

## An Algorithm of Nonparametric Maximum Likelihood Estimation for Bivariate Censored Data

**Mohamad Fatekurohman**

*Department of Mathematics, Jember University*

*Ph.D. student, Department of Mathematics, Gadjah Mada University*

*m\_fatkur@yahoo.com*

**Subanar, Danardono**

*Department of Mathematics, Gadjah Mada University*

*{subanar, danardono}@ugm.ac.id*

### Abstract

Censored data, univariate or bivariate and left/right censored data, and interval censored data can be represented by its intersection graph (Gentleman and Vandall, 2002). Studies on the censored data, especially bivariate censored data, have been discussed by several experts. Maathuis (2003) discussed the nature of the maximum likelihood estimator (MLE) for bivariate censored the data computational and algorithmic aspects while estimates of the bivariate distribution function for the censored data have been discussed by Campbell and Foldes (1982), Burke (1988), Prentice and Cai (1992), Pruitt (1993) and Prentice (1999). Some researchers argue that nonparametric maximum likelihood estimator (NPMLE) is difficult to quantify and is not unique so new methods are needed.

The case of bivariate interval censored data consists of two parts. The first involves the determination of the regions of possible support and the second is the maximization of the likelihood, i.e. the rectangles with non-zero mass are calculated. In this paper we use rooted tree for finding the estimator. As the results, the rooted tree level has at most  $n$  then there are  $n$  iteration, so that the total complexity time is  $O(n^3)$ .

*Keywords: bivariate censored data, intersection graph, NPLME, rooted tree.*

### 1. Introduction

Three algorithms that can be applied to determine the NPMLE of  $S$ ,  $F$  or  $p$ . The first is the self-consistency algorithm that was developed by Turnbull (1976) and can be regarded as an application of the EM algorithm (Dempster et al., 1977). The second algorithm is the ICM algorithm, first introduced by Groeneboom and Wellner (1992) and later modified by Jongbloed (1998). The third algorithm is a hybrid algorithm proposed by Wellner and Zhan (1997), which is referred to as the EM-ICM algorithm in the following. It basically combines the self-consistency algorithm and the ICM algorithm. Pada paper ini akan disajikan The computation is simplified through the use of a so-called reduction tree, which limits the need to detect maximal intersections to leaf-sets only (these can be thought of as "local" data sets).

### 2. Optimal Condition MLE

Suppose  $R_i$  is an observation in the form of a rectangle or line segment on  $(X_i; Y_i)$  and if  $\mathfrak{R} = \{R_1, R_2, \dots, R_n\}$  denote the set of observations, then the function *log likelihood*

$$l_n(F) = \sum_{i=1}^n \log([P_F(R_i)]) \quad (1)$$

Suppose  $\mathcal{F}$  the classes of all bivariate distribution functions, the MLE is a convex optimization problem solving infinite dimensional:

$$\hat{F}_n = \arg \max_{F \in \mathcal{F}} l_n(F) = \arg \max_{F \in \mathcal{F}} \sum_{i=1}^n \log(P_F(R_i)) \quad (2)$$

The following theorem will be presented and the entry relating to the MLE and NPMLE

**Theorem 2.1** (Maathuis, 2003)

An MLE  $\hat{\mathbf{p}}$  always exists.

**Proof:**

The set  $\left\{ p \in R^m : p_j \geq 0, j = 1, 2, 3, \dots, m, \sum_{j=1}^m p_j = 1 \right\}$  over which (2), is maximized is a

bounded and closed set in  $\mathbb{R}^m$  and therefore compact. The object function in (2) is continuous in  $p$ , and a continuous function on a compact set attains its maximum. ♦

**Lema 2.1** (Genteman & Vandall, 2002)

Any CDF which increases outside the set  $\mathcal{H}$  cannot be a MLE of  $F$ .

Based on Lema 2.1: allows us to parameterize  $F$  by setting  $p_i = P_F(H_i)$ . The optimization problem can than be stated in term of the vector  $\mathbf{p} = \{p_i\}_{i=1}^m$ , subjected to the constraints that  $\sum p_i = 1$  and  $p_i \geq 0$  for all  $i$ . Now, the likelihood can easily be rewritten in term of  $\mathbf{p}$ , since

$$P_F(R_j) = \sum_{i=1}^m \alpha_{ij} p_{ij}$$

we see that,

$$L(F) = L(\mathbf{p}) = \prod_{j=1}^n \left[ \sum_{i=1}^m \alpha_{ij} p_i \right] \text{ or}$$

$$L(p | R_1, R_2, \dots, R_n) = \prod_{j=1}^n (A' p)_j,$$

with  $\mathbf{A}$  clique matrix.

$L(\mathbf{p} | R_1, R_2, R_3, \dots, R_m)$  is some what informal in the sense that the dimension  $m$  of  $\mathbf{p}$  depends on the data  $R_1, R_2, R_3, \dots, R_m$ .

So that  $\mathbf{p}$  restricted to the simplex and with Theorem 2.1,

$$\left\{ p \in R^m : p_j \geq 0, j = 1, 2, 3, \dots, m, \sum_{j=1}^m p_j = 1 \right\}$$

$$\Delta_m = \{ \mathbf{p} = [p_1, p_2, \dots, p_m]'; p_j \geq 0, j = 1, 2, 3, \dots, m, \sum_{j=1}^m p_j = 1 \} \text{ or } \Delta_m = \{ p; p \geq 0 \text{ and } e' p = 1 \}$$

and  $l(p) \stackrel{def}{=} e' \log(A' p)$ ,  $e$  a vector of all elements is 1 and  $p \in \Delta_m$ .

obtained MLE of  $F$  is  $\mathbf{p}$ , denote  $\hat{\mathbf{p}}$ .

### 3. Rooted tree to calculate the NPMLE

For the calculation of the data censored NPMLE can be done with a rooted tree approach. The idea is based on the fact that the data can be partitioned into censored groups disjoint set.

**Definition** (Chartrand and Oellermann, 1993)

A directed graph is called a directed tree if the graph becomes a tree if ignored him . A directed tree is called a rooted tree if there is exactly one vertex with incoming degree 0 and all other vertices have degree entry 1. In a rooted tree , a vertex is called the out 0 leaves, while the vertex degrees do not zero is called a branch.

**Rooted Tree**

We call an observation *universal* if it intersects all other observations. For data set  $\mathcal{R}$  denote by  $\mathcal{R}^{(r)}$  the data set  $\mathcal{R}$  in which aU universal observations are removed. When data set  $\mathcal{R}$  can be partitioned into  $G$  groups  $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \dots, \mathcal{R}_G$ , such that observations from different groups do not intersect, we say that the data set is dividable.

The clique matrix  $\mathbf{A}_{m \times n}$  of a dividable data set  $\mathcal{R}$  can be written in the form  $\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_G \end{bmatrix}$ , where  $[\mathbf{A}_1]_{n_1 \times m_1}, \dots, [\mathbf{A}_G]_{n_G \times m_G}$  is *cliques* matrix, for  $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \dots, \mathcal{R}_G$ ,

and  $m = m_1 + m_2 + m_3 + \dots + m_G$ , and  $n = n_1 + n_2 + n_3 + \dots + n_G$ .

Suppose  $\mathbf{p}$  partitioned into groups such as  $\mathbf{p} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_G \end{bmatrix}$ , *self consistent condition*

$n\mathbf{p} = \mathbf{D}_p \mathbf{A} (\mathbf{A}' \mathbf{p})^{-1}$ , is equivalent to  $n\mathbf{p}_g = \mathbf{D}_{p_g} \mathbf{A} (\mathbf{A}' \mathbf{p}_g)^{-1}$ ,  $g=1,2,3,\dots,G$ . Therefore if

$\hat{\mathbf{p}} = \begin{bmatrix} \hat{\mathbf{p}}_1 \\ \hat{\mathbf{p}}_2 \\ \vdots \\ \hat{\mathbf{p}}_G \end{bmatrix}$  is the NPMLE for  $\mathbf{A}$ , for each  $g = 1,2,3,\dots,G$ , then

$$\begin{aligned} n\mathbf{e}' \hat{\mathbf{p}}_g &= \mathbf{e}' \mathbf{D}_{\hat{p}_g} \mathbf{A}_g (\mathbf{A}'_g \hat{\mathbf{p}}_g)^{-1} \\ &= \hat{\mathbf{p}}_g' \mathbf{A}_g (\mathbf{A}'_g \hat{\mathbf{p}}_g)^{-1} \\ &= (\mathbf{A}_g \hat{\mathbf{p}}_g)' (\mathbf{A}'_g \hat{\mathbf{p}}_g)^{-1} \\ &= n_g. \end{aligned}$$

Here the probability mass allocated to the group  $g$  oleh NPMLE adalah  $n_g/n$ .

**Principles of Rooted Trees**

The above discussion leads to the establishment of the following two simplifying principles in the computation of the NPMLE.

1. Universal observations play no role in finding NPMLE and we can simply ignore them for estimation purposes. Thus we need only consider data sets in which all universal observations are removed.
2. If a data set is dividable into groups  $\mathcal{R}_g$ ,  $g = 1,2,3,\dots,G$ , with  $G \geq 2$ , maka, then we can find the NPMLE for each group  $\mathcal{R}_g$  separately with group total mass  $|\mathcal{R}_g|/|\mathcal{R}|$ .

We can recursively apply the above two principles to form the reduction tree of the data, as described below. The tree's nodes are subsets of the original data set  $\mathcal{R}$ , while the root of the tree is set to be  $\mathcal{R}$  itself.

Case 1:  $\mathcal{R}$  dividable, then we let the groups for  $\mathcal{R}$ 's, say  $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \dots, \mathcal{R}_k$  be the children of  $\mathcal{R}$ .

Case 2:  $\mathcal{R}$  has a universal observation (and hence is not dividable), then we let  $\mathcal{R}^{(r)}$  be the only child of  $\mathcal{R}$ .

Case 3:  $\mathcal{R}$  neither contains universal observations nor is dividable, Then we let  $\mathcal{R}$  be a leaf.

Recursive approach for each  $\mathcal{R}_j$  (case 1) and  $\mathcal{R}^{(r)}$  produce a rooted tree, the leaves of the tree rooted can not be divided and free from universal observation, because the set of leaves of a single tree is rooted.

NPMLE mass divided into each leaf can be obtained in the formation of the tree rooted with the root, put on the root mass is 1. Suppose  $\mathcal{R}^*$  is a collection of observations in a vertex in an arbitrary level in the tree roots. Waste mass will not change between  $\mathcal{R}^*$  and  $\mathcal{R}_*^{(r)}$ . When  $\mathcal{R}^*$  (or  $\mathcal{R}_*^{(r)}$  same may case) be shared, the roots of the sub-tree  $\mathcal{R}^*$  (or  $\mathcal{R}_*^{(r)}$ ) will be assigned to the mass proportional to the number of observations in each vertex, with equal mass total  $\mathcal{R}^*$ . Rooted tree construction process, that total mass NPMLE always occupying every single leaf.

To get NPMLE given by the original data, focused only on the leaves of the tree rooted. Suppose  $\mathcal{R}_1, \mathcal{R}_2, \mathcal{R}_3, \dots, \mathcal{R}_L$ , data set corresponding to the leaves of the tree rooted with mass

$m_1, m_2, m_3, \dots, m_L$ , and if  $\hat{\mathbf{p}}_k$  NPMLE is a probability vector for a given set of data  $\mathcal{R}_k$

with  $k=1,2,3,\dots,L$ , the NPMLE obtained from the original data set is  $\hat{\mathbf{p}} = \begin{bmatrix} m_1 \hat{\mathbf{p}}_1 \\ m_2 \hat{\mathbf{p}}_2 \\ \vdots \\ m_L \hat{\mathbf{p}}_L \end{bmatrix}$ , with

$\hat{\mathbf{p}}_k$  NPMLE is obtained from each of the original data set  $\mathcal{R}_k$ , for  $k=1,2,3,\dots,L$ .

**Example:**

Consider the right-censored data set  $\mathcal{R}=\{R_1, R_2, R_3, R_4, R_5\}$ , where  $R_1=6^+, R_2=7, R_3=8^+, R_4=10$ , and  $R_5=20^+$ .

**Level 0 (root).** The adjacency matrix of  $\mathcal{R}_1$ 's intersection graph is

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{matrix} R_1 \\ R_2 \\ R_3 \\ R_4 \\ R_5 \end{matrix}$$

$R_1$  is universal observation in  $\mathcal{R}_1$  and  $\mathcal{R}_1^{(r)} = \{R_2, R_3, R_4, R_5\}$ . The adjacency matrix of  $\mathcal{R}_1^{(r)}$  intersection graph is

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{matrix} R_2 \\ R_3 \\ R_4 \\ R_5 \end{matrix}$$

$\mathcal{R}_1^{(r)}$  is dividiable  $\mathcal{R}_{1.1} \stackrel{def}{=} \{R_2\}$ ,  $\mathcal{R}_{1.2} \stackrel{def}{=} \{R_3, R_4, R_5\}$ .

mass allocation  $\mathcal{R}_2 \stackrel{def}{=} \{R_2\} = \frac{1}{4}$ , and  $\mathcal{R}_3 \stackrel{def}{=} \{R_3, R_4, R_5\} = \frac{3}{4}$

**Level 1.**  $\mathcal{R}_{1.1}$  is a leaf , The adjacency matrix for  $\mathcal{R}_{1.2}$  intersection graph is

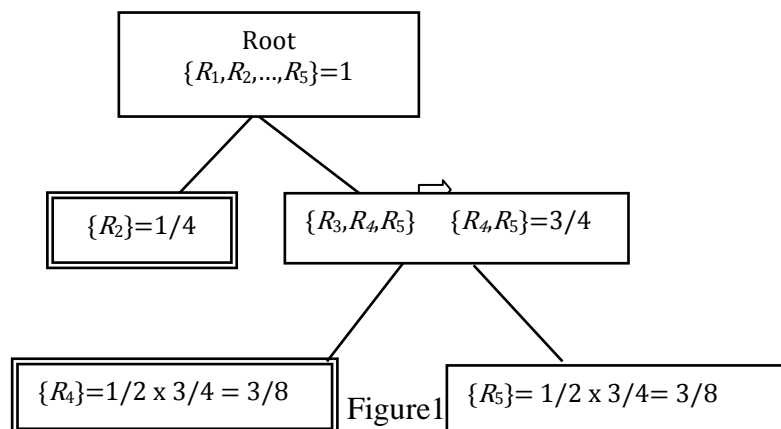
$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{matrix} R_3 \\ R_4 \\ R_5 \end{matrix}$$

Then  $\mathcal{R}_{1.2}$  has universal observation  $R_3$ .  $\mathcal{R}_{1.2}^{(r)} \stackrel{def}{=} \{R_4, R_5\} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{matrix} R_4 \\ R_5 \end{matrix}$

$\mathcal{R}_{1.2}^{(r)}$  is dividiable  $\mathcal{R}_{1.2.1} \stackrel{def}{=} \{R_4\}$  dan  $\mathcal{R}_{1.2.2} \stackrel{def}{=} \{R_5\}$  , and mass allocation  $R_4 = \frac{1}{2} \times \frac{3}{4} = \frac{3}{8}$  and  $R_5 = \frac{1}{2} \times \frac{3}{4} = \frac{3}{8}$ , respectively.

**Level 2.**  $\mathcal{R}_{1.2.1}$  and  $\mathcal{R}_{1.2.2}$  leaves.

**Illustrastrated  $\mathcal{R}^*$  and  $\mathcal{R}_*^{(r)}$  .**



We obtain 3 leaves  $\{R_2\}$ ,  $\{R_4\}$  and  $R_5\}$  with masses  $1/4$ ,  $3/8$  and  $3/8$ . In this case, the 3 leaves are maximal cliques and the NPMLE associated with data set  $\mathcal{R}$  is

$$\hat{p} = \begin{bmatrix} \frac{1}{4} & \frac{3}{8} & \frac{3}{8} \end{bmatrix}.$$

NPMLE obtained with a rooted tree approach leads to Kaplan - Meier estimates for each right- censored data sets. This approach is identical to the redistribution algorithm Efron (1967 ) univariate right censored data. While the data for multivariate censored , generally, not necessarily a single leaf as in Example.

**Complexity analysis of the implementation of a rooted tree**

The first step is to construct a rooted tree graph intersection . Intersection graph algorithms have time complexity  $O(n^2)$ , Lee (1983). While the time complexity for the rooted tree is  $O(n^3)$ . At the root vertex, get a universal observation (by checking the degree of each vertex in the graph intersection ), has obtained a linear time. Determining connectedness components have time complexity  $O(n^2)$ . At the root level , the execution time is not more than  $Cn^2$ , for a constant  $C$ . At the first level vertex,  $k$  contains a component that is  $n_1, n_2, \dots, n_k$ . each observation, where  $n_1 + n_2 + \dots + n_k \leq n$ . Regarding the first level of each vertex , such as the new root, has less than

$$\begin{aligned} Cn_1^2 + Cn_2^2 + \dots + Cn_k^2 &= C(n_1^2 + n_2^2 + \dots + n_k^2) \\ &\leq C(n_1 + n_2 + \dots + n_k)^2 \leq Cn^2 \end{aligned}$$

Because tree roots have roots level most  $n$ , then there are as many as  $n$  recursion, so that the total execution time is  $O(n^3)$ .

#### Remarks:

Univariate and bivariate to the data, the performance of the algorithm requires more study detailed scheme for the different sensors. The main advantage of finding a rooted tree is the data univariate and bivariate NPMLE simplification of the calculation

#### Steps Rooted Tree Algorithm

- Getting Started (0) is the root, which enter the intersection graph matrix results (whether or not having a universal observation)
- Step 1, a leaf, at this step contains observations are not universal, if contains the probability mass is 0, otherwise the probability mass can be sought from the rest of the matrix. (This step will be repeated until all the probability mass is obtained)

#### References

- [1] Burke, M. D. 1988. Estimation Distribution Function under Random Censorship. *Biometrika*. Vol **75**. 379-382.
- [2] Campbell, G and Foldes, A. 1982. Large-sample Properties of Nonparametric Bivariate Estimators with Censored Data. In *Nonparametric Statistical Inference*, Volume 32 of *Collog. Math.Soc. Jnos Bolyai*, pages 103-12. North-Holland, Amsterdam-New York.
- [3] Chartrand, G. and Oellermann, O.R., 1993, *Applied and Algorithmic Graph Theory*, McGraw-Hill, Inc., New York.
- [4] Efron, B. 1967. The two-sample problem with censored data. *Proc. Fifth Berkeley Symp. Mat. Statist. Probab.* 4. 831-853.
- [5] Gentleman, R. and Vandal, A.C. 2002. Graph-Theoretical Aspects of Bivariate Censored Data. *Can. J Statist* 10. 557-571.
- [6] Lee, D.T. (1983). Maximum clique problem of rectangle graphs. *Advances in Computing Research* 1, 91-107
- [7] Liu, X. 2005. *Nonparametric estimation bivariate censored data a discrete of the approach*. A thesis submitted to McGill University in partial fulfilment of the requirements degree of Doctor of Philosophy Department of Mathematics and Statistics McGill University, Montreal, Canada
- [8] Maathuis, M.H. 2003. Nonparametric Maximum Likelihood Estimation for Bivariate Censored Data. Master's thesis, Delft University of Technology, Delft.
- [9] Prentice, R.L and Cai, J. 1992. Covariance and Survivor Function Estimation Using Censored Multivariate Failure Time Data. *Biometrika*. Vol. **79**. 495-512.
- [10] Prentice, R.L. 1999. On Nonparametric Maximum Likelihood Estimation of The Bivariate Survivor Function. *Statist. Med.* **18**. 2517-2527.
- [11] Pruitt, R.C. 1993. Small Sample Comparison of Six Bivariate Survival Curve Estimators. *Journal of Statistical and Computation and Simulation* Vol **45**, 147-167.