

# Estimation of the False Discovery Rate

USSES Conference, June, 2006  
Lexington, KY

Jason A. Osborne, [osborne@stat.ncsu.edu](mailto:osborne@stat.ncsu.edu)  
Department of Statistics, North Carolina State University

(These slides available at [www4.stat.ncsu.edu/~jaosborn/research/microarray/usses/index.html](http://www4.stat.ncsu.edu/~jaosborn/research/microarray/usses/index.html).)

## Outline

- A discussion of the False Discovery Rate, FDR.
- Three approaches:
  - Benjamini-Hochberg (1995) control of FDR.
  - Storey's (2002) `qvalue()` package from Bioconductor.
  - Allison's (2002) posterior probability of a false lead.
- SAS code for all three approaches
- Illustration with timecourse experiment to investigate fungal pathogen in Norway Spruce.

## References

- Allison, D.B. et al (2002) A mixture model approach for the analysis of microarray gene expression data. *Comp. Stat. & Data Analysis*, **39**:1-20.
- Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *JRSSB*, **57**: 289-300.
- Osborne, J.A. (2006) Estimating the False Discovery Rate with SAS. *Proceedings of the Thirty-first Annual SUGI Conference*. Cary, NC: SAS Institute Inc.
- Storey J.D. (2002) A direct approach to false discovery rates. *JRSSB*, **64**: 479-498
- Storey JD, et al (2004) Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *JRSSB*, **66**: 187-205.

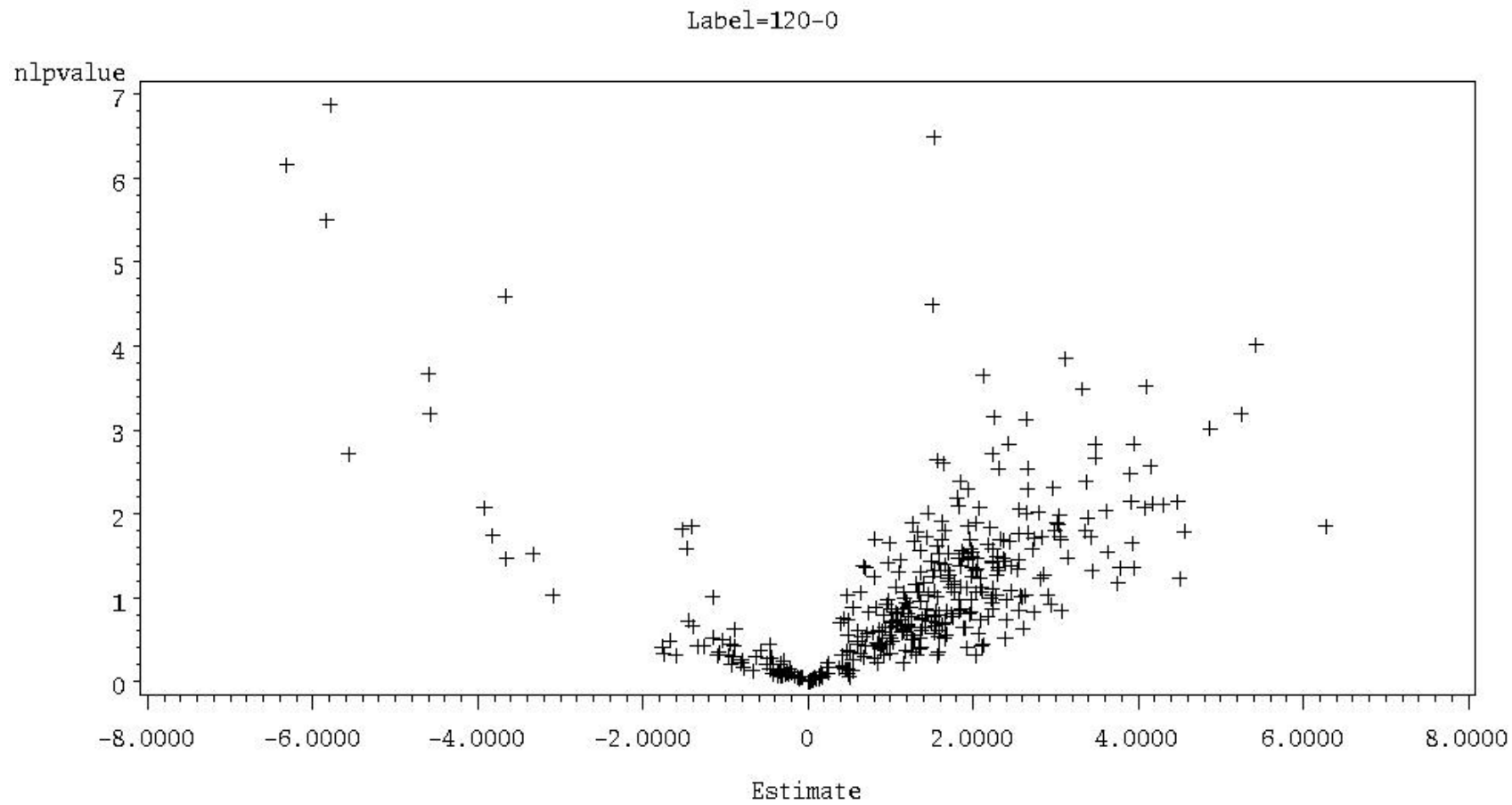
## False Discovery Rates

- Ordered, unadjusted  $p$ -values from an experiment with many tests of significance:

$$p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$$

- In microarray, “volcano plots”, with  $(-\log_{10}(p))$  on the vertical axis implicitly involve many tests of significance.

# A volcano plot



Truth table: Outcome from multiple tests

Truth	Declared Significant	Not significant	Total
Null	$FP$	$TN = m_0 - FP$	$m_0$
Alternative	$TP$	$FN = m_1 - TP$	$m_1$
Total	$S = FP + TP$	$m - S$	$m$

Some compound error measures: comparisonwise (CER), familwise (FWE) and false discovery (FDR):

$$E\left(\frac{FP}{m_0}\right) \leq CER$$

$$\Pr(FP > 0) \leq FWE$$

$$E\left(\frac{FP}{S} | S > 0\right) \leq FDR$$

Interpretation of FDR in microarray: if these genes investigated further (e.g. by PCR), FDR is proportion that will result in a dead-end.

## Benjamini-Hochberg (1995) step-up procedure

To “control” FDR at  $\alpha$ ,

1. Order the raw  $p$ -values:  $p_{(1)} \leq \dots \leq p_{(m)}$
2. Find  $\hat{k} = \max\{k : \frac{m}{k}p_{(k)} \leq \alpha\}$
3. If  $\hat{k}$  exists, reject tests corresponding to  $p_{(1)}, \dots, p_{(\hat{k})}$

Equivalently, the BH-adjusted  $p$ -values are defined as

$$\begin{aligned}\tilde{p}_{(1)} &= \min\{\tilde{p}_{(2)}, mp_{(1)}\} \\ &\vdots \\ \tilde{p}_{(m-1)} &= \min\{\tilde{p}_{(m)}, \frac{m}{m-1}p_{(m-1)}\} \\ \tilde{p}_{(m)} &= p_{(m)}\end{aligned}$$

- FDR option in PROC MULTTEST with variable “raw\_p” in dataset. Here,  $m = 10$ . (Taken from Westfall, et al, (1999))

The SAS System  
The Multtest Procedure

Test	p-Values (10/test)			False Discovery
	Raw	*Raw	Bonferroni	Rate
1	0.0001		0.0010	0.0010
2	0.0058		0.0580	0.0290
*3*	0.0132	> .0440	0.1320	0.0440
4	0.0289	< .0723	0.2890	0.0723
5	0.0498	< .0996	0.4980	0.0996
6	0.0911	< .1518	0.9110	0.1518
7	0.2012	< .2874	1.0000	0.2874
8	0.5718	< .7148	1.0000	0.7148
9	0.8912	< .9902	1.0000	0.9011
10	0.9011	< .9011	1.0000	0.9011

## A different approach to multiple testing

- The stepup BH procedure estimates the rejection region, i.e.  $\hat{k}$ , so that on average,  $FDR < \alpha$ .
- Alternatively, Storey (2002) considers fixing the critical region, and then estimating the FDR.
- The sample of  $m$   $p$ -values contains information about

$$\pi_0 = m_0/m$$

and may be used for estimation of FDR

- In cases where  $m_0 \ll m$ , Storey's procedure is less conservative than BH, which takes  $m_0 = m$ .

## Estimation of FDR

Consider the fixed critical region that rejects hypotheses with  $p$ -values  $\leq t$ . From the truth table

$$\begin{aligned} FDR(t) &\approx \frac{E[FP(t)]}{E[S(t)]} = \frac{tm_0}{E[\#\{p_i \leq t\}]} \quad (p_i \stackrel{H_0}{\sim} U(0, 1)) \\ \widehat{FDR}(t) &\approx \frac{t\hat{m}_0}{\#\{p_i \leq t\}} = \frac{t\hat{\pi}_0 m}{\#\{p_i \leq t\}} \end{aligned}$$

where

$$\pi_0 = \frac{m_0}{m}$$

## Estimation of $\hat{\pi}_0$

- Introduce a tuning parameter,  $0 < \lambda < 1$ .
- Use information in  $\frac{\#\{p_i > \lambda\}}{m}$  about  $\pi_0$ : for  $\lambda$  not close to 0,

$$E \left( \frac{\#\{p_i > \lambda\}}{m} \right) \approx (1 - \lambda)\pi_0$$

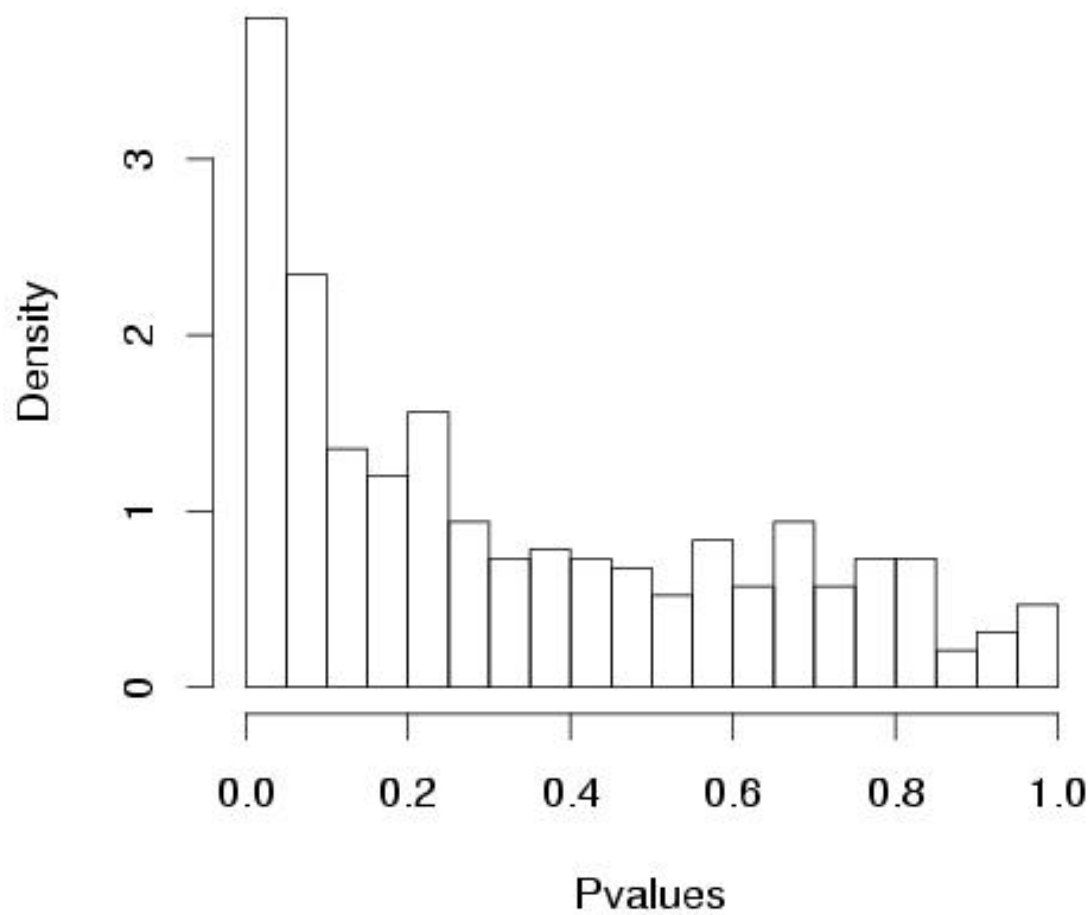
$$\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda\}}{m(1 - \lambda)}$$

- Choice of  $\lambda$  an issue.

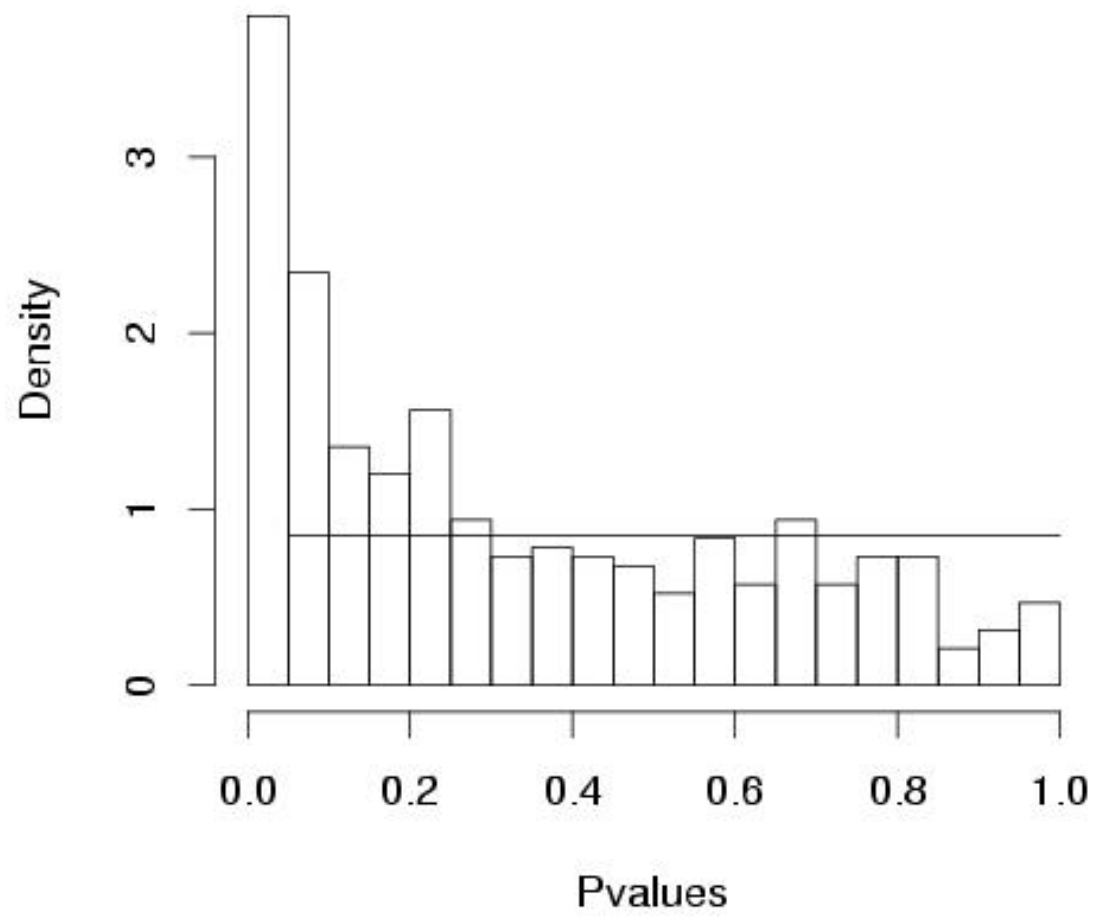
An example to illustrate estimation of  $\pi_0$ ,  $FDR$

- An experiment investigates gene expression during germination of *H. parviporum*, a fungal pathogen causing root rot
- mRNA amplified from tissue harvested at five timepoints: 0, 18, 36, 72 and 120 hours post-germination.
- Expression msmts on  $m = 384$  genes for each array
- $N = 15$  biologically independent mRNA samples,  $n = 3$  from each timept., i.e. all biological repl'n, none technical
- $H_0 : \mu_0 = \mu_{18} = \mu_{36} = \mu_{72} = \mu_{120}$  vs  $H_1 : \text{not all equal}$
- Investigators are G. Li and F. Asiegbu, Swedish U. of Ag. Sci.

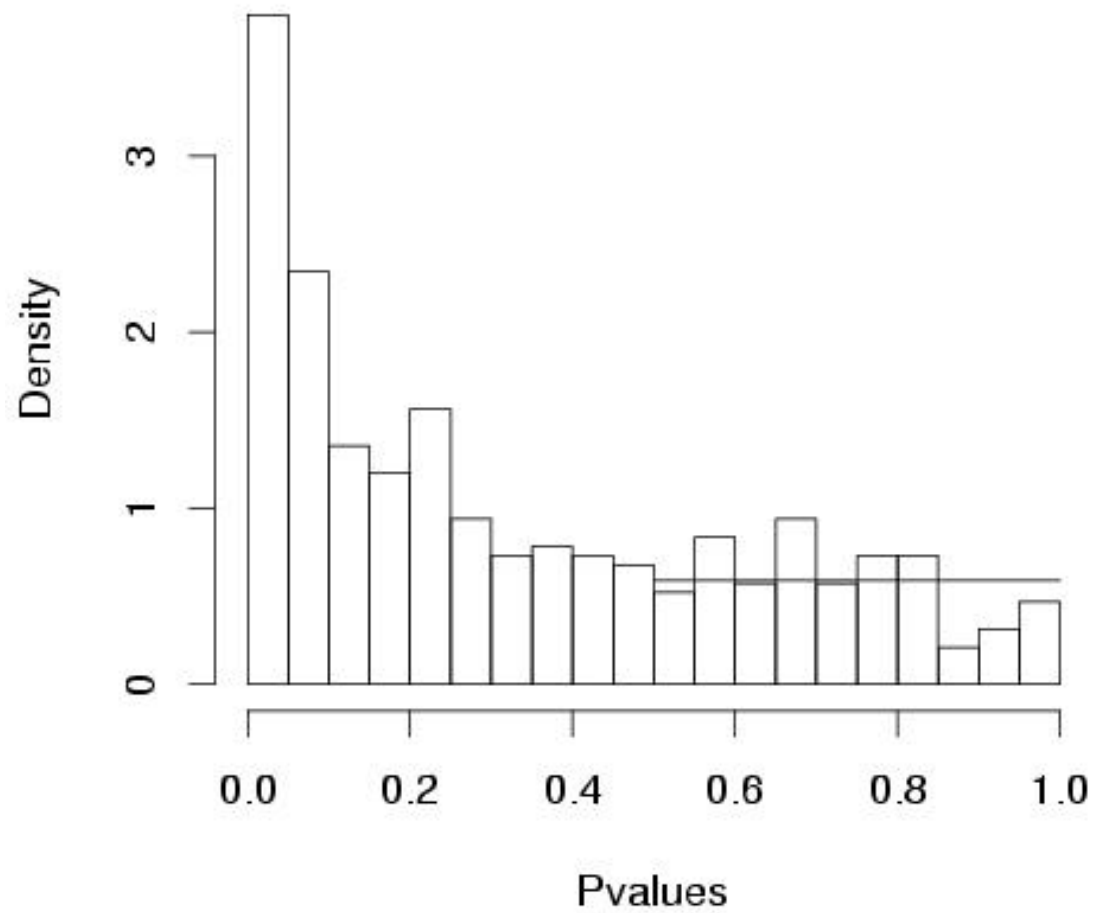
**Pvalues from 384 F-tests (df=4,10)  
(H. Parviporum germination over time)**



**Pvalues from 384 F-tests (df=4,10)**  
**lambda= 0.05 ,pi0hat= 0.852521929824561**



**Pvalues from 384 F-tests (df=4,10)**  
**lambda= 0.5 ,pi0hat= 0.588541666666667**



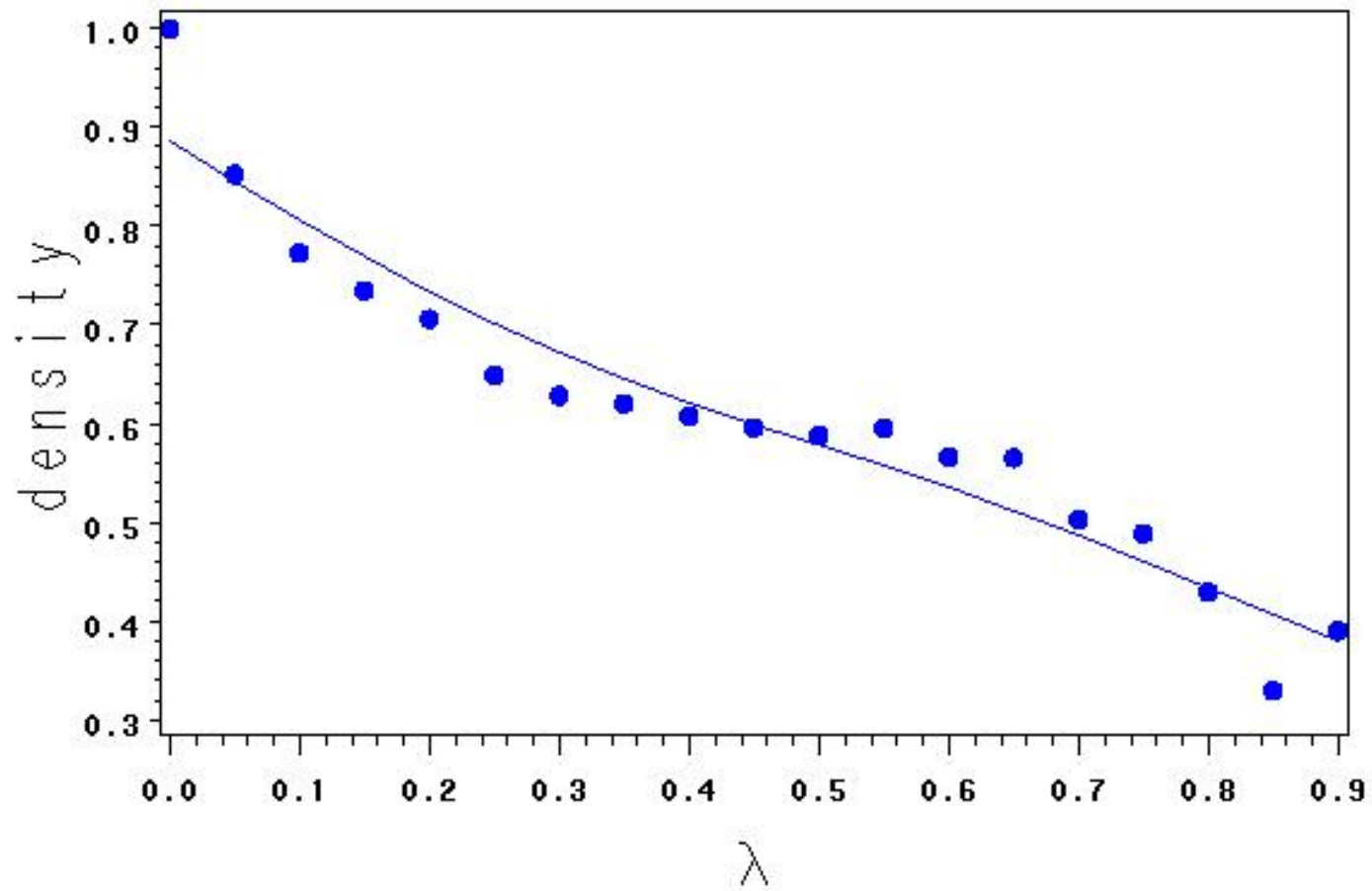
### Estimation of $\pi_0$ continued

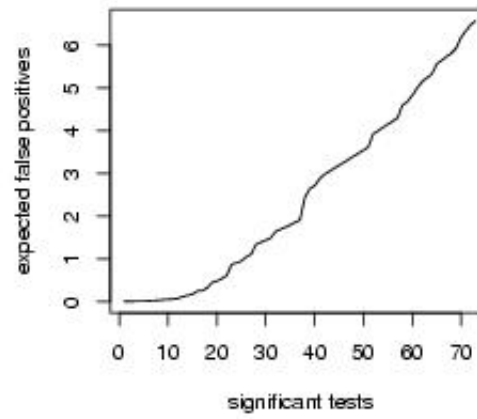
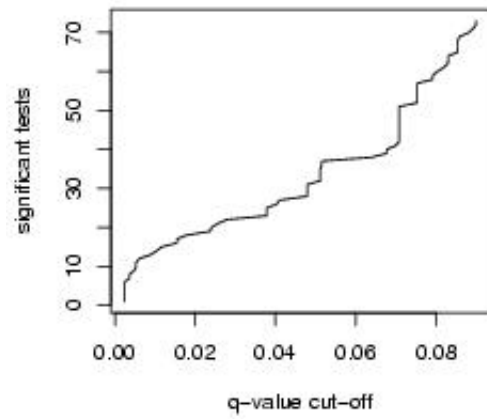
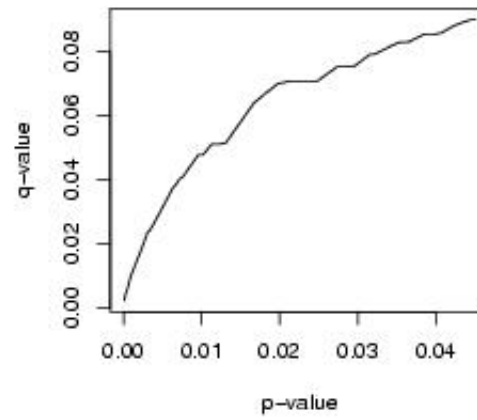
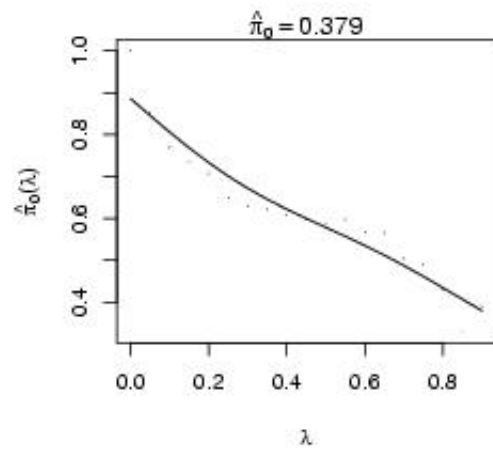
$\lambda$	$\#\{p_i > \lambda\}$	$\hat{\pi}_0(\lambda)$
0.05	311	$\frac{311}{384(1-.05)} = .8547$
0.5	113	$\frac{113}{384(1-.5)} = .5901$
0.95	9	$\frac{9}{384(1-.95)} = .4688$

- Positive bias of  $\hat{\pi}_0$  for  $\lambda$  near 0, high variance for  $\lambda$  near 1.
- `qvalue()` procedure in R fits smooth function,  $\tilde{\pi}(\lambda)$  and considers fitted value near  $\lambda = 1$ .

Estimation of  $\pi_0$ , continued

Plot 1





## Estimation of FDR, continued

Consider a critical region that declares genes with  $p$ -values  $\leq .01$  significant in the *H. parvivorum* experiment with  $m = 384$ .

$$\hat{\pi}_0 = 0.379 \text{ (smoother estimate from software)}$$

$$S(.01) = 30 \text{ (number declared significant)}$$

$$\widehat{FDR}(0.01) = \frac{\hat{\pi}_0 m t}{\#\{p_i < t\}} = \frac{0.379(384)(0.01)}{30} = 0.049$$

Bootstrap ( $B = 1000$ ) estimation of  $FDR$  and  $\pi_0$ :

Parameter	Mean (SE)	Median	95% c.i.
$\pi_0$	.380(.079)	.378	(.237,.529)
$FDR(.01)$	.050(.015)	.049	(.026,.083)

## Comparison and interpretation

- Compound error rates at .05
  - Declaring significance for  $p_i \leq .01$ , the FDR for the list of **30** significant tests is about about 5.0%, or 1.5 false leads.
  - BH-adjusted  $p$ -value  $\leq .05$  yields **18** significant tests, or about .9 false leads.
  - Bonferroni correction with  $\alpha = 0.05$  leads to **6** significant tests, and we can say that  $\Pr(\geq 1 \text{ false lead}) \leq 0.05$ .
  - If  $CER = 0.05$  (no multiplicity adjustment), **73** tests are rejected, and type I error among the  $m = 384$  tests is 5% (Corresponding FDR is about 9%, or about 6.5 false leads.)

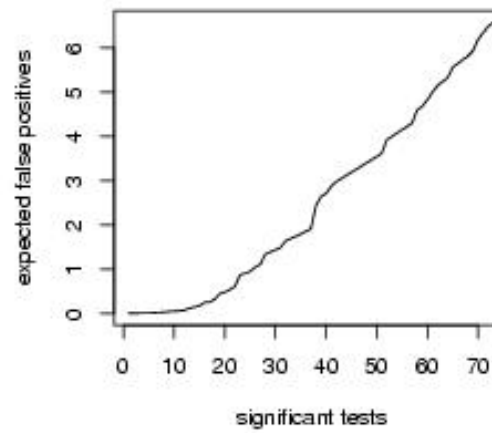
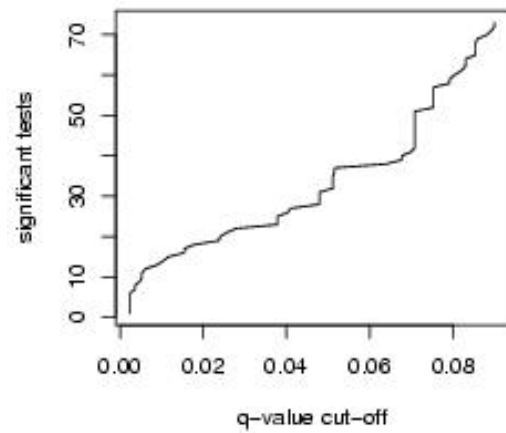
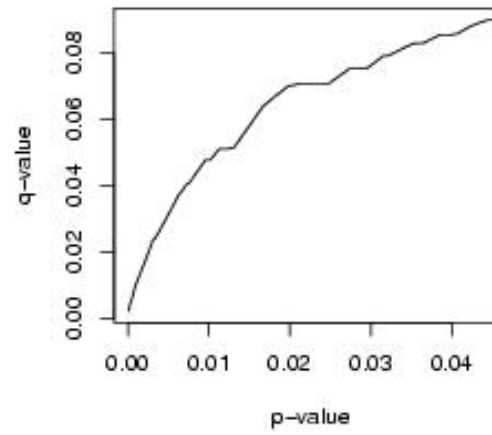
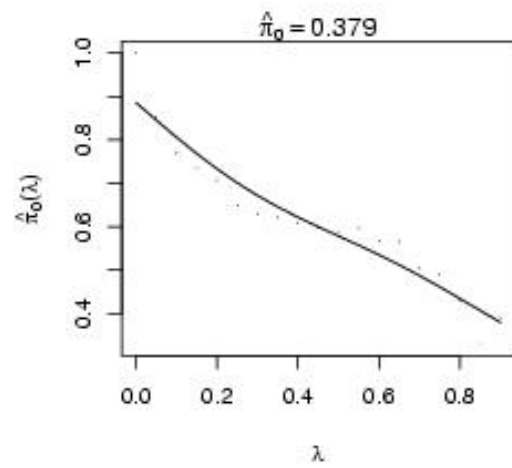
## Comparison and interpretation at $\alpha = .05$

Multiplicity Adjustment	Rule	List length	Interpretation
FDR estimation	$p_i \leq .01$	30	$FDR \approx .05$ , 1.5 false leads
BH	$\text{adj } p_i \leq .05$	18	$FDR \approx .05$ , 0.9 false leads
Bonferroni	$p_i \leq \frac{.05}{384}$	6	$\Pr(\geq 1 \text{ false lead}) \leq .05$
None	$p_i \leq .05$	73	Type I error is 5%, $F\hat{D}R = .09$ , $\sim 6.5$ false leads

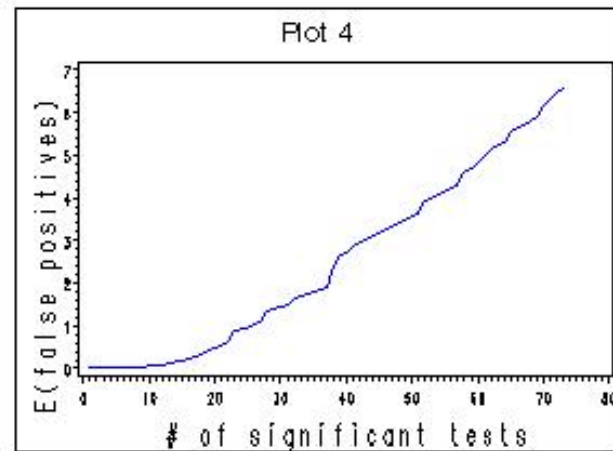
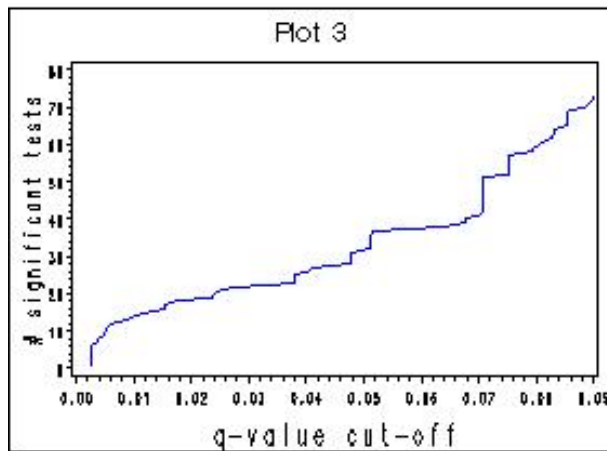
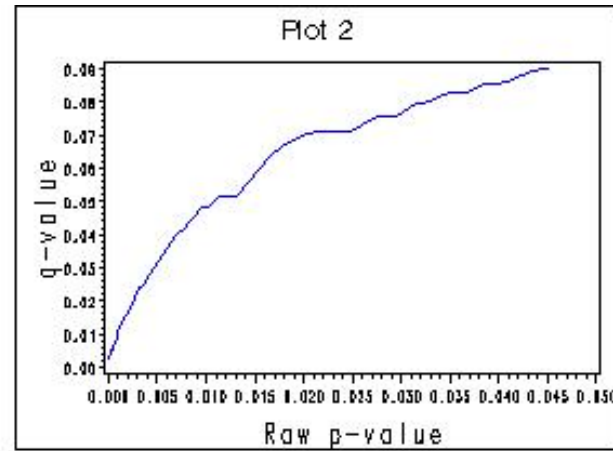
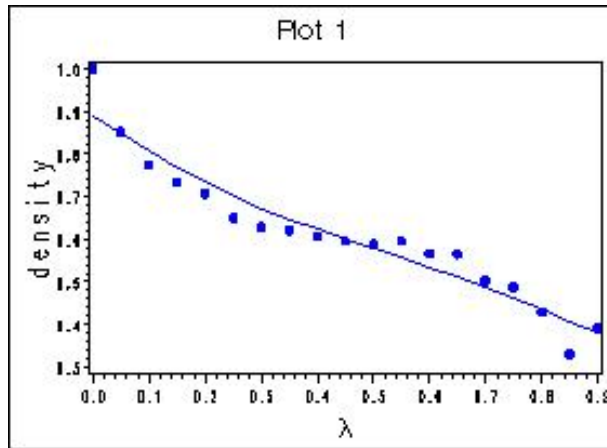
## $q$ -values and their interpretation

$$q\text{-value}(p_i) = \min_{t \geq p_i} \widehat{FDR}(t)$$

- A measure of significance in terms of the FDR.
- The smallest FDR at which the statistic may be declared significant.
- Obtained via step-up procedure just like BH



# Using SAS ...



## Reconciliation of BH and method based on $\widehat{FDR}$

- For  $m$   $p$ -values, method of BH finds  $\hat{k}$  such that

$$\hat{k} = \max\{k : \frac{m}{k}p_{(k)} \leq \alpha\}$$

and rejects  $p_{(1)}, \dots, p_{(\hat{k})}$  to control  $FDR < \alpha$

- Method based on  $\hat{\pi}_0$  finds  $\hat{l}$  such that

$$\hat{l} = \max\{l : \widehat{FDR}(p_{(l)}) \leq \alpha\}$$

But

$$\widehat{FDR}(t = p_{(l)}) = \frac{m}{l}\hat{\pi}_0 p_{(l)}$$

With  $\hat{\pi}_0 = 1$  this is equivalent to

$$\hat{l} = \max\{l : \frac{m}{l}p_{(l)} \leq \alpha\}$$

- If  $\hat{\pi}_0 < 1$ , then  $\hat{l} > \hat{k}$  with high probability. (For *H. parvaporum* data with  $\alpha = 0.05$ ,  $\hat{k} = 18$ ,  $\hat{l} = 31$ .)

## A parametric model for $p$ -values

- $p$ -values a random sample from two-component mixture:

$$f(p; a, b) = \pi_0 + (1 - \pi_0) \frac{\Gamma(a + b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1 - p)^{b-1}$$

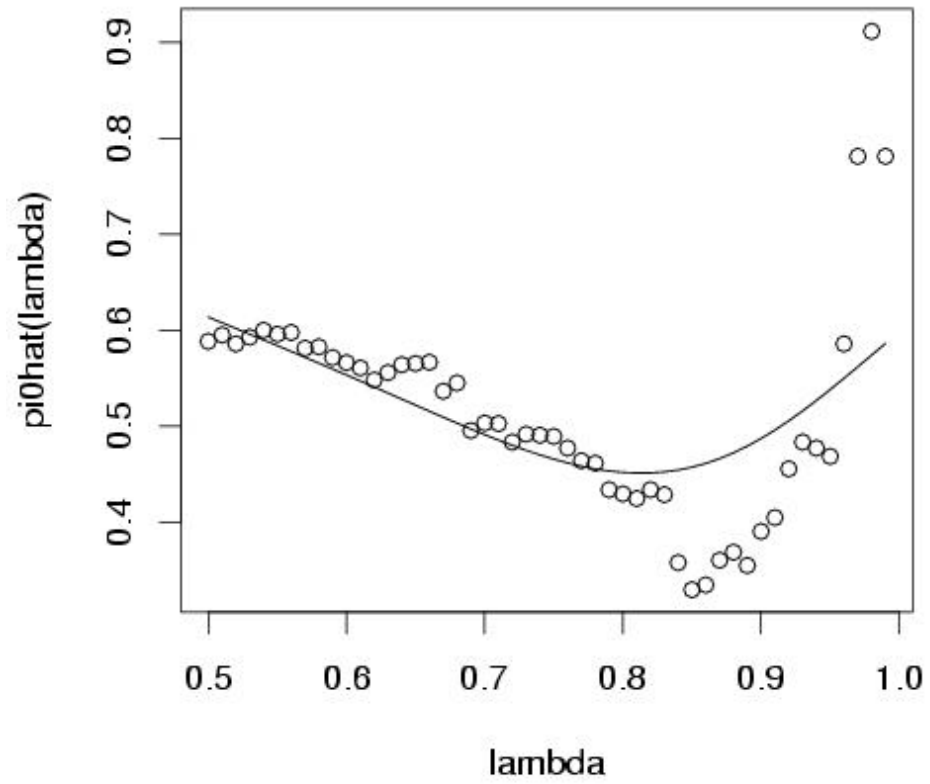
for  $0 < p < 1, a > 0, b > 0$

or

$$P \stackrel{\text{dist}}{=} \pi_0 U(0, 1) + (1 - \pi_0) \beta(a, b)$$

- Two-component mixture likelihood not hard to maximize.
- Beta distributions accomodate a variety of shapes, more than two components not needed (Allison et al, 2002).
- Choice of  $\lambda$  in `qvalue()` smoother unclear.

wacky choice of lambda, pi0hat= 0.586



## Using PROC NLMIXED to obtain MLEs

```
PROC NLMIXED DATA=pvalues;  
  PARAMETERS pi0=.5 a=2 b=2;  
  pi1=1-pi0;  
  loglikelihood=LOG(pi0+pi1*PDF('BETA',raw_p,a,b));  
  MODEL raw_p ~ GENERAL(loglikelihood);  
RUN;
```

### The NLMIXED Procedure Parameter Estimates

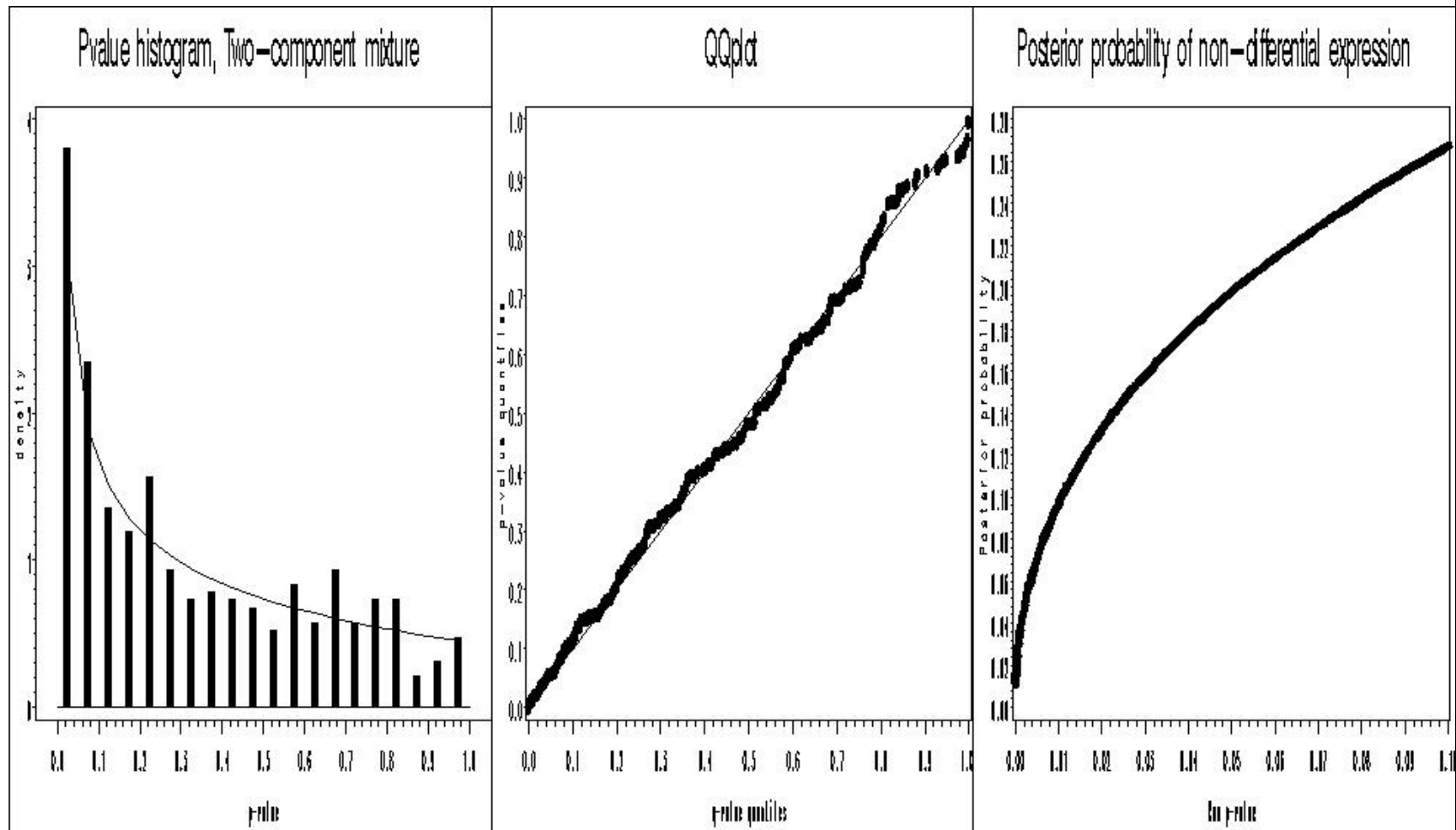
Parameter	Estimate	Standard Error	DF	t Value	Pr >  t
pi0	0.4403	0.1130	384	3.90	0.0001
a	0.5123	0.04683	384	10.94	<.0001
b	2.0308	0.6998	384	2.90	0.0039

### Posterior probability of $H_0$

If  $H_0 : \mu_0 = \mu_{18} = \mu_{36} = \mu_{72} = \mu_{120}$  and  $\pi_0 = \Pr(H_0)$ , then

$$\Pr(H_0|p_i) = \frac{\Pr(p_i \cap H_0)}{\Pr(p_i)} = \frac{\pi_0}{\pi_0 + (1 - \pi_0) \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} p^{a-1} (1 - p)^{b-1}}$$

For the *parviporum* data,  $(\hat{\pi}_0, \hat{a}, \hat{b}) = (0.44, 0.51, 2.03)$ .



[www4.stat.ncsu.edu/~jaosborn/research/microarray/software/index.html](http://www4.stat.ncsu.edu/~jaosborn/research/microarray/software/index.html)