Response to the Reviewers of "robslopes: Efficient Computation of the (Repeated) Median Slope" (RJournal 2021-48)

To Professor Simon Urbanek
Executive editor of The R Journal

Dear Professor Urbanek,

Thank you for the opportunity to revise our paper. We would also like to thank the other reviewer for the remarks, which have strengthened the quality of the package. Our revision addresses all the comments made by the reviewers, as detailed below. We have copied the reviewers' questions in black and written our answers in blue.

Answer to Reviewer 1

This is a very nicely written paper. I have no suggestions to improve the presentation.
Specific comments.

1. My only serious suggestion for the package would be to set it forth as a pair of functions robslope and robslope.fit, in the same fashion as glm/glm.fit, coxph/coxph.fit and many others. The first function has the standard 5 arguments of formula, data, subset, weights and na.action (along with your beta and verbose options), the second has x, y, etc. This makes it more accessible to users. I am speaking as someone who works entirely with clinical data, for which data.frames and formulas are much more natural.

Thank you very much for this suggestion. We have updated the package (see Raymaekers [2022]), which now contains the functions `robslope` and `robslope.fit`. They take the standard arguments, and the computation can thus now be executed on data.frames using formulas. We have added a small paragraph on page 5 of the main text explaining these wrapper functions, and have replaced the sample code on the same page so that it now uses the `robslope` function.

2. I had not appreciated before the connection between the median of slopes and concordance statistics. Using your dual space representation, chose a target value u, and compute values $v = ux - y$. The concordance statistic for x (slopes) and v is based on sorting them into concordant, discordant, and tied pairs. It turns out that concordant pairs are exactly those that have an intersection somewhere to the left of u, discordant pairs an intersection somewhere to the right of u, and tied pairs correspond to parallel lines. The concordance problem also has $\mathcal{O}(n \log(n))$ algorithms, see for instance `survival::concordance`.

Indeed, it is exactly this connection that allows for the fast computation of the Theil-Sen estimator and the repeated median slope. The connection appears in different ways. For example, Kendall's tau correlation is based on concordance statistics and the Theil-Sen regression line is such that the Kendall's tau correlation between the predictor and residuals is zero (as opposed to the Pearson correlation for OLS). The precise algorithms for the computation of the regression lines are somewhat more complicated than those for concordance statistics (especially for the repeated median), but they are very similar in spirit. With this paper, we hope to (re-)establish the connections between these concepts, in addition to providing an implementation of fast algorithms.

3. You do not mention what the solution is with case weights. It might be worthwhile to add a sentence, pointing to what the definition would be (assuming that one of the references mentions the topic).

Thank you for this interesting suggestion. The current references of the text don't mention a weighted version of the (repeated) median slope. In general, it seems that applications of weighted median slopes are (so far) rather rare. The most natural definition is the intuitive replacement of the median by a weighted counterpart. For the weighted repeated median, there is an application to time-series smoothing and filtering in Fried et al. [2007]. The weighted Theil-Sen was implemented in the `theilsen()` function of the `deming` package on CRAN (using a standard $\mathcal{O}(n^2)$ calculation of the slope). The papers on $\mathcal{O}(n \log(n))$ time algorithms for the (repeated) median slope do not mention any weighted version. It is not immediately obvious whether this can be done and how one would achieve this in $\mathcal{O}(n \log(n))$ time complexity, but we do consider it an interesting direction of future work as intuitively we feel this should be possible.

References

Roland Fried, Jochen Einbeck, and Ursula Gather. Weighted repeated median smoothing and filtering. *Journal of the American Statistical Association*, 102(480):1300–1308, 2007. ISSN 01621459. URL `http://www.jstor.org/stable/27639980`.

Jakob Raymaekers. *robslopes: Fast Algorithms for Robust Slopes*, 2022. R package version 1.1.2.