UNIVERSITA' DI TORINO
DIPARTIMENTO DI MATEMATICA
"Giuseppe Peano"
Via Carlo Alberto, 10
10123 TORINO – ITALY
Tel. 011 6702820/2809

May 28, 2022

Dear Editor,

this is the document that includes all the reviewer's comments (in italic) mixed with our responses.

We would like to thank the reviewers and the editor for having carefully read the manuscript and for her/his constructive comments. We address these comments point-by-point below, and their implementation can be found in the attached revised manuscript.

We believe that this could improved the final manuscript and its readability for readers.

Response to the Editor

> *The title is a bit awkward as is, and perhaps slight re-wording could make it more consistent with other R Journal articles: "kStatistics: Unbiased estimates of joint cumulant products from the multivariate Faà di Bruno's formula"*

Based on this suggestion, we have changed the title of the paper.

> *Most readers of the R Journal would not be familiar with this area, and it would be helpful if you can provide more motivation at the start of the paper. Provide some context of where this would be applied and used, signal processing? spectral analysis?*

Based on this suggestion, we have added more motivation in the Introduction. We have started by introducing cumulants, along with some references to fields in which they are applied. Then we have explained the usefulness of k-statistics and their generalizations by providing some application references. We have underlined that the proposed approach producing k-statistics in `R` relies on the use of suitable polynomial families which can be recovered from the multivariate Faà di Bruno formula's. This formula is a very general-purpose tool with many applications in different fields, for which we have provided some references. We have

also emphasized that in order to set up an efficient computation of this formula, the proposed approach in the package is of combinatorial type, contextualizing its use and novelty with respect to the existing literature.

> *Also think about migrating your paper to the new format for the R Journal, using the rjtools package*

We have migrated the manuscript to the new format. In the zipped folder, you will find all the necessary *.tex files, as well as the .html and .pdf files generated running *.Rmd.

Response to Reviewer 1

We have corrected all the typos suggested by the reviewer.

Response to Reviewer 2

> *Since R is employed by the vast majority of R users for data analysis rather than for symbolic algebra (in the manner of Maple or Mathematica etc.), this manuscript should focus on the numerical outputs which are suitable for data analysis, rather than on the symbolic outputs.*

Quite recently one of the authors of the manuscript would need to work in `R` with suitable families of polynomials, both for density approximation problems and for the computation of coefficients of cumulant generating functions having rather complicated expressions. A multivariate Faà di Bruno's formula in `R` would have been very useful. Indeed we had to stop processing our computations in `R`, use `Maple`, import the results into `R` and then continue. So we filled this gap by implementing the formula in `R`. The idea to update the k-statistics package with "symbolic" procedures came from this need, that maybe could be the same for others users too. We agree with the reviewer that the vast majority of `R` users works with data analysis but there is no doubt that, in recent years, the employment of symbolic methods has substantially extended both the theory of the applications of statistics and probability. We therefore find useful to show that it is possible to obtain and to use successfully symbolic outputs also in `R`.

> *As the authors have named the package kStatistics, there should be an emphasis on these k-statistics. Thus the sections 'k-statistics' and*

> *'Polykays' should be moved to be immediately after the introduction [...] The section 'Special families of polynomials', as they are symbolic results, can be moved to after the 'k-statistics' and 'Polykays' sections; and the former should be retitled as 'Bell polynomials' as this is more informative. The section 'Multivariate Faà di Bruno's formula' can be placed after all these, as it has a less direct application for data analysis, as it provides as the underlying mathematical framework, including the key step of the enumeration of scompositions, for the computations in the previous sections.*

We agree with the reviewer that a greater emphasis on k-statistics and their generalizations would increase the appeal of the manuscript. So we moved the 'k-statistics' and 'Polykays' sections after the Introduction, but first we put the section on multi-index partitions. The generation of multi-index partitions is the core of the package and a novelty in the combinatorial literature implemented in `R`. As most readers would be unfamiliar with these tools, we find it necessary to clarify their definitions at the beginning of the manuscript with examples as well. Otherwise the readers would not understand the expressions of the polynomials used in the rest of the manuscript. According to the suggestions of the reviewer, after the sections on 'k-statistics' and 'Polykays', we have inserted the section on Bell polynomials and their generalizations, also changing the name of the section. As requested, the section 'Multivariate Faà di Bruno's formula' is the last of the paper. Due to this change in the manuscript lineup, we have reconsidered the entire layout of the manuscript, but we believe that its overall readability is now improved.

> *As 'joint cumulant products' are mentioned in the manuscript title, there should be a more comprehensive treatment of them to emphasize the added value of the package outputs: a clear definition of joint cumulants, as well as their usage for data analysis, especially involving higher order k-statistics, beyond reporting their numerical values without a wider analysis context.*

As we have already written in the responses to the Editor, in this new version of the manuscript we have included a more comprehensive treatment of cumulants, recalling their main properties and referring to their applications in various fields, together with their usage in data analysis trough k-statistics.

In Examples 17, the sample mean, the sample variance, skewness and kurtosis are computed, but there is no explicit mathematical statement of how the k-statistics of orders 1 to 4 computed in kStatistics::nKS are related to these quantities. Similarly in Example 18 for the numerical values of the joint cumulants, the information that these joint cumulants provide about the data set should be elaborated. Likewise in Example 19 where the population variance is computed using the second order polykay (or product of mixed joint cumulants): since the variance can be conceptually understood without the polykay superstructure, another example of polykays should be added.

In this new version of the manuscript, Examples 17, 18 and 19 correspond to Examples 5, 6 and 7. According to these suggestions, we have made explicit the relations between k-statistics - of orders from 1 to 4 - and sample mean, sample variance, skewness and kurtosis (example 5). We have discussed the meaning of joint cumulant estimates (example 6). We have clarified the difference between the square of the sample variance and the polykay of order (2,2) (example 7).

In The nilde package [1] enumerates the scompositions of a integer i in nilde::nlde. Whilst Step (i) in kStatistics::MFB is more general since kStatistics::mkmSet enumerates the scompositions of a vector i, it is still important to cite and to make a brief comparison to nilde::nlde since it was first released on CRAN before kStatistics..

Based on these suggestions, at the end of the section on 'Partitions of a multi-index' we have added a remark on the `parts` function of the `partitions` package and on the `get.partitions` function of the `nilde` package, listing all the partitions of a given integers. We have pointed out the similarities and the differences with the `mkmSet` function of the `kStatistics` package. The differences in computational times are not significant (the `mkmSet` function is faster for a few seconds) and we have decided not to comment.

Looking forward to hearing from you as soon as possible.

Yours Sincerely,
Elvira Di Nardo
On behalf of the authors