Published in final edited form as:

Biometrics. 2013 September ; 69(3): 796. doi:10.1111/biom.12087.

Response to the Letter to the Editor "Comment on Cai and Betensky (2003), On the Poisson approximation for hazard regression"

Tianxi Cai and

Department of Biostatistics, Harvard University, Boston, MA, USA, 02115

Rebecca Betensky

Department of Biostatistics, Harvard University, Boston, MA, USA, 02115

Tianxi Cai: tcgai@hsph.harvard.edu; Rebecca Betensky: betensky@hsph.harvard.edu

The main focus of Cai and Betensky (2003) was for the analysis of interval censored data under the Cox model. They proposed to weakly parameterize the baseline hazard using a piecewise linear spline and maximize the likelihood function with a PQL approximation. Theoretical and numerical results demonstrated the validity of the proposed point and interval estimators for the general case. For the special case with right censored data, Cai and Betensky (2003) suggested the use of Poisson mixed model to approximate the likelihood for computational ease. However, as pointed out by Bové and Held (2013), the Poisson approximation given in Cai and Betensky (2003) was incorrect. Bové and Held (2013) suggested to create pseudo observations $y_{ij} = I(i = j)\delta_i$ and fit a Poisson mixed model to $\{y_{ij} : i = 1, \dots, n, j = 1, \dots, i\}$. This approach will indeed lead to a valid approximation for the point estimator. On the other hand, since their new proposal involves a pseudo data with $n(n+1)/2$ observations, there might be some computational burden in fitting such a Poisson mixed model even with moderate sample sizes such as $n = 500$. Furthermore, one may not directly obtain correct standard error estimates from the fitting since the y_{ij} 's are not independent observations.

An alternative strategy to obtain the maximizer of

$$\ell(\theta; \sigma_b) = \delta^T (\mathbf{X}_-^T \theta_- + \mathbf{Z}^T \beta) - \sum_{i=1}^n \Lambda_i(t_i) - \frac{1}{2\sigma_b^2} \mathbf{b}^T \mathbf{b} - \frac{K}{2} \sigma_b^2$$

is to consider an iterative procedure based on

$$\sum_{i=1}^n \Lambda_i(t_i) \approx \sum_{i=1}^n \sum_{j=1}^i q_{ij} \lambda_0(T_j) \exp(\mathbf{Z}_i^T \beta) = \sum_{i=1}^n \sum_{j=1}^i \exp\{\eta_0(T_j)\} q_{ij} \exp(\mathbf{Z}_i^T \beta) \quad (1)$$

using the same notation as Bové and Held, where $\theta_- = (\alpha_0, \alpha_1, \mathbf{b}^T)^T$ and $\mathbf{X}_- = [1, T_i, (T_i - \kappa)_+]_{i=1, \dots, n}$. Specifically, obtain an initial estimate of $\beta, \hat{\beta}^{(0)}$, say as the standard maximum

partial likelihood estimator. Then one may iterate via the following steps starting from $m = 1$:

1. Let $q_{ij}^{(m*)} = q_{ij} \exp(\mathbf{Z}_i^T \hat{\boldsymbol{\beta}}^{(m-1)})$ and $\bar{\mathbf{q}}^{(m*)} = [\sum_{i=j}^n q_{ij}^{(m*)}]_{j=1, \dots, n}$. Maximize

$$\ell_p^{(m*)}(\boldsymbol{\theta}_-; \sigma_b) \approx \boldsymbol{\delta}^T \mathbf{X}_-^T \boldsymbol{\theta}_- - \mathbf{1}^T \exp\{\mathbf{X}_-^T \boldsymbol{\theta}_- + \log(\bar{\mathbf{q}}^{(m*)})\} - \frac{1}{2\sigma_b^2} \mathbf{b}^T \mathbf{b} - \frac{K}{2} \sigma_b^2$$

with respect to $\{\boldsymbol{\theta}_-, \sigma_b\}$ to obtain $\{\hat{\boldsymbol{\theta}}_-^{(m)}, \hat{\sigma}_b^{(m)}\}$ and the corresponding $\hat{\eta}_0^{(m)}(T_j)$.

2. Let $q_{ij}^{(m\dagger)} = q_{ij} \exp\{\hat{\eta}_0^{(m)}(T_j)\}$ and $\bar{\mathbf{q}}^{(m\dagger)} = [\sum_{j=1}^i q_{ij}^{(m\dagger)}]_{i=1, \dots, n}$. Maximize

$$\ell_p^{(m\dagger)}(\boldsymbol{\beta}) \approx \boldsymbol{\delta}^T \mathbf{Z}^T \boldsymbol{\beta} - \mathbf{1}^T \exp\{\mathbf{Z}^T \boldsymbol{\beta} + \log(\bar{\mathbf{q}}^{(m\dagger)})\}$$

with respect to $\boldsymbol{\beta}$ to obtain $\hat{\boldsymbol{\beta}}^{(m)}$.

3. Let $m = m + 1$ and go back to Step (1) until convergence.

The maximizations can be achieved by fitting a Poisson mixed model with offset in Step (1) and a standard Poisson model with offset in Step (2).