A tutorial dbEmpLikeNorm R package

Lori A. Shepherd†◇, Wan-Min Tsai*, Albert Vexler*, and Jeffrey C. Miecznikowski*†

April 19, 2013

Department of Biostatistics
University at Buffalo*
Department of Biostatistics and Bioinformatics
Roswell Park Cancer Institute†

◇las65@buffalo.edu

1 Introduction

The **dbEmplikeNorm** R package provides a function `dbELnorm` to be used for joint assessment of normality of k independent samples with varying means and standard deviations. The function provides the test statistic and associated p-values. The p-value can be calculated by Monte-Carlo methods or estimated based on precalculated tables of selected sample sizes and alpha (Type I error) values. For details and algorithms:

> Tsai WM, Shepherd LA, Miecznikowski J, Hutson A, Vexler A. (2013). An EL based test for normality in multiple groups. Department of Biostatistics. University at Buffalo. Report 1204.

2 Examples

The following performs an assessment of normality and calculates the p-value based on Monte-Carlo methods. The null hypothesis is that the data is normally distributed with, perhaps, different means and standard deviations.

```
> library(dbEmpLikeNorm)
> normData1 = rnorm(25)
> normData2 = rnorm(40,5,2)
> normData3 = rnorm(15,7,1)
> Lst = list(normData1, normData2, normData3)
> dbELnorm(x=Lst, pvl.Table=FALSE)

...Working on teststat
...Working on p-value
    This may take a few minutes
$teststat
[1] 14.66337

$pvalue
[1] 0.726

>
```

The p-value can be estimated based on precalculated tables rather than preforming Monte-Carlo methods. This is controlled by the argument 'pvl.Table'. To estimate based on tables set the 'pvl.Table' argument to TRUE, which is the default setting.

```
> dbELnorm(x=Lst, pvl.Table=TRUE)

...Working on teststat
Table calculated on a delta = 0.5
```

```
    To use delta= 0.05  please change pvl.Table to FALSE
estimating pvalue based on table
Tables based on sets of equal sample size
    Continuing with sample size of first element in list
$teststat
[1] 14.66337

$pvalue
[1] 0.7232562

>
```

The **dbEmplikeNorm** package also provides a function `returnCutoffValue` to be used for return the cutoff for significance for the statistic (log scale) obtained from the empirical likelihood based test described in [Tsai 2013]. The cutoff can be calculated by Monte-Carlo methods, estimated based on precalculated tables of selected sample sizes and alpha values, or by a Bayesian approach specified by 'MC.Method', 'Table.Method', and 'Bayes.Method, respectively . The following returns a 0.05 cutoff for a joint assessment of 4 groups with sample sizes 15,25,25,30 calculated via Monte Carlo methods.

```
> returnCutoffValue(numberOfgroups=4, sample.size=c(15,25,25,30),
+                   MC.Method=TRUE,
+                   targetalpha=0.05, num.mc=1000)

MC.Method.rslt
      26.50766

>
```

Note the significance level for the associated cutoff is specified by the user in 'targetalpha' and the number of simulations used to estimate the cutoff is controlled by 'num.mc'. The Monte-Carlo method is the default calculation. To use the table interpolation or the Bayesian method (see details in [Tsai 2013], user should set Table.Method=TRUE or Bayes.Method=TRUE). The following returns the cutoff for joint assessment of 2 groups each with sample size 20 calculated by each method.

```
> returnCutoffValue(numberOfgroups=2, sample.size=20, targetalpha=0.05,
+                   MC.Method=TRUE, Table.Method=TRUE, Bayes.Method=TRUE)

estimating cutoff based on table
Tables are based on equal length groups. Continuing using the first sample size
   For a more accurate calculation use monte carlo method [pvl.Table=FALSE]
estimating cutoff based on bayesian method
$MC.Method.rslt
[1] 13.69252
```

```
$Table.Method.rslt
[1] 13.25156

$Bayes.Method.rslt
[1] 13.45758

>
```