In this document, we answer to the referee's comments about our paper entitled "ordinalClust: An R Package to Analyze Ordinal Data". The responses are highlighted in red and the changes in the revised manuscript too. We thank the referee for their valuable remarks that helped us to improve our paper and package.

1 General remarks

The authors introduce their R package "ordinalClust", which allows the use of several methods of clustering methods for ordinal data and classification using ordinal covariate data, which are in part introduced in this paper for the first time. In the first part of the paper the authors describe the statistical approaches and in the second part they use data examples to illustrate how their package can be used to apply the approaches described in the first part. Clustering of ordinal data is a relevant application case (consider, for example, the widespread use of ordinal scales in questionnaires) and the implemented approaches appear sensible. The presented application examples are easy to follow and enable the reader to apply the considered approaches to his/her own data. A literature review of related approaches is presented in the introduction. There do not seem to be other R packages dedicated to the task of clustering ordinal data. The package uses C++ code to speed up the time-consuming computations and I did not spot problems with the R code. Nevertheless, when playing around with the examples, I occasionally received error messaged that resulted from the algorithms producing empty clusters. These errors are, however, not due to errors in the implementation, but they are associated with the approaches themselves. See below for detailed comments.

2 Specific comments requiring revisions

1) The language of the manuscript is currently not yet acceptable. The authors should have the paper be proof-read by a native English speaker in order to fix the language problems, ideally someone roughly familiar with the topics presented in the paper. It was not possible for me to point them (all) out in the review. Some examples include "... the mode of the distributed and the other ones describes ...", "Other contributions implements algorithms ...", and "H is chosen thanks to a training data set and a validation data set:". The package documentation also suffers from language problems, which is why it should also be revised, at least in the long run. But the paper certainly needs language revision. There are also a number of careless mistakes.

The paper and the documentation were proof-read by native English speaker.

2) Abstract: "The latest stable package version is available in source and

binary form on the Comprehensive R Archive Network (CRAN)" CRAN packages are always provided in binary form and as a tar.gz file. Therefore, it is enough to just state that it is available on CRAN.

Thank you, this was changed.

3) Page 1: "The cumulative linked model (CLM) which assumes that:" The model is called "cumulative link model" not "cumulative linked model". Moreover, the sentence is missing a verb. The authors should also mention that:"$\beta_0(1) <= \beta_0(2) <= ...$".

Thank you, this was changed.

4) Page 1: "For instance, the ordinalForest package (Hornung, 2019) uses random forests" The ordinalForest package implements ordinal forests as presented in Hornung (2019); ordinal forests are different from classical random forests. Please write "ordinal forests" instead of "random forests" and, in addition to the R package ordinalForest that you already cited, please also cite the paper presenting ordinal forests: R. Hornung. Ordinal forests. Journal of Classification, 1-14, 2019.

Thank you, this was changed, the citation was added.

5) Page 1: "the other variables are of different types." Better "various" instead of "different", because the other variables may, of course, also include ordinal variables.

This was changed.

6) Page 2: "Multinomial distribution" Should be "multinomial distribution".

This was changed.

7) The references to the sections are done using numbers (e.g., Section 2.3, Section 2.4, Section 2.2.3), but the sections themselves are not numbered. This is not appropriate. The authors should use the section titles, when referencing (e.g. , "See Section "Statistical Methods"").

Thank you, the section titles are now used.

8) Page 4: "The E-step would require to calculate $G^N \times H^J$."$G^N \times H^J$ *terms*? Please specify that.

This was changed.

9) Page 6: Equation (2): The index of the sum should be written in bold font.

Thank you, the index is now written in bold font.

10) Page 7: “to take variables with different numbers of levels m into account” instead of ”to take variables with different levels m into account”

Thank you, this was changed.

11) Page 7: “Although it does not make possible to gather ordinal features with different m in a same column-cluster, it is able to take into account the fact that there are several m and therefore to perform a co-clustering on more diverse data sets.” ”the fact that there are several m” is not very clear. It should be expressed more clearly, what is meant here.

“different m” was changed to “different numbers of levels” to clarify this part.

12) Page 8: “it indicates how many cells at least must be present in a block” What does ”cells” mean in this context, clusters? That term has not yet been used before in the paper. Please clarify.

“Cells” referred to elements of the matrix. The sentence was rephrased.

13) Page 8: “v < − as.vector(dataqol.classif$death)” The use of as.vector is unnecessary here, because dataqol.classif$death is already a vector.

Thank you we changed it.

14) Page 9:

```
row.names <- c()
     for(kc in kcol){
     name <- paste0(”kc = ”,kc)
     row.names <- c(row.names,name)
}
```

While these lines produce the right result, using a loop here is not necessary and makes the code more complex than needed. This whole part can just be replaced by:

```
row.names <- paste0(”kc = ”, kcol)
```

Thank you, the loop was removed.

15) Page 9: "rownames(preds)=row.names "=" should be " ¡- " (for reasons of consistency)

Thank you, this was changed.

16) The authors set a seed of the random number generator to make their results reproducible. However, they seem to have run their code using an R version older than R 3.6.0, which is why users of more recent R versions will get different results, because the function "sample" works differently since R version R 3.6.0. The authors should state in the paper, which R version they used to run their code . They may also re-run their code using a more recent R version to make their code reproducible with respect to newer R versions. A different option would also be to use the line "RNGkind(sample.kind="Rounding")" in order to make sure that the older functionality of the "sample" function is used.

Thank you for you remark. Indeed, when running the code with R 3.6.0, the results are different. Not to change the results of the paper, we decided to state in the paper (at the beginning of Section "Application on the patients quality of life analysis in oncology") that the code was run with R 3.5.3. We also indicate that if the user wants to use a version $\geq$ 3.6.0, they can get the same results by running the command "RNGkind(sample.kind="Rounding")" before running the script. We also added this command at the beginning of the script for submission.

17) Page 9: "The two parsimonious models kc = 1 and kc = 3 obtains the best results. This illustrates the interest of introducing parsimonious models in a supervised context." The training data set consists of 28 observations and the validation data set of 12 observations. These numbers of observations are too small to draw reliable conclusions from these analyses. A warrant should be added that the numbers of observations used here are too small to draw definitive conclusions. Apart from this, using such small numbers of observations is not a problem here, because the code presented in the paper should serve illustrative purposes merely.

Thank you for your remark. We added a warrant so that the user knows the number of observations is too small to draw definitive conclusions.

18) When seeking to illustrate that the numbers of observations are too small to draw reliable conclusions from the results, I wanted to repeat the authors' workflow using different seeds than the seed 1 used by the authors. I tried the following seeds: 12, 123, 1234, 12345, 123456. However, for each of these I got error messages for at least one value of kc. These were: "Error in prediction(classif, x, seed) : Expecting a single value: [extent=0]. In addition: Warning message: In bosclassif(x = x.train, y = v.train, kr = kr, kc = kcol[kc], : Error: probably empty clusters"

Can you explain, what is going wrong here and possibly fix these issues?

The warning message “bosclassif(x = x.train, y = v.train, kr = kr, kc = kcol[kc], : Error: probably empty clusters” happens when the algorithm finds a spurious solution with empty column-clusters. We agree that the message is confusing: it says ’error’ when it is a warning, and it does not indicate what to do in this case. There are two types of solution to overcome the issue when getting this warning:

- Choose another initialisation. If the chosen intialisation was “kmeans”, the user can change the argument init to “random” or “randomBurnin”. If the init argument was already “random” or “randomBurnin”, the user can re-run the algorithm and the initialization will be different.
- Choose a smaller parameter kc. When the number of column-classes is high with respect to the number of features, the algorithm is more likely to get empty clusters. Choosing a smaller argument kc can also be a solution.

The warning message is now set to: “The algorithm found a spurious solution with empty clusters. You can: 1) Run the algorithm with another type of initialisation, 2) If you run the algorithm with init to “random” or “randomBurnin”, running it again will change the initialisation , 3) Run the algorithm with a smaller argument kc.”

The error “prediction(classif, x, seed) : Expecting a single value: [extent=0].” is closely related the warning message above. When the function bosclassif finds a spurious solution with empty clusters, it returns an object of class ResultClassifOrdinal, but with empty slots. The function predict does not check this. We added a function to check that the object of class ResultClassifOrdinal given to the function predict is not empty. When the object is empty, it returns a warning message saying: ”bosclassif returned an empty object, due to empty clusters. Predictions cannot be run.”

19) Page 11: “cluser@params is a list:” ”clust@params” instead of ”cluser@params”

Thank you, this was changed.

20) Page 11: “Thanks to the command object@icl,” ”clust@icl” instead of ”object@icl”

Thank you, this was changed.

21) Page 12:

```
\end{CodeInput}
\begin{CodeOutput}
```

This seems to be an artifact and should be removed.

Thank you, this was removed.

22) Page 14: “In Figure 7, it is easily observed that between the 100th iteration and the 150th iteration (corresponding to nbSEM=100 and nbSEMburn=150), the parameters have reached on their stationary state. Therefore, nbSEM=150 and nbSEMburn=100 were well defined.” Shouldn’t ”corresponding to nbSEM=100 and nbSEMburn=150” be “corresponding to nbSEMburn=100 and nbSEMburn=150”? “nbSEMburn” is the number of burn-in iterations discarded and “nbSEM” is the number of iterations used to approximate the estimates.

Thank you for your remark. Indeed, the part “(corresponding to nbSEM=100 and nbSEMburn=150)” is a mistake, this was changed in the manuscript. However, “nbSEM” is the total number of iterations. So the number of iterations used to approximate the estimates is “nbSEM-nbSEMburn”.

I was also confused by: “Therefore, nbSEM=150 and nbSEMburn=100 were well defined.”. “nbSEMburn = 100” was clear to me, as the plots suggest that the estimates become stable after 100 iterations. But it was not clear to be, how the choice “nbSEM = 150” would be a consequence (suggested by “Therefore,”) of the content of the preceding sentence. I agree that “nbSEM = 100” are enough, because, even when not throwing away the burn-in observations (i.e., setting “nbSEMburn = 0”), the curves are basically horizontal after 100 iterations. But the way the authors phrased this seems confusing, probably mainly because of the use of “Therefore,”.

The number of iterations used to approximate the estimates is “nbSEM-nbSEMburn”, (50 in this case). We see on the plots that after 100 iterations (the burn-in period), the parameters are very stable. Therefore in this case, we can use nbSEM=150 with nbSEMburn=100 the number of iterations for the burn-in period, and 50 iterations to estimate the parameters. The sentence was rephrased.

23) Page 14: In Section “Conclusion” the authors should take care that they only declare those approaches as new that were actually introduced in this paper; for example, ”First, it proposes a clustering and co-clustering framework based on the Latent Block Model, coupled with a SEM-Gibbs algorithm and the BOS distribution.” seems to suggest that clustering and co-clustering would have been introduced in this paper for the first time, which is not the case.

Thank you for your comment. We changed the term “proposes” by “implements”