Supplementary material for constrained nonparametric maximum likelihood estimation of stochastically ordered survivor functions

Yongseok Park*, John D. Kalbfleisch and Jeremy M.G. Taylor

Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109-2029.

APPENDIX

Algorithm 4 (Bounded Below Constraint (modified from Dykstra, 1982))**.**

1. Set $i_0 = 0$, $\ell = 1$ and $m' = \max\{i : n_{1i} > 0\}$.
2. If there exists $b > i_{\ell-1}$ such that $K(i_{\ell-1}+1, b) > 0$, then let $i_\ell = \min\left\{\arg\max_{b>i_{\ell-1}} K(i_{\ell-1}+1, b)\right\}$ and $k_\ell = K(i_{\ell-1}+1, i_\ell)$. Otherwise, let $i_\ell = m'$, $k_\ell = 0$.
3. Let $\hat{h}_{1j} = \log\{1 - d_{1j}/(n_{1j} + k_\ell)\}, i_{\ell-1}+1 \le j \le i_\ell$.
4. If $i_\ell = m'$, stop. Otherwise, set $\ell = \ell + 1$ and go to step 2.

Note that this algorithm gives a KM type estimator in which the number at risk is potentially modified at each potential failure time. It can be shown that $k_1 \ge k_2 \ge \cdots$, so that this estimate is essentially a KM estimate based on modified data where k_1 more subjects are placed at risk at time 0, and at time $a_{i_\ell}, k_\ell - k_{\ell+1}, \ell = 1, 2, \ldots$ of these additional subjects are censored.

Step 2 in Algorithm 4 is looking for the next active constraint in the solution. A root finding procedure is needed to calculate $K(i_{\ell-1}+1, b)$. To find a root with high precision is computationally intensive, so it is inefficient to calculate $K(i_{\ell-1}+1, b)$ for all $b > i_{\ell-1}$ to find the index of the next active constraint. Instead we propose another algorithm that is equivalent to Algorithm 4 but only calculates $K(i_{\ell-1}+1, b)$ when necessary.

Algorithm 5 (Bounded Below Constraint)**.**

1. Set $i_0 = 0$, $\ell = 1$ and $m' = \max(i : n_{1i} > 0)$.
2. Let $i_\ell = \min_{b>i_{\ell-1}} \{b : H(i_{\ell-1}+1, b, 0) < 0\}$, then set $k_\ell = K(i_{\ell-1}+1, i_\ell)$. If no such i_ℓ exists, set $i_\ell = m'$ and $k_\ell = 0$ and go to step 4.
3. Let $I = \min_{b>i_\ell} \{b : H(i_\ell+1, b, k_\ell) < 0\}$. If no such I exists, then go to step 4. Otherwise, set $i_\ell = I$ and repeat step 3.
4. Let $\hat{h}_{1j} = \log\{1 - d_{1j}/(n_{1j} + k_\ell)\}, i_{\ell-1}+1 \le j \le i_\ell$.
5. If $i_\ell = m'$, stop. Otherwise, set $\ell = \ell + 1$ and go to step 2.

The two Algorithms are equivalent because steps 2 and 3 in Algorithm 5 are looking for $\min\left\{\arg\max_{b>i_{\ell-1}} K(i_{\ell-1}+1, b)\right\}$ as in Algorithm 4. However, Algorithm 5 implements a root finding procedure only when it finds a position b, where $K(i_{\ell-1}+1, b)$ is larger than the previously found maximum $K(i_{\ell-1}+1, i_\ell)$. This significantly improves the efficiency of the calculations.

Proof of Theorem 4. First, we show that $\tilde{S}_1(t)$ is a C-NPMLE. Let $\bar{S}_1(t)$ be a C-NPMLE subject to $T_1 \le_{st} T_2$. Note that $\hat{S}_1(t)$ is the maximum likelihood estimator subject to fewer constraints (only at times X_i') compared to $\bar{S}_1(t)$, we have that $L(\hat{S}_1(t)) \geq L(\bar{S}_1(t))$. Further $L(\bar{S}_1(t)) \geq L(\tilde{S}_1(t))$ since $\tilde{S}_1(t) = \min(\hat{S}_1(t), S_2(t)) \leq S_2(t)$. Note that at every time X_i', $i = 1, \cdots, n_{tot}$, $\hat{S}_1(X_i') \leq S_2(X_i')$, the difference between $\hat{S}_1(t)$ and $\tilde{S}_1(t)$ may only occur in time interval (X_i', X_{i+1}') for some i. The five possible time intervals are (C, C), (C, X^-), (X, C) (X^-, X) and (X, X^-), where C represents censoring time, X event time and X^- time just before X. None of these intervals includes C, X or X^-, the three elements that determine likelihood (3). So $L(\tilde{S}_1(t)) = L(\hat{S}_1(t)) = L(\bar{S}_1(t))$, which implies that $\tilde{S}_1(t)$ is a C-NPMLE subject to $T_1 \leq_{st} T_2$.

Then, we show that $\tilde{S}_1(t)$ is the MC-NPMLE. Suppose it is not, we must be able to find a time x^* where $\bar{S}_1(x^*) > \tilde{S}_1(x^*) = \min\{\hat{S}_1(x^*), S_2(x^*)\}$. Then $\bar{S}_1(x^*) > \hat{S}_1(x^*)$ since $\bar{S}_1(x^*) \leq S_2(x^*)$. Consider another survivor function $S_1'(t)$ with jumps only at the times X_i' and $S_1'(X_i') = \bar{S}_1(X_i')$ for all i, $S_1'(t)$ is constrained estimator of $S_1(t)$ subject to discrete constraint at all X_i's. Since $S_1'(x^*) = S_1'(\max(X_i' : X_i' \leq x^*)) = \bar{S}_1(\max(X_i' : X_i' \leq x^*)) \geq \bar{S}_1(x^*) > \hat{S}_1(x^*)$ and $\hat{S}_1(t)$ is the MC-NPMLE with discrete constraint, $S_1'(t)$ is not a C-NPMLE subject to the discrete constraint. So $L(\bar{S}_1(t)) = L(S_1'(t)) < L(\hat{S}_1(t)) = L(\tilde{S}_1(t))$, which is a contradiction. Thus, $\tilde{S}_1(t)$ is the MC-NPMLE. ■

Proof of Theorem 5. To fix notation, let $a_0 = 0, a_1, a_2, \cdots a_m$ be the complete ordered observed event times of any given data in the two sample case, $i_0 = 0, i_1, i_2, \cdots i_L$ the index of active constraint times, and $k_1, k_2, \cdots, k_L$ be the corresponding k values from Algorithm 2.

The last active constraint time from Algorithm 2 satisfies $a_{i_L} \leq \tau$. $\hat{S}_1(t)/S_1^*(t)$ is non-decreasing and $\hat{S}_2(t)/S_2^*(t)$ is non-increasing in t in any sample. At the last active constraint a_{i_L}, $S_1^*(a_{i_L}) \leq \hat{S}_1(a_{i_L}) = \hat{S}_2(a_{i_L}) \leq S_2^*(a_{i_L})$.

So for any $x \leq a_{i_L}$,

$$\begin{aligned} S_1^*(x) \leq \hat{S}_1(x) \leq S_1^*(x)\frac{\hat{S}_1(a_{i_L})}{S_1^*(a_{i_L})} &\leq S_1^*(x)\frac{S_2^*(a_{i_L})}{S_1^*(a_{i_L})} \\ &\leq \left[\sup_{t \leq a_{i_L}} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(x) \leq \left[\sup_{t \leq \tau} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(x) \end{aligned}$$

Similarly,

$$S_2^*(x) \geq \hat{S}_2(x) \geq \left[\inf_{t \leq \tau} \frac{S_1^*(t)}{S_2^*(t)}\right] S_2^*(x).$$

For any $x > a_{i_L}$ in the same sample,

$$\begin{aligned} S_1^*(x) \leq \hat{S}_1(x) &= \hat{S}_1(a_{i_L}) \times \frac{S_1^*(x)}{S_1^*(a_{i_L})} \\ &\leq \left[\sup_{t \leq \tau} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(a_{i_L}) \times \frac{S_1^*(x)}{S_1^*(a_{i_L})} = \left[\sup_{t \leq \tau} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(x). \end{aligned}$$

So, regardless of where a_{i_L} is, in any sample, for any $x \leq \tau$, we always have

$$S_1^*(x) \leq \hat{S}_1(x) \geq \left[\sup_{t \leq \tau} \frac{S_2^*(t)}{S_1^*(t)}\right] S_1^*(x) \text{ and } S_2^*(x) \geq \hat{S}_2(x) \geq \left[\inf_{t \leq \tau} \frac{S_1^*(t)}{S_2^*(t)}\right] S_2^*(x)$$

As n_1, n_2 go to ∞, from (4), for any $x \leq \tau$, $S_1^*(x) \to S_1(x)$ and $S_2^*(x) \to S_2(x)$. So

$$\frac{S_2^*(x)}{S_1^*(x)} \to \frac{S_2(x)}{S_1(x)} \Rightarrow \sup_{x \leq \tau} \frac{S_2^*(x)}{S_1^*(x)} \to \sup_{x \leq \tau} \frac{S_2(x)}{S_1(x)} = 1,$$

in probability if indeed $S_1(t) \geq S_2(t)$ for all t. Thus $P\left\{\sup_{x \leq \tau} |\hat{S}_1(x) - S_1^*(x)| > \epsilon\right\} \to 0$ for every $\epsilon > 0$. Then using Meier's result (4), we obtain the desired result for $\hat{S}_1(t)$.

Similarly, we can show that $P\{\sup_{x\leq\tau}|\hat{S}_2(x)-S_2^*(x)|>\epsilon\}\to 0$ for every $\epsilon>0$ and hence we obtain the desired result for $\hat{S}_2(t)$.

To simplify the proof, we only show consistency of the C-NPMLE in the case of iid data. However, it can be shown that the estimators are consistent in the more general situation as discussed in Dykstra (1982). ■

Proof of Theorem 3. Recall that $m_g=\max\{i:n_{gi}>0\}, g=1,2$ and $m'=\min(m_1,m_2)$. In this section, we prove that the $\hat{h}_{gi}, g=1,2, i=1,\ldots,m_g$ in Theorem 3 and Algorithm 2 is the C-NPMLE. We use $\hat{h}_{gi}^A$ and $\hat{h}_{gi}^T$ to distinguish the results from Algorithm 2 and Theorem 3 before we prove that they are the same. In Lemma 2, we show that $\hat{h}_{gi}^A=\hat{h}_{gi}^T$, and in Lemma 3, we show that $\hat{h}_{1i}$ is the C-NPMLE and $\hat{h}_{2i}$ is the MC-NPMLE.

Characteristics of Results From Theorem 3 and Algorithm 2

Steps 2,3 and 4 in Algorithm 2 are used to calculate k_ℓ. Suppose the algorithm has R_ℓ iterations with initial step 2 to 3 (0^{th} iteration) and step 4 to 3 ($1^{st},\ldots,R_\ell^{th}$ iteration) before it finds i_ℓ. Here, $i_\ell^{(r)}$ and $k_\ell^{(r)}, r=0,\ldots,R_\ell$ are results from iteration r. Let $i_\ell^{(-1)}=i_{\ell-1}$ and $k_\ell^{(-1)}=0$ for convenience. Note that $H_2(a,b,k)$ is a non-decreasing function in k and strictly increasing if there is at least one observed event between a and b.

Lemma 1. *The result for any data from Algorithm 2 has the following properties:*
(a). $\sum_{j=1}^{i_\ell}(\hat{h}_{1j}^A-\hat{h}_{2j}^A)=0,\ \ell=1,\ldots,L;$
(b). $k_\ell=\min\{K_2(i_{\ell-1}+1,i_\ell),n_{2i_\ell}\}$ *and* $H_2(i_{\ell-1}+1,i_\ell,k_\ell)\leq 0;$
(c). $k_\ell^{(r)}>k_\ell^{(r-1)},\ r=1,\ldots,R_\ell;$
(d). $\sum_{j=1}^{i}\hat{h}_{1j}^A\geq\sum_{j=1}^{i}\hat{h}_{2j}^A,\ i=1,\ldots,m';$
(e). $k_1\geq k_2\geq\cdots\geq k_L>0.$

Proof. (a) It is obvious since $\sum_{j=i_{\ell-1}+1}^{i_\ell}\hat{h}_{2j}^A=\sum_{j=i_{\ell-1}+1}^{i_\ell}\hat{h}_{2j}^A$, $\ell=1,\ldots,L$.

(b) From step 3 in Algorithm 2, k_ℓ can take two values: $k_\ell=n_{2i_\ell}$ if $H_2(i_{\ell-1}+1,i_\ell,n_{2i_\ell})\leq 0$ (in this case $K_2(i_{\ell-1}+1,i_\ell)\geq n_{2i_\ell}$), or $k_\ell=K_2(i_{\ell-1}+1,i_\ell)$.

(c) Show by contradiction. Suppose there exists r' such that $k_\ell^{(r')}\leq k_\ell^{(r'-1)}$. It follows that

$$\begin{aligned}&H_2(i_{\ell-1}+1,i_\ell^{(r')},k_\ell^{(r')})\leq H_2(i_{\ell-1}+1,i_\ell^{(r')},k_\ell^{(r'-1)})\\&=H_2(i_{\ell-1}+1,i_\ell^{(r'-1)},k_\ell^{(r'-1)})+H_2(i_\ell^{(r'-1)}+1,i_\ell^{(r')},k_\ell^{(r'-1)})\\&=H_2(i_\ell^{(r'-1)}+1,i_\ell^{(r')},k_\ell^{(r'-1)})<0\qquad(\text{ step 2 or 4 in Algorithm 2 }).\end{aligned}\tag{6}$$

However, from step 3 in Algorithm 2, $k_\ell^{(r')}$ must either satisfy:

$K_2(i_{\ell-1},i_\ell^{(r')})$. Then $H_2(i_{\ell-1}+1,i_\ell^{(r')},k_\ell^{(r')})=0$, which contradicts (6); or
$n_{2i_\ell^{(r')}}$. Then $k_\ell^{(r'-1)}\geq k_\ell^{(r')}=n_{2i_\ell^{(r')}}$, which contradicts the condition $n_{2i_\ell^{(r')}}>k_\ell^{(r'-1)}$ that is required to reach iteration r' in step 4 of Algorithm 2.

(d) Suppose there exists i' such that $\sum_{j=1}^{i'}\hat{h}_{1j}^A<\sum_{j=1}^{i'}\hat{h}_{2j}^A$, equivalently $\sum_{j=i_{\ell'-1}+1}^{i'}\hat{h}_{1j}^A<\sum_{j=i_{\ell'-1}+1}^{i'}\hat{h}_{2j}^A$. Then each of the three possible valid ranges of i' leads to a contradiction. That is either:

$i'=i_\ell,\ell\leq L$. Then $\sum_{j=1}^{i'}\hat{h}_{1j}^A=\sum_{j=1}^{i'}\hat{h}_{2j}^A$, which contradicts Lemma 1 (a); or
$i'>i_L$. Then $H_2(i_L+1,i',0)=\sum_{j=i_L+1}^{i'}(\hat{h}_{1j}-\hat{h}_{2j})<0$, which contradicts the condition

$H_2(i_L+1,b,0) \geq 0$ for all $b > i_L$ in step 2 of Algorithm 2; or

$i_{\ell-1} < i' < i_\ell$ for a $\ell \leq L$. Then for $r' = \max\{r : i_\ell^{(r)} < i'\}$,

$$H_2(i_{\ell-1}+1, i', k_\ell) = \sum_{j=i_{\ell-1}+1}^{i'} (\hat{h}_{1j} - \hat{h}_{2j}) < 0.$$

It follows that $H_2(i_{\ell-1}+1, i', k_\ell^{(r')}) < 0$ since $k_\ell^{(r')} < k_\ell^{(R_\ell)} = k_\ell$ from (c) and $H_2(i_{\ell-1}+1, i_\ell^{(r')}, k_\ell^{(r')}) + H_2(i_\ell^{(r')}+1, i', k_\ell^{(r')}) < 0$. Thus $H_2(i_\ell^{(r')}+1, i', k_\ell^{(r')}) < 0$, which implies that there exists r^* with $r^* > r'$ and $i_\ell^{(r^*)} \leq i'$. This is impossible because if so, then $i' = i_\ell^{(r'+1)}$ and for any $r'+1 < R_\ell$, we have

$$0 = H_2(i_{\ell-1}+1, i_\ell^{(r'+1)}, k_\ell^{(r'+1)}) < H_2(i_{\ell-1}+1, i_\ell^{(r'+1)}, k_\ell) = \sum_{j=i_{\ell-1}+1}^{i'} (\hat{h}_{1j}^A - \hat{h}_{2j}^A) < 0.$$

(e) Suppose there exists ℓ such that $k_{\ell+1} > k_\ell > 0$. Then $k_\ell < k_{\ell+1} \leq n_{2i_{\ell+1}}$. Moreover, $H_2(i_\ell+1, i_{\ell+1}, k_\ell) < H_2(i_\ell+1, i_{\ell+1}, k_{\ell+1}) \leq 0$. It follows that the algorithm must not have stopped at R_ℓ^{th} iteration in step 4 of Algorithm 2, which is a contradiction. ■

Lemma 2. *Based on the same data, the results from Algorithm 2 and Theorem 3 satisfy:*
(a) $\hat{k}^i = k_\ell$ *if* $i = i_{\ell-1}+1, \ldots, i_\ell$, $\ell = 1, \ldots, L$ *and* $\hat{k}^i = 0$ *if* $i > i_L$*;*
(b) $\hat{h}_{gi}^T = \hat{h}_{gi}^A, g = 1, 2, i = 1, \ldots, m_g$.

Proof. (a) If $i_{\ell-1} < i \leq i_\ell$ for $\ell \leq L$, then for any $a \leq i$, there exists $\ell' \leq \ell$ such that $i_{\ell'-1} < a \leq i_{\ell'}$. Then from Lemma 1 (b),

$$H_2(i_{\ell'-1}, i_{\ell'}, k_{\ell'}) \leq 0 = H_2(a, i_{\ell'}, K_2(a, i_{\ell'})).$$

It follows that $H_2(a, i_{\ell'}, k_{\ell'}) \leq H_2(a, i_{\ell'}, K_2(a, i_{\ell'}))$, since $H_2(i_{\ell'-1}, a-1, k_{\ell'}) \geq 0$ from Lemma 1 (d) and so $K_2(a, i_{\ell'}) \geq k_{\ell'} \geq \cdots \geq k_\ell$ from Lemma 1 (e). Thus

$$H_2(a, i_\ell, K_2(a, i_\ell)) = 0 = H_2(a, i_{\ell'}, K_2(a, i_{\ell'})) + \sum_{j=\ell'+1}^{\ell} H_2(i_{j-1}, i_j, k_j)$$

$$\geq H_2(a, i_{\ell'}, k_\ell) + \sum_{j=\ell'+1}^{\ell} H_2(i_{j-1}, i_j, k_\ell) = H_2(a, i_\ell, k_\ell).$$

It follows $K_2(a, i_\ell) \geq k_\ell > 0$ and $\min(K_2^+(a, i_\ell), n_{2i_\ell}) \geq k_\ell$ since $k_\ell \leq n_{2i_\ell}$ from Lemma 1 (b). Therefore,

$$\hat{k}^i = \min_{a \leq i} \max_{b \geq i} \min\{K_2^+(a,b), n_{2b}\} \geq \min_{a \leq i} \min\{K_2^+(a, i_\ell), n_{2i_\ell}\} \geq k_\ell.$$

However, obtaining $\hat{k}_i$ is a minimization problem and its lower bound can be reached when $a = i_{\ell-1}+1$ and $b = i_\ell$. Thus $\hat{k}^i = k_\ell$.

If $i > i_L$, then $K_2(i_L+1, b) \leq 0$ for all $b > i_L$ because $H_2(i_L+1, b, 0) \geq 0$ from step 2 of Algorithm 2. So $K_2^+(i_L+1, b) = 0$ for all $b > i_L$. Hence, $0 \leq \hat{k}^i \leq \max_{b \geq i} \min\{K_2^+(i_L+1, b), n_{2b}\} = 0$, i.e. $\hat{k}^i = 0$.

(b) For population 1, $\hat{h}_{1i}^T = \log[1 - d_{1i}/(n_{1i} + \hat{k}^i)]$, $\hat{h}_{1i}^A = \log[1 - d_{1i}/(n_{1i} + k_\ell)]$ and $\hat{k}^i = k_\ell$ if $i_{\ell-1} < i \leq i_\ell, i = 1, \ldots, L$, so $\hat{h}_{1i}^T = \hat{h}_{1i}^A$ for all $i \leq i_L$. If $i > i_L$, then $\hat{h}_{1i}^T = \log[1 - d_{1i}/n_{1i}] = \hat{h}_{1i}^A, i_L < i \leq m_1$.

For population 2, we use induction.

Let $\hat{h}_{20}^A = \hat{h}_{20}^T = 0$, then the result holds for $j = 0$;

Assume for all $i \le j$, $\hat{h}^A_{2i} = \hat{h}^T_{2i}$;
For the next index $j+1$, there are possible cases (i), (ii), (iii):

(i). $d_{2(j+1)} > 0$. Then

$$\hat{h}^T_{2(j+1)} = \log[1 - d_{2(j+1)}/(n_{2(j+1)} - \hat{k}^i)]$$

$$= \log[1 - d_{2(j+1)}/(n_{2(j+1)} - k_\ell)] = \hat{h}^A_{2(j+1)}, \quad \text{if } i_{\ell-1} < j+1 \le i_\ell, i \le L$$

$$\hat{h}^T_{2(j+1)} = \log[1 - d_{2(j+1)}/n_{2(j+1)}] = \hat{h}^A_{2(j+1)}, \quad \text{if } i_L < j+1 \le m_2.$$

(ii). $d_{2(j+1)} = 0$ and $j+1 \ne i_\ell, \ell = 1, \cdots, L$. Then $\hat{h}^A_{(j+1)} = 0$. So $\sum_{i=1}^{(j+1)} \hat{h}^A_{1i} - \sum_{i=1}^{j} \hat{h}^A_{2i} = \sum_{i=1}^{j+1} (\hat{h}^A_{1i} - \hat{h}^A_{2i}) \ge 0$. It follows that

$$\hat{h}^T_{(j+1)} = \min\{\sum_{i=1}^{j+1} \hat{h}^T_{1i} - \sum_{i=1}^{j} \hat{h}^T_{2i}, 0\} = \min\{\sum_{i=1}^{j+1} \hat{h}^A_{1i} - \sum_{i=1}^{j} \hat{h}^A_{2i}, 0\} = 0 = \hat{h}^A_{(j+1)};$$

(iii). $d_{2(j+1)} = 0$ and $j+1 = i_\ell$. Then $\sum_{i=1}^{j+1} (\hat{h}^A_{1i} - \hat{h}^A_{2i}) = 0$. So

$$\hat{h}^T_{(j+1)} = \min\{\sum_{i=1}^{j+1} \hat{h}^T_{1i} - \sum_{i=1}^{j} \hat{h}^T_{2i}, 0\} = \min\{\sum_{i=1}^{j+1} \hat{h}^A_{1i} - \sum_{i=1}^{j} \hat{h}^A_{2i}, 0\} = \hat{h}^A_{2(j+1)}. \quad \blacksquare$$

Optimization Problem for the Two-sample Case

Consider a general nonlinear optimization problem with inequality constraint

$$\text{minimize } f(\boldsymbol{\theta})$$

$$\text{subject to } g_j(\boldsymbol{\theta}) \le 0, \ j = 1, 2, \ldots, m,$$

for $\boldsymbol{\theta} \in \mathbb{R}^n$. Define the Lagrangian as

$$Lagr(\boldsymbol{\theta}, \boldsymbol{\lambda}) = f(\boldsymbol{\theta}) + \sum_{j=1}^{m} \lambda_j g_j(\boldsymbol{\theta}).$$

For continuously differentiable functions f and g_j, Karush (1939) and Kuhn & Tucker (1951) independently derived the necessary conditions at the solution $\boldsymbol{\theta}^*$. Assume the existence of Lagrange multipliers $\boldsymbol{\lambda}^* \in \mathbb{R}^m$, then at the solution $\boldsymbol{\theta}^*$, the following conditions must be satisfied:

$$\frac{\partial f}{\partial \theta_i}(\boldsymbol{\theta}^*) + \sum_{j=1}^{m} \lambda_j \frac{\partial g_j(\boldsymbol{\theta}^*)}{\partial \theta_i} = 0, i = 1, \ldots, n$$

$$g_i(\boldsymbol{\theta}^*) \le 0, i = 1, \ldots, m$$

$$\lambda_i^* g_i(\boldsymbol{\theta}^*) = 0, i = 1, \ldots, m$$

$$\lambda_i^* \ge 0, i = 1, \ldots, m.$$

These conditions, known as KKT conditions, also constitute sufficient conditions if $f(\boldsymbol{\theta})$ and $g_i(\boldsymbol{\theta})$ are also convex functions. For more details, see Snyman (2005).

In the two-sample problem, we maximize the log likelihood (2) subject to a stochastic ordering constraint. As seen in Theorem 3 and Algorithm 2, we estimate $\hat{h}_{gi}$ for $i \le m_g$, $g = 1, 2$, since for $i > m_g$, there is no data available in population g. Further, if we set $\hat{h}_{1i} = 0$ for all $i > m_1$ and $\hat{h}_{2i} = -\infty$ for all $i > m_2$, the stochastic ordering constraint is automatically satisfied given $\sum_{j=1}^{m'} (\hat{h}_{1j} - \hat{h}_{2j}) \ge 0$ for all $i > m'$. So the log likelihood (2) can be maximized separately for $i \le m'$ and $i > m'$. Like the KM estimator, the log likelihood is maximized by $\hat{h}_{gi} = \log(1 - d_{gi}/n_{gi})$, $m'+1 \le i \le m_g$. So in Lemma 3, we only consider maximizing

$\sum_{g=1}^{2}\sum_{i=1}^{m'}\left\{d_{gi}\log\left(1-e^{h_{gi}}\right)+\left(n_{gi}-d_{gi}\right)h_{gi}\right\}$ under the stochastic ordering constraint.

Lemma 3. *The result* $\{\hat{h}_{gi}, g=1,2, i=1,\dots,m_g\}$ *from Theorem 3 and Algorithm 2 is the solution of maximizing the log likelihood* (2) *under stochastic ordering constraint* $\sum_{j=1}^{i}(h_{2j}-h_{1j})\le 0$, $i=1,\dots,m'$, *and* $h_{1i}, h_{2i}\le 0$, $i=1,\dots,m_g$.

Proof. The optimization problem here is:

$$\text{minimize } -\sum_{g=1}^{2}\sum_{i=1}^{m'}\left\{d_{gi}\log\left(1-e^{h_{gi}}\right)+\left(n_{gi}-d_{gi}\right)h_{gi}\right\} \tag{7}$$

$$\text{subject to } \begin{cases} g_i(\boldsymbol{h}_1,\boldsymbol{h}_2)=\sum_{j=1}^{i}\left(h_{2j}-h_{1j}\right)\le 0 \\ g_{m'+i}(\boldsymbol{h}_1,\boldsymbol{h}_2)=h_{2i}\le 0 \\ g_{2m'+i}(\boldsymbol{h}_1,\boldsymbol{h}_2)=h_{1i}\le 0 \end{cases}$$

and the corresponding Lagrangian is

$$Lagr(\boldsymbol{h}_1,\boldsymbol{h}_2,\boldsymbol{\lambda})=-\sum_{g=1}^{2}\sum_{i=1}^{m'}\left\{d_{gi}\log\left(1-e^{h_{gi}}\right)+\left(n_{gi}-d_{gi}\right)h_{gi}\right\}+\sum_{j=1}^{3m'}\lambda_j g_j(\boldsymbol{h}_1,\boldsymbol{h}_2).$$

Thus the KKT conditions are:

$$\frac{d_{1i}e^{\hat{h}_{1i}}}{1-e^{\hat{h}_{1i}}}-(n_{1i}-d_{1i})-\sum_{j=i}^{m'}\hat{\lambda}_j+\hat{\lambda}_{i+2m'}=0 \tag{8a}$$

$$\frac{d_{2i}e^{\hat{h}_{2i}}}{1-e^{\hat{h}_{2i}}}-(n_{2i}-d_{2i})+\sum_{j=i}^{m'}\hat{\lambda}_j+\hat{\lambda}_{i+m'}=0 \tag{8b}$$

$$\sum_{j=1}^{i}(\hat{h}_{2j}-\hat{h}_{1j})\le 0 \tag{8c}$$

$$\hat{\lambda}_i\sum_{j=1}^{i}(\hat{h}_{2j}-\hat{h}_{1j})=0 \tag{8d}$$

$$\hat{\lambda}_i,\lambda_{i+m'},\lambda_{i+2m'}\ge 0 \tag{8e}$$

$$\hat{h}_{1i}\le 0, h_{2i}\le 0 \tag{8f}$$

$$\hat{\lambda}_{i+m}\hat{h}_{2i}=0 \tag{8g}$$

$$\hat{\lambda}_{i+2m}\hat{h}_{1i}=0 \tag{8h}$$

We define $\hat{\lambda}_i$, $\hat{\lambda}_{i+m'}$ and $\hat{\lambda}_{i+2m'}$, $i=1,\cdots,m'$ as follows :

$$\hat{\lambda}_i=\begin{cases} k_L & \text{if } i=i_L \\ k_\ell-k_{\ell+1} & \text{if } i=i_\ell,\ \ell=1,\dots,L-1 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$\hat{\lambda}_{i+m'}=\begin{cases} 0 & \text{if } d_{2i}>0, \\ n_{2i}-k_\ell & \text{if } d_{2i}=0 \text{ and } i_{\ell-1}<i\le i_\ell,\ \ell=1,\dots,L \\ n_{2i} & \text{if } d_{2i}=0 \text{ and } i>i_L \end{cases} \tag{10}$$

$$\hat{\lambda}_{i+2m'} = \begin{cases} 0 & \text{if } d_{1i} > 0 \\ n_{1i} + \sum_{j=i}^{m'} \hat{\lambda}_j & \text{if } d_{1i} = 0 \end{cases} \tag{11}$$

Conditions (8c) and (8f) are satisfied by Algorithm 2. Condition (8e) is also satisfied since $k_1 \geq \cdots \geq k_L > 0$ (Lemma 1 (e)) and $k_\ell \leq n_{2i_\ell} \leq n_{2i}$ (Lemma 1 (b)) if $i_{\ell-1} < i \leq i_\ell$. Algorithm 2 always gives $\hat{h}_{1i} = 0$ if $d_{1i} = 0$ and $\hat{\lambda}_{i+2m'} = 0$ if $d_{1i} > 0$ from (11), so condition (8h) is satisfied. If $d_{2i} = 0$, $\hat{h}_{2i} < 0$ only when $i = i_\ell$ for some ℓ and $k_\ell = n_{2i_\ell}$. So in this case, $\hat{\lambda}_{i+m'} = 0$ from (10), which can lead to (8g). From (9), $\hat{\lambda}_i \neq 0$ when $i = i_\ell$. However, $\sum_{j=1}^{i_\ell}(\hat{h}_{2j} - \hat{h}_{1j}) = 0$, so condition (8d) is also satisfied. From (9), we also know that $\sum_{j=i}^{m'} \hat{\lambda}_j = k_\ell$ and $\sum_{j=i}^{m'} \hat{\lambda}_j = 0$ if $i \geq i_L$. If $d_{1i} = 0$, then the condition (8a) is $-n_{1i} - \sum_{j=i}^{m'} \hat{\lambda}_j + \hat{\lambda}_{i+2m'} = 0$, which is satisfied with definition of $\hat{\lambda}_{i+2m'}$ in (11). If $d_{1i} > 0$, then $\hat{\lambda}_{i+2m'} = 0$ from (11), so the condition (8a) is also satisfied. Similarly, the condition (8b) is satisfied.

All KKT conditions are satisfied at the solution from Algorithm 2, and (7) reaches the global minimum since the optimization function and all constraints are convex. ■

Lemma 4. *From Theorem 3 and Algorithm 2, $\{\hat{h}_{1i}, i = 1, \ldots, m_1\}$ is the unique C-NPMLE of h_{1i} and $\{\hat{h}_{2i}, i = 1, \ldots, m_2\}$ is the unique MC-NPMLE of h_{2i} under the stochastic ordering constraint.*

Proof. In this proof, we first remove some unnecessary stochastic ordering constraints; then we show that $\hat{k}^i = \sum_{j=i}^{m'} \hat{\lambda}_j$ are unique; last we discuss the uniqueness of the C-NPMLE of h_{1i} and the MC-NPMLE of h_{2i}.

For any C-NPMLE, $\hat{h}_{1i} = 0$ if $d_{1i} = 0$, because $\hat{\lambda}_{i+m'} = n_{1i} + \sum_{j=i}^{m'} \hat{\lambda}_j > 0$ if $i \leq m_1$. So $\sum_{j=1}^{i}(\hat{h}_{2j} - \hat{h}_{1j}) \leq 0$ for $d_{1i} = 0$ will be automatically satisfied given $\sum_{j=1}^{i}(\hat{h}_{2j} - \hat{h}_{1j}) \leq 0$ for $d_{1i} > 0$ and $\hat{h}_{1i}, \hat{h}_{2i} \leq 0$.

Thus for $d_{1i} = 0$, the condition (8d) is not necessary, or we can simply set $\hat{\lambda}_i = 0$. Based on this setting of $\hat{\lambda}_i$, we show that $\hat{k}^i = \sum_{j=i}^{m'} \hat{\lambda}_j$ is unique.

Suppose $\hat{k}_i$ is not unique, then we can find two sets of $\{\hat{k}^i\}$ and $\{\tilde{k}^i\}$ from $\{\hat{\lambda}_i\}$ and $\{\tilde{\lambda}_i\}$ that satisfy the KKT conditions with corresponding solutions $\{\hat{h}_{gi}\}$ and $\{\tilde{h}_{gi}\}$. Let $i^* = \min(i : \hat{k}^i \neq \tilde{k}^i)$. Without loss of generality, assume $\hat{k}^{i^*} > \tilde{k}^{i^*} \geq 0$. Then

$$\sum_{j=1}^{i^*-1}(\tilde{h}_{2j} - \tilde{h}_{1j}) = 0. \tag{12}$$

Because $\tilde{\lambda}_{i^*-1} = \tilde{k}^{(i^*-1)} - \tilde{k}^{i^*} = \hat{k}^{(i^*-1)} - \tilde{k}^{i^*} > \hat{k}^{(i^*-1)} - \hat{k}^{i^*} = \hat{\lambda}_{i^*-1} \geq 0$.

Let $\hat{i} = \min\{i \geq i^* : \sum_{j=1}^{i}(\hat{h}_2 - \hat{h}_1) = 0, d_{1i} > 0\}$. Then $\hat{\lambda}_j = 0$ for $j = i^*, \ldots, \hat{i}$ from condition (8d). So

$$\hat{k}^{i^*} = \cdots = \hat{k}^{\hat{i}} > \tilde{k}^{i^*} \geq \cdots \geq \tilde{k}^{\hat{i}} \tag{13}$$

and $\tilde{\lambda}_{i+m'} = n_{2i} - \tilde{k}^i > n_{2\hat{i}} - \hat{k}^{\hat{i}} \geq 0$ from condition (8b) if $d_{2i} = 0, i = i^*, \ldots, \hat{i}$. So

$$\tilde{h}_{2i} = 0 \qquad \text{if } d_{2i} = 0, i = i^*, \ldots, \hat{i} \tag{14}$$

from condition (8g). Therefore,

$$\sum_{j=1}^{\hat{i}}(\tilde{h}_{2j} - \tilde{h}_{1j}) = \sum_{j=1}^{i^*-1}(\tilde{h}_{2j} - \tilde{h}_{1j}) + \sum_{j=i^*}^{\hat{i}}(\tilde{h}_{2j} - \tilde{h}_{1j}) = \sum_{j=i^*}^{\hat{i}}(\tilde{h}_{2j} - \tilde{h}_{1j}) \qquad \text{(from (12))}$$

$$= \sum_{j=i^*}^{\hat{i}} \{\log(1 - d_{2j}/(n_{2j} - \tilde{k}^j) - \log(1 - d_{1j}/(n_{1j} + \tilde{k}^j)\} \qquad \text{(from (14))}$$

$$> \sum_{j=i^*}^{\hat{i}} \{\log(1 - d_{2j}/(n_{2j} - \hat{k}^j) - \log(1 - d_{1j}/(n_{1j} + \hat{k}^j)\} \qquad \text{(from (13))}$$

$$\geq \sum_{j=i^*}^{\hat{i}} (\hat{h}_{2j} - \hat{h}_{1j}) \geq \sum_{j=1}^{i^*-1} (\hat{h}_{2j} - \hat{h}_{1j}) + \sum_{j=i^*}^{\hat{i}} (\hat{h}_{2j} - \hat{h}_{1j}) = 0.$$

which contradicts condition (8c). Thus $\hat{k}^i$ is unique and $\hat{h}_{1i}$ is unique, because $\hat{h}_{1i} = \log\{1 - d_{1i}/(n_{1i} + \hat{k}^i)\}$ if $d_{1i} > 0$ and $\hat{h}_{1i} = 0$ if $d_{1i} = 0$.

Also $\hat{h}_{2i} = \log\{1 - d_{2i}/(n_{2i} - \hat{k}^i)\}$ if $d_{2i} > 0$, and $\hat{h}_{2i} = 0$ if $d_{2i} = 0$ and $\hat{k}^i < n_{2i}$ (because $\hat{\lambda}_{i+m'} = n_{2i} - \hat{k}^i > 0$). Therefore, all C-NPMLEs may only differ from each other in population 2 when $d_{2i} = 0$ and $\hat{k}^i = n_{2i}$.

If we sequentially set $\hat{h}_{2i} = \min(\sum_{j=1}^{i} \hat{h}_{1j} - \sum_{j=1}^{i-1} \hat{h}_{2j}, 0)$ if $d_{2i} = 0$ as in Theorem 3, then $\sum_{j=1}^{i} \hat{h}_{2j}$ is maximized. Because if $\sum_{j=1}^{i-1} \hat{h}_{2j}$ is maximized, the maximum possible value of $\sum_{j=1}^{i} \hat{h}_{2j}$ is $\min(\sum_{j=1}^{i-1} \hat{h}_{2j}, \sum_{j=1}^{i} \hat{h}_{1j})$, which can be obtained by setting $\hat{h}_{2i} = \min(\sum_{j=1}^{i} \hat{h}_{1j} - \sum_{j=1}^{i-1} \hat{h}_{2j}, 0)$ if $d_{2i} = 0$. Since $\hat{h}_{2i}$ sequentially takes a unique value, the MC-NPMLE is also unique. ■

The results of Theorem 1 (Theorem 2) can be directly obtained from Theorem 3 by treating the sample size of stochastically smaller (larger) group goes to infinity and so the proofs are not presented here. ■

BIBLIOGRAPHY

R. L. Dykstra (1982), Maximum likelihood estimations of the survival functions of stochastically ordered random variables. *Journal of the American Statistical Association*, 77,621–628.

W. Karush (1939), *Minima of Functions of Several Variables with Inequalities as Side Constraints*, M.Sc. dissertation, University of Chicago, Chicago, Illinois, Department of Mathematics.

H. W. Kuhn & A. W. Tucker (1951), Nonlinear programming, in J. Neyman, editor, *Proceedings of 2nd Berkeley Symposium*, Berkeley: University of California Press, 481–492.

J. A. Snyman (2005), *Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms*, Springer, New York.