Supplementary Material: A fast and stable algorithm for calculating the non-parametric maximum likelihood estimate of 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. 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.

This method requires the R function `optim` - a generic optimization function and requires the likelihood from Equation 1 and the gradient of the likelihood from

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

Because `optim` minimizes objective functions by default, we supply it with the negative likelihood and gradient. We set the `optim` function to call the L-BFGS-B algorithm (Byrd et al., 1995), a memory limited 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 increase 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 6, and the conditions for all values of h_j, 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 6, 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 $b_j(1 - h_j) = 0$. 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 $h_j = 0$, $b_j = 0$ 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 the maximum likelihood, the following checks ensure 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.

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