Dwueng Chwuan Jhwueng <djhwueng@umail.iu.edu>

Decision on your PeerJ submission: "An improved tree-based statistical method for genome-wide association study" (#2018:08:30503:0:3:REVIEW)

PeerJ <peer.review@peerj.com> Mon, Sep 10, 2018 at 3:58 AM
Reply-To: PeerJ <peer.review@peerj.com>
To: Tony Jhwueng <djhwueng@umail.iu.edu>

PeerJ

Dear Tony,

Thank you for your submission to PeerJ.

It is my opinion as the Academic Editor for your article - An improved tree-based statistical method for genome-wide association study - that it requires a number of **Major Revisions**.

My suggested changes and reviewer comments are shown below and on your article '*Overview*' screen.

If you address these changes and resubmit, there's a good chance your article will be accepted (although this isn't guaranteed).

Resubmission

1. **Use the line numbers in your review PDF** when reading the comments from your editor and reviewers, and when writing your rebuttal letter:

Download review PDF

2. Download your resubmission checklist:

Download checklist

3. Edit and resubmit when ready:

Edit and resubmit

Although not a hard deadline, we expect you to submit your revision within the next 55 days.

With kind regards,

Joseph Gillespie
Academic Editor, PeerJ

If you're on twitter, you may want to share:

> **I just received first decision on my article in 15 days! I highly recommend #OpenAccess publishing at the Life and Environment Journal @thePeerJ**

Send Tweet

(you can edit before sending)

Editor comments (Joseph Gillespie)

MAJOR REVISIONS

Dear Dr. Jhwueng:

Thanks for submitting your manuscript to PeerJ. I have now received two independent reviews of your work, and as you will see, the reviewers raised some concerns about the research. Neither reviewer is at all optimistic with the current version of your manuscript. However, if you choose to do so, you may address the issues raised by both reviewers and submit a revision. I cannot guarantee that these reviewers will be willing to evaluate your work again, but a resubmission will again be fully reviewed.

Your manuscript appears to lack a through literature cited section. Very important works relevant to you study are missing and/or not adequately discussed. The English is also difficult to follow at times. Most importantly, it is difficult to follow much of your methodology since it is poorly described. Please think about how the overall presentation of your methodology can be made clearer. Maybe a descriptive flow chart will help? The presentation of results also needs much improvement. As in all modeling studies, your work will greatly benefit from analysis of a real biological dataset.

Please also ensure that you thoroughly distinguish your new methodology from that of others, especially the work of Thompson and Kubatko's.

I look forward to seeing your revision, and thanks again for submitting your work to PeerJ.

Good luck with your revision,

-joe

Reviewer 1 (Anonymous)

Basic reporting

- The English needs to be improved.
- A large part of the literature is missing.

Experimental design

- Both the simulated and real data are inappropriately described. It is difficult to follow.
- The model described in (7) is a Gaussian Process (GP) model. These kind of models have been used frequently in genetics, usually under the name of RKHS regression (see Morota and Gianola, Front. Gen. 2014, 5:363 for a review). I doubt that that the clustering matrix D improves inference, but this needs to be shown.

Validity of the findings

- I doubt that the analysis of the simulated data is relevant. The p-values in Figure 4 spans the whole range from 0 to 1. I suspect that this means that we are looking at

random data which makes no sense at all. You can definitely not draw the conclusion that BM and OU are different.
- How do you get to Figure 7 from model (7)?

which

Reviewer 2 (Anonymous)

Basic reporting

This paper proposed a method which to associate SNPs with a continuous trait that evolves according to an Ornstein-Uhlenbeck process. The method is a direct extension of one proposed by Thompson and Kubatko (2013).

I think that the author didn't explain the background sufficiently. Specifically, the author only explained the method proposed by Thompson and Kubatko (2013). However, there should be many more papers on this topic. I suggest the author do a more thorough literature review. Also, the author didn't explain clearly how to obtain the tree and only briefly mentioned that he used SVDquatets method in the Simulation section.

The author should provide more details regarding the proposed method. In this paper, the OU model is explained but the interaction between SNPs and the trait isn't. As a consequence, I have a hard time to understand why the proposed method will work.

I don't understand why in Algorithm 1, Y is simulated again in line 6 and 9.

Experimental design

I think the simulation studied is not sufficient to support that the new method is better than Thompson and Kubatko's method. The author focused on quantile of sigma^2 (Table 2) and quantile of alpha (Table 3) and the increasing of p-value (Figure 4 and 5). These are not the indication of the performance of the method. The author needs to focus on the percentage of detection and localization distance (just like the simulation study in Thompson and Kubatko 2013).

Validity of the findings

The explanation in the real-data analysis is lacking. At least, the author should discuss how many signals are detected in the analysis.

Comments for the Author

Some typos:

Line 28: Thompson and Kubatko (2013) -> (Thompson and Kubatko, 2013)

Line 33: categoy -> category

Line 90 and later: y_{i,t} -> y_t

Line 118: appraoch -> approach

Line 126: phylgogenetic -> phylogenetic

Line 134: non-ultrametric tree -> untrametric tree

Article ID: 30503

Need help? Just reply to this email.

Publisher of: *PeerJ – Life & Environment*, *PeerJ Computer Science* and *PeerJ Preprints*

Follow us on Twitter, Facebook and our Blog

© 2018, PeerJ, Inc. PO Box 910224 San Diego, CA 92191, USA