Discussion of: predictive comparison of joint longitudinal-survival modeling: a case study illustrating competing approaches

Radu V. Craiu

I congratulate the authors for a veritable tour-de-force in which they review an impressive amount of literature and consider a wide variety of Bayesian joint models for the medfly data. The authors (henceforth referred to as HBJ) also propose a model selection criterion based on predictive power. I have greatly enjoyed reading the paper and, although a number of issues have raised my interest, my discussion will be mainly centered on two parts: the first one contains comments on the model choice criterion and the second summarizes an alternative analysis. Before I go forward I would like to emphasize that my somewhat critical remarks below should not diminish all the positive aspects of the paper: bringing to the forefront the important issue of model selection and comparison, a thorough discussion of popular survival models used in the literature for joint modeling and a careful data analysis.

1 Bayesian model and model choice criterion

HBJ propose the cross-validated log-pseudo-marginal likelihood (LPML) as the vehicle for model choice. While I agree with the authors that an important aspect of the analysis has to do with model comparison and model choice, I also believe that a useful model choice strategy should aim at determining which specifications are damaging the analysis. For such complex models as the ones considered by HBJ, I am skeptical about LPML's ability to determine which elements of the model impede the "identification of a predictively viable model". For instance, if the model is deemed inadequate then one may wonder whether the longitudinal or the survival specification requires further tuning—can LPML offer an answer? An additional complication is that a high LPML value can be due to a fly with an unusual long life and not necessarily to a poor

R. V. Craiu (✉)
Department of Statistics, University of Toronto, Toronto, Canada
e-mail: craiu@utstat.toronto.edu

model fit. In a spatial analysis of relative risk where they use a criterion similar to LPML, Stern and Cressie (2000) mention that "there is a potential problem in that we may end up always blaming the statistical assumptions of the model for the existence of extreme values."

A model comparison based solely on LPML would be hard put to detect situations in which *all* models considered happen to be unsuitable for prediction. The authors are aware of this shortcoming and show for some of the flies in the sample the raw trajectories along with model fits. However, such diagnostics can be difficult to take in if the number of trajectories are in the hundreds or thousands (as in the original medfly data). Defining suitable statistics on which one could detect poor predictive power, possibly in the absence of an alternate model, via posterior predictive checks as proposed by Gelman et al. (1996) remains an important issue in joint modeling.

Lack of predictive power may be also due to poor prior specification and the methods of Evans and Moshonov (2006) and Evans and Jang (2010) for detecting prior-data conflict may be adaptable to this setting. Alternatively, a sensitivity analysis can be carried over in which, the statistical model is kept fixed but alternative priors are used in the analysis and the effects of the change on relevant aspects of the posterior (both inferential and predictive) are summarized. For instance, when predictive models are considered one generally needs to keep in check the number of parameters in the model. Have the authors considered the influence of decreasing the level M in the MFPT?
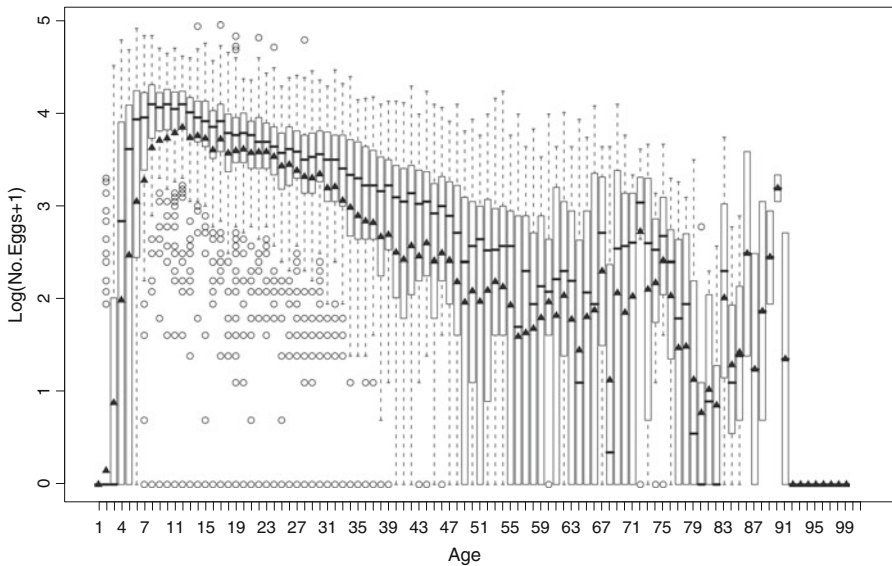
2 An alternative analysis using median regression

Figure 1 shows the boxplots of $Y_i = \log(N_i + 1)$, where N_i is the number of eggs laid on day i, $1 \le i \le 99$. The mean values for each day are represented by triangles. The plot suggests that the normal approximation (5) proposed by HBJ may not be appropriate as the distribution of $\log(N_i + 1)$ is clearly skewed.

For such skewed distributions which are also largely influenced by the number of zeros observed in each day, the median for each day may be more informative than the mean (Koenker 2005) so I consider an alternative Bayesian joint analysis which uses a quantile regression model for the longitudinal data observed. The Bayesian quantile regression models proposed by Yu and Moyeed (2001) and Geraci and Bottai (2007) are adapted here for the joint analysis setting.

More precisely, we base the longitudinal model on the asymmetric Laplace distribution with density $f_\tau(y|\mu, \sigma) = \tau(1 - \tau)/\sigma \exp\left\{-\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right\}$, where $\rho_\tau(v) = v(\tau - 1_{\{v \le 0\}})$ and $\tau \in (0, 1)$. When $\tau = 1/2$ this model corresponds to median regression. The counterpart of model (5) considered by HBJ is $y_i(t)|x_i(t), b_i, \sigma \sim f_{1/2}(y_i|x_i(t), b_i, \sigma)$, where $x_i(t) = b_{i1} + b_{i2} \log t + b_{i3}(t - 1)$ and $b_i = (b_{i1}, b_{i2}, b_{i3})$ is the vector of random effects. In our approach we compare hierarchical models with normal and non-central Student t_3 priors for the random effects. The latter prior can reduce the influence of outliers, as discussed by West (1984).

The survival model I use is more parsimonious than MFPT, assuming the Cox model with a piecewise constant baseline hazard function h_0 i.e., $h(t|x(t)) = \exp(\beta x(t))h_0(t)$ with $h_0(t) = \sum_{j=1}^{K} \lambda_j 1_{[a_j, a_{j+1})}(t)$ with $K = 6$ (see also Brown and Ibrahim 2003).

Fig. 1 Boxplots of samples $Y_i = \log(N_i + 1)$ for each day $i \in \{1, \ldots, 99\}$. Means are marked by *triangles*

A priori each λ_i is independently distributed with Gamma (0.1,0.1). An interesting alternative which allows dependence between hazard values in consecutive intervals is the gamma process prior of Nieto-Barajas and Walker (2002) (see also Craiu and Duchesne 2004).

The model is fitted via data augmentation (Wong and Tanner 1987) and Metropolis-within-Gibbs. In this case the simulation involves 251 tridimensional vectors of random effects and 14 parameters of interest. The tuning of the proposal distributions for such high-dimensional problems is a challenge which can be addressed using recent developments in adaptive Markov chain Monte Carlo (AMCMC). The latter is a class of MCMC samplers in which the transition kernel is allowed to change while the simulation proceeds (for details see Roberts and Rosenthal 2007; Craiu et al. 2009) and is particularly suitable in large dimensional problems where we have little idea about the shape of the target distribution.

Due to time constraints, I could not analyze the data using the MFPT prior so the comparison between the mean model (5) of HBJ and median regressions with Student and Gaussian priors for the random effects is done only for the Cox survival model described above. Table 1 contains the LPML values obtained for the models

Table 1 LPML values across different models

Regression model	Random effects prior	LPML
Median	Student t_3	−778
Median	Gaussian	−782
Mean	Gaussian	−823

considered. The results seem to suggest that using a median regression model may increase the predictive power of the model and it would be interesting to see what would be the the LPML score when the MFPT prior is considered.

In conclusion, I would like to thank the authors for a very stimulating paper.

3 Acknowledgment

References

Brown E, Ibrahim J (2003) A Bayesian semiparametric joint hierarchical model for longitudinal and survival data. Biometrics 59:221–228

Craiu RV, Duchesne T (2004) Using EM and data augmentation for the competing risks model. In: Gelman A, Meng XL (eds) Applied Bayesian modeling and causal inference from an incomplete-data perspective. Wiley, London

Craiu RV, Rosenthal JS, Yang C (2009) Learn from thy neighbor: parallel-chain and regional adaptive MCMC. J Am Stat Assoc 104:1454–1466

Evans M, Jang G-H (2010) Invariant p-values for model checking and checking for prior-data conflict. Ann Stat 38:511–525

Evans M, Moshonov H (2006) Checking for prior-data conflict. Bayesian Anal 1:893–914

Gelman A, Meng X-L, Stern H (1996) Posterior predictive assessment of model fitness via realized discrepancies (with discussion). Stat Sinica 6:733–807

Geraci M, Bottai M (2007) Quantile regression for longitudinal data using the asymmetric Laplace distribution. Biostatistics 8:140–154

Koenker R (2005) Quantile regression. Cambridge University Press, New York

Nieto-Barajas L, Walker S (2002) Markov beta and gamma processes for modeling hazard rates. Scand J Stat 29:413–424

Roberts GO, Rosenthal JS (2007) Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. J Appl Probab 44:458–475

Stern H, Cressie N (2000) Posterior predictive model checks for disease mapping models. Stat Med 19:2377–2397

West M (1984) Outlier models and prior distributions in Bayesian linear regression. J R Stat Soc B 46:431–439

Wong WH, Tanner MA (1987) The calculation of posterior distributions by data augmentation. J Am Stat Assoc 82:528–540

Yu K, Moyeed R (2001) Bayesian quantile regression. Stat Probab Lett 54:437–447