BIOS 791 Miniproject 3

Taeim Kwon

The paper chosen for this miniproject was Maximum Likelihood Estimation for Semiparametric Regression Models with Panel Count Data[1]. The authors developed an EM algorithm and an asymptotic theory to study how one or more types of recurrent events are affected by time-dependent covariates. One limitation was that they assumed independence of the examination times on the recurrent event process conditional on covariates. However, there are some situations where the examination times depend on the patient's health status, such as hospital visits for chronic diseases. That is, informative observation process may exist in panel count data. Although some methods were proposed to analyze such data, little research has been done to incorporate time-dependent covariates into the model or to consider cases where examination times are independent of recurrent event process given covariates. Our first research question, therefore, can extend the EM algorithm and the asymptotic theory given that the examination times depend on the recurrent event process conditional on covariates.

One of the approaches that can be proposed is to employ a joint model for panel count data with informative observation process [2]. Another approach is to use a frailty effect to account for correlation between recurrent event process and observation times [3]. These two methods are similar in that they constructed models for two processes, one for recurrent event process and another for observation process. It would be recommended to consider the two processes, as the original paper modeled the recurrent event process only [1].

Recall a nonhomogeneous Poisson process with intensity function was specified to be $N_{ki}(t)$ for subject $i = 1, ..., n$, type of recurrent events $k = 1, ..., K$ and time t with a set of some time-dependent covariates $X_i(t)$ where the intensity function was given by

$$\lambda_{ki}(t) = \lambda_k(t) exp(\beta_k^T X_i(t) + b_{ki}^T Z_i(t) + \xi_i^T \tilde{Z}_i(t)).$$

The model for observation process can be constructed so that a joint model can address dependence between recurrent event process and examination times. Since the authors pointed out that having a frailty term like the Gamma frailty would be restrictive, constructing a joint model would be better than introducing the frailty term. If the proposed research were successful, the potential impacts include showing the robustness of the EM algorithm and flexibly applying the asymptotic theory to more practical scenarios, where recurrent event process is related to examination times. In addition, it would result in more accurate individual risk predictions.

Furthermore, we can generalize the EM algorithm for the proportional intensity model to that for a proportional means model, as suggested in the paper [1]. They noted that the EM algorithm ignoring the integration over random effects and not updating variance parameters can be adapted for a proportional means model, which explains the population-average effects.

The original likelihood given from the paper is

$$\prod_{i=1}^{n}\left(\int_{\xi_i}\phi(\xi_i;\Psi)\prod_{k=1}^{K}\int_{b_{ki}}\phi(b_{ki};\Sigma_k)\prod_{j=1}^{m_{ki}}\frac{\left[\int_{U_{ki,j-1}}^{U_{kij}}\exp\left\{\beta_k^\top X_i(t)+b_{ki}^\top Z_i(t)+\xi_i^\top \tilde{Z}_i(t)\right\}\,d\Lambda_k(t)\right]^{\Delta_{kij}}}{\Delta_{kij}!}\right.$$
$$\left.\times\exp\left[-\int_0^{C_{ki}}\exp\left\{\beta_k^\top X_i(t)+b_{ki}^\top Z_i(t)+\xi_i^\top \tilde{Z}_i(t)\right\}\,d\Lambda_k(t)\right]\,db_{ki}\,d\xi_i\right).$$

Then, the likelihood used for the EM algorithm for a proportional means model is proportional to

$$\prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{j=1}^{m_{ki}}\frac{\left[\int_{U_{ki,j-1}}^{U_{kij}}\exp\left\{\beta_k^\top X_i(t)\right\}\,d\Lambda_k(t)\right]^{\Delta_{kij}}}{\Delta_{kij}!}\quad\times\exp\left[-\int_0^{C_{ki}}\exp\left\{\beta_k^\top X_i(t)\right\}\,d\Lambda_k(t)\right].\quad(1)$$

Since the proportional means model generally assumes that observation process is non-informative, developing the likelihood from the proportional intensity model that handles informative observation may help reduce bias in results for the proportional means model and be applied to more various datasets, although the likelihood would change from (1).

Lastly, the authors compared their work on interval-censored data with that on panel count data [4]. The general approach appeared to be similar, such as the construction of likelihood models, the development of the EM algorithm, and the application of asymptotic theory. However, as they noted, panel count data differs from interval-censored data in that any nonnegative integer can represent the number of a subject's observation within each time interval, whereas interval-censored data takes at most one event throughout the study. Thus, in their proof of Theorem 1 for identifiability, they used different Δ_{ij} compared to their previous work [4] to handle nonnegative integers.

Applying the results to other types of data can become a research question, as there are some data similar to panel count data, such as interval-censored data. Another example can be recurrent event data which measures observations continuously, while the panel count data measures the event of interest intermittently. For example, it can be of interest to analyze mixed recurrent event and panel count data. One study from the paper was the type of data, as some subjects provided their information continuously, while others were observed periodically [5]. Therefore, if the method were successfully adapted to recurrent event data, it would be applied to more complex data. Overall, the EM algorithm and asymptotic theory from the original paper [1] can be improved, although some details in the proofs should be adjusted, depending on the settings mentioned above.

References

[1] Zeng D. and Lin D. Maximum likelihood estimation for semiparametric regression models with panel count data. *Biometrika*, 108(4):947–960, 2020.

[2] Haixiang Zhang, Hui Zhao, Jianguo Sun, Dehui Wang, and KyungMann Kim. Regression analysis of multivariate panel count data with an informative observation process. *Journal of Multivariate Analysis*, 119:71–80, 2013.

[3] Chia-Hui Huang. Determination of correlations in multivariate count data with informative observation times. *Statistical Methods in Medical Research*, 33(2):273–294, 2024.

[4] Donglin Zeng, Fei Gao, and DY Lin. Maximum likelihood estimation for semiparametric regression models with multivariate interval-censored data. *Biometrika*, 104(3):505–525, 2017.

[5] Liang Zhu, Hui Zhao, Jianguo Sun, Wendy Leisenring, and Leslie L Robison. Regression analysis of mixed recurrent-event and panel-count data with additive rate models. *Biometrics*, 71(1):71–79, 2015.