Biostatistics (2014), **0**, 0, *pp.* 1–8
doi:10.1093/biostatistics/SupplementaryData

Supplementary materials to "Standard error estimation in the EM algorithm when joint modeling of survival and longitudinal data"

Cong Xu, Paul Baines and Jane-Ling Wang*

Department of Statistics, University of California, Davis, California 95616, U.S.A

jlwang.ucdavis@gmail.com

Appendix

A.1 Computation of I_{oc}

Denote the ordered uncensored event times by $U^1 < U^2 < \cdots < U^K$ and the corresponding jump sizes of the cumulative baseline hazard function by $\Lambda^1, \Lambda^2, \cdots, \Lambda^K$. Consider the complete-data log-likelihood:

$$\begin{aligned}\ell_n(\boldsymbol{\eta}|\mathbf{C}) = \quad & -\frac{N}{2}\log\left(2\pi\sigma_e^2\right) - \frac{\sum_i\sum_j\left(Y_{ij} - \mathbf{X}_{ij}^\top\boldsymbol{\beta} - \mathbf{Z}_{ij}^\top\mathbf{b}_i\right)^2}{2\sigma_e^2} \\ & + \sum_i \Delta_i\left[\log\lambda(V_i) + \mathbf{W}_i^\top(V_i)\boldsymbol{\gamma} + \alpha m_i(V_i)\right] \\ & - \sum_i \int_0^{V_i} \lambda(t)\exp\left\{\mathbf{W}_i^\top(t)\boldsymbol{\gamma} + \alpha m_i(t)\right\}\mathrm{d}t \\ & - \frac{n}{2}\log\left(|2\pi\Sigma_b|\right) - \frac{1}{2}\sum_i \mathbf{b}_i^\top\Sigma_b^{-1}\mathbf{b}_i.\end{aligned} \tag{A.1.1}$$

Take the derivative of $\ell_n(\boldsymbol{\eta}|\mathbf{C})$ w.r.t. Λ^k:

$$\frac{\partial\ell_n(\boldsymbol{\eta}|\mathbf{C})}{\partial\Lambda^k} = \frac{\sum_i\Delta_i I(V_i = U^k)}{\Lambda^k} - \sum_i I(V_i \geqslant U^k)\exp\left\{\mathbf{W}_i^\top(U^k)\boldsymbol{\gamma} + \alpha m_i(U^k)\right\},$$

*To whom correspondence should be addressed.

therefore,

$$\frac{\partial \ell_n(\boldsymbol{\eta}|\mathbf{C})}{\partial \Lambda^k} = 0 \Longrightarrow \hat{\Lambda}^k(\boldsymbol{\theta}) = \frac{\sum_i \Delta_i I(V_i = U^k)}{\sum_i I(V_i \geqslant U^k) \exp\left\{\mathbf{W}_i^\top(U^k)\boldsymbol{\gamma} + \alpha m_i(U^k)\right\}}.$$

Let $\dot{\ell}_n(\boldsymbol{\theta}) = \frac{\partial \ell_n(\boldsymbol{\eta}|\mathbf{C})}{\partial \boldsymbol{\theta}}|_{\Lambda^k = \hat{\Lambda}^k(\boldsymbol{\theta})}$, then $I_{oc} = E\left[-\frac{\partial^2 \ell_n(\boldsymbol{\eta}|\mathbf{C})}{\partial \boldsymbol{\theta} \cdot \partial \boldsymbol{\theta}}|\mathbf{O}, \boldsymbol{\eta}^*\right]$ is a $p \times p$ (p is the length of the finite dimensional parameter $\boldsymbol{\theta}$) matrix with $\frac{\partial^2 \ell_n(\boldsymbol{\eta}|\mathbf{C})}{\partial \boldsymbol{\theta} \cdot \partial \boldsymbol{\theta}}$ substituted by $\frac{\partial \dot{\ell}_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$.

A.2 Computation of $D^{10}Q_{\boldsymbol{\theta}}(\boldsymbol{\eta}', \boldsymbol{\eta})$ and $S_{\boldsymbol{\theta}}(\boldsymbol{\eta})$

The Q function is given in (2.6) of the paper and the $D^{10}Q$ function before profiling is

$$D^{10}Q(\boldsymbol{\eta}', \boldsymbol{\eta}) = \frac{\partial E\left[\log f_n(\boldsymbol{\eta}'|\mathbf{C})|\mathbf{O}, \boldsymbol{\eta}\right]}{\partial \boldsymbol{\eta}'} = \frac{\partial E\left[\ell_n(\boldsymbol{\eta}'|\mathbf{C})|\mathbf{O}, \boldsymbol{\eta}\right]}{\partial \boldsymbol{\eta}'}.$$

From (A.1.1) we obtain:

$$\frac{\partial Q(\boldsymbol{\eta}', \boldsymbol{\eta})}{\partial \Lambda^{k'}} = \frac{\sum_i \Delta_i I(V_i = U^k)}{\Lambda^{k'}} - \sum_i I(V_i \geqslant U^k) E\left[\exp\left\{\mathbf{W}_i^\top(U^k)\boldsymbol{\gamma}' + \alpha' m_i(U^k)\right\}|\mathbf{O}, \boldsymbol{\eta}\right],$$

and

$$\frac{\partial Q(\boldsymbol{\eta}', \boldsymbol{\eta})}{\partial \Lambda^{k'}} = 0 \Longrightarrow \hat{\Lambda}^k(\boldsymbol{\theta}') = \frac{\sum_i \Delta_i I(V_i = U^k)}{\sum_i I(V_i \geqslant U^k) E\left[\exp\left\{\mathbf{W}_i^\top(U^k)\boldsymbol{\gamma}' + \alpha' m_i(U^k)\right\}|\mathbf{O}, \boldsymbol{\eta}\right]}.$$

Then the profile version of the $D^{10}Q$ function is defined as

$$D^{10}Q_{\boldsymbol{\theta}}(\boldsymbol{\eta}', \boldsymbol{\eta}) = \frac{\partial E\left[\ell_n(\boldsymbol{\eta}'|\mathbf{C})|\mathbf{O}, \boldsymbol{\eta}\right]}{\partial \boldsymbol{\theta}'}|_{\Lambda^{k'} = \hat{\Lambda}^k(\boldsymbol{\theta}')},$$

which is a vector of length p instead of length $p + n_u$ (n_u is the number of uncensored survival time points). The corresponding S function is naturally $S_{\boldsymbol{\theta}}(\boldsymbol{\eta}) = D^{10}Q_{\boldsymbol{\theta}}(\boldsymbol{\eta}', \boldsymbol{\eta})|_{\boldsymbol{\eta}'=\boldsymbol{\eta}}$.

A.3 Additional simulation study

The following is another simulation study to examine the performance of different SE estimation procedures. Again, consider two time-independent covariates X_{1i} (from Bernoulli distribution with success probability 0.5) and X_{2i} (from $\mathcal{N}(2, 0.4^2)$), $i = 1, 2, \cdots, n$. The longitudinal model is the same with that in Section 4.1:

$$Y_{ij} = Y_i(t_{ij}) = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 t_{ij} + \beta_4 X_{1i} t_{ij} + \beta_5 X_{2i} t_{ij} + b_{1i} + b_{2i} t_{ij} + e_{ij},$$

with $\mathbf{b}_i = (b_{1i}, b_{2i})^\top \sim \mathcal{N}(0, \Sigma_b)$, $e_{ij} \sim \mathcal{N}(0, \sigma_e^2)$ and $t_{ij} = 0.2(j-1)$ for $i = 1, 2, \ldots, n$. The survival model only shares the same random effects with the longitudinal model:

$$\lambda(t|\mathbf{b}_i, X_{1i}, X_{2i}) = \lambda(t) \exp\left\{\gamma_1 X_{1i} + \gamma_2 X_{2i} + \gamma_3 X_{1i} t + \gamma_4 X_{2i} t + \alpha(b_{1i} + b_{2i} t)\right\}.$$

The true baseline hazard function is $\lambda(t) \equiv 1$ and the true values of the parameters are chosen as below:

$$\beta_1 = 2.0,\ \beta_2 = -1.0,\ \beta_3 = 2.0,\ \beta_4 = -2.0,\ \beta_5 = 1.5,\ \sigma_e^2 = 0.1,\ \Sigma_b = \begin{pmatrix} 0.5 & -0.1 \\ -0.1 & 0.25 \end{pmatrix},$$
$$\gamma_1 = -1.0,\ \gamma_2 = -1.5,\ \gamma_3 = 2,\ \gamma_4 = 1,\ \alpha = 0.5.$$

Again, the simulation is repeated 500 times with sample size $n = 200$. The simulation results are presented in Table 3 below.

A.4 Comparison with the `JM` package

Both the piecewise constant and the spline-based approaches, as implemented in the R package "`JM`", are examples of the method of sieves, where the sieve spaces are seemingly parametric after a proper sieve dimension has been selected. Hsieh *and others* (2013) discusses properties of the method of sieves approach in the context of joint modeling and the issue of sieves biases as the true baseline hazard function may not belong to the sieve spaces. Typically, a moderate or large number of parameters is needed to model the sieve space so the sieve bias can be contained. In this regard, the profiling method proposed in Section 3.2 and 3.3 can be useful for standard error estimation under such flexible hazard models where the number of parameters is large.

Table 4 below shows the results of our joint model with completely unspecified baseline hazard function along with those of the joint models with piecewise constant and spline-based baseline hazard functions fitted by the `JM` package, where the follow-up period is divided into 7 intervals with knots placed at the corresponding quantiles. As illustrated by the table, comparing with the semiparametric joint model, the joint models with piecewise constant and spline-based baseline hazard yield satisfactory parameter estimates but the biases for α are 0.02306 and 0.02803, respectively, which are 188% and 228% of 0.01229, the bias under our semiparametric approach. For the empirical standard errors ("MCSE") and estimated standard errors ("Est. SE"), the results in Table 4 are somewhat unexpected as the semiparametric approach led to comparable standard errors as parametric ones. Thus, there is no disadvantage in using the semiparametric approach. This simulation also demonstrates that parametric approaches can provide good approximations but the bias in the estimates can be improved by a nonparametric approach.

A.5 Another model for the HIV clinical trial data

The following is another model fitted for the HIV clinical trial data: (results are given in Table 5 below)

$$Y_{ij} = m_i(t_{ij}) + e_{ij} = \beta_0 + \beta_1 t_{ij} + \beta_2 \text{drug}_i t_{ij} + b_{1i} + b_{2i} t_{ij} + e_{ij},$$
$$\lambda(t|\mathbf{b}_i, \text{drug}_i) = \lambda(t) \exp\{\gamma \text{drug}_i + \alpha m_i(t)\}.$$

Additional tables

Additional tables for the simulation results in Section 4 are presented below.

References

HSIEH, F., DING, J. AND WANG, J.L. (2013). Method of sieves to joint model survival and longitudinal data. *Statistica Sinica* **23**, 1181-1213.

Table 1. *Additional SE estimates for case I. The results for $h = 10^{-3}, 10^{-4}$ are provided for the PFDM/PREM, PFDS/PRES and PL methods.*

θ	MCSE	PFDM ($h = 10^{-3}$)	PREM ($h = 10^{-3}$)	PFDS ($h = 10^{-3}$)	PRES ($h = 10^{-3}$)	PL ($h = 10^{-3}$)
β_1	0.09939	0.09413	0.09529	0.09527	0.09527	0.09526
β_2	0.11760	0.11164	0.11908	0.11939	0.11821	0.11813
β_3	0.12354	0.12002	0.12008	0.12030	0.11968	0.11968
β_4	0.10917	0.11681	0.11498	0.11498	0.11482	0.11481
β_5	0.18441	0.19139	0.18310	0.18326	0.18311	0.18232
γ_1	0.24130	0.20559	0.24144	0.24677	0.23849	0.21644
γ_2	0.37139	0.30929	0.35430	0.36708	0.35264	0.30828
α	0.13989	0.09840	0.13595	0.14233	0.13585	0.11895
N.	Posit.	489	500	499	500	500
θ	MCSE	PFDM ($h = 10^{-4}$)	PREM ($h = 10^{-4}$)	PFDS ($h = 10^{-4}$)	PRES ($h = 10^{-4}$)	PL ($h = 10^{-4}$)
β_1	0.09939	0.09424	0.09528	0.09548	0.09527	0.09524
β_2	0.11760	0.11230	0.11908	0.11645	0.11821	0.11798
β_3	0.12354	0.11999	0.12008	0.11614	0.11968	0.11963
β_4	0.10917	0.11642	0.11498	0.11431	0.11482	0.11476
β_5	0.18441	0.19053	0.18310	0.17992	0.18311	0.18089
γ_1	0.24130	0.20895	0.24144	0.20139	0.23849	0.17305
γ_2	0.37139	0.31326	0.35430	0.26674	0.35264	0.21645
α	0.13989	0.10335	0.13595	0.10198	0.13585	0.08609
N.	Posit.	500	500	417	500	500

Table 2. *Simulation results for case II in Section 4.1.*

θ	β_1	β_2	β_3	β_4	β_5	γ_1	γ_2	α	
θ_0	-1.0	-1.5	1.0	-0.5	0.5	-0.5	1.5	0.5	N.
Mean	-0.99861	-1.50279	1.00077	-0.50080	0.49397	-0.49678	1.54872	0.51314	Posit.
MCSE	0.10840	0.12984	0.15929	0.14415	0.23427	0.25420	0.37959	0.15970	
PFDM(10^{-2})	0.10465	0.13650	0.16131	0.14876	0.22435	0.30367	0.43448	0.20554	404
PFDM(10^{-3})	0.10514	0.13443	0.15701	0.15742	0.23093	0.23320	0.32174	0.12332	317
PFDM(10^{-4})	0.10474	0.13399	0.15334	0.15719	0.23241	0.20654	0.29096	0.09787	354
PFDM(10^{-5})	0.10473	0.13389	0.15256	0.15666	0.23333	0.20587	0.28974	0.09733	366
PREM(10^{-2})	0.10469	0.13186	0.15498	0.15168	0.23170	0.25054	0.36675	0.15421	500
PREM(10^{-3})	0.10469	0.13186	0.15499	0.15168	0.23171	0.25054	0.36675	0.15421	500
PREM(10^{-4})	0.10469	0.13186	0.15499	0.15168	0.23171	0.25054	0.36675	0.15421	500
PREM(10^{-5})	0.10469	0.13186	0.15499	0.15168	0.23171	0.25054	0.36675	0.15421	500
PFDS(10^{-2})	0.10458	0.13188	0.15512	0.15151	0.23126	0.25094	0.36670	0.15062	500
PFDS(10^{-3})	0.10457	0.13217	0.15570	0.15141	0.23165	0.25899	0.37874	0.15739	497
PFDS(10^{-4})	0.10489	0.13465	0.16351	0.15339	0.23094	0.27439	0.39983	0.16535	399
PFDS(10^{-5})	0.10507	0.13077	0.14623	0.14983	0.22293	0.19493	0.25200	0.10130	468
PRES(10^{-2})	0.10455	0.13191	0.15511	0.15179	0.23181	0.25037	0.36683	0.15232	500
PRES(10^{-3})	0.10455	0.13191	0.15511	0.15178	0.23181	0.25037	0.36683	0.15232	500
PRES(10^{-4})	0.10455	0.13191	0.15511	0.15178	0.23181	0.25037	0.36683	0.15232	500
PRES(10^{-5})	0.10455	0.13191	0.15511	0.15178	0.23181	0.25037	0.36683	0.15232	500
PL(10^{-2})	0.10460	0.13188	0.15540	0.15165	0.23149	0.25034	0.36606	0.15038	500
PL(10^{-3})	0.10457	0.13183	0.15481	0.15135	0.23043	0.23937	0.34522	0.14130	500
PL(10^{-4})	0.10455	0.13170	0.15462	0.15110	0.22791	0.19425	0.25924	0.10495	499
PL(10^{-5})	0.10454	0.13164	0.15456	0.15100	0.22686	0.17323	0.21870	0.08785	497
BT($B = 50$)	0.10440	0.13296	0.15964	0.15611	0.23938	0.25709	0.37621	0.15463	500
BT($B = 100$)	0.10426	0.13306	0.15984	0.15772	0.24005	0.25904	0.37680	0.15589	500
Computation Time for the Bootstrap method: 8456s ($B = 50$), 16639s ($B = 100$)									

Table 3. *Simulation results for the additional simulation study in Section A.3.*

θ	β_1	β_2	β_3	β_4	β_5	γ_1	γ_2	γ_3	γ_4	α	
θ_0	2.0	-1.0	2.0	-2.0	1.5	-1.0	-1.5	2	1	0.5	N.
Mean	2.0025	-1.0008	1.9916	-2.0033	1.5037	-1.0403	-1.4722	2.0519	0.9829	0.5057	Posit.
MCSE	0.1072	0.0375	0.2297	0.0904	0.1111	0.6153	0.6622	0.4999	0.4878	0.1137	
PFDM(10^{-2})	0.1099	0.0336	0.2925	0.0630	0.1297	0.6434	0.6491	0.5367	0.5061	0.1284	388
PFDM(10^{-3})	0.0930	0.0366	0.1276	0.1086	0.0818	0.6332	0.6374	0.5122	0.4392	0.1024	384
PFDM(10^{-4})	0.0953	0.0368	0.1365	0.1072	0.0833	0.6364	0.6361	0.5149	0.4427	0.1038	449
PFDM(10^{-5})	0.0954	0.0368	0.1380	0.1071	0.0838	0.6365	0.6366	0.5151	0.4434	0.1040	463
PREM(10^{-2})	0.1022	0.0353	0.2245	0.0903	0.1079	0.6384	0.6417	0.5243	0.4732	0.1158	500
PREM(10^{-3})	0.1022	0.0353	0.2244	0.0903	0.1079	0.6384	0.6417	0.5244	0.4732	0.1158	500
PREM(10^{-4})	0.1022	0.0353	0.2244	0.0903	0.1079	0.6384	0.6417	0.5244	0.4732	0.1158	500
PREM(10^{-5})	0.1022	0.0353	0.2243	0.0903	0.1079	0.6384	0.6417	0.5244	0.4732	0.1158	500
PFDS(10^{-2})	0.1022	0.0353	0.2248	0.0902	0.1080	0.6383	0.6419	0.5243	0.4733	0.1159	500
PFDS(10^{-3})	0.1032	0.0353	0.2285	0.0900	0.1083	0.6489	0.6628	0.5266	0.4811	0.1195	498
PFDS(10^{-4})	0.1003	0.0342	0.2174	0.0879	0.1071	0.6053	0.5646	0.5193	0.4506	0.1090	486
PFDS(10^{-5})	0.0999	0.0336	0.2164	0.0868	0.1064	0.5968	0.5546	0.5171	0.4440	0.1088	497
PRES(10^{-2})	0.1022	0.0353	0.2245	0.0903	0.1079	0.6384	0.6417	0.5243	0.4732	0.1157	500
PRES(10^{-3})	0.1022	0.0353	0.2244	0.0903	0.1079	0.6384	0.6417	0.5243	0.4732	0.1158	500
PRES(10^{-4})	0.1022	0.0353	0.2244	0.0903	0.1079	0.6384	0.6419	0.5244	0.4734	0.1158	500
PRES(10^{-5})	0.1022	0.0353	0.2244	0.0903	0.1079	0.6384	0.6419	0.5244	0.4733	0.1158	500
PL(10^{-2})	0.1022	0.0353	0.2250	0.0903	0.1083	0.6379	0.6410	0.5242	0.4731	0.1159	500
PL(10^{-3})	0.1012	0.0352	0.2237	0.0899	0.1075	0.6211	0.6060	0.5214	0.4611	0.1128	500
PL(10^{-4})	0.1000	0.0351	0.2222	0.0895	0.1067	0.5983	0.5573	0.5173	0.4439	0.1093	500
PL(10^{-5})	0.0999	0.0351	0.2219	0.0894	0.1065	0.5970	0.5548	0.5171	0.4430	0.1092	498
BT($B = 50$)	0.1015	0.0348	0.2270	0.0908	0.1092	0.6701	0.6833	0.5505	0.5106	0.1213	500
BT($B = 100$)	0.1021	0.0351	0.2271	0.0910	0.1090	0.6750	0.6848	0.5543	0.5126	0.1208	500

Table 4. *Estimates from the EM algorithm for case I in Section 4.1. θ_0: true value of the parameters. "Mean" and "MCSE": empirical means and standard errors of the parameter estimates from the 500 simulations. "CR": coverage ratios of the 95% confidence interval constructed using the "MCSE" column. "Est. SE: estimated standard errors (for the unspecified baseline hazard, obtained by using PRES with $h = 10^{-5}$; for the piecewise constant and spline-based hazard, obtained from the "JM" package).*

θ	β_1	β_2	β_3	β_4	β_5	γ_1	γ_2	α
θ_0	-1.0	-1.5	1.0	-0.5	0.5	-0.5	1.5	0.5
With Unspecified Baseline Hazard Function								
Mean	-0.99922	-1.50637	1.00145	-0.50416	0.49678	-0.49800	1.54721	0.51229
Bias	0.00078	0.00637	0.00145	0.00416	0.00322	0.00200	0.04721	0.01229
MCSE	0.09939	0.11760	0.12354	0.10917	0.18441	0.24130	0.37139	0.13989
CR	0.942	0.950	0.952	0.958	0.940	0.948	0.940	0.950
Est. SE	0.09527	0.11821	0.11968	0.11482	0.18311	0.23849	0.35264	0.13585
With Piecewise Constant Baseline Hazard Function								
Mean	-0.99920	-1.50711	1.00316	-0.50514	0.49655	-0.49118	1.56189	0.52306
Bias	0.00080	0.00711	0.00316	0.00514	0.00345	0.00882	0.06189	0.02306
MCSE	0.09942	0.11758	0.12367	0.10918	0.18429	0.23876	0.36891	0.13948
CR	0.942	0.950	0.952	0.958	0.940	0.950	0.944	0.948
Est. SE	0.09527	0.11820	0.11967	0.11481	0.18309	0.23391	0.34903	0.13223
With Spline-based Baseline Hazard Function								
Mean	-0.99841	-1.50683	1.00594	-0.50420	0.49653	-0.49117	1.58066	0.52803
Bias	0.00159	0.00683	0.00594	0.00420	0.00347	0.00883	0.08066	0.02803
MCSE	0.09940	0.11730	0.12325	0.10906	0.18395	0.24321	0.37447	0.14406
CR	0.942	0.950	0.952	0.960	0.942	0.956	0.950	0.960
Est. SE	0.09525	0.11818	0.11976	0.11482	0.18313	0.23821	0.35206	0.13497

Table 5. *Results for the HIV clinical trail data analysis under the model in Appendix A.3. "Est. SE" are the estimated SEs obtained using the PRES method with $h = 10^{-5}$.*

θ	β_0	β_1	β_2	γ	α
Est. Value	2.5198	-0.0417	0.0053	0.3557	-1.0749
Est. SE	0.04261	0.00453	0.00635	0.15343	0.11509
P-value	< 0.0001	< 0.0001	0.4039	0.0204	< 0.0001