Differences to earlier versions of *twilight*

Stefanie Scheid[*]

Max Planck Institute for Molecular Genetics
Computational Diagnostics
Department for Computation Molecular Biology
Ihnestr. 63-73, D-14195 Berlin, Germany

1 Changes in version 1.2.0

We added the argument `filtering = FALSE` to function `twilight.pval` which (if set to `TRUE`) invokes the filtering for class label permutations that produce uniform p-value distributions. The set of admissible permutations is found using function `twilight.filtering` which is called internally in function `twilight.pval`. However, it can also be used directly.

We changed the local FDR estimation in function `twilight` slightly. Instead of estimating both densities f_0 and f from the output of SEP, we rely on the uniform assumption such that $f_0(p) = 1$ for all $p \in [0, 1]$. Hence, the 10 runs of SEP lead to 10 estimates $\widehat{\pi_0}$. The average of these is taken as the final value which is mutliplied with the density estimate of the mixture density f.

The mixture density estimation of f also changed slightly. Still, the estimates are based on smoothed histogram counts. To improve the estimation for very small p-values, the histogram bins were changed from equidistant to quantile bins.

2 Changes in version 1.1.0

The computation of p-values in `twilight.pval` changed from gene-wise to pooled p-values. For the computation of a gene-wise p-value for gene i, only the permutation scores of gene i are taken into account. For pooled p-values, all permutation scores of all genes are taken as null distribution. This change has several advantages: First, gene-wise p-values were not monotonically

---

[*]Contact: `stefanie.scheid@molgen.mpg.de`

increasing with scores because each gene had its own null distribution. Thus, two genes with almost equal scores might get quite different p-values. Now, the null distribution is the same for all genes, that is the union set of all permutation scores. Second, pooled p-values are less granular than gene-wise p-values. Gene-wise p-values are computed from `B` permutation scores whereas pooled p-values are computed from `B` · (number of genes) scores.

These two important features gave rise to further changes: The ordering of the `result` table is now more intuitive because the most significant genes on top have the highest scores, the lowest p- and q-values and are candidates (if there are any). In addition, the default value of the number of permutations `B` is lowered to 1000 permutations. Computation of pooled p-values is slower than for gene-wise p-values. On the other hand, changing to pooled p-values increases the number of values in the null distribution by the factor of number of genes. Hence, even with less permutations, the number of null values is larger than before.

We integrated Pearson and Spearman correlation coefficients into `twilight.pval`. Each gene is correlated to an numerical input vector. Expected scores are computed from random permutations of the input vector.

The `result` table contains an additional `index` column with genes indices which comes in handy for sorting back to original ordering.

All output matrices of the permutation functions `twilight.combi`, `twilight.permute.pair` and `twilight.permute.unpair` have the original labeling vector as first row. This is also the case if balanced permutations are wanted, although the input vector is not balanced. Hence, the permutation matrix within `twilight.pval` now includes the original labeling even for balanced permutations implying that the smallest *possible* p-value is 1/(number of permutations).

3   Changes in version 1.0.3

A `print.twilight` function was added which produces a short information about the contents stored in the `twilight` object.

4   Changes in version 1.0.2

The `which` argument of the plot command changed from `plot1` style to more intuitive labels like `scores` or `fdr`.