Response to Reviewers: AdequacyModel: An R Package for Probability Distributions and General Purpose Optimization

Manuscript ID: PONE-D-19-10117

Enclosed herewith is the pdf file of the revised version of our paper "AdequacyModel: An R Package for Probability Distributions and General Purpose Optimization" (co-authored by Rodrigo B. Silva, Marcelo Bourguignon, Gauss M. Cordeiro and myself), which we hope you will now find satisfactory. Firstly, we would like to thank the reviewers and the editor for your time and very constructive comments. We now answer all the comments made by the referees in the order they appeared in their reports and outline some other important changes made in the manuscript.

Reviewer #1: *Beacuse this paper provided a package in r for a general purpose optimization method based on the Particle Swarm Optimization approach and a set of statistical measures for the assessment of the adequacy of lifetime models for a given dataset. It seems to be suitable for Journal of Statistcial Software. No new idea is found in this paper.*

Reviewer #2: *The paper describes an R package (AdequacyModel) that uses particle swarm optimization (PSO) as a tool for general purpose optimization and for deriving maximum likelihood estimates for the parameters of probability distributions. The package also presents some goodness-of-fit tests. Such a computational library is of great interest for the statistical, mathematical, engineering communities, the paper is generally well written and organized. However, the results and examples provided in the manuscript do not clearly demonstrate the differences (advantages / disadvantages) of AdequacyModel when compared to other available packages (e.g. "pso") and PSO methods. Paper contribution should be clearly stated and supported by the results. The following issues should be addressed by the authors:*

1. *Lines 49-50: "We introduce a variation of PSO method in the AdequacyModel package for the R statistical computing environment..." and Lines 131-134: "The default optimization did not address the optimization problem restricted to a region R. In the course of the iterations, it is common for several particles to fall outside the search region R. The strategy of eliminating these particles and randomly relocating them in the search region increases the variability of the algorithm by preventing all particles from converging to a local minimum."*

   Response: Thanks for the opportunity to better explain such points. The main idea behind the proposed R package (AdequacyModel) is to provide a set of tools for the assessment of the adequacy of lifetime models for a given data set. Our contribution to the PSO algorithm consists to replace the particles that eventually fall outside the search region R, which is a subtle variation of the original approach. With this, we expect to keep the initial variability of the algorithm and prevent all particles from converging to a local optimum. We think that this strategy allows to find good solutions for problems of global function optimization with box-constrained. Note that this modification of the PSO algorithm is not our focus or the main contribution of this work. Therefore, the performance of the PSO algorithm implemented here is not compared with other PSO-based approaches in R.

2. *Ref. [1] suggests a standard PSO with the "let particles fly" strategy, in which a particle does not have its fitness evaluated if it is at an infeasible position and, by means of the social interaction with other particles, it tends to return to the feasible space in the subsequent iterations. Thus, what is the gain of using the proposed approach over the "let particle fly" approach? The numerical experiments should support the answer to this question.*

Response: Good point. Indeed, one possible method is not assess the fitness of the particles outside the search region and expect that these particles return to the search region according some social interaction with other particles as we can see in Bratton and Kennedy (2007). However, many problems involving likelihood-based inference require numerical constrained optimization. For example, the log-likelihood function is maximized subject to the constraint that the parameter of interest takes on the null-hypothesized value in the likelihood ratio test. In such problems, replacing the particles outside the feasible search region is a way to keep the initial variability of the algorithm.

3. *The functions considered in the AdequacyModel could be successfully optimized by the pso R package?*

   Response: We believe that the `pso` and **AdequacyModel** packages have similar performance for the functions considered in the paper. However, we would like to make clear that is not the point of the paper to build a variation of the PSO algorithm that provides a improvement over the **pso** R package, even because the **pso** package was developed after ours with a more general purpose. Our contribution to the PSO is to provide more control over some aspects of the algorithm, such as the number of particles and a conditional stop criterion, which is based on a minimum number of iterations and the variance of a given proportion of optimal values. Further, our proposed package allows to easily enter with a data set for which the objective function makes use. Rather than focusing in the PSO itself, what we intend to do is to introduce an easy-to-use set of statistical measures to assess the adequacy of lifetime models for a given dataset, using the PSO as the underlying optimization method. In addition to maximum likelihood estimation, the package provides some useful statistics to assess the goodness-of-fit of probabilistic models, including Cramér-von Mises and Anderson-Darling statistics. These statistics are often used to compare non-nested models. You can also calculate other goodness-of-fit measures such as the AIC, CAIC, BIC, HQIC statistics and the KS test, all this through the `goodnes.fit()` function. The usefulness of the package yielded a quotation from Nadarajah and Zhang (2017), published in the Plos One: *"there are other packages for fitting univariate distributions, for example, the R package `fitdistrplus` due to [79]. But none of these packages give as much output as [78]* (**AdequacyModel**) *gives."*

4. *It seems that each example of the paper considered a single initial point, but random initializations should be tested to assess their influence on the ability of PSO in providing accurate and precise optimal points / estimates.*

   Response: Thanks for the observation. The Section 5 of the paper presents two Monte Carlo simulations (20,000 replicas), in which for each of the simulations, an objective function was considered. Both simulations proved to be consistent regardless of the initial kicks.

5. *Lines 215-231 and 245-261 present the same extract of code.*

   Response: Thanks for the observation. Done.

6. *Marker sizes in Fig. 1 should be increased. For example, PSO answer is difficult to be observed, mainly in grayscale.*

   Response: Thanks for the suggestion. Done.

7. *Line 337: Ref. "[3]" is in the beginning of the sentence.*

   Response: Thanks for the observation. We deleted the reference.

8. *Line 391: Is S = 500 the number of particles or of iterations? By default, the number of particles is 150 (Lines 175-176).*

   Response: Indeed, we made a mistake here typing `S` (number of particles) instead `N` (minimum number of iterations). Therefore, we changed the sentence to: "An interesting fact is that the `pso` function also failed to get good estimates for `N = 500` iterations. However, the problem is easily circumvented by increasing the number of iterations."

9. *Line 409: What is the idea of the passage "However, not always the assumption that F is adequate."?*

   Response: We apologize for this. The sentence does not make sense. We rewrite the sentence as follows: "Some statistics are commonly used to verify the adequacy of the cdf F_θ to fit the observations."

10. *Lines 468-469: The authors claim that two examples of the use of the goodness.fit function are provided. However, these two examples are not clear. Are they in subsections 4.1 and 4.2? If the answer is yes, where is the goodness.fit function used to obtain the TTT plot?*

    Response: Thank you for the observation. In this new version of the paper, we added a new example. Now, the paper has two examples of the use of the `goodness.fit()` function. Please, see Section 4. Furthermore, to obtain the TTT curve, we use the TTT function, not the `goodness.fit()` function.

11. *Only one probability distribution is considered (Exponentiated Weibull). Examples with other probability distributions that are related to likelihood functions with complex search spaces should be given.*

    Response: Thank you. Done. We added one more empirical example considering the Kumaraswamy Beta distribution, which is well-known for presenting a likelihood function with complex search space. Please, see the Section 4.

12. *The paper does not have a Conclusion section. What are the main findings of the work, what limitations and new perspectives should be investigated?*

    Response: We added a conclusion section according to the reviewer's suggestion.

Reviewer #3: *In this paper, the authors develop an R Package AdequacyModel for Probability Distributions and General Purpose Optimization. The test results are also provided. In general, this paper is technically sound. But it is not well-organized. Here there are the comments of this reviewer:*

1. *The abstract requires revisions to reflect the key ideas of this paper as well as the details of the proposed approach.*

   Response: we rewrote the abstract according to the reviewer's suggestion.

2. *The original contributions are not clearly presented. In addition, the deficiencies of existing works need to be further summarized.*

   Response: Many sentences of the manuscript have been carefully rewritten.

3. *The simulation results should be extended to validate the presented method sufficiently. Authors need to add more results to thoroughly support the main findings.*

   Response: Thanks for the observation. Several improvements to the text have been added and another example of using the `goodness.fit()` function has been added. In addition, Monte Carlo simulations (two simulations) were performed to test the efficiency of the `pso()` function over two objective functions (Rastrigin and Himmelblau's functions).

   Also attached was the code of the simulations so that the results can be reproduced. The code makes use of parallel computing (multicore), allowing results to be reproduced in a shorter time.

   The function `pso()` also provides good results when submitted to the optimization of the Himmelblau's function. In this case, particles are attracted to the valleys containing the four global optimum points. In addition, it has been realized that the `pso()` function may be useful in detecting cavities on a surface (valleys detection). This is most likely due to the fact that particles are randomly replaced in the search space, which allows the particles to divide into groups that will not necessarily be attracted to the same cavity.

   The PSO algorithm given in Section 2.2 and encoded in the `pso()` function of the **AdequacyModel** package presented satisfactory results in obtaining global minimums in both cases. For the case of the Rastringin function, not always the best solution is obtained, but good solutions are reached. The dot cloud, as shown Figure 11(a) in paper, has been concentrated in regions with good candidates to the point of global minimum.

4. *The conclusion should be added. This reviewer strongly suggest to improve the flow of conclusion section. Start with a brief explanation of the paper goal (like the abstract), but make sure that the conclusion should be different from the abstract. Provide the main findings/claims. Explain the numerical findings of the simulations. Clearly explain, what are the significant findings and why your paper is really important. Then finalize your conclusion by providing one or two suggestions for future work.*

   Response: Thanks for the observation. We added a conclusion section according to the reviewer's suggestion. As with all changes made, the text is in red in the paper. We mention that one of the future modifications to the package will be in the `pso()` function which will be rewritten in C++ to try to improve the computational performance of the function.

5. *There are some expression errors. Pleased double check it.*

   Response: Thank you for the observation. We fully improve the text.

Reference

NADARAJAH, Saralees; ZHANG, Yuanyuan. Wrapped: An R package for circular data. PloS one, v. 12, n. 12, 2017.