Supplementary Material: A fast and stable algorithm for non-parametric maximum likelihood estimation of survival functions for left-truncated and interval-censored data

Zachary Waller, Adele H. Marshall, Frank Kee, Felicity Lamrock

1 Other NPMLE algorithms

Here are the forms of the NPMLE algorithms used for comparison in the simulations in Section 5. Each requires the inner-intervals calculated as in Section 2.1. The EM algorithms are all parameterized to allow for $O(n)$ implementation (or $O(n_1)$ in the Breslow method case).

1.1 Turnbull's Algorithm

From Turnbull (1976).

Step 1. For $j = 1, ..., m$, let $s_j^{(0)} = 1/m$.

Step $g+1$ ($g \geq 1$). For $1 \leq j \leq m$, update s_j by

$$s_j^{(g+1)} = s_j^{(g)} \frac{M_j^{(g)}}{\sum_{k=1}^{m} M_k^{(g)}}$$

where

$$M_j^{(g)} = \sum_{i=1}^{n} \left[\frac{\alpha_{ij}}{S_{L_i^*}^{(g)} - S_{R_i^*}^{(g)}} + \frac{(1-\beta_{ij})}{S_{T_i^*}^{(g)}} \right] \tag{1}$$

and

$$S_k^{(g)} = 1 - \sum_{l=1}^{k} s_l^{(g)}$$

with $S_0^{(g)} = 1$.

Stop when $|\log \mathcal{L}^{(g+1)} - \log \mathcal{L}^{(g)}| < \epsilon$ where ϵ is some small positive number.

1.2 Yu's Algorithm

From Yu (2023).

Step 1. For $j = 1, ..., m$, let $s_j^{(0)} = 1/m$.

Step $g+1$ ($g \geq 1$). For $1 \leq j \leq m$, update s_j by

$$s_j^{(g+1)} = s_j^{(g)} \frac{1}{n} \sum_{i=1}^{n} \left[\frac{\alpha_{ij}}{S_{L_i^*}^{(g)} - S_{R_i^*}^{(g)}} + \delta_i \left(1 - \frac{\beta_{ij}}{S_{T_i^*}^{(g)}} \right) \right]$$

where

$$\delta_i = I(S_{T_i^*}^{(g)} < 1)$$

and

$$S_k^{(g)} = 1 - \sum_{l=1}^{k} s_l^{(g)}$$

with $S_0^{(g)} = 1$.

Stop when $|\log \mathcal{L}^{(g+1)} - \log \mathcal{L}^{(g)}| < \epsilon$ where ϵ is some small positive number.

1.3 Shen's Algorithm

From Shen (2020).

Step 1. For $j = 1, ..., m$, let $s_j^{(0)} = 1/m$.

Step $g + 1$ $(g \geq 1)$. For $1 \leq j \leq m$, update s_j by

$$s_j^{(g+1)} = s_j^{(g)} \left\{ 1 + \frac{\sum_{i=1}^{n} \left[\frac{\alpha_{ij}}{S_{L_i^*}^{(g)} - S_{R_i^*}^{(g)}} - \frac{\beta_{ij}}{S_{T_i^*}^{(g)}} \right]}{\sum_{i=1}^{n} 1/S_{T_i^*}^{(g)}} \right\}$$

where

$$S_k^{(g)} = 1 - \sum_{l=1}^{k} s_l^{(g)}$$

with $S_0^{(g)} = 1$.

Stop when $|\log \mathcal{L}^{(g+1)} - \log \mathcal{L}^{(g)}| < \epsilon$ where ϵ is some small positive number.

1.4 Breslow method

Adapted from Li et al. (2020) and Appendix A of Gao and Chan Gao and Chan (2018). This parameterizes the likelihood using the cumulative hazard $\Lambda(t)$ where $S(t) = e^{-\Lambda(t)}$ with $\lambda_j = \Lambda(r_j) - \Lambda(l_j)$.

Step 1. For $j = 1, ..., m$, let $\lambda_j^{(0)} = 1/m$.

Step $g + 1$ $(g \geq 1)$. For $1 \leq j \leq m$, update λ_j by

$$\lambda_j^{(g+1)} = \frac{\sum_{i=n_0+1}^{n} w_{ij}^{(g)}}{\sum_{i=n_0+1}^{n} I(Q_j \subseteq [T_i, \tilde{R}_i])}$$

where

$$\tilde{R}_i = I(R_i < \infty)R_i + I(R_i = \infty)L_i\,,$$

$$w_{ij}^{(g)} = I(Q_j \subseteq (L_i, R_i], R_i < \infty) \frac{\lambda_j^{(g)}}{1 - \exp\left(-\Lambda_{R_i^*}^{(g)} + \Lambda_{L_i^*}^{(g)}\right)}$$

and

$$\Lambda_k^{(g)} = \sum_{l=1}^{k} \lambda_l^{(g)}$$

with $\Lambda_0 = 0$.

Stop when $|\log \mathcal{L}^{(g+1)} - \log \mathcal{L}^{(g)}| < \epsilon$ where ϵ is some small positive number.

1.5 Quasi-Newton method

Adapted from a description of a process recommended by Hudgens (2005). The exact details are not given there, but this method should be similar.

The implementation of this method uses the R function `optim` – a generic optimization function and takes as arguments the likelihood from Equation 2 and its gradient

$$\frac{\partial \log \mathcal{L}}{\partial s_j} = \sum_{i=1}^{n} \left[\frac{\alpha_{ij}}{\sum_{k=1}^{m} \alpha_{ik} s_k} - \frac{\beta_{ij}}{\sum_{k=1}^{m} \beta_{ik} s_k} \right]. \tag{2}$$

Because `optim` minimizes objective functions by default, we supply it with the negative likelihood and gradient. We set the `optim` function to use the L-BFGS-B algorithm (Byrd et al., 1995) for maximization, a limited-memory version of the Broyden–Fletcher–Goldfarb–Shanno algorithm that allows box constraints for each parameter. To improve stability, we found it necessary to limit each parameter to $0 + tol \leq s_j \leq 1 - tol$ for $1 \leq j \leq m$ where tol is some small tolerance value. For the simulations we chose $tol = 10^{-10}$, and found that smaller values did little to improve accuracy.

The values $\tilde{s}_j$ given as output by `optim` are normalized to equal one $\hat{s}_j = \tilde{s}_j / \sum_{k=1}^{m} \tilde{s}_k$.

2 Karush-Kuhn-Tucker Conditions

From the Lagrangian in Equation 9, and the conditions for on $\mathbf{h}$, some conditions for convergence of the algorithm can be formulated. This follows closely the approach of Gentleman and Geyer (1994).

Let $D_j = \frac{\partial \log \mathcal{L}}{\partial h_j}$. From Equation 9 it is clear that, at the maximum likelihood,

$$D_j + a_j - b_j = 0$$

where $a_j, b_j \geq 0$, $a_j h_j = 0$ and $b_j(1 - h_j) = 0$ for $j = 0, \ldots, m$. These conditions must all be met at for a candidate solution to be the maximum likelihood. This implies that when $0 < h_j < 1$, $a_j = b_j = 0$ and $D_j = 0$.

If $h_j = 0$ then $b_j = 0$ and therefore $a_j = -D_j$. Given, $a_j \geq 0$ this implies that $D_j \leq 0$.

Conversely, if $h_j = 1$, $a_j = 0$ and therefore $b_j = D_j$ and $D_j \geq 0$.

Therefore, at algorithm convergence, the following checks ensure the estimate is a maximum likelihood estimate

$$D_j \begin{cases} \leq 0 \text{ if } h_j = 0 \\ = 0 \text{ if } 0 < h_j < 1 \\ \geq 0 \text{ if } h_j = 1\,. \end{cases}$$

For LTIC data, the Hessian of the likelihood need not be negative-definite and so the Karush-Kuhn-Tucker conditions only guarantee a local (potentially non-unique) maximum.

3 Other Simulations

These simulation studies have the same setup as those in Section 5 of the main paper, but with a different number of resulting observations after truncation.

Table 1: 500 observations 0% truncation

Method	log-likelihood	Iterations	CPU time (s)	% converged
Breslow	-401.52(19.36)	2,281(329)	0.069(0.012)	100%
PL	**-401.52(19.35)**	**1,501(235)**	**0.031(0.006)**	**100%**
Shen	-401.52(19.36)	1,692(251)	0.044(0.007)	100%
Turnbull	-401.52(19.36)	1,692(251)	0.048(0.007)	100%
Yu	-401.52(19.36)	1,692(251)	0.039(0.006)	100%
Breslow-ICM	-401.51(19.36)	77(11)	0.002(0.000)	100%
PL-ICM	**-401.51(19.36)**	**77(22)**	**0.002(0.000)**	**100%**
Turnbull-ICM	-401.51(19.36)	88(22)	0.002(0.001)	100%
Yu-ICM	-401.51(19.36)	88(22)	0.002(0.000)	100%
ICM	-401.51(19.36)	18(3)	0.001(0.000)	100%
Quasi-Newton	-401.52(19.36)	108(2)	0.151(0.023)	100%

Median and inter-quartile range (in parentheses) of log-likelihood, run times (in iterations and seconds) and percentage converged before 100,000 iterations or with finite log-likelihood for 0% truncated experiment with 500 observations after truncation. The proposed PL and PL-ICM methods are highlighted in bold.

Table 2: 500 observations 25% truncation

Method	log-likelihood	Iterations	CPU time (s)	% converged
Breslow	-397.89(19.38)	2,505(475)	0.083(0.018)	100%
PL	**-397.89(19.38)**	**1,980(428)**	**0.050(0.012)**	**100%**
Shen	-397.97(19.41)	5,818(9,983)	0.181(0.293)	100%
Turnbull	-397.97(19.41)	5,818(9,983)	0.199(0.327)	100%
Yu	-406.77(∞)	2,667(3,328)	0.080(0.099)	68%
Breslow-ICM	-397.88(19.38)	544(671)	0.013(0.016)	100%
PL-ICM	**-397.88(19.38)**	**88(25)**	**0.002(0.001)**	**100%**
Turnbull-ICM	-397.89(19.39)	24,585(24,998)	0.776(0.804)	100%
Yu-ICM	-398.28(19.80)	1,996(2,571)	0.055(0.074)	99%
ICM	-397.94(19.41)	3,150(8,380)	0.229(0.584)	100%
Quasi-Newton	-398.54(19.04)	112(3)	0.475(0.060)	100%

Median and inter-quartile range (in parentheses) of log-likelihood, run times (in iterations and seconds) and percentage converged before 100,000 iterations or with finite log-likelihood for 25% truncated experiment with 500 observations after truncation. The proposed PL and PL-ICM methods are highlighted in bold.

Table 3: 500 observations 50% truncation

Method	log-likelihood	Iterations	CPU time (s)	% converged
Breslow	-340.81(24.59)	2,042(378)	0.062(0.011)	100%
PL	**-340.81(24.59)**	**1,527(304)**	**0.037(0.009)**	**100%**
Shen	-340.99(24.61)	6,564(12,603)	0.197(0.378)	100%
Turnbull	-340.99(24.61)	6,564(12,603)	0.216(0.396)	100%
Yu	-352.74(∞)	2,802(3,357)	0.081(0.104)	73%
Breslow-ICM	-340.80(24.59)	517(660)	0.011(0.014)	100%
PL-ICM	**-340.80(24.59)**	**88(22)**	**0.002(0.001)**	**100%**
Turnbull-ICM	-340.82(24.59)	26,675(28,619)	0.804(0.897)	100%
Yu-ICM	-343.95(26.21)	1,358(5,734)	0.037(0.149)	93%
ICM	-340.88(24.59)	2,336(8,789)	0.168(0.632)	100%
Quasi-Newton	-341.65(25.09)	113(4)	0.418(0.045)	100%

Median and inter-quartile range (in parentheses) of log-likelihood, run times (in iterations and seconds) and percentage converged before 100,000 iterations or with finite log-likelihood for 50% truncated experiment with 500 observations after truncation. The proposed PL and PL-ICM methods are highlighted in bold.

Table 4: 2000 observations 0% truncation

Method	log-likelihood	Iterations	CPU time (s)	% converged
Breslow	-1654.24(43.06)	6,302(729)	0.781(0.089)	100%
PL	**-1654.23(43.06)**	**4,274(472)**	**0.347(0.041)**	**100%**
Shen	-1654.23(43.06)	4,716(490)	0.480(0.053)	100%
Turnbull	-1654.23(43.06)	4,716(490)	0.517(0.055)	100%
Yu	-1654.23(43.06)	4,716(490)	0.420(0.046)	100%
Breslow-ICM	-1654.21(43.06)	99(22)	0.008(0.002)	100%
PL-ICM	**-1654.21(43.06)**	**88(11)**	**0.007(0.001)**	**100%**
Turnbull-ICM	-1654.21(43.06)	110(22)	0.011(0.002)	100%
Yu-ICM	-1654.21(43.06)	110(22)	0.009(0.002)	100%
ICM	-1654.21(43.06)	21(2)	0.007(0.001)	100%
Quasi-Newton	-1655.58(42.86)	108(2)	1.912(0.070)	100%

Median and inter-quartile range (in parentheses) of log-likelihood, run times (in iterations and seconds) and percentage converged before 100,000 iterations or with finite log-likelihood for 0% truncated experiment with 2000 observations after truncation. The proposed PL and PL-ICM methods are highlighted in bold.

Table 5: 2000 observations 25% truncation

Method	log-likelihood	Iterations	CPU time (s)	% converged
Breslow	-1650.46(42.17)	6,574(720)	0.912(0.119)	100%
PL	**-1650.45(42.17)**	**5,228(678)**	**0.541(0.082)**	**100%**
Shen	-1650.48(42.19)	10,696(8,392)	1.365(0.930)	100%
Turnbull	-1650.48(42.19)	10,696(8,392)	1.497(1.041)	100%
Yu	-1658.22(48.69)	7,393(2,546)	0.926(0.268)	88%
Breslow-ICM	-1650.43(42.18)	170(610)	0.018(0.059)	100%
PL-ICM	**-1650.43(42.18)**	**99(33)**	**0.011(0.003)**	**100%**
Turnbull-ICM	-1650.43(42.18)	8,498(49,420)	1.090(6.388)	100%
Yu-ICM	-1650.43(42.18)	4,218(4,746)	0.472(0.576)	100%
ICM	-1650.43(42.22)	4,000(8,272)	1.225(2.483)	100%
Quasi-Newton	-1654.34(41.73)	108(0)	6.016(0.652)	100%

Median and inter-quartile range (in parentheses) of log-likelihood, run times (in iterations and seconds) and percentage converged before 100,000 iterations or with finite log-likelihood for 25% truncated experiment with 2000 observations after truncation. The proposed PL and PL-ICM methods are highlighted in bold.

Table 6: 2000 observations 50% truncation

Method	log-likelihood	Iterations	CPU time (s)	% converged
Breslow	-1438.19(51.27)	5,381(550)	0.754(0.090)	100%
PL	**-1438.18(51.27)**	**4,156(390)**	**0.431(0.046)**	**100%**
Shen	-1438.31(51.01)	15,006(11,842)	1.915(1.577)	100%
Turnbull	-1438.31(51.01)	15,006(11,842)	2.096(1.714)	100%
Yu	-1452.73(61.37)	7,686(2,704)	0.928(0.311)	85%
Breslow-ICM	-1438.17(51.27)	588(646)	0.057(0.063)	100%
PL-ICM	**-1438.17(51.27)**	**99(36)**	**0.011(0.004)**	**100%**
Turnbull-ICM	-1438.17(51.23)	54,654(53,768)	7.357(7.256)	100%
Yu-ICM	-1438.17(51.22)	6,022(8,344)	0.742(0.999)	100%
ICM	-1438.18(51.11)	4,296(14,555)	1.374(4.672)	100%
Quasi-Newton	-1442.53(51.27)	108(1)	6.576(0.381)	100%

Median and inter-quartile range (in parentheses) of log-likelihood, run times (in iterations and seconds) and percentage converged before 100,000 iterations or with finite log-likelihood for 50% truncated experiment with 2000 observations after truncation. The proposed PL and PL-ICM methods are highlighted in bold.

References

Byrd, R.H., Lu, P., Nocedal, J., Zhu, C.: A limited memory algorithm for bound constrained optimization. SIAM Journal on scientific computing **16**(5), 1190–1208 (1995). Publisher: SIAM

Gao, F., Chan, K.C.G.: Semiparametric Regression Analysis of Length-Biased Interval-Censored Data. Biometrics **75**(1), 121–132 (2018) https://doi.org/10.1111/biom.12970

Gentleman, R., Geyer, C.J.: Maximum likelihood for interval censored data: Consistency and computation. Biometrika **81**(3), 618–623 (1994). Publisher: Oxford University Press

Hudgens, M.G.: On Nonparametric Maximum Likelihood Estimation with Interval Censoring and Left Truncation. Journal of the Royal Statistical Society Series B: Statistical Methodology **67**(4), 573–587 (2005) https://doi.org/10.1111/j.1467-9868.2005.00516.x

Li, C., Pak, D., Todem, D.: Adaptive lasso for the Cox regression with interval censored and possibly left truncated data. Statistical Methods in Medical Research **29**(4), 1243–1255 (2020) https://doi.org/10.1177/0962280219856238

Shen, P.-S.: Nonparametric estimators of survival function under the mixed case interval-censored model with left truncation. Lifetime Data Analysis **26**(3), 624–637 (2020) https://doi.org/10.1007/s10985-020-09493-2

Turnbull, B.W.: The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data. Journal of the Royal Statistical Society: Series B (Methodological) **38**(3), 290–295 (1976) https://doi.org/10.1111/j.2517-6161.1976.tb01597.x

Yu, Q.: The generalised MLE with truncated interval-censored data. Journal of Nonparametric Statistics **35**(2), 266–282 (2023) https://doi.org/10.1080/10485252.2022.2147173