Reviewer's report

Title: Permutation-based variance component test in generalized linear mixed model with application to multilocus genetic association study

Version: 2
Date: 16 February 2015
Reviewer: Lucie Biard

Reviewer's report:

Zeng et al. propose a test to investigate simultaneously the effect of multiple markers within a gene on a binary outcome, in the framework of generalized linear mixed models. In the settings of multiple independent random effects, they present a procedure based on a quasi-likelihood ratio and a permutation test.

Permutations tests for random effects in mixed models have been discussed previously, in various settings of hierarchical data and genetic analysis. However, Zeng et al.'s work presents the advantages of such a procedure (straightforwardness, ease of implementation, and most of all good statistical properties), in contrast of available methods. This paper may help bring attention to this tool.

Major Compulsory Revisions

1. Authors could briefly present the rationale for choosing random effect modelling (and a global test for the effect of a group of markers) instead of relying on more traditional fixed effect modelling and testing. It would provide a useful reminder for readers that may not be familiar with genetic studies and their issues.

2. In the same manner, in the interest of readers unfamiliar with mixed-effect models, variables X and especially Z, could be clearly identified as \((n \times p)\) and \((n \times K)\) design matrices respectively, from the beginning (P5L32), consistently with the upcoming notations and matrices dimensions on page 6 (cf. P6L23 for instance).

3. To provide a useful comparative review of the field, they could compare their method to existing permutation procedures, such as Goeman’s globaltest. This comparison may add informative simulations results to their work and provide ground for discussion regarding the power limitations of the proposed test.

4. The use of the 0.65:0.35 \(\#^2\) mixture could be argued and discussed by the authors. It was suggested by Pinheiro and Bates, in a linear mixed model for the analysis a specific dataset of longitudinal data. This approximation of the LRT distribution under the null hypothesis may well not apply to the present settings.

5. Simulation study. Data generation should be described so that the simulation study is reproducible; please clarify the simulation sequence for Z and X1.

Overall, simulation \#2 seems to address 2 (if not 3) issues at the same time,
which may be confusing in interpreting results: power of the test in the case of mis-specified estimation model regarding the distribution of the random effects, power of the test when not all biomarkers (k<K) have a significant effect and power of the test with #² variations (how #² is different from simulation#1?). These issues could be addressed separately, first:

Effect of a “shape mis-specification”: a simulation set, under the alternative hypothesis, with K # sampled from a distribution with the same mean and variance (0.15) but a different shape as simulation 1,

Effect of #² value on the power of the test, if appropriate,

Effect of the presence of non-significant markers among the K modelled markers.

6. The motivating example dataset is a “hybrid of simulated and real data”, as stated by Almasy et al. The structure of these datasets amounts to 200 simulated datasets generated from a common distribution. For this GAW17 analysis, could you provide the proportion of k<K significant markers within each gene, as you did for simulation #2?

Table 4 is a bit confusing since counts are reported instead of proportions, contrary to previous tables. Also, proportions are not reported in the text either. Proportions would allow comparing the results to those in the other simulations more easily, for instance regarding the influence of small K on power. Authors should either provide the proportion of p-values less than #, as mentioned in the table footnote along with counts, or else change “proportion of p-values” to “number of p-values” in the footnote.

P10L13: “the non-causal SNPs included in the gene lead to the reduction of power” is contradictory to P9L7 “more simulations (data not shown) indicated that other choices of k do not change the conclusions”. Could you discuss in more details your reason #(ii) for the observed low power ?

P10L13-20 addresses some potential limitations of the proposed test; it could be part of the discussion section instead of results section.

7. Overall, the discussion section could be slightly enriched, to address possible extensions of the method: for instance, testing for a subset of markers? correlation between markers?, interaction with fixed effect variables (predictive factors such as age or interaction with a treatment in the settings of clinical trials); also to address alternatives to LRT (score tests?), along with existing similar procedures in other frameworks (e.g. Goeman J et al. Bioinformatics, 2004;20:93, Michiels S et al. Statistics in Medicine 2011;30:1502).

One of the assets of this work is the straightforwardness of the procedure, despite non negligible computing time. It would be helpful for the reader to provide more details on the implementation: software platform, package/command, etc. Also, as a way to illustrate the computational issues of the permutation procedure, an example of the needed time for the analysis of a dataset could be reported (with the number of permutations performed).

Minor Essential Revisions:
1. P6L8: “b=0" --> “#=0”?
2. P6 equation (3): + instead of –, between the 2 terms in square brackets? (. Cf. Breslow & Clayton
3. P6 L23: in your notation (# for the fixed effects and # for the random effects), the 2nd equation presented on line23 gives # # rather than # # (Breslow & Clayton equation (10))
4. P7 Algorithm table: This table should be understandable by itself, without the text. Therefore, it may be helpful to specify how new datasets are generated (step 2). Also, title: no need to enumerate algorithm “1”, since there is only one presented in the manuscript. Homogenize cases: “step1”, “Step2”, etc.
5. P8L15: Is it common in genetic analyses to assume a coding {0,1,2} for genetic markers (of minor alleles) and corresponding underlying hypotheses regarding the effect of such a marker?
6. P8L27: Strictly speaking, in order to estimate the type I error rate, samples are generated under the null hypothesis that is #²=0 here, not #=0.
7. P8L28, add “for power” to the second part of your sentence (as you specified “for type I error rate” in the first part).
8. P9 Data analysis: could you please confirm which set of GAW17 was used (unrelated indiviuals?) and which fixed-effect variables were included in your analysis, if any?
9. Spelling:
   P3L24: “guaranteed in practice;” --> “guaranteed in practice,”
   P5L3: “where multiple makers”--> “where multiple markers”
   P9L3: “casual" --> “causal”
   P9L13-14: “based on” is repeated twice
   P9L15: “smoke”--> “smoking status”
   P9L16: Upper case for “the details of this…”; “details... were” instead of was
   P9L19: the proportion of p-values, plural
   P9L28: “the tests based mixtures” --> based on mixtures?
   P9L16: “the number of … become” --> becomes
   P12L2: p values, plural
   P12L11: “utilize” --> “use”
   P12L18: “Whereas” --> “Still” ?

Discretionary Revisions
If available, a nice addition would be the analysis of a real dataset.

**Level of interest:** An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests.