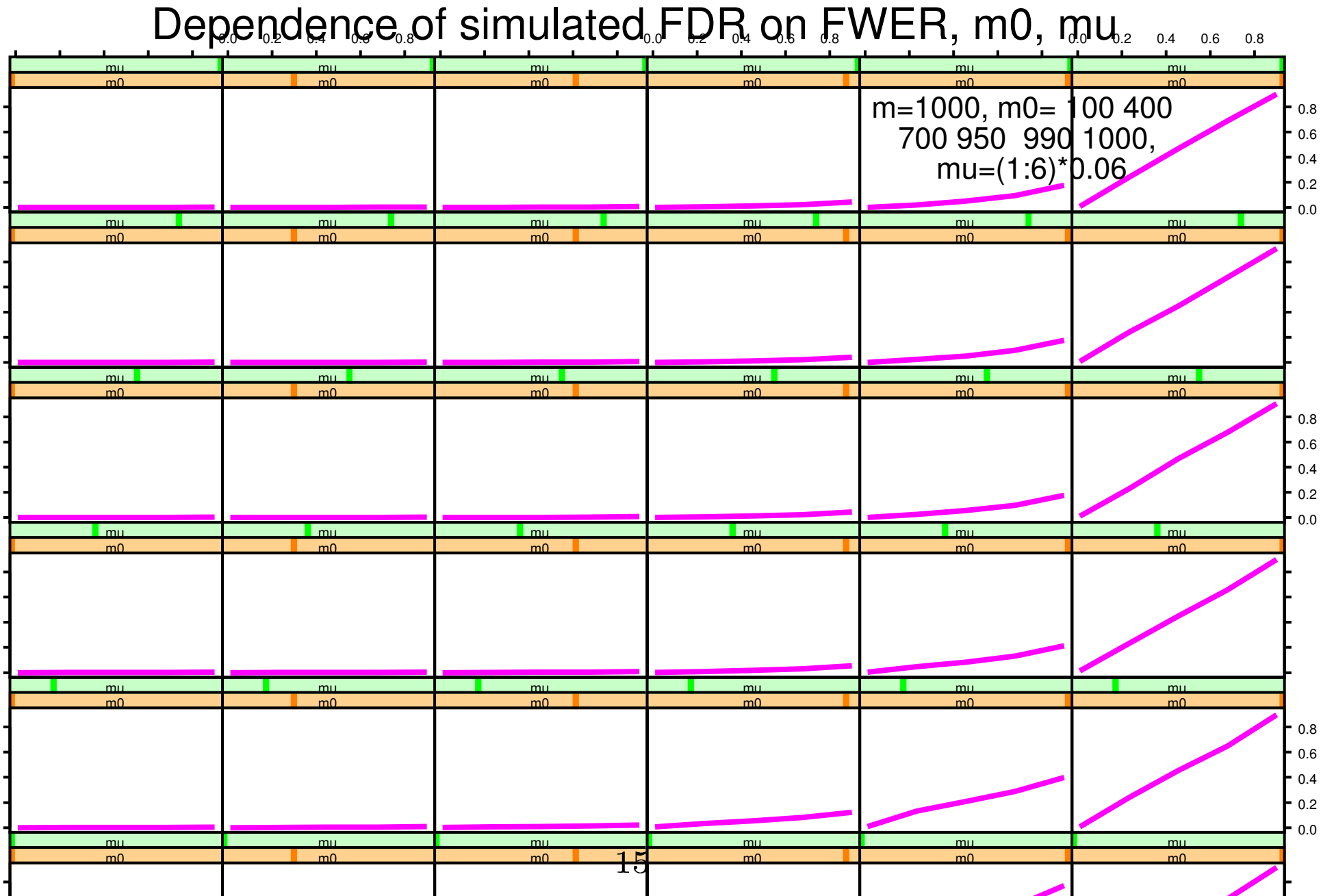
Consider the following independent tests case:

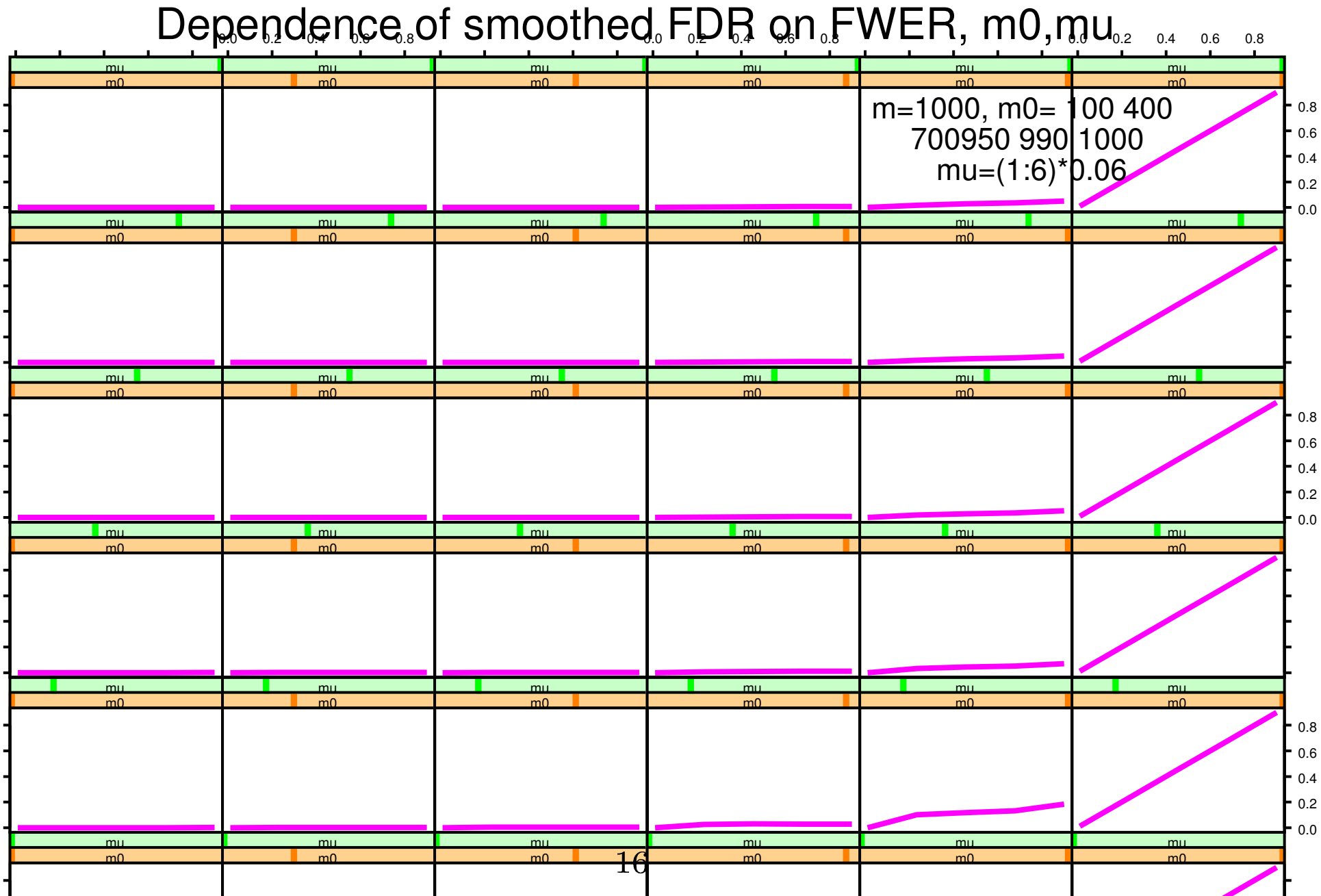
$P_1, \ldots, P_{m_0} \overset{iid}{\sim} \mathcal{U}(0,1)$, $P'_1, \ldots, P'_{m-m_0} \overset{iid}{\sim} \mathcal{F}$

$$FWER(c) = P\{\max_{i=1,\ldots,m_0} T_i > c\} = 1 - P(t_{n-1} \leq c)^{m_0} = \alpha$$

$$FDR(c) = E(\frac{V}{V+S} 1_{\{V \geq 1\}})$$

where $V \sim B(m_0, p_0)$ and $S \sim B(m_0, p_1)$ are indp with $p_0 = 1 - (1-\alpha)^{1/m_0}, p_1 = 1 - p(t_{n-1} \leq c - \sqrt{n}\mu)$.


Dependence of simulated FDR on FWER, m0, mu
m=1000, m0= 100 400
700 950 990 1000,
mu=(1:6)*0.06
mu
m0
0.0
0.2
0.4
0.6
0.8


Dependence of smoothed FDR on FWER, m0,mu
m=1000, m0= 100 400
700950 990 1000
mu=(1:6)*0.06
mu
m0
0.0
0.2
0.4
0.6
0.8

New Procedures

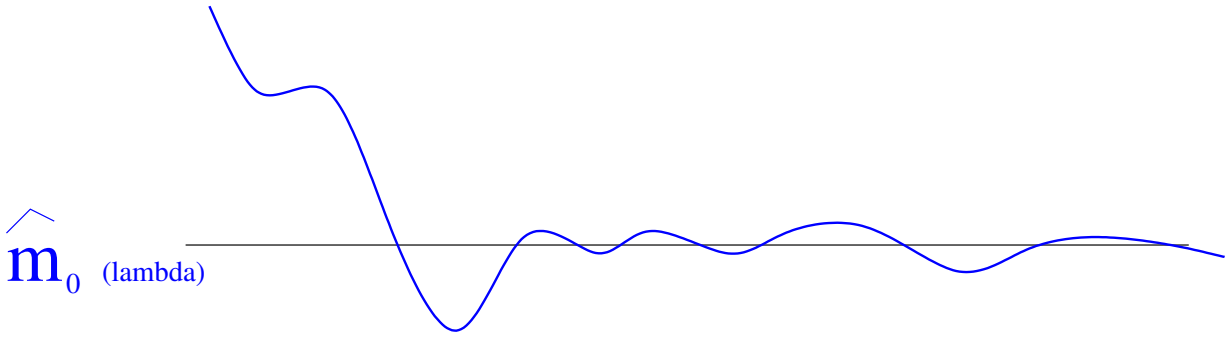
Idea: Estimate m_0 in the first step and then use the BH procedure with $\alpha' = \frac{m}{\hat{m}_0}\alpha$.

- Benjamini, Kriegar and Yekutieli (01): In the 2nd stage, use $k = \max\{i : p_{(i)} \leq \frac{i}{m-\hat{k}_1}\alpha\}$. *(Compare with p6)*

- Storey et al (02): In the 2nd stage, use $k = \max\{i : p_{(i)} \leq \frac{i}{\hat{m}_0}\alpha\}$

where $\hat{m}_0 = \hat{m}_0(\lambda^*)$ minimizes an (over estimated) MSE (by bootstrap) of

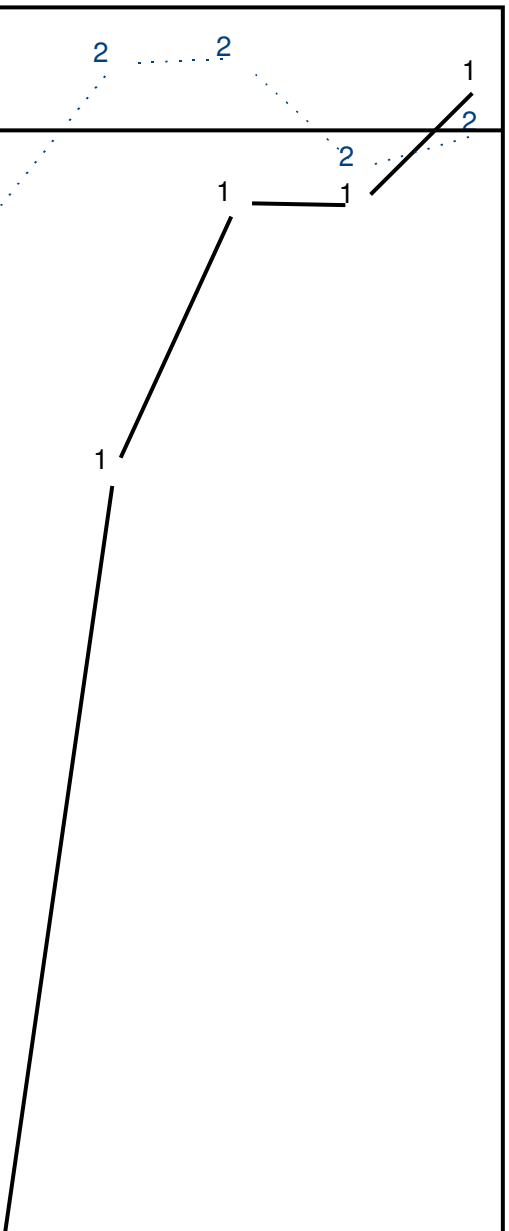
$$\hat{m}_0(\lambda) = \frac{\sum_{i=1}^{m} 1_{\{P_i \geq \lambda\}}}{1-\lambda} \quad (\approx \frac{m_0(1-\lambda)}{1-\lambda})$$



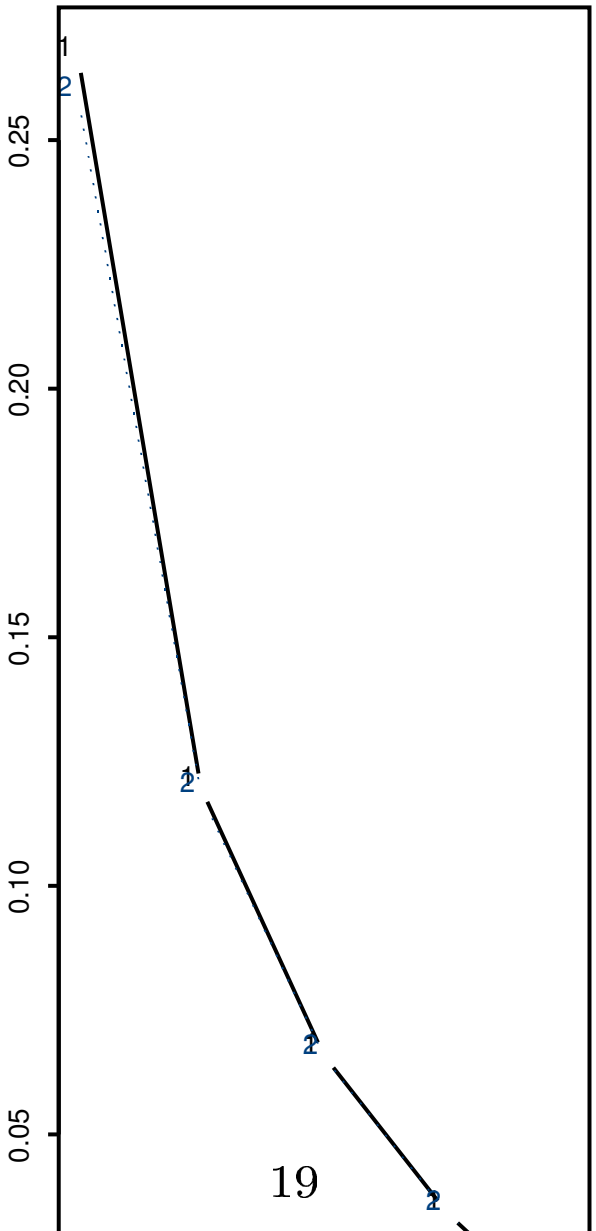
- Our new $\hat{m}_0$ is an appropriate average of $\hat{m}_0(\lambda_i)$'s.

Computationally faster, smaller variance!

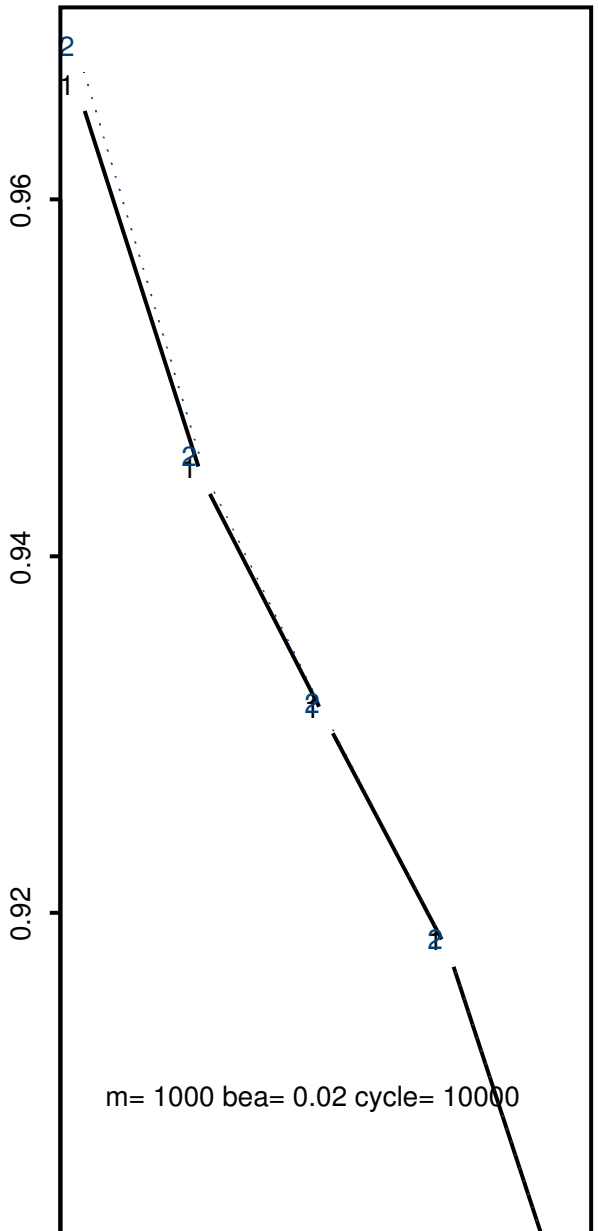
veragefdr by 1.storey,2 PP



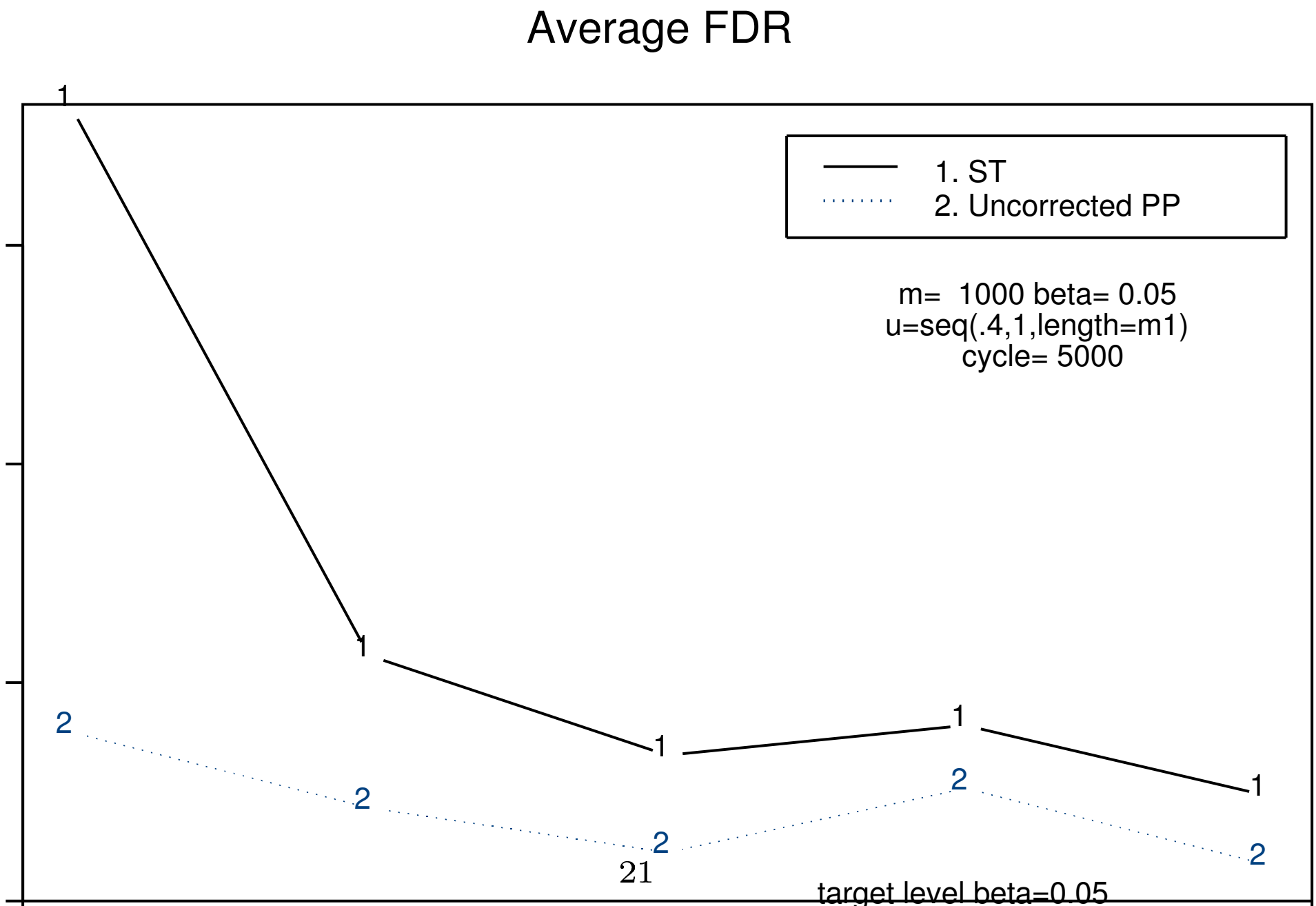
Average fnr by 1.storey 2 PP



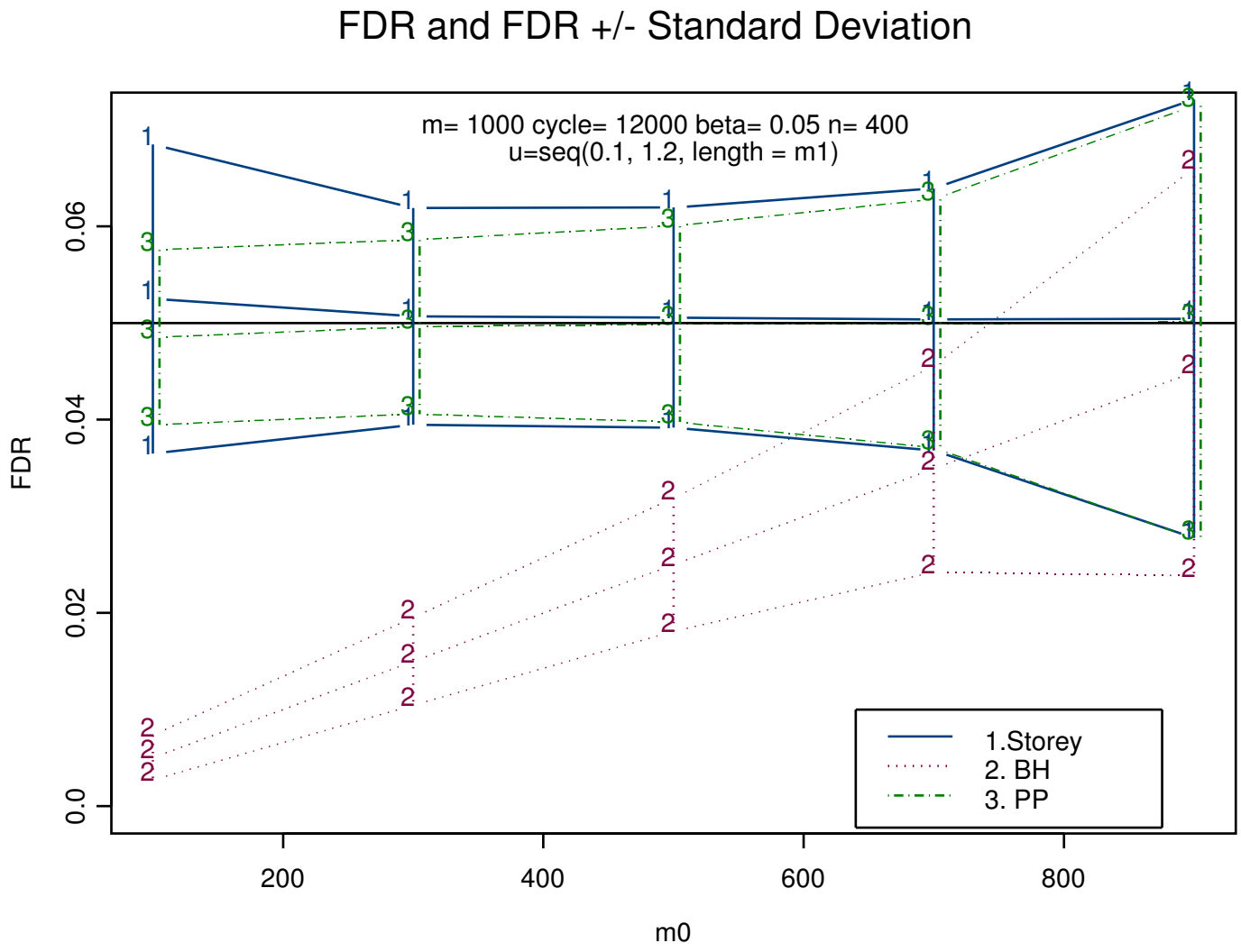
Average Power by 1 storey 2 PP



Notes: Average FDR FNR, POWER by STS', Corrected PP procedures. 10000 replications were used, and number of tests is m=1000. Correction term was put in with $\delta = 0.035$.


Average FDR
1. ST
2. Uncorrected PP
m= 1000 beta= 0.05
u=seq(.4,1,length=m1)
cycle= 5000
1
2
21
target level beta=0.05

Notes: STS' procedure is conservative when signals are moderately strong. Average FDR's by STS' and our uncorrected PP procedures were plotted. 5000 replications were used and number of tests is m=1000.


FDR and FDR +/- Standard Deviation
m= 1000 cycle= 12000 beta= 0.05 n= 400
u=seq(0.1, 1.2, length = m1)
FDR
0.0
0.02
0.04
0.06
200
400
600
800
m0
1.Storey
2. BH
3. PP

Notes: Variance comparison of false discovery ratio of 3 procedures. Averaged FDR's by STS', PP, BH's procedures were plotted together with curves of one standard deviation added on the plot. 12000 replications were used for average, and $m = 1000$. (Back to page 10)

Microarray data application

We applied our procedure to the training data set of leukemia bone marrow:

`www-genome.wi.mit.edu/mpr/publications/projects/Leukemia/Files_descriptions.txt`

`www-genome.wi.mit.edu/mpr/publications/projects/Leukemia/Golub_et_al_1999.pdf`

There are $m = 7129$ genes and $n = 38$ samples, 27 of them come from AML (acute myelocytic leukemia) patients while the remaining 11 come from ALL (acute lymphoblastic leukemia) patients.

Goal: construct a classification rule based on the

expression training data, then to use the rule to predict a future sample's class.

Key: build a classification rule based on preselected genes that are most significant to the class differentiation.

Now we will try our method to find these genes by our FDR controlling procedure.

Several steps are involved in this procedure:

1. calculate the statistics T_i for each gene i by:

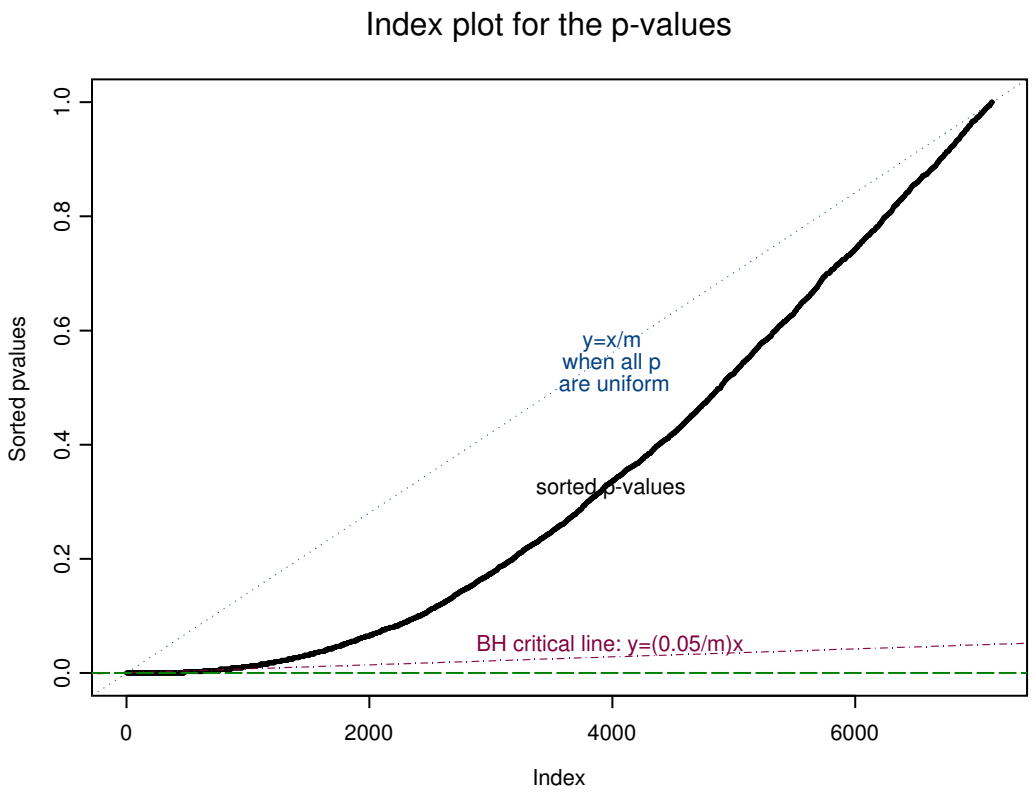
$$T_i = \frac{\bar{X}_{i,1} - \bar{X}_{i,0}}{\sqrt{S_{i,1}^2/n_1 + S_{i,0}^2/n_0}};$$

2. find the null distribution for each T_i by randomly permuting the labels of samples repeatedly B times and calculate the statistic T_i^b with the same manner for $b = 1, 2, \cdots, B$, and compute the p-value p_i by

$$p_i = \frac{\sum_1^B 1_{\{|T_i^b| \geq |T_i|\}}}{B};$$

3. feed these p-values into 3 procedures: ours, STS' and BH's, to find out the most significant genes.

We made a total $B = 450$ permutations, plotted the sorted p-values against index,



We see that the signal does not seem to be strong. Now Apply the three procedures at FDR level $\alpha = 0.02$, we find out that there are 637 genes rejected for our

proposed procedure and STS' procedure while the direct application of BH procedure will only reject 476 genes.

If the above process was running at different levels α, the total numbers of rejected hypotheses by running these 3 procedures are summarized in the following table.

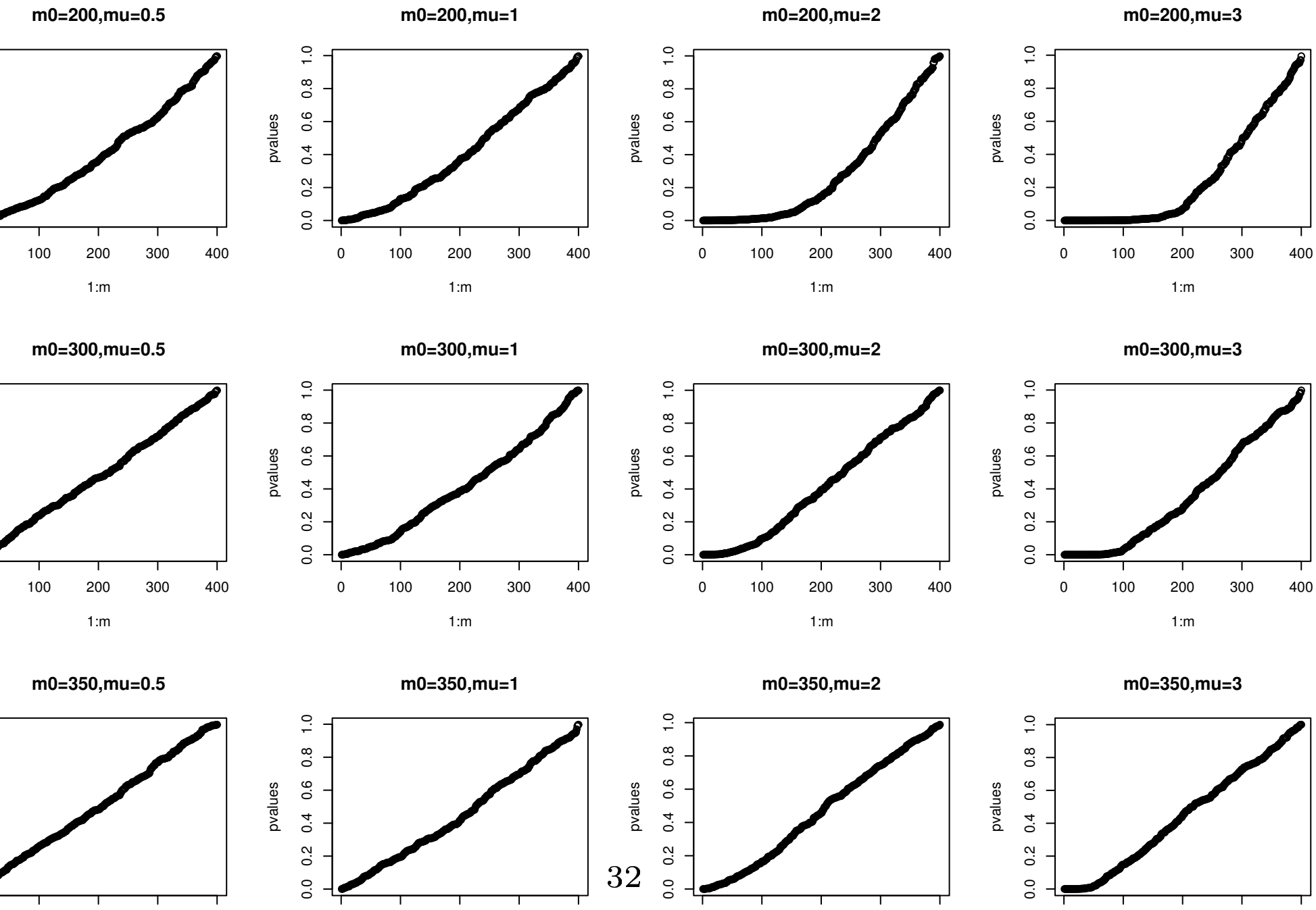
Table 1: Total Number of Rejected Hypotheses on AML-ALL cancer data set

β	ST	PP	BH
0.01	476	476	476
0.02	637	637	476
0.03	767	767	637
0.04	862	862	637
0.05	1024	1024	767

Connections and Discussions

Mixtures: $\pi U + (1 - \pi)F$

Change point analyses: segment regression; one with a reflection point; change of fitted gradient, all going backwards from top.


m0=200,mu=0.5
m0=200,mu=1
m0=200,mu=2
m0=200,mu=3
m0=300,mu=0.5
m0=300,mu=1
m0=300,mu=2
m0=300,mu=3
m0=350,mu=0.5
m0=350,mu=1
m0=350,mu=2
m0=350,mu=3
pvalues
1:m
0.0
0.2
0.4
0.6
0.8
1.0
0
100
200
300
400

Discussion: boundary, correlation, criterion, ...

References

[Benjamini, Hochberg(1991)] Benjamini, Hochberge, (1995) *Controlling the false discovery rate: A practical and powerful approach to multiple testing.* JSTOR-B, 1995(57)

[Benjamini, Yekutieli(2001)] Benjamini, Yekutieli(2001). *The Control of the false discovery rate in multiple testing under dependency.* Tech Report, School of Mathematical Sciences, Tel Aviv University. April 27, 2001. Available at http://www.math.tau.ac.il/ benja/

[Storey JD. (2002)] Storey John A direct approach to false discovery rates. Journal of the Royal Statistical Society, Series B, 64: 479-498

[Genovese Christopher, Wasserman, L] enovese, C. R. and Wasserman, L. (2002). Operating Characteristics and Ex-

tensions of the False Discovery Rate Procedure, J. Royal Statist. Soc. B, 64, 499–518.

[Storey JD, Taylor JE, and Siegmund D] Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. Written in June 2002. Journal of the Royal Statistical Society, Series B,

[T. R. Golub et al] olecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring SCIENCE VOL 286 15 OCTOBER 1999.