# Estimation of the False Discovery Rate

Coffee Talk, Bioinformatics Research Center, Sept, 2005

Jason A. Osborne, osborne@stat.ncsu.edu
Department of Statistics, North Carolina State University

## Outline

- A discussion of the False Discovery Rate, FDR.
- Storey's (2002) `qvalue()` freeware available from Bioconductor.
- An equivalent SAS macro for its estimation.
- Posterior probability of a false lead.
- Concepts illustrated using an example using gene expression data from a microarry experiment to investigate a fungal pathogen that causes root rot in Norway Spruce.

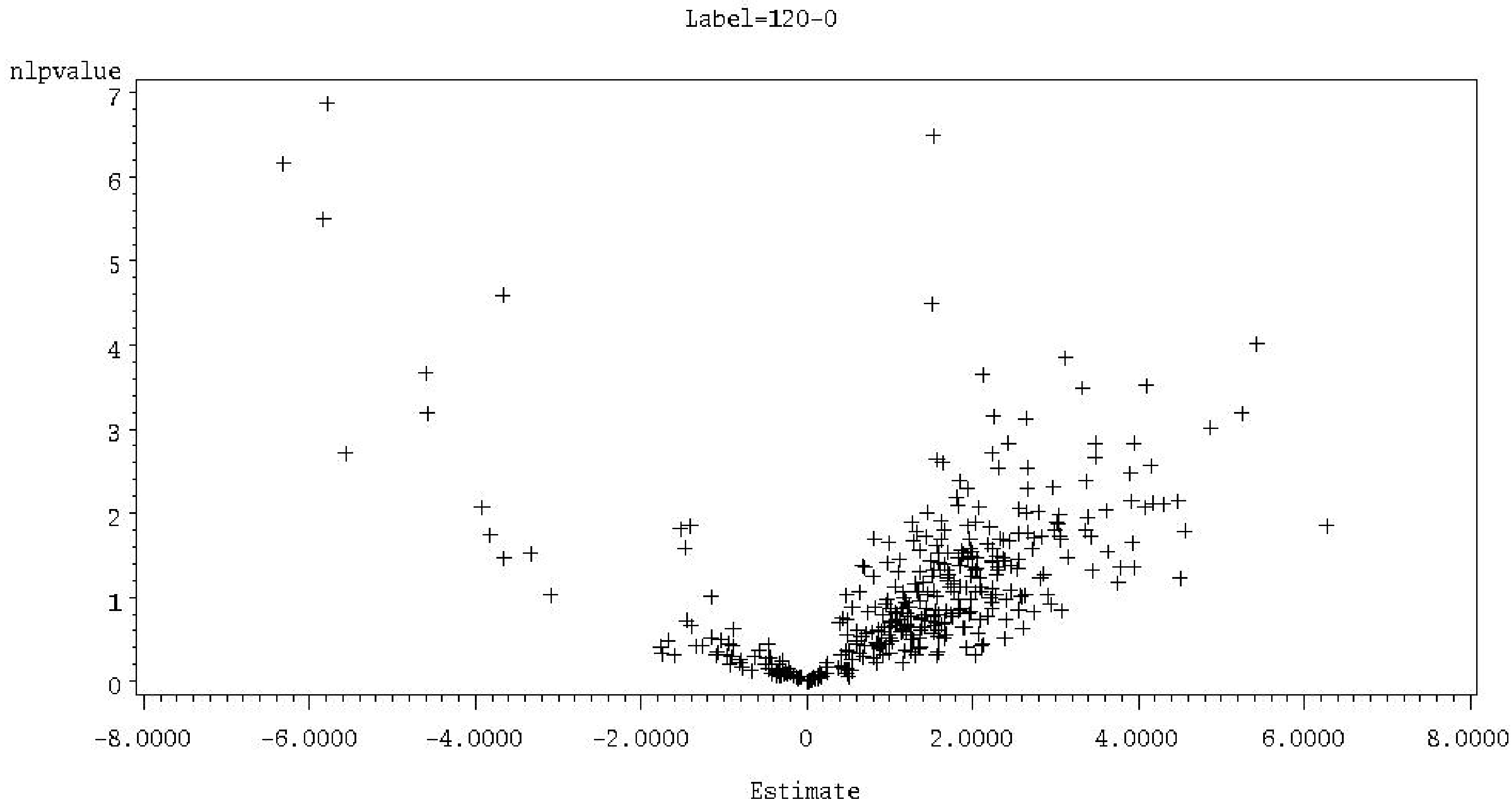
## Microarray experiments

- Use emerging technology to simultaneously observe expression intensity for thousands of genes across different experimental conditions.
- One of many challenges: the search for differentially expressed genes and the identification/declaration of "significance."

## False Discovery Rates

- Consider an expt with many tests of significance ($m = 400$ or $4000$)
- $p_{(1)} \leq p_{(2)} \leq \cdots \leq p_{(m)}$ denote ordered, unadjusted $p$-values.
- 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 is true | $F$ | $m_0 - F$ | $m_0$ |
| Alternative is true | $S$ | $m_1 - S$ | $m_1$ |
| Total | $R = F + S$ | $m - R$ | $m$ |

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

$$\begin{aligned} E\left(\frac{F}{m_0}\right) &\leq CER \\ \Pr(F > 0) &\leq FWE \\ E\left(\frac{F}{R}|R > 0\right) &\leq FDR \end{aligned}$$

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

## How the Benjamini-Hochberg (1995) step-up procedure works

To "control" FDR at $\alpha$,

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

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

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

- FDR option in PROC MULTTEST with variable “raw_p” in dataset. (Taken from Westfall, et al, (1999))

```
                        The SAS System
                    The Multtest Procedure
                          p-Values                    False
                                                  Discovery
Test            Raw          Bonferroni                Rate
   1         0.0001           *   0.0010             0.0010
   2         0.0058           *   0.0580             0.0290
   3         0.0132     <.015*    0.1320             0.0440
   4         0.0289     >.02      0.2890             0.0723
   5         0.0498     >.025     0.4980             0.0996
   6         0.0911     >.03      0.9110             0.1518
   7         0.2012     >.035     1.0000             0.2874
   8         0.5718     >.04      1.0000             0.7148
   9         0.8912     >.045     1.0000             0.9011
  10         0.9011     >.05      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.
- Information in the $p$-values about $m_0$ may be used to obtain an estimator and to construct a more powerful procedure that may still be used to control or estimate FDR.
- BH procedure uses $m_0 = m$.

## Estimation of FDR

Consider fixing the critical region by rejecting hypotheses with $p$-values less than $t$. From the truth table

$$FDR(t) \approx \frac{E[F(t)]}{E[R(t)]} = \frac{tm_0}{E[\#\{p_i \leq t\}]}$$

$$\widehat{FDR}(t) \approx \frac{t\hat{m}_0}{\#\{p_i \leq t\}} = \frac{t\hat{\pi}_0 m}{\#\{p_i \leq t\}}$$

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,

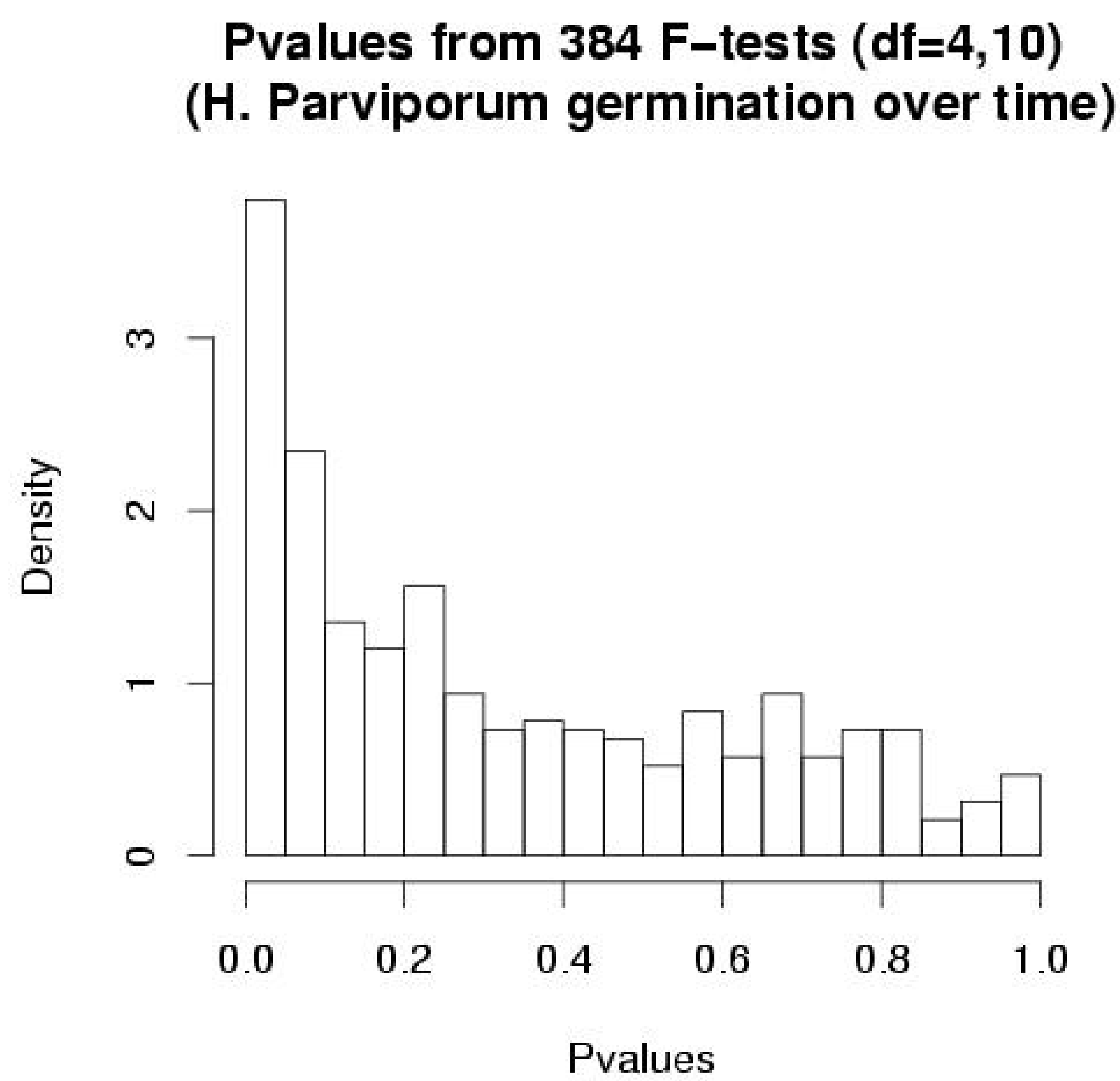
$$E\left(\frac{\#\{p_i > \lambda\}}{m}\right) \approx (1 - \lambda)\pi_0$$

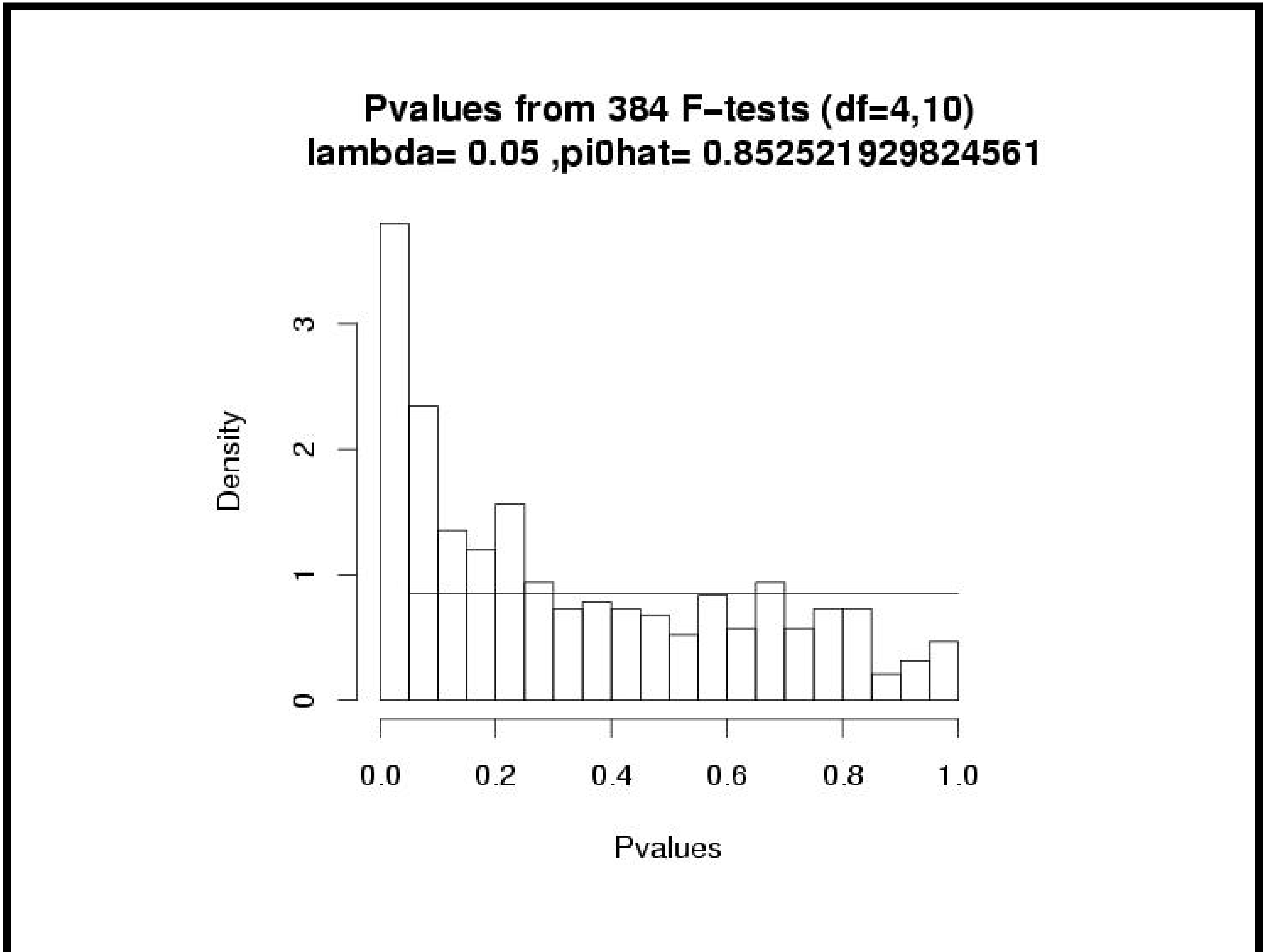
$$\hat{\pi}_0(\lambda) = \frac{\#\{p_i > \lambda\}}{m(1 - \lambda)}$$

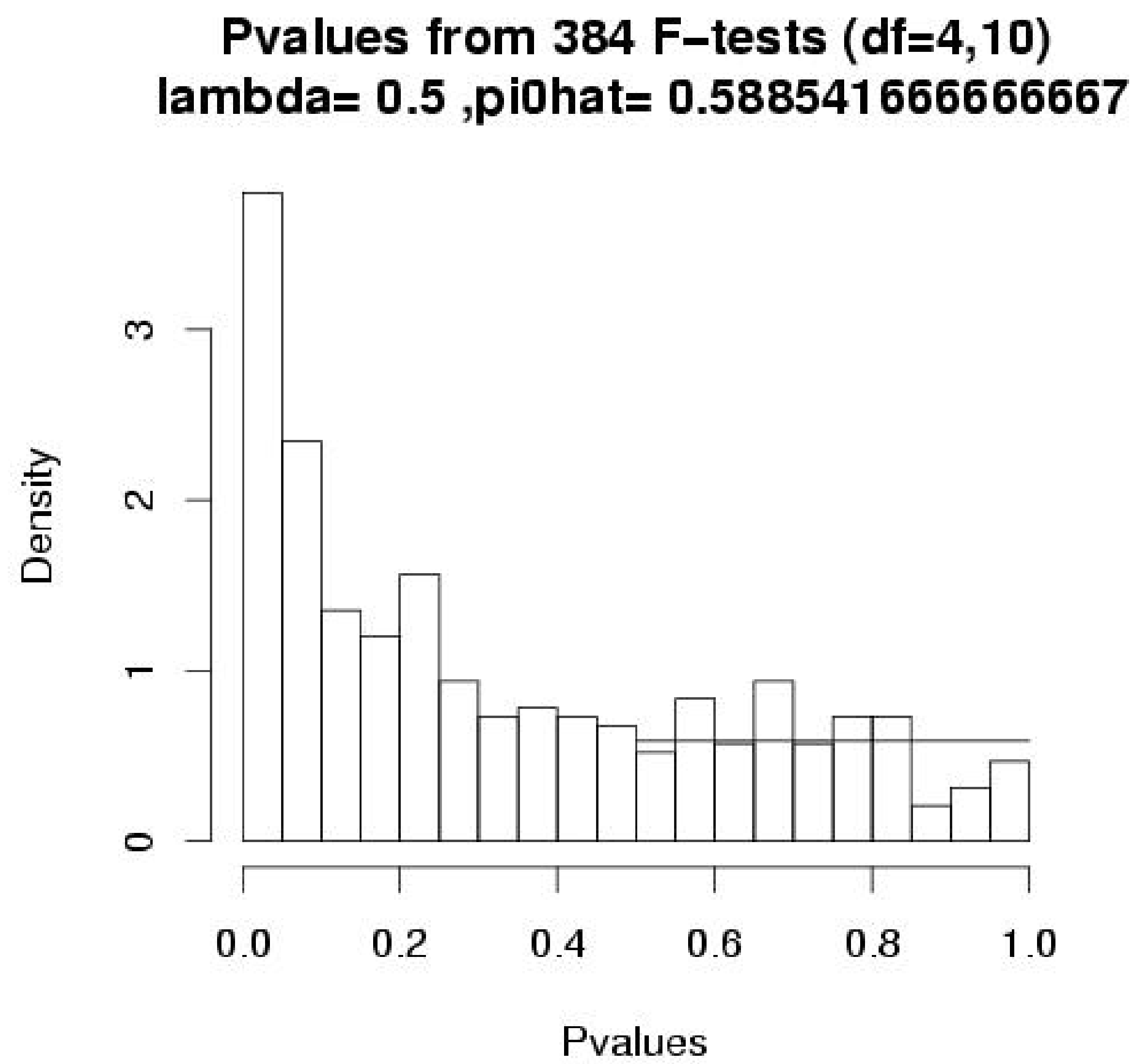
- Substitute $\hat{\pi}_0$ into expression for $\widehat{FDR}(t)$

## 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 timepoints 0, 18, 36, 72 and 120 hours post-germination. ($t = 5$ "treatments")
- Expression measurements on 384 cDNA genes for each array
- $N = 15$ independent microarrays, $n = 3$ for each timepoint.
- 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)
Density
0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
Pvalues

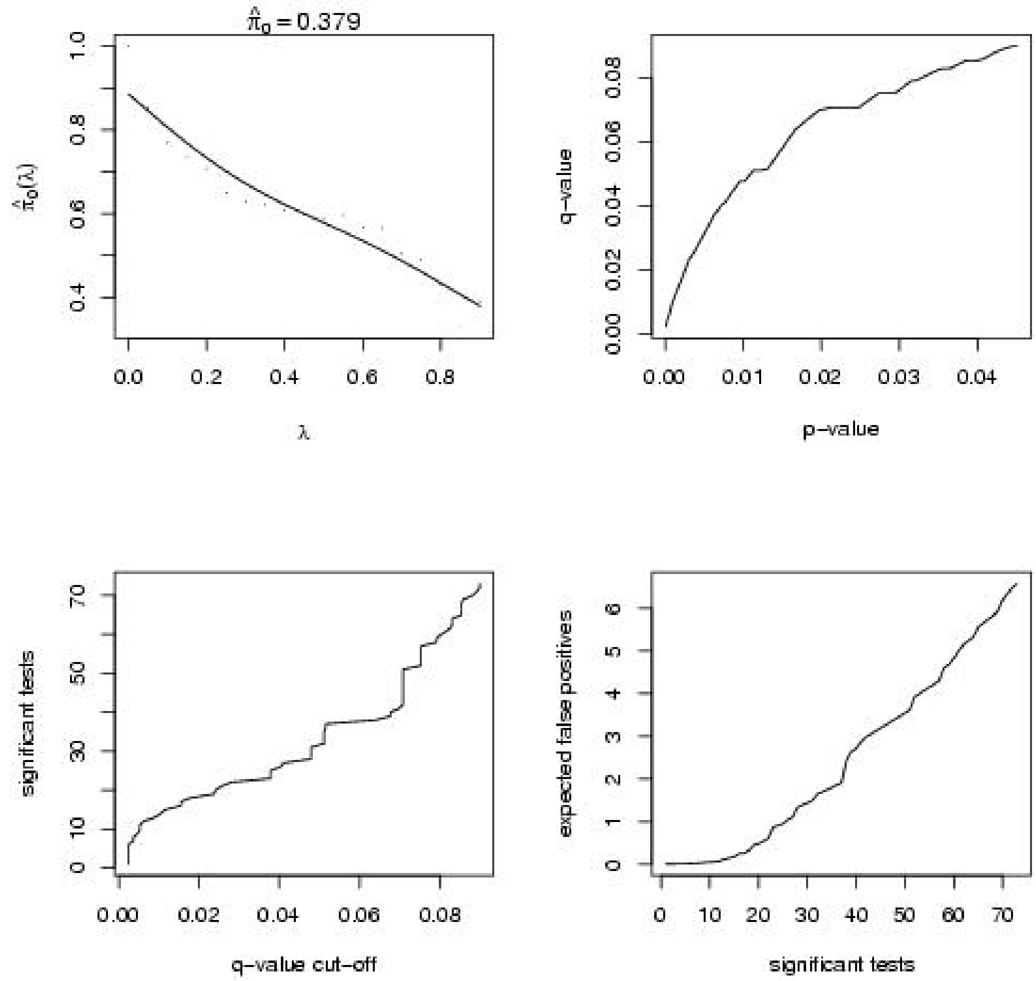

Pvalues from 384 F−tests (df=4,10)
lambda= 0.05 ,pi0hat= 0.852521929824561
Density
0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
Pvalues


Pvalues from 384 F−tests (df=4,10)
lambda= 0.5 ,pi0hat= 0.588541666666667
Density
0
1
2
3
0.0
0.2
0.4
0.6
0.8
1.0
Pvalues

## 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, $\pi(\lambda)$ and considers fitted value near $\lambda = 1$.


$\hat{\pi}_0 = 0.379$
$\hat{\pi}_0(\lambda)$
1.0
0.8
0.6
0.4
0.0
0.2
0.4
0.6
0.8
λ
q−value
0.08
0.06
0.04
0.02
0.00
0.00
0.01
0.02
0.03
0.04
p−value
significant tests
70
50
30
10
0
0.00
0.02
0.04
0.06
0.08
q−value cut−off
expected false positives
6
5
4
3
2
1
0
0
10
20
30
40
50
60
70
significant tests

## Estimation of FDR, continued

Consider a rejection region of $(0, .01)$ for the $m = 384$ genes in the *H. parviporum* experiment.

$$\begin{aligned}
\hat{\pi}_0 &= 0.379 \text{ (smoother estimate from software)} \\
R(.01) &= 30 \quad \text{(number of tests rejected)} \\
\widehat{FDR}(0.01) &= \frac{\hat{\pi}_0 m t}{\#\{p_i < t\}} = \frac{0.379(384)(0.01)}{30} = 0.049
\end{aligned}$$

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) |

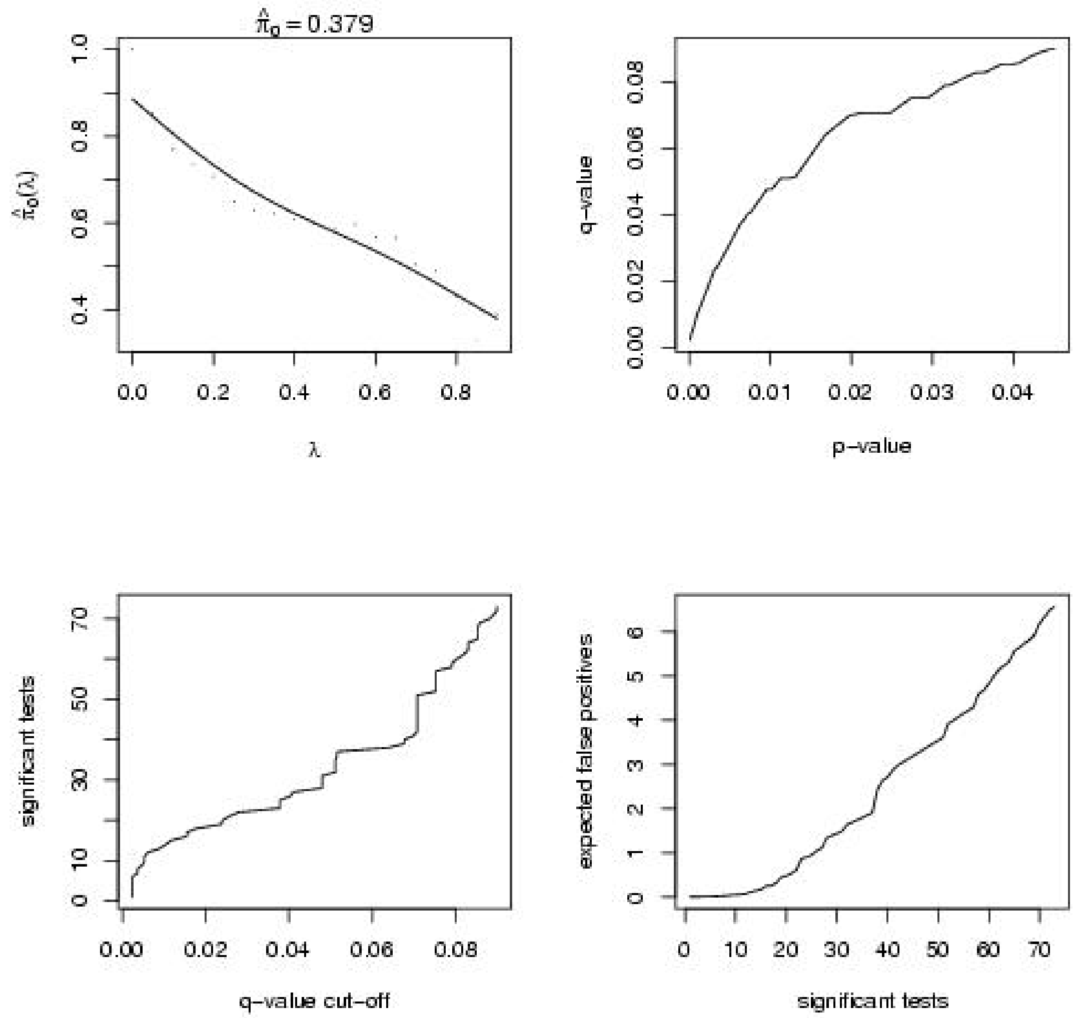
## Interpretation

- Compound error rates at .05
  - The proportion of the 30 rejected tests that are false discoveries is estimated to be about 5.0%, or 1.5 false leads.
  - BH-adjusted $p$-value $< .05$ yields 18 rejections.
  - Bonferroni correction with $\alpha = 0.05$ leads to 6 rejected tests, and we're able to say that $\Pr(\geq 1 \text{ false lead}) \leq 0.05$. (With larger $m$, it is not uncommon for Bonferroni to lead to no rejections, despite departure of $p$-value histogram from $U(0, 1)$.)
  - If $CER = 0.05$ (no multiplicity adjustment), 73 tests are rejected, and type I error among the $m = 384$ tests is 5%

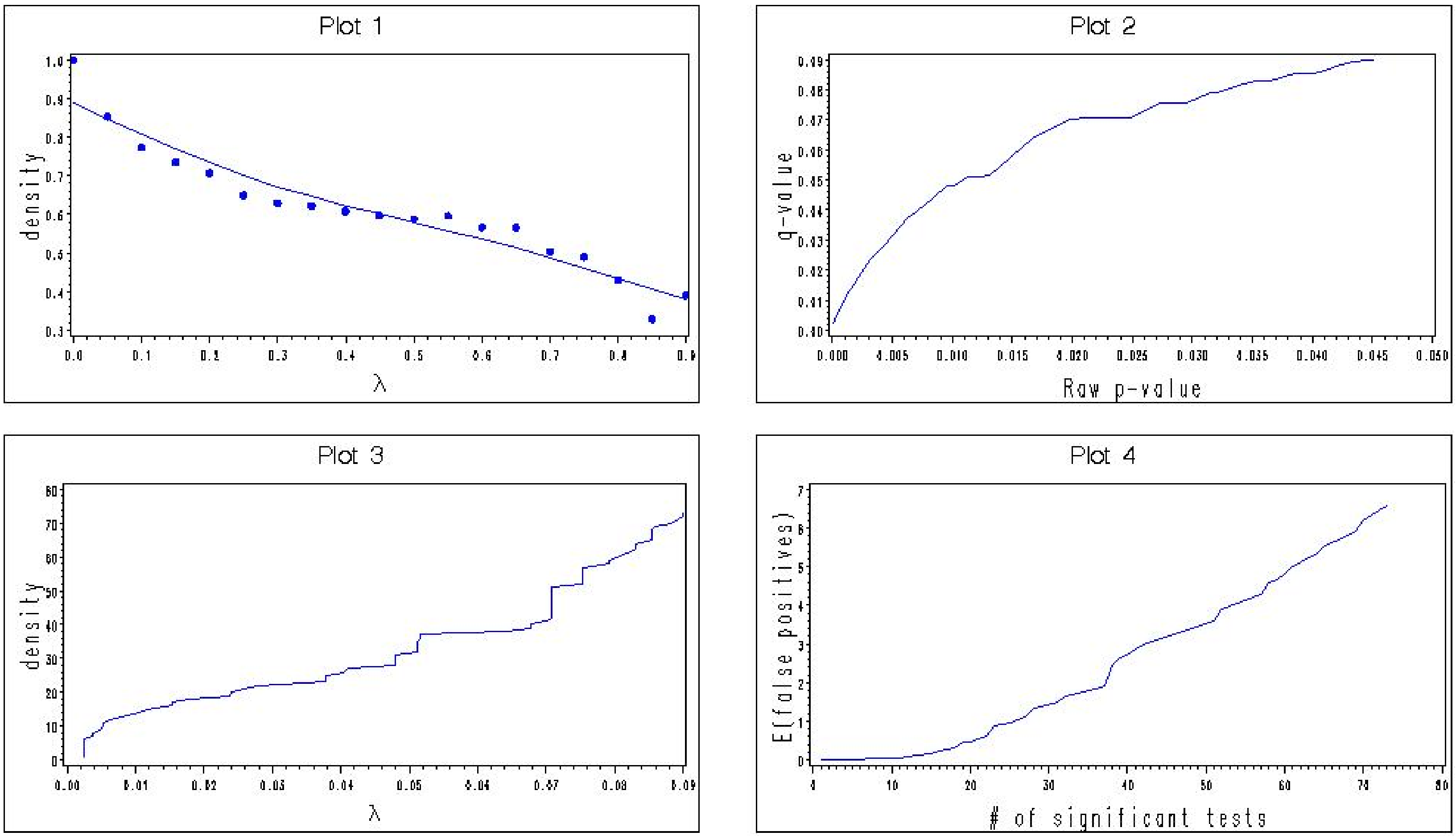
## $q$-values and their interpretation

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

- A measure of significance in terms of the FDR.
- The smallest FDR at which the statistic may be declared significant.


$\hat{\pi}_0 = 0.379$
$\hat{\pi}_0(\lambda)$
1.0
0.8
0.6
0.4
0.0
0.2
0.4
0.6
0.8
λ
q−value
0.08
0.06
0.04
0.02
0.00
0.00
0.01
0.02
0.03
0.04
p−value
significant tests
70
50
30
10
0
0.00
0.02
0.04
0.06
0.08
q−value cut−off
expected false positives
6
5
4
3
2
1
0
0
10
20
30
40
50
60
70
significant tests

# SAS macro can also be used



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

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

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

  and rejects $p_{(1)}, \ldots, p_{(\hat{k})}$ 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

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

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

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

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

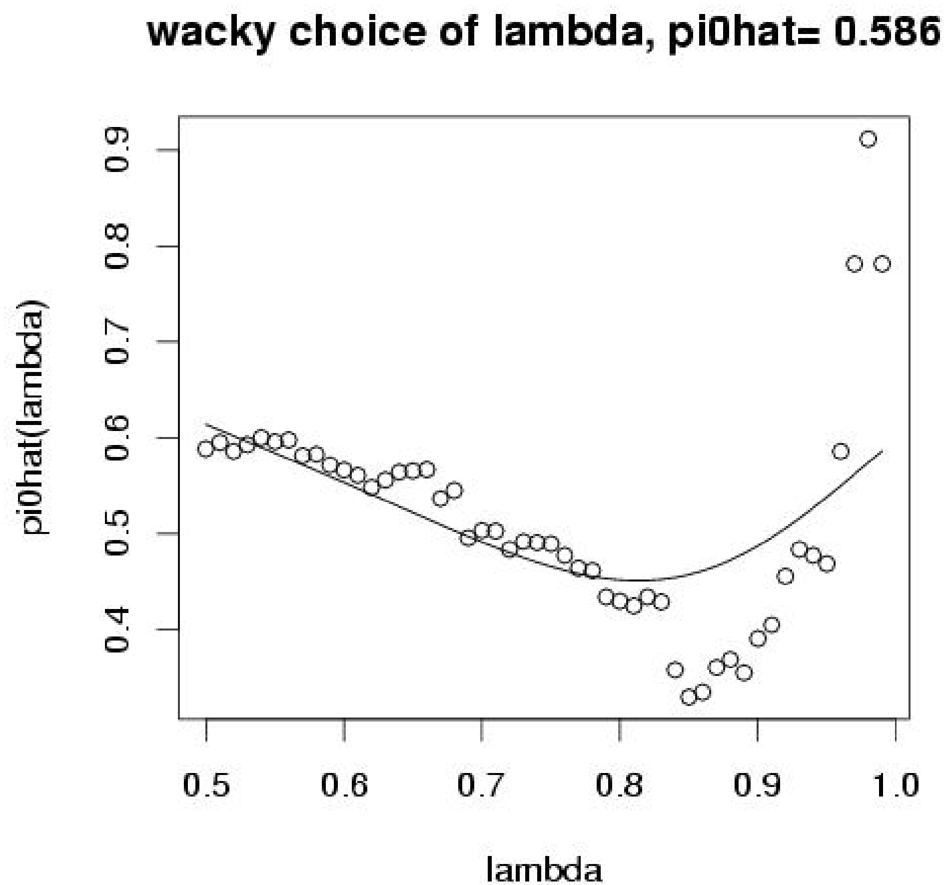
## A parametric model for $p$-values

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

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

$$\text{for } 0 < p < 1, a > 0, b > 0$$

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


wacky choice of lambda, pi0hat= 0.586
pi0hat(lambda)
0.4
0.5
0.6
0.7
0.8
0.9
0.5
0.6
0.7
0.8
0.9
1.0
lambda

## 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

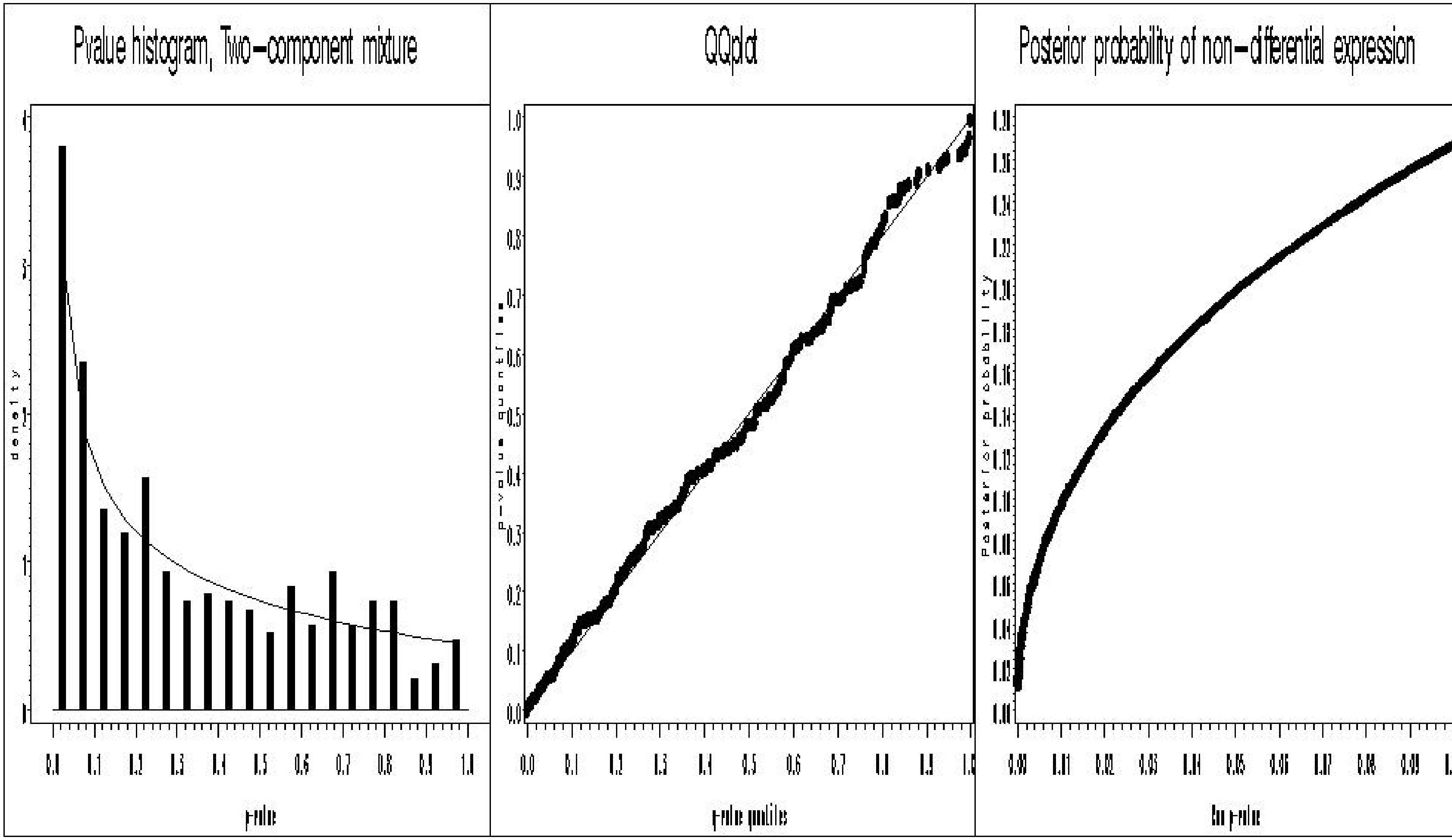
| Parameter | Estimate | Standard Error | DF | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| pi0 | 0.4403 | 0.1130 | 384 | 3.90 | 0.0001 |
| a | 0.5123 | 0.04683 | 384 | 10.94 | <.0001 |
| b | 2.0308 | 0.6998 | 384 | 2.90 | 0.0039 |

## 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

## 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.
- Storey JD. (2002) A direct approach to false discovery rates. *JRSSB*, **64**: 479-498
- Storey JD and Tibshirani R. (2003) Statistical significance for genome-wide studies. *PNAS*, **100**: 9440-9445.
- Storey JD. (2003) The positive false discovery rate: A Bayesian interpretation and the q-value. *Annals of Statistics*, **31**: 2013-2035.
- 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.