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 microarray 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

Label = 120 - 0

nlpvalue

Estimate
Truth table: Outcome from multiple tests

<table>
<thead>
<tr>
<th>Truth</th>
<th>Declared Significant</th>
<th>Not significant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null is true</td>
<td>$F$</td>
<td>$m_0 - F$</td>
<td>$m_0$</td>
</tr>
<tr>
<td>Alternative is true</td>
<td>$S$</td>
<td>$m_1 - S$</td>
<td>$m_1$</td>
</tr>
<tr>
<td>Total</td>
<td>$R = F + S$</td>
<td>$m - R$</td>
<td>$m$</td>
</tr>
</tbody>
</table>

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

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

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

$$
E\left(\frac{F}{R} | R > 0\right) \leq FDR
$$

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

$$\tilde{p}(m) = p(m)$$
$$\tilde{p}(m-1) = \min\{\tilde{p}(m), \frac{m}{m-1}p(m-1)\}$$
$$\vdots$$
$$\tilde{p}(1) = \min\{\tilde{p}(2), mp(1)\}$$
- FDR option in PROC MULTTEST with variable “raw_p” in dataset. (Taken from Westfall, et al, (1999))

The SAS System
The Multtest Procedure

<table>
<thead>
<tr>
<th>Test</th>
<th>Raw</th>
<th>Bonferroni</th>
<th>False Discovery</th>
<th>Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0001</td>
<td>* 0.0010</td>
<td>0.0010</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.0058</td>
<td>* 0.0580</td>
<td>0.0290</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.0132</td>
<td>&lt;.015* 0.1320</td>
<td>0.0440</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.0289</td>
<td>&gt;.02 0.2890</td>
<td>0.0723</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.0498</td>
<td>&gt;.025 0.4980</td>
<td>0.0996</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.0911</td>
<td>&gt;.03 0.9110</td>
<td>0.1518</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.2012</td>
<td>&gt;.035 1.0000</td>
<td>0.2874</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.5718</td>
<td>&gt;.04 1.0000</td>
<td>0.7148</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.8912</td>
<td>&gt;.045 1.0000</td>
<td>0.9011</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>0.9011</td>
<td>&gt;.05 1.0000</td>
<td>0.9011</td>
<td></td>
</tr>
</tbody>
</table>
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 \( \hat{FDR}(t) \)
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.
P-values from 384 F-tests (df=4,10)
(H. Parviporum germination over time)
Pvalues from 384 F-tests (df=4,10)
lambda = 0.05 , \hat{\pi_0} = 0.852521929824561
Pvalues from 384 F-tests (df=4,10)
lambda = 0.5, \( \hat{\pi}_0 = 0.5885416666666667 \)
Estimation of $\pi_0$ continued

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$#{p_i &gt; \lambda}$</th>
<th>$\hat{\pi}_0(\lambda)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>311</td>
<td>$\frac{311}{384(1-.05)} = .8547$</td>
</tr>
<tr>
<td>0.5</td>
<td>113</td>
<td>$\frac{113}{384(1-.5)} = .5901$</td>
</tr>
<tr>
<td>0.95</td>
<td>9</td>
<td>$\frac{9}{384(1-.95)} = .4688$</td>
</tr>
</tbody>
</table>

- 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 $\lambda = 1$. 
Estimation of FDR, continued

Consider a rejection region of \((0, .01)\) for the \(m = 384\) genes in the \textit{H. parviporum} experiment.

\[\hat{\pi}_0 = 0.379\] (smoother estimate from software)

\[R(.01) = 30\] (number of tests rejected)

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

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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean (SE)</th>
<th>Median</th>
<th>95% c.i.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\pi_0)</td>
<td>.380(.079)</td>
<td>.378</td>
<td>(.237,.529)</td>
</tr>
<tr>
<td>(FDR(.01))</td>
<td>.050(.015)</td>
<td>.049</td>
<td>(.026,.083)</td>
</tr>
</tbody>
</table>
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 $\Pr(\geq 1$ 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%
\textit{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.
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:

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

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

• Two-component mixture likelihood not hard to maximize.

• Beta distributions accommodate a variety of shapes (Allison et al, 2002).

• Choice of $\lambda$ in `qvalue()` smoother unclear.
wacky choice of lambda, \( \hat{\pi} = 0.586 \)
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 | 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 |
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


