Published in final edited form as:

Lifetime Data Anal. 2017 January ; 23(1): 102–112. doi:10.1007/s10985-016-9374-z.

Acceleration of Expectation-Maximization algorithm for length-biased right-censored data

Kwun Chuen Gary Chan¹

¹Department of Biostatistics, University of Washington, Campus, Box 357232, Seattle, WA 98195-7232, USA

Abstract

Vardi's Expectation-Maximization (EM) algorithm is frequently used for computing the nonparametric maximum likelihood estimator of length-biased right-censored data, which does not admit a closed-form representation. The EM algorithm may converge slowly, particularly for heavily censored data. We studied two algorithms for accelerating the convergence of the EM algorithm, based on iterative convex minorant and Aitken's delta squared process. Numerical simulations demonstrate that the acceleration algorithms converge more rapidly than the EM algorithm in terms of number of iterations and actual timing. The acceleration method based on a modification of Aitken's delta squared performed the best under a variety of settings.

Keywords

Aitken's delta squared; Expectation-Maximization; Iterative convex minorant; Isotonic regression; Multiplicative censoring

1 Introduction

Length-biased survival data are frequently observed when data are sampled from a group of individuals who have experienced disease incidence but not failure event before the sampling time. Prevalent sampling is often considered a more focused and economical study design (Brookmeyer and Gail 1987; Wang 1991). The observed data from prevalent sampling is typically left truncated and right censored, where truncation time is defined as the time between disease onset and the recruitment time. When disease incidence is stationary over calendar time, left-truncated survival data are length-biased. Length-biased survival data exhibits unique statistical challenges. For example, the NPMLE for left-truncated right-censored data (Tsai et al. 1987) is inefficient for length-biased survival data because information for stationary disease incidence is not utilized. Vardi (1989) discussed an EM algorithm for computing the NPMLE for a general class of multiplicative censoring problem in which length-biased right-censored data is a special case. See Wang (1991) and Asgharian et al. (2002) for related discussions. The NPMLE does not admit a closed-form estimator, and recently Huang and Qin (2011) studied a closed form estimator which is more efficient than the truncation product limit estimator. Qin et al. (2011) extended the EM algorithm and studied NPMLE for more general models.

Despite the lack of closed-form representation, estimation based on NPMLE is desirable because of optimal estimation efficiency. The lack of closed-form expression for the point estimator also affects the estimation of asymptotic variance. In general, no simple plug-in estimator is available and bootstrapping is needed. However, the speed of the EM algorithm can be slow, and the problem compounds when bootstrapping is performed. Therefore, improvement in the speed of the computation of NPMLE will be useful in practice.

Computation of NPMLE is central to many survival analysis problems. A prominent alternative to the EM algorithm is the iterative convex minorant algorithm (Jongbloed 1998), which has been studied extensively for current status data and interval censoring (Song 2004; Zhang and Sun 2010), double censoring (Wellner and Zhan 1997) and panel count data (Wellner and Zhang 2000).

The idea of ICM can also be used to accelerate the convergence of EM algorithms. To improve the computation speed of NPMLE for doubly censored data, Wellner and Zhan (1997) proposed a hybrid algorithm that adds a gradient-type proposal before each EM iteration is performed. To ensure the proposal is a survival distribution, an iterative convex minorant algorithm is used. Wellner and Zhan (1997) proved that the algorithm converges globally under general conditions. They studied doubly censoring in detail, and briefly mentioned that the algorithm is applicable to multiplicative censoring as well. In this paper we first study a version of the hybrid algorithm that specializes for the problem of length-biased survival data in detail. Although the hybrid EM algorithm based on ICM leads to accelerated convergence, each iteration requires additional computations of the first two derivatives of the log-likelihood function and a weighted isotonic least square problem. To circumvent these expensive computations, we also explore a different acceleration method based on Aitken (1926) known as the delta squared process.

The paper is organized as follows. Vardi's EM algorithm, and acceleration algorithms based on iterative convex minorant and Aitken's delta squared process is given in Sect. 2. Simulation results are given in Sect. 3 to demonstrate the improvement of the acceleration algorithms. Concluding remarks are given in Sect. 4.

2 Accelerated EM algorithms

2.1 Overview of Vardi's EM algorithm

Let y_i , $i = 1, \dots, n$ be observed survival data and δ_i be the indicator of failure event. Let $F(t)$ be the distribution function of the survival time of interest. Under length-biased sampling, the likelihood for (y_i, δ_i) , $i = 1, \dots, n$ is proportional to

$$\mathcal{L} = \left\{ \prod_{i=1}^n \frac{[dF(y_i)]^{\delta_i} [1-F(y_i)]^{1-\delta_i}}{\mu} \right\}$$

where $\mu = \int_0^\infty [1-F(t)]dt$. From Vardi (1989), the nonparametric maximum likelihood estimator can only allocate positive masses at $y_i = 1, \dots, n$. Unlike usual right-censored data, positive masses may be assigned to censored observations, and the NPMLE is still uniquely

defined when all observations are censored. See Vardi (1989) for detailed discussions. Let $t_1, \dots, t_h$ denote the distinct and ordered values of $y_1, \dots, y_n$ such that $0 \equiv t_0 < t_1 < \dots < t_h$, and let ξ_j and ζ_j be the multiplicity of uncensored and censored events at t_j , that is,

$$\xi_j = \sum_{i=1}^n I(y_i = t_j, \delta_i = 1) \text{ and } \zeta_j = \sum_{i=1}^n I(y_i = t_j, \delta_i = 0).$$

Furthermore, to simplify the notations for the ICM algorithm, we consider the following parametrization such that $x_j = F(t_j)$, $j = 1, \dots, h$. By definition, we have an ordering constraint:

$$0 \equiv x_0 \leq x_1 \leq \dots \leq x_h \leq 1. \quad (1)$$

Moreover, μ can be expressed as

$$\mu = \sum_{j=1}^h (1 - x_{j-1})(t_j - t_{j-1}) = t_h - \sum_{j=1}^h x_{j-1}(t_j - t_{j-1}).$$

Therefore, the likelihood for the observed data is proportional to

$$\mathcal{L} \propto \prod_{j=1}^h \frac{(x_j - x_{j-1})^{\xi_j} (1 - x_j)^{\zeta_j}}{[t_h - \sum_{l=1}^h x_{l-1}(t_l - t_{l-1})]}. \quad (2)$$

Vardi (1989) derived his EM algorithm using a different parametrization of the likelihood, which is equivalent to (2) upon reparametrization. We use the current parametrization so that the parameters $x_1, \dots, x_h$ satisfy the shape constraint (1) which is crucial for the ICM algorithm.

For completeness, we state Vardi's EM algorithm using our current parametrization:

1. Initialize x_j^{old} , $j = 1, \dots, h$ such that $0 < x_1^{old} < x_2^{old} < \dots < x_h^{old}$.
2. Replace x_j^{old} with

$$x_j^{new} = \sum_{l=1}^j \frac{t_l^{-1} [\xi_l + (x_l^{old} - x_{l-1}^{old}) \sum_{k=1}^l \zeta_k (1 - x_{k-1}^{old})^{-1}]}{\sum_{l'=1}^h t_{l'}^{-1} [\xi_{l'} + (x_{l'}^{old} - x_{l'-1}^{old}) \sum_{k=1}^{l'} \zeta_k (1 - x_{k-1}^{old})^{-1}]}.$$

2.2 Acceleration based on iterative convex minorant

Speeding up the EM algorithm using Newton-type methods has been studied extensively in the literature, see for example Meilijson (1989). However, the proposed parameter value from a Newton-step may not be a distribution function, that is, (1) may not be satisfied. Moreover, the Hessian matrix for NPMLE is high-dimensional and can be prohibitively expensive to compute. In order to address these two problems, Wellner and Zhan (1997)

proposed the use of an ICM algorithm first proposed by Jongbloed (1998). The ICM algorithm, similar to the Newton's method, involves a quadratic approximation of the log-likelihood function at the current estimate. The major difference is that the quadratic approximation together with the shape constraint (1) defines an isotonic regression problem (Barlow et al. 1972) and the solution can be computed by the pool-adjacent-violator algorithm (Ayer et al. 1955).

Details for the length-biased right-censored problem is given as follows. The log-likelihood function based on (2) is

$$\phi(x) = \sum_{j=1}^h [\xi_j \log(x_j - x_{j-1}) + \zeta_j (1 - x_j)] - \log \left[t_h - \sum_{l=1}^h x_{l-1} (t_l - t_{l-1}) \right].$$

The maximization problem defining the NPMLE is given by:

$$\text{Maximize } \phi(x) \text{ over } x \in \mathcal{C}_x = \{x: 0 \leq x_1 \leq \dots \leq x_h \leq 1\}.$$

Let $\nabla^2 \phi$ be the Hessian matrix of ϕ . For an arbitrary real vector $\alpha = (\alpha_1, \dots, \alpha_h)^T$, we can follow the proof in Vardi (1989) and Chan and Qin (2016) to show that

$$\begin{aligned} \alpha^T (\nabla^2 \phi) \alpha = & - \sum_{i=1}^h (\alpha_i - \alpha_{i-1})^2 \frac{\xi_i}{(x_i - x_{i-1})^2} - \sum_{i=1}^h (\alpha_h - \alpha_i)^2 \frac{\zeta_i}{(1 - x_i)^2} \\ & - \frac{h}{\mu^2} \left[\left\{ \sum_{i=1}^h t_i (x_i - x_{i-1}) \right\} \left\{ \sum_{i=1}^h t_i (x_i - x_{i-1}) a_i^2 \right\} - \left\{ \sum_{i=1}^h t_i (x_i - x_{i-1}) a_i \right\}^2 \right] \end{aligned}$$

where $\alpha_0 = x_0 = 0$ and $a_i = (\alpha_i - \alpha_{i-1}) / (x_i - x_{i-1})$. Since the last term is non-positive by Cauchy-Schwarz inequality and $\xi_i \geq 0$, $\zeta_i \geq 0$ with $\xi_i + \zeta_i > 0$, the above quadratic form is strictly negative unless $\alpha \equiv 0$. Therefore, ϕ is strictly concave.

Let $\nabla \phi_j(x) = \phi(x) / x_j$ and $d_j(x) = -\partial^2 \phi(x) / \partial x_j^2$, that is

$$\nabla \phi_j(x) = \frac{\xi_j}{x_j - x_{j-1}} - \frac{\xi_{j+1}}{x_{j+1} - x_j} - \frac{\zeta_j}{1 - x_j} + \frac{n(t_{j+1} - t_j)}{t_h - \sum_{l=1}^h x_{l-1} (t_l - t_{l-1})},$$

and

$$d_j(x) = \frac{\xi_j}{(x_j - x_{j-1})^2} + \frac{\xi_{j+1}}{(x_{j+1} - x_j)^2} + \frac{\zeta_j}{(1 - x_j)^2} - \frac{n(t_{j+1} - t_j)^2}{[t_h - \sum_{l=1}^h x_{l-1} (t_l - t_{l-1})]^2},$$

where we define $\xi_{h+1} = 0$, $t_{h+1} = t_h$ and $0/0 = 0$. Following Wellner and Zhan (1997), let $r_j = x_j + \nabla \phi_j(x)/d_j$, the maximization problem of the ICM algorithm is equivalent to the following isotonic regression problem:

$$\text{Maximize } -\frac{1}{2} \sum_{j=1}^h d_j (w_i - r_i)^2 \text{ over } w \in \mathcal{C}_w.$$

The solution for the above problem can be computed by the pooled adjacent violator algorithm (Ayer et al. 1955) that attains the solution in $O(n)$ time (Grotzinger and Witzgall 1984). The solution can be represented as the left derivative of the convex minorant of the cumulative sum diagram consisting the following points:

$$P_0 = (0, 0), \quad P_j = (G_j(x), V_j(x)), j=1, \dots, h$$

where $G_j(x) = \sum_{l=1}^j d_l(x)$ and $V_j(x) = \sum_{l=1}^j [x_l d_l(x) + \nabla \phi_l(x)]$.

Similar to any gradient-type method, the ICM step does not guarantee that the likelihood increases at every iteration. To guarantee the ascent property enjoyed by the EM algorithm, a line search is typically performed on the direction defined as the difference between the proposed value and the last value. Different types of line search algorithms can be used, for example, step-halving or backtracking, see Lange (2013) for a detailed discussion. In particular, Jongbloed (1998) and Wellner and Zhan (1997) used backtracking with Armijio's rule. The hybrid EM algorithm is given as follows:

1. Initialize $x^{old} = (x_1^{old}, \dots, x_h^{old})$ such that $0 < x_1^{old} < x_2^{old} < \dots < x_h^{old}$.
2. Compute a proposal value $\tilde{x}$ which is the left derivative of the convex minorant of the cumulative sum diagram consisting of the following points:

$$P_0 = (0, 0), \quad P_j = (G_j(x^{old}), V_j(x^{old})), j=1, \dots, h$$

3. If $\phi(\tilde{x}) > \phi(x^{old})$, proceed to the next step. Otherwise, replace $\tilde{x} = x^{old} + \varepsilon(\tilde{x} - x^{old})$ where $\varepsilon \in [0, 1)$ such that $\phi(\tilde{x}) > \phi(x^{old})$. This can be found by step-halving or backtracking.
4. Replace x_j^{old} with

$$x_j^{new} = \sum_{l=1}^j \frac{t_l^{-1} [\xi_l + (\tilde{x}_l - \tilde{x}_{l-1}) \sum_{k=1}^l \zeta_k (1 - \tilde{x}_{k-1})^{-1}]}{\sum_{l'=1}^h \sum_{l''=1}^h t_{l''}^{-1} [\xi_{l''} + (\tilde{x}_{l''} - \tilde{x}_{l''-1}) \sum_{k=1}^{l''} \zeta_k (1 - \tilde{x}_{k-1})^{-1}]}.$$

2.3 Acceleration based on Aitken's delta squared process

Although the ICM-based acceleration discussed in the previous subsection can lead to substantial reduction in the number of iterations as will be shown in the simulations, the actual time saved was not as substantial as the author initially expected. The main reason is

because each iteration requires additional computations of the first two derivatives of the log-likelihood function and an isotonic least-square problem. To circumvent these difficulties, we study a variant of Aitken's delta squared process (Aitken 1926) proposed by Steffensen (1933) which is specifically designed for accelerating convergence of fixed-point algorithms with linear rates of convergence, for which the EM algorithm is a particular example.

Aitken's delta squared process is an extrapolation algorithm based on three points in a sequence. Suppose that a scalar sequence $\{z_k\}_{k=0}^{\infty}$ converges at a linear rate to a limit z^* , we have

$$z_{k+1} - z^* = K(z_k - z^*) + o(|z_k - z^*|)$$

where $|K| < 1$. As shown in Appendix D of Traub (1964),

$$z_k - \frac{(z_{k+1} - z_k)^2}{z_{k+2} - 2z_{k+1} + z_k} = z^* + o(|z_k - z^*|). \quad (3)$$

Let

$$\hat{z}_k = z_k - \frac{(z_{k+1} - z_k)^2}{z_{k+2} - 2z_{k+1} + z_k},$$

it follows from (3) that $\hat{z}_k$ converges to z^* at a faster rate than z_k .

Aitken's delta squared process has been widely used in fixed-point algorithms, defined by

$$z_{n+1} = f(z_n), \quad n=0, 1, 2, \dots$$

for some iteration function f . Given the original sequence z_n , the transformed sequence $\hat{z}_n$ can be calculated by three successive values z_n , z_{n+1} and z_{n+2} in the original sequence. The computation of the transformed sequence $\hat{z}_n$ only requires evaluating the iteration function f but not its derivatives. A slight variation called Steffensen's Method (Steffensen 1933) redefines the iteration function as follows:

$$f^S(x) = x - \frac{(f(x) - x)^2}{(f \circ f(x) - 2f(x) + x)},$$

and the corresponding sequence is defined as

$$y_{n+1} = f^S(y_n), \quad n=0, 1, 2, \dots$$

Comparing Aitken's transformed sequence $\hat{z}_n$ and Steffensen's iterations y_n , the computation of $\hat{z}_n$ requires the original sequence z_n to be computed, and can be regarded as an extraction of extra information from a given sequence. The Steffensen's Method, on the other hand, alternates between two fixed-point iterations and one Aitken extrapolation, so that the values of acceleration steps are used as initial values in subsequent steps.

However, Steffensen's Method cannot be directly applied to Vardi's EM algorithm component wise, because the Aitken extrapolation step does not guarantee that the order restriction (1) is satisfied.

To circumvent this problem, we consider a reparametrization in terms of hazards:

$$\lambda_j = \frac{x_j - x_{j-1}}{1 - x_{j-1}}, \quad j=1, \dots, h.$$

It is required that $\lambda_j \geq 0, j=1, \dots, h$ which can be easily imposed component wise. A similar transformation is considered in Kuroda et al. (2008) for log-linear models with partially classified categorical data.

The modified EM algorithm implementing Steffensen's variation of Aitken's delta process is given as follows:

1. Initialize $x^{old} = (x_1^{old}, \dots, x_h^{old})$ such that $0 < x_1^{old} < x_2^{old} < \dots < x_h^{old}$.
2. Compute two EM steps:

$$x_j^{EM,1} = \sum_{l=1}^j \frac{t_l^{-1} [\xi_l + (x_l^{old} - x_{l-1}^{old}) \sum_{k=1}^l \zeta_k (1 - x_{k-1}^{old})^{-1}]}{\sum_{l'=1}^h t_{l'}^{-1} [\xi_{l'} + (x_{l'}^{old} - x_{l'-1}^{old}) \sum_{k=1}^{l'} \zeta_k (1 - x_{k-1}^{old})^{-1}]},$$

$$x_j^{EM,2} = \sum_{l=1}^j \frac{t_l^{-1} [\xi_l + (x_l^{EM,1} - x_{l-1}^{EM,1}) \sum_{k=1}^l \zeta_k (1 - x_{k-1}^{EM,1})^{-1}]}{\sum_{l'=1}^h t_{l'}^{-1} [\xi_{l'} + (x_{l'}^{EM,1} - x_{l'-1}^{EM,1}) \sum_{k=1}^{l'} \zeta_k (1 - x_{k-1}^{EM,1})^{-1}]},$$

and transform the distribution functions $x^{old}, x^{EM,1}$ and $x^{EM,2}$ into hazards $\lambda^{old}, \lambda^{EM,1}$ and $\lambda^{EM,2}$.

3. Compute the Aitken's iteration:

$$\lambda^* = \max \left(0, \lambda_j^{old} - \frac{(\lambda_j^{EM,1} - \lambda_j^{old})^2}{\lambda_j^{old} - 2\lambda_j^{EM,1} + \lambda_j^{EM,2}} \right)$$

and back transform $x_j^* = 1 - \prod_{k=1}^{j-1} (1 - \lambda_k^*) \lambda_j^*$, where $\prod_{j=1}^0 x_j = 1$ by convention.

4. Replace x^{old} with x^* if $\phi(x^*) > \phi(x^{old})$. Otherwise, replace x^{old} with $x^{EM,2}$.

Similar to the ICM algorithm, Steffensen's method does not guarantee that the likelihood increases at every iteration. Step 4 serves as a monotone correction since the EM algorithm has a monotone convergence property.

3 Numerical examples

We performed simulation studies to evaluate the performance of Vardi's EM algorithm and the acceleration algorithms discussed in Sect. 2. Independent data sets are generated 1000 times for each scenario. Survival times T are generated from an exponential distribution with mean 3 times units. We also simulated data from Weibull distributions and the results are similar and are omitted. Here, we studied the performance of the algorithms under a variety of sample sizes and length of study periods. To obtain length-biased samples, we generate random truncation times A^0 from a $U(0, 30)$ distribution; an observation is in the cross-sectional sample if $-A^0 + T \geq 0$. Data are generated until the cross-sectional samples have $n = 100, 200, 500$ and 1000 observations. The survival endpoint is censored if an individual in the cross-sectional cohort survives past C' time units after recruitment, where C' is generated from $U(0, \tau)$ distribution, $\tau = 0, 1, 2, 3$. The maximum length of prospective follow-up is τ . Note that when $\tau = 0$, there is no follow-up after recruitment and all observations are censored. Unlike right censored data where NPMLE does not exist when all observations are censored, NPMLE for length-biased survival data without follow-up exists (Vardi 1989). The reason is that partial survival information is available from the time between disease onset and recruitment. We compared Vardi's EM algorithm, the acceleration based on ICM and Steffensen's method. Convergence criterion is based on maximum coordinate wise distance between two iterations, and the tolerance level is set to 10^{-6} .

The results are shown in Table 1. Compare to Vardi's EM algorithm, the two acceleration based on ICM and Steffensen's method substantially decrease the average number of iterations. When prospective follow-up is present, Aitken's acceleration is the fastest. Although the number of iterations for ICM is much smaller than Vardi's EM algorithm, the actual timing decrease is not as substantial mainly due to additional computations required in each iteration. We also performed simulations for the Louis' method as discussed in Sect. 4.8 of McLachlan and Krishnan (2008). Louis' method requires additional computations of derivatives and the performance is worse than the other acceleration algorithms. When there is no prospective follow-up, ICM performed the best among the three algorithms. It is because ICM is particularly designed for current status data (Jongbloed 1998) and the statistical structure of length-biased data without follow-up is similar to current status data. The difference between EM algorithm and ICM acceleration decreases with increasing follow-up time, which is similar to the results in Wellner and Zhan (1997). We also studied the performance of NPMLE computed by different algorithms. Table 2 shows the results for $\tau = 3$. It can be seen that the bias and variability of the estimates computed by different algorithms have negligible differences.

4 Concluding remarks

Vardi's EM algorithm is very simple to implement, and is numerically stable. However, convergence can be slow particularly for heavily censored data. Acceleration algorithms discussed in this paper can substantially reduce the number of iterations and the time to compute the nonparametric maximum likelihood estimator.

Theoretical properties for the hybrid ICM-EM algorithm has been vigorously developed by Wellner and Zhan (1997), but we found that the Aitken's delta squared process can compute the NPMLE faster than the ICM-EM algorithm because calculation of derivatives and isotonic regression estimates are not required. Aitken's algorithm also retains stability and simplicity of the EM algorithm. We found that the ICM-EM algorithm is typically quite effective in increasing the likelihood in the initial iterations, and the Aitken's delta squared process is particularly effective near convergence. Therefore, a combination of ICM at initial iterations and Aitken's extrapolation at later iterations would further decrease the number of iterations needed. In limited simulations we found, however, the decrease in iterations was generally outweighed by the increase in computation time needed for the ICM steps, unless the data is heavily (or totally) censored. Therefore, we recommend the use of Aitken's acceleration in practice when prospective follow-up is present.

Acknowledgments

The author is partially funded by the National Institute of Health Grant R01 HL122212.

References

- Aitken AC. On bernoulli's numerical solution of algebraic equations. *Proc R Soc Edinb.* 1926; 46:289–305.
- Asgharian M, M'Lan CE, Wolfson DB. Length-biased sampling with right censoring: an unconditional approach. *J Am Stat Assoc.* 2002; 97(457):201–209.
- Ayer M, Brunk HD, Ewing GM, Reid W, Silverman E, et al. An empirical distribution function for sampling with incomplete information. *Ann Math Stat.* 1955; 26(4):641–647.
- Barlow, RE., Bartholomew, DJ., Bremner, J., Brunk, HD. Statistical inference under order restrictions: the theory and application of isotonic regression. Wiley; New York: 1972.
- Brookmeyer R, Gail M. Biases in prevalent cohorts. *Biometrics.* 1987; 43(4):739–749. [PubMed: 3427161]
- Chan KCG, Qin J. Nonparametric maximum likelihood estimation for the multi-sample wicksell corpuscle problem. *Biometrika.* 2016; 103(2):253–271. [PubMed: 27279656]
- Grotzinger S, Witzgall C. Projections onto order simplexes. *Appl Math Optim.* 1984; 12(1):247–270.
- Huang CY, Qin J. Nonparametric estimation for length-biased and right-censored data. *Biometrika.* 2011; 98(1):177–186. [PubMed: 23049126]
- Jongbloed G. The iterative convex minorant algorithm for nonparametric estimation. *J Comput Graph Stat.* 1998; 7(3):310–321.
- Kuroda M, Sakakihara M, Geng Z. Acceleration of the em and ecm algorithms using the aitken δ^2 method for log-linear models with partially classified data. *Stati Probab Lett.* 2008; 78(15):2332–2338.
- Lange, K. Optimization. Springer; New York: 2013.
- McLachlan, G., Krishnan, T. The EM algorithm and extensions. Wiley; New York: 2008.
- Meilijson I. A fast improvement to the em algorithm on its own terms. *J R Stat Soc Ser B.* 1989; 51(1): 127– 138.
- Qin J, Ning J, Liu H, Shen Y. Maximum likelihood estimations and em algorithms with length-biased data. *J Am Stat Assoc.* 2011; 106(496):1434–1449. [PubMed: 22323840]
- Song S. Estimation with univariate “mixed case” interval censored data. *Stat Sin.* 2004; 14:269–282.
- Steffensen J. Remarks on iteration. *Scand Actuar J.* 1933; 1933(1):64–72.
- Traub, JF. Iterative methods for the solution of equations. Prentice Hall; Englewood Cliffs: 1964.
- Tsai WY, Jewell NP, Wang MC. A note on the product-limit estimator under right censoring and left truncation. *Biometrika.* 1987; 74(4):883–886.

- Vardi Y. Multiplicative censoring, renewal processes, deconvolution and decreasing density: non-parametric estimation. *Biometrika*. 1989; 76(4):751–761.
- Wang MC. Nonparametric estimation from cross-sectional survival data. *J Am Stat Assoc*. 1991; 86(413):130–143.
- Wellner JA, Zhan Y. A hybrid algorithm for computation of the nonparametric maximum likelihood estimator from censored data. *J Am Stat Assoc*. 1997; 92(439):945–959.
- Wellner JA, Zhang Y. Two estimators of the mean of a counting process with panel count data. *Ann Stat*. 2000; 28(3):779–814.
- Zhang Z, Sun J. Interval censoring. *Stat Methods Med Res*. 2010; 19(1):53–70. [PubMed: 19654168]

Simulation results comparing the average number of iterations and time in milliseconds, for Vardi's EM algorithm (EM), acceleration based on iterative convex minorant (ICM) and Steffensen's method

Table 1

$\tau = 0$		$\tau = 1$		$\tau = 2$		$\tau = 3$		
Iterations	Time	Iterations	Time	Iterations	Time	Iterations	Time	
$n = 100$								
EM	1578	141	585	48	230	18	144	11
ICM	456	97	45	18	32	13	27	11
Steffensen	249	45	50	8	20	3	14	2
$n = 200$								
EM	2406	217	579	50	240	20	140	12
ICM	357	74	45	18	33	14	25	12
Steffensen	503	112	61	12	21	4	14	3
$n = 500$								
EM	4088	737	495	89	203	36	121	23
ICM	118	91	47	38	35	30	30	26
Steffensen	809	432	68	27	23	9	15	6
$n = 1000$								
EM	5628	2396	433	132	170	51	105	34
ICM	102	135	48	67	36	47	28	35
Steffensen	1256	924	64	45	23	15	16	11

Simulation results comparing the mean and standard deviations of the estimators at p-th percentile of the true distribution, for Vardi's EM algorithm (EM), acceleration based on iterative convex minorant (ICM) and Steffensen's method

Table 2

		$p = 0.2$		$p = 0.4$		$p = 0.6$		$p = 0.8$	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
$n = 100$									
	EM	0.209	0.129	0.396	0.117	0.597	0.086	0.797	0.054
	ICM	0.208	0.127	0.396	0.115	0.597	0.085	0.797	0.053
	Steffensen	0.207	0.125	0.395	0.115	0.597	0.084	0.797	0.053
$n = 200$									
	EM	0.188	0.099	0.387	0.084	0.591	0.063	0.796	0.038
	ICM	0.188	0.098	0.387	0.084	0.591	0.063	0.796	0.038
	Steffensen	0.189	0.097	0.387	0.083	0.591	0.062	0.796	0.038
$n = 500$									
	EM	0.195	0.069	0.396	0.057	0.596	0.042	0.798	0.025
	ICM	0.195	0.069	0.396	0.057	0.596	0.042	0.798	0.025
	Steffensen	0.195	0.069	0.396	0.057	0.596	0.042	0.798	0.025
$n = 1000$									
	EM	0.198	0.053	0.398	0.044	0.598	0.032	0.800	0.018
	ICM	0.198	0.053	0.398	0.044	0.598	0.032	0.800	0.018
	Steffensen	0.198	0.053	0.398	0.044	0.599	0.032	0.800	0.018