fFDR Package Vignette (Version 0.1)

David G. Robinson

July 15, 2013

1 Introduction

This is a vignette for the `fFDR` package, which performs functional false discovery rate control.

2 Example Usage

Simulated data can be found in the `simtests` dataset.

```
library(fFDR)
data(simtests)

head(simtests)

##      pvalue sample.size   mean oracle
## 1 6.748e-02          25 0.3639   TRUE
## 2 2.729e-01          10 0.5518   TRUE
## 3 1.482e-01           2 0.3301   TRUE
## 4 2.406e-01           3 0.3093   TRUE
## 5 1.088e-01          25 0.3923   TRUE
## 6 1.527e-10          89 0.5138   TRUE
```

This dataset contains 4000 one-sample t-tests of different means and sample sizes. The means (while provided in the table) can be assumed to be unknown, while the sample sizes are known. Thus, we need to adjust the p-values from the t-tests based on the information we have about the sample sizes.

To do this, we use the `fqvalue` function, which takes at least two arguments: the p-values and the values it should be controlling for, which in this case is the sample size.

```
fq = fqvalue(simtests$pvalue, simtests$sample.size)
```

This produces a `data.table` with 6 columns, and the same number of rows as the number of pvalues given:

```r
fq
```

```
##        pvalue   X      qX    pi0 cumulative.density fqvalue
##    1: 0.06748  25 0.71525 0.5680             0.3448 0.11116
##    2: 0.27285  10 0.52400 0.7062             0.4438 0.43417
##    3: 0.14820   2 0.04688 0.8235             0.1631 0.74829
##    4: 0.24056   3 0.16112 0.8222             0.3074 0.64343
##    5: 0.10882  25 0.71525 0.5680             0.4226 0.14625
##   ---
## 3996: 0.20611   7 0.43188 0.8017             0.3028 0.54567
## 3997: 0.64535 263 0.96500 0.4936             0.8021 0.39712
## 3998: 0.04598  84 0.89300 0.5031             0.4629 0.04996
## 3999: 0.34497  12 0.56588 0.6369             0.5408 0.40628
## 4000: 0.22800   4 0.26700 0.8431             0.3257 0.59011
```

The first two, `pvalue` and `X`, are the original inputs, and `qX` is the quantile of `X` that was used in the functional FDR computations. `pi0` is the estimate of π_0 for each hypothesis, which varies depending on `X`.

The `fqvalue` is the functional q-value. This is designed so that rejecting the null hypothesis for all tests with q-value less than q should lead to a false discovery rate of q. The cumulative density is a value used in the computation of the functional q-value: it represents the estimated probability that a p-value at the given X_i would be less than p_i.

2.1 Analysis

The `fq$fqvalue` column shows us the computed q-values:

```r
sum(fq$fqvalue < 0.05)
```

```
## [1] 519
```

```r
sum(fq$fqvalue < 0.01)
```

```
## [1] 359
```

We can also perform a comparison to a traditional q-value, in this case using the `qvalue` package, and see how many we find:

```r
library(qvalue)

q = qvalue(fq$pvalue)
sum(q$qvalues < 0.05)
```

```
## [1] 478
```

```r
sum(q$qvalues < 0.01)
```

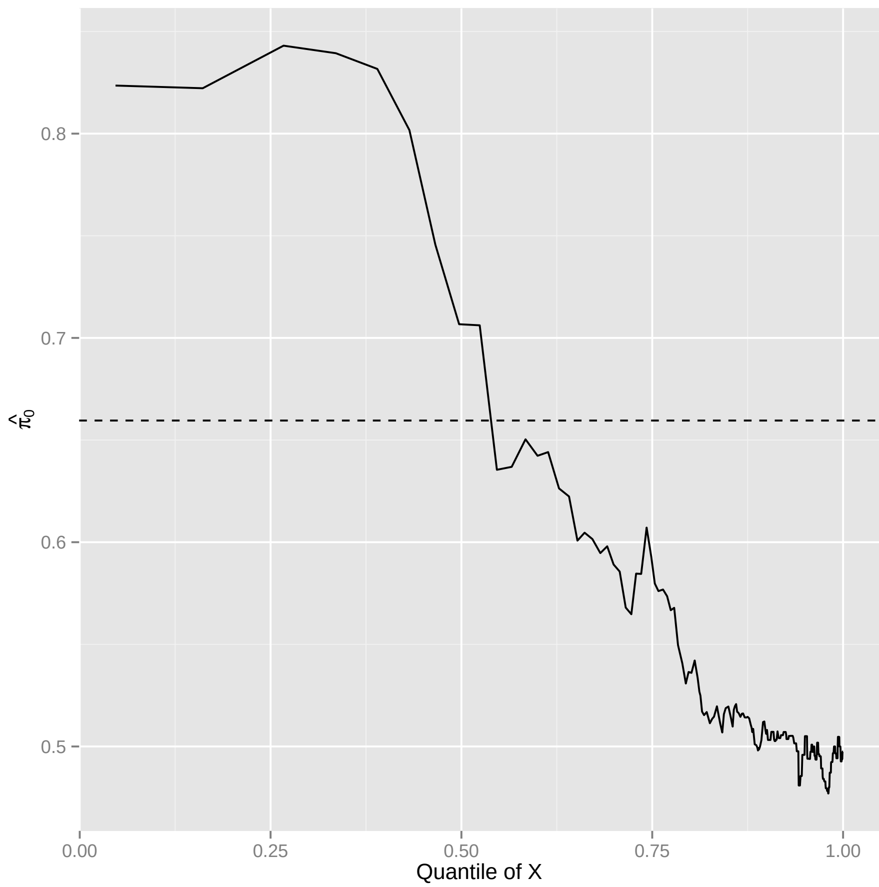
```
## [1] 333
```

3 Plotting

The `fFDR` package provides functions for plotting various useful graphs based on the output of`fqvalue`.

One of `fFDR`'s functionalities is to calculate how π_0 varies based on X. Plotting this π_0 based on the quantile of X can show how the

```
plot.pi0(fq, horizontal.line = TRUE)
```



Another question is to what extent the functional q-value differs from the traditional q-value, which does not take X into account.

Another way of comparing traditional q-values to functional q-values is to scatter the p-values against the quantiles of the X distribution.

```
compare.qvalue(fq)
```

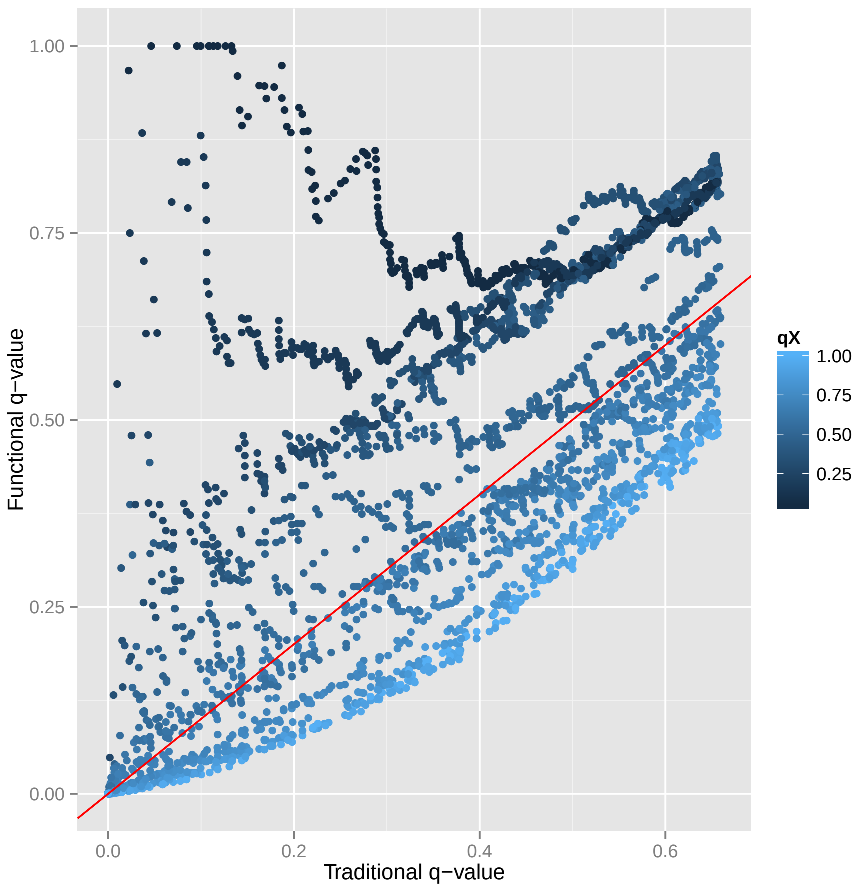


Figure 1: Comparison between the traditional q-value and the functional q-value, colored based on the quantile of the X distribution. Notably, values with high X have functional q-values that are lower than their traditional q-values, while values with low X have higher functional q-values.

```
plot.pvalue.qX(fq)

## Loading required package:  plyr
##
## Attaching package:  'reshape'
## The following object(s) are masked from 'package:plyr':
##
##     rename, round_any

##    qvalue    pvalue
## 1  0.005 0.0005095
## 2   0.01 0.0012332
## 3   0.05 0.0089293
## 4    0.1 0.0238818
```

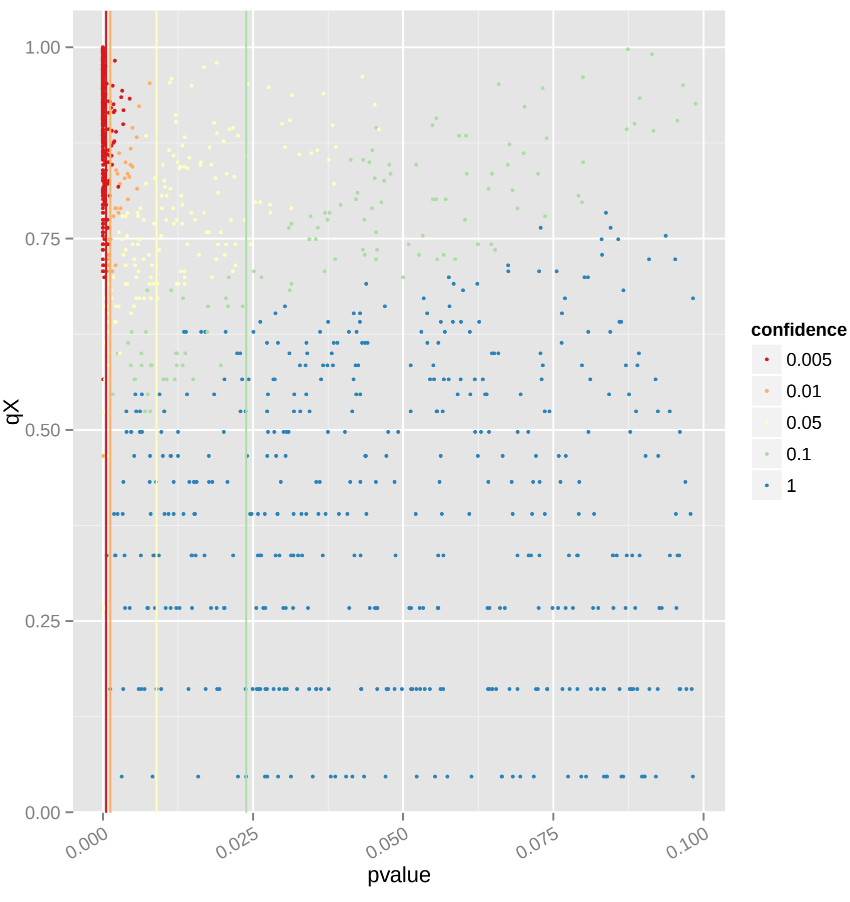


Figure 2: The vertical lines represent the q-value significance thresholds using traditional methods, in which the q-value depends only on the p-value and not X. The colors of the points represent the significance thresholds of the functional q-value.

```
plot.pvalue.qX(fq, doublelog = TRUE)

##   qvalue    pvalue
## 1  0.005 0.0005095
## 2   0.01 0.0012332
## 3   0.05 0.0089293
## 4    0.1 0.0238818
```

Figure 3: Equivalent to Figure 2, but places the p-value on a $\log(-\log(p))$ scale, which more clearly shows the effects on very low p-values.