Response to Reviewers' Comments

Maroufy, V. and Marriott, P.

July 3, 2018

We would like to thank the Editor and the two referees for their hard work and their extremely helpful comments. We feel we have fully addressed all the comments and that this has considerable improved the paper and added clarity to the presentation. This letter details the responses. For clarity we have included the relevant parts of the referees' comments, in italics, before detailing our response.

1 Referee 1

1.1 Major Comments

Referee 1 writes in Point 1

> *1. Here is a naive idea for extending a gamma frailty model in which the frailty distribution is modeled by a finite mixture of gamma distributions, cf. Hougaard, P. (1995). Frailty models for survival data. Lifetime data analysis, 1(3), 255-273. What is a difference between such modeling and the local mixture modeling?*

Modeling unknown frailty using discrete mixture of gamma distribution in Hougaard, P. (1995) is without a doubt a great idea and is mentioned at the start of §2.1. However, finite mixture models in general are difficult to work with because of unidentifiability, dimensionality issues (unknown number of component) and non-concave likelihood functions. Indeed it is precisely to address these issues that we advocate using local mixture models when the mixing is 'small'. We do note that the three parameter model created from a gamma and an Inverse Gaussian model, although useful and flexible, it is still a specific model and does not relax the distribution assumption. We do think this is a useful approach and the paper is now cited in the manuscript, in the introduction.

> *2. Authors claim in the real data analysis that that the gamma frailty model has the effect of considerably under-estimating standard errors when compared to our more general assumptions. Is there any theoretical support for the assertion, building any asymptotics for investigating the consistency and asymptotic normality, cf. Zucker, D. M., Gorfine, M., & Hsu, L. (2008). Pseudo-full likelihood estimation for prospective survival analysis with a general semiparametric shared frailty model: Asymptotic theory. Journal of Statistical Planning and Inference, 138(7), 1998-2016.*

This is a very interesting question. Thanks for this comment, in fact (standard) asymptotic approaches are surprising subtle for the following reasons. The cost to our approach is that our model includes boundaries. Boundaries may cause maximum likelihood to not be turning points and asymptotic methods in sample size break down, ultimately, irregardless of their order. *A*naya-Izquierdo et al (2014) give a diagnostic approach to identify when a first order asymptotic is appropriate, depending on the square distance of the MLE from the boundary, measured using the Fisher information.

We have included this discussion in §2.1 and referenced the paper that you mentioned which we found very interesting.

> *3. The number of component is fixed to four in a prescribed manner. Is there any other option to select the component number based on data?*

The model with order $k = 4$ is flexible enough for our analysis and, as has been illustrated in *Marriott (2002)*, simply increasing the order of LMMs does not significantly increase flexibility. Nevertheless, all the results and algorithms can be generalized to higher dimensions. This is discussed on page 6.

1.2 Minor Comments

Many thanks for your thorough read. All the typos and errors have be fixed in the revised file.

2 Referee 2

Referee 2 asks

> *1 In the abstract you say 'we examine if there is a cost to having the gamma as a default option'. This could be examined in a simulation study that generated data using something other than the gamma. The simulation study generates data using the gamma and so shows the cost of using the LMM approach.*

This is a great point, thanks for that. Although, the theory of LMM and discrete mixture of LMMs is much broader, in the application here we only address one specific question: 'how are estimates and inference influenced when a specific model is postulated within a larger space of feasible models?'

Assuming a relatively small variation for the frailty term, the space of one LMM components includes *any* mixture model with a specifically selected mixing distribution. Hence, the simulation study leads to two observations: first, LMM model is capable of returning the same bias as the true model; second, the cost to this achievement and model flexibility is paid by a larger standard deviation. We have chosen to compare with the Gamma model because it is widely accepted in the literature and has a closed form solution.

To clarify this we have considerably added to the simulation study in Section 3.1 adding other commonly used frailty models. This nicely confirms our results.

2. Also in the abstract, 'We show that the gamma frailty assumption has the effect of considerably under-estimating standard errors'. This would require examining a model where the frailty is not described by a gamma distribution.

The extended simulation study does this now.

Under the assumptions of the example in this paper, the LMM model includes gamma frailty as a specific model with gamma mixing model. Hence, one would expect the same bias under both models, and a bigger variation under the LMM model, because it is more flexible. The same argument will apply to any specific frailty model which satisfies the conditions of being a local mixture.

3. 'which can be used as a bench-marking tool.' This is not clear. Just because a method makes fewer assumptions does not mean it is an appropriate benchmark. Parametric models continue to be useful even when assumptions do not hold exactly. Parametric models fail when there are egregious violations that are nevertheless of scientific/practical interest. Simulations using such models would be instructive.

Great comment, and we agreed in general. In general of course wrong specific assumptions would lead to bias, and less flexibility leads to smaller variation. Hence, between two models which both return the same bias, the more flexible model would be a better choice as it would return a less restrictive and more representative variation when the more specific model has been chosen purely for mathematical convenience. This is what we mean by 'bench-marking'. As mentioned we have added more simulation studies (Tables 2 and 3) in the manuscript, with frailty model generated form different models, confirming this intuition.

4.
– belongs to a one parameter exponential family, would be clearer
– does this depend on the parameterization θ? If so, the definition should include the parameterization. If not, make it clear that this is a geometric quantity, i.e., parameter invariant.
– Why $\geq$ and not $>$ in the definition of Λ_ϑ. For discrete spaces using $\geq$ means the support changes with some parameter values.

Respectively:

– LMMs are also defined for higher dimensions, specifically we have initial results on LMM of a Normal distribution $N(\mu, \sigma)$ with respect to both parameter, $\theta = (\mu, \sigma)$.
– Clarified in the definition, It does not depend on the choice of parameter, see Marriott (2002) and but has a clear interpretation when the mean is chosen as the parameter to be mixed over. These comments have been added on page 5.
–We chose $\geq$ to be consistent with the general definition of density and distribution functions and, as in related work by Marriott this reflects the geometry of extended exponential families which we feel is the appropriate one for general information geometry structures.

5. Example 1. "... gives excellent L1 approximations for all mixtures Q with support in [0.3,5]."
– Quantifying this would be good.

Quantification is easily given by direct calculation of the Talyor's expansion error.

6. "we only make the comparison to a one component local mixture with $k = 4$."
– It would seem the claim of excellent L1 approximations no longer applies. Please clarify.

The application of this paper focuses on small frailty variation, for which one LMM would be suitable. The order $k = 4$ has been shown flexible in applications (Marriott, 2007), and higher orders do not add significant flexibility. Furthermore, we are working on the generalization to relax the restriction of small variation in the application. One open problem is finding a good estimate for the cumulative hazard function when a L1 approximation to a general mixture is used. If we can work around this restriction then we will be able to run more general simulation studies and tackle more general questions related to frailty. Nevertheless, the theory is all well-developed and we have applied that in other areas, such as sensitivity analysis and sparsity models in a recently accepted paper (*Maroufy and Marriott, 2018*).

We have added comments to the discussion to clarify these issues.

Typically, the gamma frailty is a one parameter family with mean fixed at 1 and the standard deviation varies; this is also reflected in the choice of models for your simulation study. This is not a one parameter exponential family; it is a one parameter curved exponential family. Please discuss how this fits with the assumptions of the LLM.

Interesting question; however, in the LMM approximation to the mixture $\int f(x, \theta)\, dQ(\theta)$, the exponential family assumption is on the component density function, $f(x, \theta)$, while the gamma distribution here is the mixing distribution, dQ.

'Second, our method automatically addresses the possibility of departures from the gamma assumption by doubling the standard errors.'
– In this particular example, the standard errors happen to be about double. – You've not looked at non gamma frailty models. These methods may compare very differently across different frailty structures both in terms of standard error and bias.

This is a great comment and we agree. In our simulation the standard deviation for the LMM method just happen to be almost twice as big as that under the gamma frailty, and might change under different frailty distributions, this has been explored in the new simulation studies.

'Furthermore, $\hat{\lambda}$ lie in the relative interior of λ_ϑ , indicating that the one component local mixture is adequate.'
– More details would be helpful.

A one component LMM is an approximation to a mixture model with a small-variation mixing distribution; hence, Λ_ϑ is a space of all LMMs close to the unmixed model, the gamma model with mean 1.

The measurement of 'closeness' is, as discussed above, is an L^1 one. When the estimate is on the boundary the two distributions are still L^1 close, but have different support and so are far apart in a KL -sense. We have added comments in the paper addressing this just after Table 4.

> *3:'It could be due to either misspecification bias or an under estimate of the standard error in the gamma frailty model.'*
> *– For these to be the only two options requires LMM to be unbiased.*

We have added a comment on bias here and the end of §3, thanks for the insight.

> *Simulation of 1 parameter regression with gamma frailty does not mean this holds more generally.*
> *– It would be illustrative to compare the two models in terms of goodness of fit analysis.*
> *The variable sex is not significant; does the difference for the WBC coefficient hold after sex is removed?*

Comments about simulation study and bias in general have been addressed above.

> *The variable sex is not significant; does the difference for the WBC coefficient hold after sex is removed?*

After removing predictor sex, all three remaining estimates stayed the same, with ignorable changes.

References

1. Marriott, P. (2007). Extending local mixture models. *A*nnals of the Institute of Statistical Mathematics, 59(1), 95-110.

2. Anaya-Izquierdo, K., Critchley, F., & Marriott, P. (2014). When are first-order asymptotics adequate? A diagnostic. *S*tat, 3(1), 17-22.

3. Maroufy, V. and Marriott, P. (2018): Local and global robustness in conjugate Bayesian analysis and sparsity models. Statistica Sinica, (In press)