Abstracts of the Ninth Workshop on Simulation (2018) 1-2

On the interval right censored data with unobserved status after censoring

Irina Yu. Malova[1], Alexandre Berred[2] and Sergey V. Malov[3]

We consider the interval censoring approach for right-censored survival data. Right-censored event times are not observed exactly, but at some random inspection times. Status of each individual becames not observable just after censoring. We discuss crucial properties of the maximum likeihood estimator and investigate asymptotic properties of the estimator.

1 Introduction

Let T and U be independent failure and censoring times respectively, having distribution functions F and G respectively. Right censored observation is given by the event time $X = T \wedge U$ and the indicator $\delta = \mathbb{I}_{\{T \le U\}}$. In practice often the event time is not observed exactly, but at some random observation times $0 = W_0 < W_1 < \ldots < W_k < \infty$, having common distribution $\boldsymbol{J}$, which divide the time line into a fixed number of r disjoint finite intervals $I_1, \ldots, I_s$: $I_1 = [0, W_1]$, $I_k = (W_{k-1}, W_k]$, $k = 2, \ldots, s$, and the infinite interval (W_k, ∞). We use notation $\boldsymbol{J}$ for the common distribution function of $\boldsymbol{W} = (W_1, \ldots, W_k)$. The intrval right censored observation is given by the set of dummy variables $\eta_i = \mathbb{I}_{\{T \in I_i; U > W_i\}}$ and $\nu_i = \mathbb{I}_{\{T > W_{i-1}; U \in I_i\}}$, $i = 1, \ldots, k$. Interval right censored data is a sample from the distribution $(\boldsymbol{W}, \nu, \eta)$.

The Kaplan–Meier estimator [3] is the nonparametric maximum likelihood estimator (NPMLE) in this case. Asymptotic properties of the Kaplan–Meier estimator are well developed [1]. The iterative convex minorant algorithm to create NPMLE from interval censored data and the asymptotic behavior of the NPMLE due to [2]. The nonparametric estimate from interval right censored data with observed status after censoring was developed by [M], the particular case of the life time data was considered in [5]. The particular case of the interval right censored data with unknown status after censoring under fixed observation times and the corresponding categorical survival tests was given in [6]. We consider NPMLE from a sample of interval right censored data with unobserved status after censoring under random observation times and discuss its asymptotic properties.

[1]Higher School of Technology and Energy, St. Petersburg State University of Industrial Technologies and Design & Department of Mathematics, St. Petersburg Electrotechnical University (LETI), St. Petersburg, Russia

[2]Université du Havre UFR Sciences et Techniques, BP 540 76058 Le Havre Cedex, France

[3]Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University & Department of Mathematics, St. Petersburg Electrotechnical University (LETI), St. Petersburg, Russia E-mail: malovs@sm14820.spb.edu

2 Formulation of the problem

Assume that (T_j, U_j) be a sample from the distribution of (T, U) and $(\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n)$: $\boldsymbol{W}_j = (W_{1j}, \ldots, W_{kj})$, $0 < W_{1j} < \ldots < W_{kj}$ be a sample from $\boldsymbol{J}$; $I_{1i} = [0, W_{1i}]$ and $I_i = (W_{i-1}, W_i]$, $j = 2, \ldots, k$, $i = 1, \ldots, n$. The observed data are W_{ij} η_{ij} and ν_{ij}, where $\eta_{ij} = \mathbb{1}_{\{T_j \in I_{ij}; U_j > W_{ij}\}}$, $\nu_{ij} = \mathbb{1}_{\{T_j > W_{j-1}; U_j \in I_{ij}\}}$, $j = 1, \ldots, s$, $i = 1, \ldots, n$. The log likelihood function can be written as the sum $LL(F, G) = LL^f(F) + LL^c(G)$ with

$$\begin{aligned} LL^f(F) = \textstyle\sum_{j=1}^n \sum_{i=1}^k \big\{ & \eta_{ij} \log(F(W_{ij}) - F(W_{i-1j})) \\ & + \nu_{ij} \log(1 - F(W_{i-1j}))\big\} + (1 - \eta_{\bullet j} - \nu_{\bullet j}) \log(1 - F(W_k)) \end{aligned}$$

and

$$\begin{aligned} LL^c(G) = \textstyle\sum_{j=1}^n \sum_{i=1}^k \big\{ & \nu_{ij} \log(G(W_{ij}) - G(W_{i-1j})) \\ & + \eta_{ij} \log(1 - G(W_{ij}))\big\} + (1 - \eta_{\bullet j} - \nu_{\bullet j}) \log(1 - G(W_k)), \end{aligned}$$

where $\eta_{\bullet j} = \sum_{i=1}^k \eta_{ij}$ and $\nu_{\bullet j} = \sum_{i=1}^k \nu_{ij}$, $j = 1, \ldots, n$. In the particular case of $k = 1$, $LL^f(F) = \sum_{j=1}^n \eta_j \log F(W_j) + (1 - \eta_j - \nu_j) \log(1 - F(W_j))$ and $LL^c(G) = \sum_{j=1}^n \nu_j \log G(W_j) + (1 - \nu_j) \log(1 - G(W_j))$ and the maximum likelihood estimators (MLE) for the distribution functions F and G can be obtained from LL^f and LL^c using convex minorant algorithm. In general case of $k > 1$, the MLE can be created using a special case of the iterative convex minorant algorithm.

Consistency and a rate of convergence as well as some problems in obtaining of weak convergence will be discussed. Some features of the case of fixed observation times will be discussed too.

References

[1] Breslow N., Crowley J. *A Large Sample Study of the Life Table and Product Limit Estimates under Random Censorship.*//Ann. Stat.,1974,v. 2, p. 437–453.

[2] Groeneboom P., Wellner J.A. *Information Bounds and Non-parametric Maximum Likelihood Estimation*, DMV Seminar Band 19 // Birkhauser, Basel, 1992.

[3] Kaplan E.L., Meier P. *Nonparametric Estimation from Incomplete Observations.* // JASA, 1958, v. 53, p. 457–481.

[4] Malov S.V. *Nonparametric estimation from case 1 interval right-censored data.* // Unpublished manuscript, submitted to Stat Neerl.

[5] Malov S.V., O'Brien S.J. *Life Table Estimator Revisited.* // Commun Stat Theory Methods, 2018, v. 47(9), p. 2126–2133.

[6] Malova I.Yu., Malov S.V. *On survival categorical methods from grouped right censored data with unobserved status after censoring.* // In Applied Methods of Statistical Analysis. Nonparametric Methods in Cybernetics and System Analysis. Proc. of the International Workshop AMSA'17, 2017, p. 136–142.