Interval censored data: a note on the nonparametric maximum likelihood estimator of the distribution function

Dankmar Böhning

Department of Epidemiology
Free University Berlin

Presentation at the XVIIIth International Biometrics Conference

2. July 1996

Overview

1. The Problem
2. Uniqueness and a Characterization of the MLE of the Distribution Function
3. Algorithmic Construction
4. A Biometric Application

Ref.: Böhning, Schlattmann, Dietz (1996, *Biometrika)*
Gentleman and Geyer (1994, *Biometrika*)

1. The Problem

T Time until a certain event occurs

Pr ($T \leq t$) = F(t) distribution function

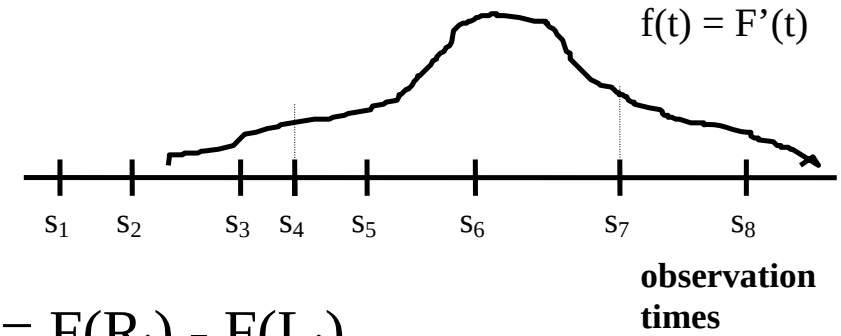
T is *intervalcensored*, e.g.

$T \in (L, R]$

Example: Repeated testing for occult events (in tumorgenesis)

contribution of i-th interval $(L_i, R_i]$ to the likelihood:

$$\Pr\{T_i \in (L_i, R_i]\} = \Pr\{L_i < T_i \leq R_i\}$$



$$= F(R_i) - F(L_i)$$

Example: $F(R_i) - F(L_i) = F(s_7)-F(s_4)$
$= F(s_7) - F(s_6)$
$+ F(s_6) - F(s_5)$
$+ F(s_5) - F(s_4)$

$$F(R_i) - F(L_i)$$

$$= \sum_{j=1}^{m} \alpha_{ij} [F(s_j) - F(s_{j-1})], \text{ with } F(s_0)=0$$

where $\alpha_{ij}= 1$, if $(s_{j-1},s_j] \subseteq (L_i\backslash,R_i]$ and 0 otherwise

full likelihood

$$\prod_{i=1}^{n} F(R_i)-F(L_i)$$

$$= \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} [F(s_j)-F(s_{j-1})]$$

$$= \prod_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} p_j$$

Log-Likelihood

$$l(p) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{m} \alpha_{ij}\, p_j \right)$$

NPMLE $\hat{\mathbf{p}}$

$$\text{maximizes } l(p) = \sum_{i=1}^{n} \log\left(\sum_{j=1}^{m} \alpha_{ij}\, p_j \right)$$

under the restrictions

$p_j \geq 0$ f.a. $j=1,...,m$ and $p_1+p_2+...+p_m=1$

analogy to the mixture problem

$$\text{maximize } \sum_{i=1}^{n} \log\left(\sum_{j=1}^{m} f_j(x_i)\quad p_j \right)$$

under the restrictions

$p_j \geq 0$ f.a. $j=1,...,m$ and $p_1+p_2+...+p_m=1$

2. Uniqueness and a Characterization of the MLE of the Distributionfunction

a) Uniqueness

Score

$$d_j = \frac{\partial l}{\partial p_j} = \sum_{i=1}^{n} \frac{\partial}{\partial p_j} \log(\eta_i), \quad \eta_i = \sum_{j=1}^{m} \alpha_{ij} p_j$$

$$= \sum_{i=1}^{n} \alpha_{ik}/\eta_i$$

Hessian

$$\frac{\partial^2 l}{\partial p_k \partial p_j} = -\sum_{i=1}^{n} \alpha_{ij}\alpha_{ik}/\eta_i^2$$

$$-\nabla^2 l(p) = \sum_{i=1}^{n} \alpha_i \alpha_i^T/\eta_i^2 = A^T W A$$

symmetric and nonnegative definite

Result

A full rank $\Rightarrow$ $A^T W A$ positive definite

$\Rightarrow$ $-l$ strictly convex or l strictly concave

$\Rightarrow$ $\hat{p}$ unique

b) Characterisation

$$\Phi(p,q) = \lim_{\alpha\to 0} \frac{1}{\alpha} l\{((1-\alpha)p+\alpha q) - l(p)\}$$

$$= \mathbf{d^T(q - p)}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} q_j / \eta_i - n$$

$$\Phi(p,e_k) = \sum_{i=1}^{n} \alpha_{ik} / \eta_i - n, \quad e_k = (0,..,0,1,0,..,0)^T$$

(k-th position)

Result

$\hat{p}$ NPMLE $\Leftrightarrow \Phi(\hat{p}, e_k) \le 0$ for all k=1,..,m

$\Leftrightarrow d_k(\hat{p}) \le n$ for all k=1,..,m

$\Leftrightarrow \frac{1}{n} d_k(\hat{p}) \le 1$ for all k=1,..,m

Moreover, $\hat{p}_k > 0 \Rightarrow \frac{1}{n} d_k(\hat{p}) = 1$

Example

$l(p) = \log(p_1)+2\log(p_1+p_2)+2\log(p_2+p_3)+\log(p_3)$

$\hat{p} = \frac{1}{3}(1,1,1)^T$

$$d_1 = \frac{1}{p_1} + \frac{2}{p_1+p_2} = 3+3 = 6 = n$$

$$d_2 = \frac{1}{p_1+p_2} + \frac{2}{p_2+p_3} = 3+3 = 6 = n$$

$$d_3 = \frac{2}{p_2+p_3} + \frac{1}{p_3} = 3+3 = 6 = n$$

and

$$\hat{F}(t) = F(\hat{s}_{j-1})$$

$$= \begin{cases} 1; \text{ if } t \geq s_m \\ \sum_{s=1}^{j} p_s = (j-1)/3 \text{ für } s_{j-1} \leq t < s_j;\ j=1 \ldots m \\ 0; \text{ if } t < s_1 \end{cases}$$

3. Algorithmic Construction of the NPMLE

one possibility: EM iteration

$$p_j^{EM} = p_j \, d_j/n$$

attractive, because

$$l(\mathbf{p}^{EM}) \geq l(\mathbf{p})$$

and $\mathbf{p}^{EM} \rightarrow \hat{\mathbf{p}}$ (strict concavity)

disadvantage: converges slowly

alternative possibility (**V**ertex-**E**xchange **M**ethod)

$$\mathbf{p}^{VEM} = \mathbf{p} + \beta p_{min} (\mathbf{e}_{max} - \mathbf{e}_{min})$$

„min“ und „max“ indeces with:

$d_{min} = \min \{d_j | \; j=1,...,m\}$

$d_{max} = \max \{d_j | \; j=1,..,m\}$

and β is *monotone* step-length

available in package C.A.MAN (Böhning, Schlattmann, Lindsay, *Biometrics*, 1992)

4. A Biometric Application

Finkelstein und Wolfe (*Biometrics* 1985)
Breast Cancer Cosmetic Data

T = Time (in Month) until breast-retraction of 94 patients with breast cancer, after radiation therapy (46), and after radiation therapie and chemotherapy (48)

follow-up at clinical visits every 4-6 months

data:

$(L_i,R_i]$ $\downarrow$ $\qquad$ $(s_{j-1},s_j]$ $\rightarrow$

	(4,5]	(6,7]	(7,8]	(11,12]	(15,16]	(17,18]	(24,25]	....	(34,35]	(40,44]	(46,48]
(45,*]	0	0	0	0	0	0	0	....	0	0	1
(6,10]	0	1	1	0	0	0	0	....	0	0	0
(0,7]	1	1	0	0	0	0	0	....	0	0	0
						.....					

complete matrix $A = (\alpha_{ij})$:

0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	1	1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	0	1	1	1	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1	1	1	1	1
0	0	1	1	0	0	0	0	0	0	0	0	0	0

0 0 0 0 0 0 0 0 0 0 0 1 1 0
1 1 1 0 0 0 0 0 0 0 0 0 0 0
1 1 1 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 1 1 1 1 1 1 1 1 1
0 0 0 1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 1 1 1 1 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1 1 1 1 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 1 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 1 1 1 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1
0 0 0 0 0 0 0 0 0 0 0 1 1 1
0 0 0 0 0 0 0 0 0 0 0 0 1 1
0 0 0 0 0 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 0 0 1
0 0 0 1 1 1 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 1 1
0 1 1 1 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 1 1
1 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 1 1 1 1 1 1 1
0 0 0 0 0 0 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1
0 1 1 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 0 1 1 1 1 0 0 0 0
0 0 0 0 0 1 1 0 0 0 0 0 0 0
0 0 0 0 0 0 1 1 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1
0 0 0 0 0 0 0 0 1 1 1 1 1 1
0 0 0 0 0 0 1 1 0 0 0 0 0 0
0 0 0 0 0 0 0 0 0 0 0 1 1 1
0 0 0 0 0 0 0 0 0 1 1 1 1 1
0 0 0 0 0 0 0 0 0 0 1 1 1 1

Weights and Gradient at NPMLE ($\max_{1\le k\le m} \bar{d}_k \le 0.000001$)

$\bar{d}_k$	p_k	k
1.0000	0.0463	1
1.0000	0.0334	2
1.0000	0.0887	3
1.0000	0.0708	4
0.4722	**0.0000**	5
0.8337	**0.0000**	6
1.0000	0.0926	7
0.7965	**0.0000**	8
1.0000	0.0818	9
0.7713	**0.0000**	10
0.9377	**0.0000**	11
1.0000	0.1209	12
0.9394	**0.0000**	13
1.0000	0.4656	14

Summary

Problem of finding the NPMLE in case of interval censored event times is equivalent to the problem finding the MLE of a mixture distribution of indicator functions

Uniqueness and a characterisation of the NPMLE is possible in a simple way

Algorithmic construction of the NPMLE is readily done with the C.A.MAN software

A demonstration was given using well-known clinical data