

Bioconductor's qvalue package

Alan Dabney and John Storey
Department of Biostatistics
University of Washington
email: jstorey@u.washington.edu

March 4, 2015

Contents

1 Overview	1
2 Case study: the breast cancer dataset of Hedenfalk et al. (2001)	1
3 The qvalue function	2
4 The qsummary function	2
5 The qplot function	3
6 The qwrite function	3
7 The qvalue Shiny app	3

1 Overview

The `qvalue` package contains functions for the computation and presentation of q-values for features from an experiment with multiple comparisons. A q-value is a measure of significance for a particular feature, analogous to a p-value. However, where the p-value assigns significance in terms of the false positive rate, the q-value assigns significance in terms of the false discovery rate. See [2] for more detailed information.

This document provides a tutorial for using the `qvalue` package. The package consists of four functions (`qvalue` for computing q-values from p-values, `qplot` for visualizing the results with graphs, `qsummary` for visualizing the results with a table, and `qwrite` for writing the output to file). As with any R package, detailed information on functions, their arguments and value, can be obtained in the help files. For instance, to view the help file for the function `qvalue` within R, type `? qvalue`.

2 Case study: the breast cancer dataset of Hedenfalk et al. (2001)

We demonstrate the functionality of this package using gene expression data from the breast cancer study of [1]. Comparison was made between two types of genetic mutation that are associated

with an increased risk of breast cancer, BRCA1 and BRCA2. There were 7 and 8 cDNA arrays for BRCA1 and BRCA2, respectively. The example considered here is restricted to 3,170 genes as described in [2]. A 2-sample t-statistic was used to compare the two mutation types, gene by gene, and p-values were assigned by a permutation-based simulation of the null distribution. The original data and code for preprocessing can be found at

<http://genomine.org/qvalue/results.html>.

The p-values from the BRCA1/BRCA2 analysis are included with the `qvalue` package as the dataset `hedenfalk`. To obtain the p-values, type `data(hedenfalk)`, and to view a description of the experiments and data, type `? hedenfalk`. We also check the length of the p-value vector and plot a histogram.

```
> library(qvalue)
> data(hedenfalk)
> length(hedenfalk)
```

```
[1] 3170
```

```
> hist(hedenfalk)
```

3 The qvalue function

The `qvalue` function computes q-values from the p-values of an experiment with multiple comparisons. We assign the output of the function call to the object `qobj` and plot a histogram of the q-values.

```
> qobj <- qvalue(hedenfalk)
> hist(qobj$qvalues)
```

4 The qsummary function

The `qsummary` function reports an estimate (π_0) of the proportion of genes for which the null hypothesis is true, and presents a table comparing p-values to q-values.

```
> qsummary(qobj)
```

Call:

```
qvalue(p = hedenfalk)
```

```
pi0:          0.6635185
```

Cumulative number of significant calls:

	<1e-04	<0.001	<0.01	<0.025	<0.05	<0.1	<1
p-value	15	76	265	424	605	868	3170
q-value	0	0	1	73	162	319	3170

5 The `qplot` function

The `qplot` function produces four plots:

- A plot of π_0 versus the tuning parameter λ (see [2]).
- A plot comparing p-values to q-values.
- A plot of the number of significant genes by q-value.
- A plot of the number of expected false positives by the number of significant genes.

```
> qplot(qobj)
```

6 The `qwrite` function

The `qwrite` function writes the output of the function `qvalue` to a file. If all defaults were chosen in the call to `qvalue`, the output file contains π_0 in the first row and a $3,071 \times 2$ matrix of p- and q-values in the rest. Note that `qwrite` writes to the file “my-qvalue-results.txt” in the working directory by default. The file name and location can be specified with the `filename` argument to `qwrite`.

```
> qwrite(qobj)
```

7 The `qvalue` Shiny app

A point-and-click version of `qvalue` can be accessed at <http://qvalue.princeton.edu/>.

References

- [1] I. Hedenfalk, D. Duggan, Y.D. Chen, M. Radmacher, M. Bittner, R. Simon, P. Meltzer, B. Gusterson, M. Esteller, O.P. Kallioniemi, B. Wilfond, A. Borg, and J. Trent. Gene expression profiles in hereditary breast cancer. *New Engl. Jour. Medicine*, 344:539–548, 2001.
- [2] J.D. Storey and R. Tibshirani. Statistical significance for genome-wide experiments. *Proceedings of the National Academy of Sciences*, 100:9440–9445, 2003.

Histogram of hedenfalk

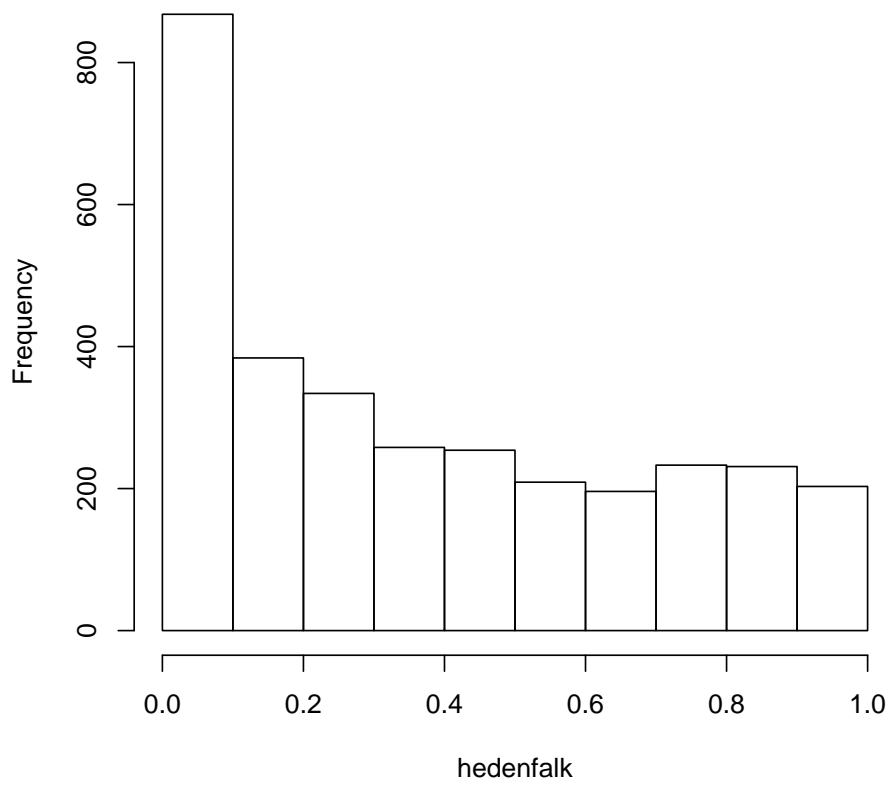


Figure 1: Histogram of p-values.

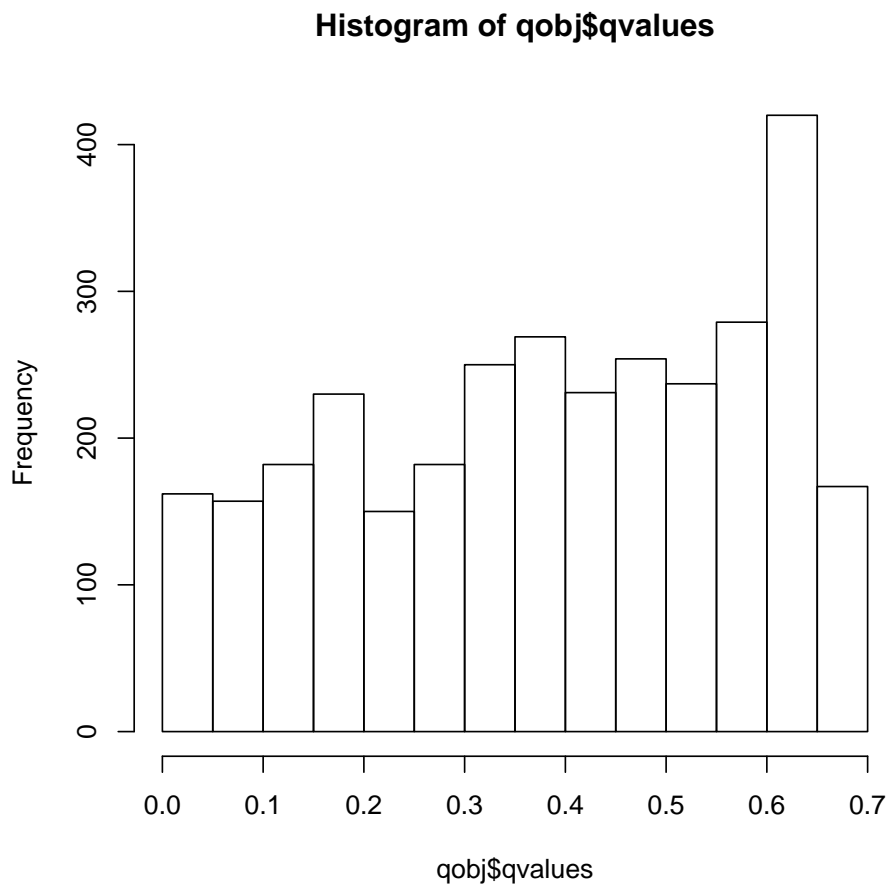


Figure 2: Histogram of q-values.

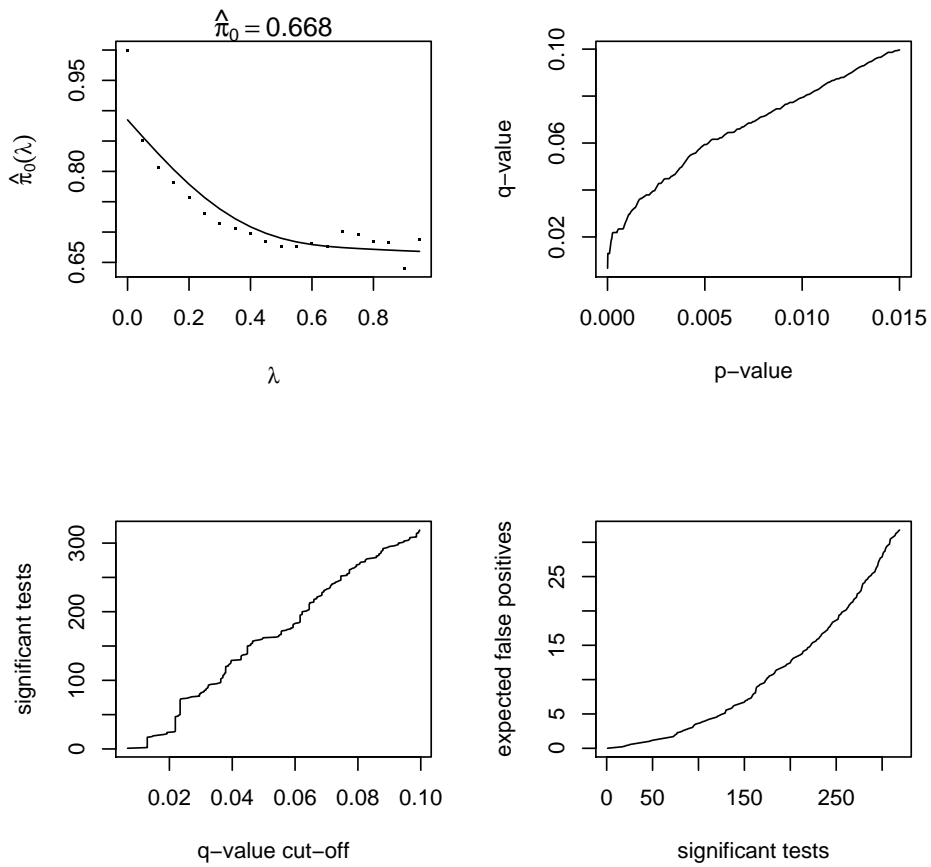


Figure 3: Miscellaneous plots.