The locfdr Package

February 16, 2008

Title Computes local false discovery rates

Version 1.1-6

Author Bradley Efron, Brit B. Turnbull and Balasubramanian Narasimhan

Description Computation of local false discovery rates

Maintainer Bradley Efron <brad@stat.Stanford.EDU>

License GPL 2.0

Depends splines

R topics documented:

`hivdata` *HIV data set*

Description

The data comprises 7680 z-values, each relating to a two-sample t-test. The test compares gene expression values for 4 HIV patients with values for 4 normal subjects; the t-score `T[i]` for gene i has been transformed to a normal scale, `z[i] = qnorm(pt(T[i], df=6))`, so that the `z[i]`'s theoretically would have a standard $N(0,1)$ distribution under the null hypothesis. The original experiment is described in van't Wout et. al. (2003).

Usage

```
data(hivdata)
```

Format

A vector containing 7680 z-values

References

van't Wout, et. al., Cellular gene expression upon human immuno-deficiency virus type 1 infection of CD4+-T-Cell lines, Journal ofVirology 77, 1392-1402.

`lfdrsim` *Simulated data set for locfdr*

Description

A simulated dataset that involves 2000 "genes", each of which has yielded a test statistic "zex", with $zex[i]$ $N(mu[i], 1)$ (independently for $i = 1, 2, ...2000.$) The data comprises 2000 μ_i values and 2000 z-values.

Usage

```
data(lfdrsim)
```

Format

A matrix of 2000 rows and 2 columns containing mu and the z-score values (`zex`)

`locfdr` *Local False Discovery Rate Calculation*

Description

Compute local false discovery rates, following the definitions and description in references listed below.

Usage

```
locfdr(zz, bre = 120, df = 7, pct = 0, pct0 = 1/4, nulltype = 1, type =
0, plot = 1, mult, mlests, main = " ", sw = 0)
```

Arguments

zz	A vector of summary statistics, one for each case under simultaneous consideration. The calculations assume a large number of cases, say length(zz) exceeding 200. Results may be improved by transforming zz so that its elements are theoretically distributed as $N(0,1)$ under the null hypothesis. See the locfdr vignette for tips on creating zz.
bre	Number of breaks in the discretization of the z-score axis, or a vector of breakpoints fully describing the discretization. If length(zz) is small, such as when the number of cases is less than about 1000, set bre to a number lower than the default of 120.
df	Degrees of freedom for fitting the estimated density $f(z)$.
pct	Excluded tail proportions of zz's when fitting $f(z)$. pct=0 includes full range of zz's. pct can also be a 2-vector, describing the fitting range.
pct0	Proportion of the zz distribution used in fitting the null density $f0(z)$ by central matching. If a 2-vector, e.g. pct0=c(0.25,0.60), the range [pct0[1], pct0[2]] is used. If a scalar, [pct0, 1-pct0] is used.
nulltype	Type of null hypothesis assumed in estimating $f0(z)$, for use in the fdr calculations. 0 is the theoretical null $N(0,1)$, 1 is maximum likelihood estimation, 2 is central matching estimation, 3 is a split normal version of 2.
type	Type of fitting used for f; 0 is a natural spline, 1 is a polynomial, in either case with degrees of freedom df [so total degrees of freedom including the intercept is df+1.]
plot	Plots desired. 0 gives no plots. 1 gives single plot showing the histogram of zz and fitted densities f and $p0 * f0$. 2 also gives plot of fdr, and the right and left tail area Fdr curves. 3 gives instead the f1 cdf of the estimated fdr curve; plot=4 gives all three plots.
mult	Optional scalar multiple (or vector of multiples) of the sample size for calculation of the corresponding hypothetical Efdr value(s).
mlests	Optional vector of initial values for (delta0, sigma0) in the maximum likelihood iteration.
main	Main heading for the histogram plot when plot>0.
sw	Determines the type of output desired. 2 gives a list consisting of the last 5 values listed under Value below. 3 gives the square matrix of dimension bre-1 representing the influence function of log(fdr). Any other value of sw returns a list consisting of the first 5 (6 if mult is supplied) values listed below.

Details

See the locfdr vignette for details and tips.

Value

fdr	the estimated local false discovery rate for each case, using the selected type and nulltype.

`fp0`	the estimated parameters delta (mean of f0), sigma (standard deviation of f0), and p0, along with their standard errors.
`Efdr`	the expected false discovery rate for the non-null cases, a measure of the experiment's power as described in Section 3 of the second reference. Overall Efdr and right and left values are given, both for the specified nulltype and for nulltype 0. If `nulltype==0`, values are given for nulltypes 1 and 0.
`cdf1`	a 99x2 matrix giving the estimated cdf of fdr under the non-null distribution f1. Large values of the cdf for small fdr values indicate good power; see Section 3 of the second reference. Set plot to 3 or 4 to see the cdf1 plot.
`mat`	A matrix of estimates of $f(x)$, $f0(x)$, $fdr(x)$, etc. at the $bre-1$ midpoints "x" of the break discretization, convenient for comparisons and plotting. Details are in the locfdr vignette.
`z.2`	the interval along the zz-axis outside of which $fdr(z) < 0.2$, the locations of the yellow triangles in the histogram plot. If no elements of zz on the left or right satisfy the criterion, the corresponding element of z.2 is NA.
`call`	the function call.
`mult`	If the argument mult was supplied, vector of the ratios of hypothetical Efdr for the supplied multiples of the sample size to Efdr for the actual sample size.
`pds`	The estimates of p0, delta, and sigma.
`x`	The bin midpoints.
`f`	The values of $f(z)$ at the bin midpoints.
`pds.`	The derivative of the estimates of p0, delta, and sigma with respect to the bin counts.
`stdev`	The delta-method estimates of the standard deviations of the p0, delta, and sigma estimates.

Author(s)

Bradley Efron, Brit B. Turnbull, and Balasubramanian Narasimhan

References

Efron, B. (2004) "Large-scale simultaneous hypothesis testing: the choice of a null hypothesis", Jour Amer Stat Assoc, **99**, pp. 96–104

Efron, B. (2006) "Size, Power, and False Discovery Rates"

Efron, B. (2007) "Correlation and Large-Scale Simultaneous Significance Testing", Jour Amer Stat Assoc, **102**, pp. 93–103

http://www-stat.stanford.edu/~brad/papers/

Examples

```
## HIV data example
data(hivdata)
w <- locfdr(hivdata)
```

Index