

## Feedback on "Survival Analysis with Semiparametric Proportional Hazards Models" by Ziang

Alex Stringer

2019/09/13

### Overall

- Very impressive work. Well written. Problem is explained clearly as well as proposed solution. Appropriate level of detail overall.
- Typesetting: make sure to use latex's `\text{}` environment for things like "exp", "log", "Poisson", etc. All text in math mode should be formatted using this command.
- Use the `{[()]}` convention for nested brackets– see the case crossover paper for how I typeset equations.  $h(g(f(x)))$  should be typeset as  $h\{g[f(x)]\}$ .

### Section 1

Split this section into two sections: one on the *model*, and one on the *inference methodology*. In the *model* section, you describe the proportional hazards model, censoring, etc, but without mentioning "INLA". In the *inference methodology* section, you can talk about why INLA can't be used to fit survival models right away, and can give the details of the data augmentation trick they use. Trying to put everything into one section sometimes results in important ideas being obscured.

- Add a brief introduction to survival analysis. What is the response variable (event time), what is a hazard function, and so on. Assume the reader is a statistician but not that they have a background in survival methods specifically.
- **Important:** I think what you are describing in this section is *not* the Cox Proportional Hazards model, but rather a survival model with a *non-parametric* hazard function. These are different, I think– can you explain?
- Make sure to point out that not all survival data is censored.
- I would start this section by describing the proportional hazards model, then talk about censoring. My favourite reference for survival data is *Statistical Models* by Anthony Davison (if you don't already know this book anyways, I highly recommend it!). He explains the model and all terms and then derives the likelihood. Your explanation starts by saying what you're *doing*– this is good, but make sure to very briefly describe *why* you're doing it.
- Are truncation and censoring the same thing, or are they different? I thought Patrick was saying they are different.
- For the proportional hazards model, write equation (2) as depending on  $\eta$ , not  $\beta$ . The reason is because we can define  $\eta$  to be whatever we want, and this part doesn't change. We don't want to restrict ourselves to linear functions only. See the case crossover paper for my definition of  $\eta$  (and  $\lambda$ , in that paper)– we should use the same one.

- State the likelihood as well as the log-likelihood. Log-likelihood is more for computation; likelihood is easier to read and interpret.
- Make sure to define all terms clearly, explicitly, and prior to their use in equations. What is  $\delta$ ,  $k_i$ ,  $h_0(t)$ , etc?
- Otherwise, your notation is good, very clear and consistent.
- Is the hessian you're deriving for the **Poisson-augmented data**, or for the original data? We'll need the one for the original data. It will have the same block-diagonal form that you wrote down, but the blocks won't be diagonal, they will themselves be block-diagonal.

## Sections 2 and 3

- Provide some background information on the datasets, even just at the level of detail that's usually seen inside the R documentation. What is the response variable, what is/are the covariate(s) of interest, why is a survival model being used, is there censoring, etc.
- What is the "frailty.gaussian()" thing? This should be fully explained.
- Why is there an intercept in the model fit with Bayesian inference, and no intercept in the model fit by maximum likelihood? Can you comment on whether the intercept is estimable in survival models fit by maximum likelihood? Does removing the intercept from the INLA call give coefficients that are closer to coxph?