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 \( m' = \max\{i: n_{i1} > 0\} \).
2. If there exists \( b > i_{\ell-1} \) such that \( K(i_{\ell-1} + 1, b) > 0 \), then let \( i_\ell = \min \{\arg \max_{b > i_{\ell-1}} K(i_{\ell-1} + 1, b)\} \) and \( k_\ell = K(i_{\ell-1} + 1, i_\ell) \). Otherwise, let \( i_\ell = m', k_\ell = 0 \).
3. Let \( \hat{h}_{ij} = \log\{1 - d_{ij}/(n_{ij} + k_i)\}, i_{\ell-1} + 1 \leq j \leq i_\ell \).
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 \( k_1 \geq k_2 \geq \cdots \), 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 \( a_{k_1}, k_\ell - k_{\ell-1}, \ell = 1, 2, \ldots \) 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 \( m' = \max\{i: n_{i1} > 0\} \).
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 \( i_\ell \) exists, set \( i_\ell = m' \) and \( k_\ell = 0 \) and go to step 4.
3. Let \( I = \min_{b > i_\ell} \{b: H(i_{\ell-1} + 1, b, k_\ell) < 0\} \). If no such \( I \) exists, then go to step 4. Otherwise, set \( i_\ell = I \) and repeat step 3.
4. Let \( \hat{h}_{ij} = \log\{1 - d_{ij}/(n_{ij} + k_i)\}, i_{\ell-1} + 1 \leq j \leq i_\ell \).
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 \( \min \{\arg \max_{b > i_{\ell-1}} K(i_{\ell-1} + 1, b)\} \) 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 \( \hat{S}_1(t) \) is a C-NPMLE. Let \( \hat{S}_i(t) \) be a C-NPMLE subject to \( T_1 \leq_{st} T_2 \). Note that \( \hat{S}_1(t) \) is the maximum likelihood estimator subject to fewer con-
strains (only at times $X_i'$) compared to $\tilde{S}_1(t)$, we have that $L(\tilde{S}_1(t)) \geq L(\tilde{S}_1(t))$. Further $L(\tilde{S}_1(t)) \geq L(\tilde{S}_1(t))$ since $\tilde{S}_1(t) = \min(S_1(t), S_2(t)) \leq S_2(t)$. Note that at every time $X_i'$, $i = 1, \ldots, n_{tot}$, $\tilde{S}_1(X_i') \leq S_2(X_i')$, the difference between $\tilde{S}_1(t)$ 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 $L(\tilde{S}_1(t)) = L(\tilde{S}_1(t)) = L(\tilde{S}_1(t))$, which implies that $\tilde{S}_1(t)$ is a C-NPMLE subject to $T_1 \leq t \leq T_2$.

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

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

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

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

$$S'_1(x) \leq \tilde{S}_1(x) \leq S'_1(x) \frac{\tilde{S}_1(a_{i_L})}{S'_1(a_{i_L})} \leq S'_1(x) \left[ \sup_{t \leq a_{i_L}} S'_2(t) \right] S'_1(x) \leq \left[ \sup_{t \leq \tau} S'_1(t) \right] S'_1(x)$$

Similarly,

$$S'_2(x) \geq \tilde{S}_2(x) \geq \left[ \inf_{t \leq \tau} S'_2(t) \right] S'_2(x).$$

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

$$S'_1(x) \leq \tilde{S}_1(x) = \tilde{S}_1(a_{i_L}) \times S'_1(x) \frac{\tilde{S}_1(a_{i_L})}{S'_1(a_{i_L})} \leq \left[ \sup_{t \leq \tau} S'_2(t) \right] S'_1(x) \leq \left[ \sup_{t \leq \tau} S'_1(t) \right] S'_1(x).$$

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

$$S'_1(x) \leq \tilde{S}_1(x) \geq \left[ \sup_{t \leq \tau} S'_1(t) \right] S'_1(x) \text{ and } S'_2(x) \geq \tilde{S}_2(x) \geq \left[ \inf_{t \leq \tau} S'_2(t) \right] S'_2(x).$$

As $n_1, n_2$ go to $\infty$, from (4), for any $x \leq \tau$, $S'_1(x) \rightarrow S_1(x)$ and $S'_2(x) \rightarrow S_2(x)$. So

$$\frac{S'_2(x)}{S'_1(x)} \rightarrow \frac{S_2(x)}{S_1(x)} \Rightarrow \frac{S'_2(x)}{S'_1(x)} \rightarrow \frac{S_2(x)}{S_1(x)} \rightarrow 1,$$

in probability if indeed $S_1(t) \geq S_2(t)$ for all $t$. Thus $P\{\sup_{t \leq \tau} |S_1(x) - S'_1(x)| > \epsilon\} \rightarrow 0$ for every $\epsilon > 0$. Then using Meier’s result (4), we obtain the desired result for $\tilde{S}_1(t)$.
Similarly, we can show that \( P\{\sup_{x \leq \tau} |\tilde{S}_2(x) - S_2^\ast(x)| > \epsilon \} \to 0 \) for every \( \epsilon > 0 \) and hence we obtain the desired result for \( S_2(t) \).

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 \( \hat{h}_{j1} \) is the C-NPMLE and \( \hat{h}_{2j} \) 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 \( k_\ell \). Suppose the algorithm has \( R_\ell \) iterations with initial step 2 to 3 (\( 0^{th} \) iteration) and step 4 to 3 (\( 1^{st} \), \( 2^{nd} \), \( 3^{rd} \) iteration) before it finds \( i_\ell \). 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). \( \sum_{j=1}^{i_{\ell}} \hat{h}_{1j}^A \geq \sum_{j=i_{\ell-1}+1}^{i_{\ell}} \hat{h}_{2j}^A, i = 1, \ldots, m'; \)

(e). \( k_1 \geq k_2 \geq \cdots \geq k_L > 0. \)

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

(b) From step 3 in Algorithm 2, \( k_\ell \) 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 \( k_\ell^{(r')} < k_\ell^{(r'-1)} \). It follows that

\[
H_2(i_{\ell-1} + 1, i_\ell^{(r')} , k_\ell^{(r'-1)}) \leq H_2(i_{\ell-1} + 1, i_\ell^{(r')} , k_\ell^{(r'-1)})
\]

\[
= H_2(i_{\ell-1} + 1, i_\ell^{(r'-1)} , k_\ell^{(r'-1)}) + H_2(i_\ell^{(r'-1)} + 1, i_\ell^{(r')} , k_\ell^{(r'-1)})
\]

\[
> H_2(i_\ell^{(r'-1)} + 1, i_\ell^{(r')}, k_\ell^{(r'-1)}) < 0 \quad \text{(step 2 or 4 in Algorithm 2).}
\]

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

\[
K_2(i_{\ell-1}, i_\ell^{(r')}). \quad \text{Then } H_2(i_{\ell-1} + 1, i_\ell^{(r')} , k_\ell^{(r')} ) = 0, \text{which contradicts (6); or}
\]

\[
n_{2i_\ell^{(r')}}. \text{Then } k_\ell^{(r'-1)} \geq k_\ell^{(r')} = n_{2i_\ell^{(r')}}, \text{which contradicts the condition } n_{2i_\ell^{(r')}} > k_\ell^{(r'-1)} \text{ that is required to reach iteration } r' \text{ in step 4 of Algorithm 2.}
\]

(d) Suppose there exists \( i' \) such that \( \sum_{j=i_{\ell-1}+1}^{i'} \hat{h}_{1j}^A < \sum_{j=i_{\ell-1}+1}^{i'} \hat{h}_{2j}^A \), 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:

\[
i' = i_\ell, \ell \leq L. \text{Then } \sum_{j=i_{\ell-1}+1}^{i'} \hat{h}_{1j}^A = \sum_{j=i_{\ell-1}+1}^{i'} \hat{h}_{2j}^A, \text{which contradicts Lemma 1 (a); or}
\]

\[
i' > i_\ell. \text{Then } H_2(i_\ell + 1, i', 0) = \sum_{j=i_\ell+1}^{i_{\ell-1}+1}(\hat{h}_{1j} - \hat{h}_{2j}) < 0, \text{which contradicts the condition}
\]

**DOI:** The Canadian Journal of Statistics / La revue canadienne de statistique
Lemma 1 (d) and so it follows that $H_2(i_{t-1} + 1, i', k_t) = \sum_{j=i_{t-1}+1}^{i'} (\hat{h}_{ij} - \hat{h}_{ij}) < 0$.

It follows that $H_2(i_{t-1} + 1, i', k^{(r')}_{t}) < 0$ since $k^{(r')}_{t} < k^{(R_t)}_{t} = k_t$ from (c) and $H_2(i_{t-1} + 1, i^{(r')}_{t}, k^{(r')}_{t}) + H_2(i^{(r')}_{t} + 1, i', k^{(r')}_{t}) < 0$. Thus $H_2(i^{(r')}_{t} + 1, i', k^{(r')}_{t}) < 0$, which implies that there exists $r^*$ with $r^* > r'$ and $i^{(r')}_{t} \leq i'$. This is impossible because if so, then $i' = i^{(r'+1)}_{t}$ and for any $r' < R_t$, we have

$$0 = H_2(i_{t-1} + 1, i^{(r'+1)}_{t}, k^{(r'+1)}_{t}) < H_2(i_{t-1} + 1, i^{(r'+1)}_{t}, k_t) = \sum_{j=i_{t-1}+1}^{i'} (\hat{h}^A_{ij} - \hat{h}^A_{ij}) < 0.$$  

(e) Suppose there exists $\ell$ such that $k_{\ell+1} > k_{\ell} > 0$. Then $k_{t} < k_{t+1} \leq n_{2t+1}$. Moreover, $H_2(i_{t-1} + 1, i_{t-1} + 1, k_{t}) \leq H_2(i_{t-1} + 1, i_{t-1} + 1, k_{t+1}) \leq 0$. It follows that the algorithm must not have stopped at $R_t$, 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) $\tilde{k}^t = k_t$ if $i = i_{t-1} + 1, \ldots, i_t$, $\ell = 1, \ldots, L$ and $\tilde{k}^t = 0$ if $i > i_t$;

(b) $\tilde{h}^*_{ij} = \hat{h}^A_{ij}, g = 1, 2, i = 1, \ldots, m_g$.

Proof. (a) If $i_{t-1} < i \leq i_t$ 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),

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

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

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

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

$$\tilde{k}^t = \min_{a \leq i \leq b} \min\{K_2^+(a, b), n_{2b}\} \geq \min_{a \leq i \leq b} \min\{K_2^+(a, i), n_{2i}\} \geq k_t.$$  

However, obtaining $\tilde{k}^t$ is a minimization problem and its lower bound can be reached when $a = i_{t-1} + 1$ and $b = i_t$. Thus $\tilde{k}^t = k_t$.

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

(b) For population 1, $\tilde{h}^*_{ij} = \log[1 - d_{ij}/(n_{ij} + \tilde{k}^t)]$, $\tilde{h}^A_{ij} = \log[1 - d_{ij}/(n_{ij} + k_t)]$ and $\tilde{k}^t = k_t$ if $i_{t-1} < i \leq i_t, i = 1, \ldots, L$, so $\tilde{h}^*_{ij} = \tilde{h}^A_{ij}$ for all $i \leq i_t$. If $i > i_t$, then $\tilde{h}^A_{ij} = \log[1 - d_{ij}/n_{ij}] = \tilde{h}^A_{ij}, i_t < i \leq m_1$.

For population 2, we use induction.

Let $\tilde{h}^A_{ij} = \tilde{h}^A_{ij} = 0$, then the result holds for $j = 0;$
Assume for all \( i \leq j \), \( \hat{h}_2^A = \hat{h}_2^F \).

For the next index \( j + 1 \), there are possible cases (i), (ii), (iii):

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

\[
\hat{h}_{2(j+1)}^T = \log[1 - d_{2(j+1)}/(n_{2(j+1)} - \hat{k}^j)] = \log[1 - d_{2(j+1)}/(n_{2(j+1)} - k_i)] = \hat{h}_{2(j+1)}^A, \quad \text{if} \; i - \ell < j + 1 \leq i, \; i \leq L
\]

\[
\hat{h}_{2(j+1)}^T = \log[1 - d_{2(j+1)}/n_{2(j+1)}] = \hat{h}_{2(j+1)}^A, \quad \text{if} \; i_L < j + 1 \leq m_2.
\]

(ii). \( d_{2(j+1)} = 0 \) and \( j + 1 \neq i, \ell = 1, \ldots, L \). Then \( \hat{h}_{(j+1)}^A = 0 \). So

\[
\sum_{i=1}^{j+1} \hat{h}_{1i}^A - \sum_{i=1}^{j} \hat{h}_{2i}^A = \sum_{i=1}^{j+1} (\hat{h}_{1i}^A - \hat{h}_{2i}^A) \geq 0. \text{ It follows that}
\]

\[
\hat{h}_{(j+1)}^T = \min \{ \sum_{i=1}^{j+1} \hat{h}_{1i}^T - \sum_{i=1}^{j} \hat{h}_{2i}^T, 0 \} = \min \{ \sum_{i=1}^{j+1} \hat{h}_{1i}^A - \sum_{i=1}^{j} \hat{h}_{2i}^A, 0 \} = 0 = \hat{h}_{(j+1)}^A;
\]

(iii). \( d_{2(j+1)} = 0 \) and \( j + 1 = i \). Then

\[
\sum_{i=1}^{j+1} (\hat{h}_{1i}^A - \hat{h}_{2i}^A) = 0. \text{ So}
\]

\[
\hat{h}_{(j+1)}^T = \min \{ \sum_{i=1}^{j+1} \hat{h}_{1i}^T - \sum_{i=1}^{j} \hat{h}_{2i}^T, 0 \} = \min \{ \sum_{i=1}^{j+1} \hat{h}_{1i}^A - \sum_{i=1}^{j} \hat{h}_{2i}^A, 0 \} = \hat{h}_{2(j+1)}^A. \]

**Optimization Problem for the Two-sample Case**

Consider a general nonlinear optimization problem with inequality constraint

\[
\text{minimize } f(\theta) \text{ subject to } g_j(\theta) \leq 0, \; j = 1, 2, \ldots, m,
\]

for \( \theta \in \mathbb{R}^n \). Define the Lagrangian as

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

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

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

\[
g_i(\theta^*) \leq 0, \; i = 1, \ldots, m
\]

\[
\lambda^*_i g_i(\theta^*) = 0, \; i = 1, \ldots, m
\]

\[
\lambda^*_i \geq 0, \; i = 1, \ldots, m.
\]

These conditions, known as KKT conditions, also constitute sufficient conditions if \( f(\theta) \) and \( g_i(\theta) \) 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 \( i \leq m_g, \; g = 1, 2 \), 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 \( \sum_{j=1}^{m'} (\hat{h}_{1ij} - \hat{h}_{2ij}) \geq 0 \) for all \( i > m' \). So the log likelihood (2) can be maximized separately for \( i \leq m' \) and \( i > m' \). Like the KM estimator, the log likelihood is maximized by \( \hat{h}_{gi} = \log(1 - d_{gi}/n_{gi}), \; m' + 1 \leq i \leq m_g \). So in Lemma 3, we only consider maximizing
Thus the KKT conditions are:

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

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

**Proof.** The optimization problem here is:

\[
\text{minimize} \quad \sum_{g=1}^{2m} \sum_{i=1}^{m'} \left\{ d_{gi} \log (1 - e^{h_{gi}}) + (n_{gi} - d_{gi}) h_{gi} \right\} \\
\text{subject to} \quad \begin{align*}
&g_i(h_1, h_2) = \sum_{j=1}^{i} (h_{2j} - h_{1j}) \leq 0 \\
&g_{m'+i}(h_1, h_2) = h_{2i} \leq 0 \\
&g_{2m'+i}(h_1, h_2) = h_{1i} \leq 0
\end{align*}
\]

and the corresponding Lagrangian is

\[
\text{Lagr}(h_1, h_2, \lambda) = \sum_{g=1}^{2m} \sum_{i=1}^{m'} \left\{ d_{gi} \log (1 - e^{h_{gi}}) + (n_{gi} - d_{gi}) h_{gi} \right\} + \sum_{j=1}^{3m'} \lambda_j g_j(h_1, h_2).
\]

Thus the KKT conditions are:

### (8a)

\[
\frac{d_{1i} e^{h_{1i}}}{1 - e^{h_{1i}}} - (n_{1i} - d_{1i}) - \sum_{j=1}^{m'} \hat{\lambda}_j + \hat{\lambda}_{i+2m'} = 0
\]

### (8b)

\[
\frac{d_{2i} e^{h_{2i}}}{1 - e^{h_{2i}}} - (n_{2i} - d_{2i}) + \sum_{j=1}^{m'} \hat{\lambda}_j + \hat{\lambda}_{i+2m'} = 0
\]

### (8c)

\[
\sum_{j=1}^{i} (h_{2j} - h_{1j}) \leq 0
\]

### (8d)

\[
\hat{\lambda}_i \sum_{j=1}^{i} (h_{2j} - h_{1j}) = 0
\]

### (8e)

\[
\hat{\lambda}_i, \lambda_{i+m' + \hat{\lambda}_{i+2m'} \geq 0}
\]

### (8f)

\[
\hat{h}_{1i} \leq 0, h_{2i} \leq 0
\]

### (8g)

\[
\hat{\lambda}_{i+m} h_{2i} = 0
\]

### (8h)

\[
\hat{\lambda}_{i+2m} h_{1i} = 0
\]

We define \( \hat{\lambda}_i, \hat{\lambda}_{i+m}, \) and \( \hat{\lambda}_{i+2m'} \) as follows:

\[
\hat{\lambda}_i = \begin{cases} 
  k_L & \text{if } i = i_L \\
  k_\ell - k_{\ell+1} & \text{if } i = i_\ell, \ \ell = 1, \ldots, L - 1 \\
  0 & \text{otherwise}
\end{cases} \quad (9)
\]

\[
\hat{\lambda}_{i+m'} = \begin{cases} 
  n_{2i} - k_\ell & \text{if } d_{2i} > 0, d_{2i} = 0 \text{ and } i_\ell < i \leq i_\ell, \ \ell = 1, \ldots, L \\
  n_{2i} & \text{otherwise}
\end{cases} \quad (10)
\]
\[
\hat{\lambda}_{1+2m'} = \begin{cases} 
0 & \text{if } d_{11} > 0 \\
 n_{11} + \sum_{j=1}^{m'} \hat{\lambda}_j & \text{if } d_{11} = 0
\end{cases}
\] (11)

Conditions (8c) and (8f) are satisfied by Algorithm 2. Condition (8c) is also satisfied since \(k_1 \geq \cdots \geq k_L > 0\) (Lemma 1 (e)) and \(k_L \leq n_{2i} \leq n_{2i}\) (Lemma 1 (b)). Algorithm 2 always gives \(\hat{h}_{11} = 0\) if \(d_{11} = 0\) and \(\hat{\lambda}_{1+2m'} = 0\) if \(d_{11} > 0\) from (11), so condition (8h) is satisfied. If \(d_{2j} = 0\), \(\hat{h}_{2j} < 0\) only when \(i = i_\ell\) for some \(\ell\) and \(k_L = n_{2i}\). So in this case, \(\hat{\lambda}_{1+2m'} = 0\) from (10), which can lead to (8g). From (9), \(\hat{\lambda}_i \neq 0\) when \(i = i_\ell\).

However, \(\sum_{j=1}^{m'} (\hat{h}_{2j} - \hat{h}_{1j}) = 0\), so condition (8d) is also satisfied. From (9), we also know that \(\sum_{j=1}^{m'} \hat{\lambda}_j = k_L\) and \(\sum_{j=1}^{m'} \hat{\lambda}_j = 0\) if \(i = i_L\). If \(d_{11} = 0\), then the condition (8a) is \(-n_{11} - \sum_{j=1}^{m'} \hat{\lambda}_j = 0\), which is satisfied with definition of \(\hat{\lambda}_{1+2m'}\) in (11). If \(d_{11} > 0\), then \(\hat{\lambda}_{1+2m'} = 0\) 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-NPML of \(h_{1i}\) and \(\hat{h}_{2i}, i = 1, \ldots, m_2\) is the unique MC-NPML of \(h_{2i}\) under the stochastic ordering constraint.

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

For any C-NPML, \(\hat{h}_{1i} = 0\) if \(d_{1i} = 0\), because \(\hat{\lambda}_{i+m'} = n_{i1} + \sum_{j=1}^{m'} \hat{\lambda}_j > 0\) 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 \(d_{1j} > 0\) and \(h_{1i}, h_{2i} \leq 0\).

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 \(\hat{k} = \sum_{j=1}^{m'} \hat{\lambda}_j\) is unique.

Suppose \(\hat{k}_i\) is not unique, then we can find two sets of \(\{\hat{k}^i\}\) and \(\{\hat{k}^i\}\) from \(\{\hat{\lambda}_i\}\) and \(\{\hat{\lambda}_i\}\) that satisfy the KKT conditions with corresponding solutions \(\{\hat{h}_{1j}\}\) and \(\{\hat{h}_{2j}\}\). Let \(i^* = \min(i : \hat{k}^i \neq \hat{k}^i)\). Without loss of generality, assume \(\hat{k}^i > \hat{k}^{i^*} \geq 0\). Then

\[
\sum_{j=1}^{i^*-1} (\hat{h}_{2j} - \hat{h}_{1j}) = 0. 
\] (12)

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

Let \(\hat{i} = \min\{i \geq i^* : \sum_{j=1}^{i} (\hat{h}_{2j} - \hat{h}_{1j}) = 0, d_{1i} > 0\}\). Then \(\hat{\lambda}_j = 0\) for \(j = i^*, \ldots, \hat{i}\) from condition (8d). So

\[
\hat{k}^{i^*} = \cdots = \hat{k}^{\hat{i}} > \hat{k}^{i} \geq \cdots \geq \hat{k}^{i} 
\] (13)

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

\[
\hat{h}_{2i} = 0 \quad \text{if } d_{2i} = 0, i = i^*, \ldots, \hat{i}
\] (14)

from condition (8g). Therefore,

\[
\sum_{j=1}^{i} (\hat{h}_{2j} - \hat{h}_{1j}) = \sum_{j=1}^{i^*-1} (\hat{h}_{2j} - \hat{h}_{1j}) + \sum_{j=i^*}^{\hat{i}} (\hat{h}_{2j} - \hat{h}_{1j}) = \sum_{j=i^*}^{\hat{i}} (\hat{h}_{2j} - \hat{h}_{1j}) \quad \text{(from (12))}
\]
\[
\sum_{j=1}^{i} \{ \log(1 - d_{2j}/(n_{2j} - \bar{k}^i)) - \log(1 - d_{1j}/(n_{1j} + \bar{k}^i)) \} \quad \text{(from (14))}
\]

\[
\sum_{j=1}^{i} \{ \log(1 - d_{2j}/(n_{2j} - \bar{k}^i)) - \log(1 - d_{1j}/(n_{1j} + \bar{k}^i)) \} \quad \text{(from (13))}
\]

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

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

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

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 \(\min(\sum_{j=1}^{i} \hat{h}_{2j}, \sum_{j=1}^{i} \hat{h}_{1j})\), 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**


