Nonparametric Maximum Likelihood Estimation of Survival Functions When They Cross

Myongsik Oh
Pusan University of Foreign Studies, Department of Statistics
55-1 Uam-Dong Nam-Gu
Pusan 608-738, Korea
moh@stat.pufs.ac.kr

1. Introduction

Estimation of functions under restriction has been one of major statistical issues for a long time. The restrictions of interest are, to name a few, monotonicity, unimodality, stochastic ordering, hazard rate ordering, likelihood ratio ordering and so on. Many different approaches for these problem have been developed. These are nonparametric maximum likilehood estimation, Bayesian nonparametric estimation, kernel estimation, spline smoothing. The key tools for attacking these problems is, however, so-called isotonic regression, which is extensively used in order restricted statistical inference.

We now here consider a crossing point of two survival functions or distribution functions. This problem is seemingly unrelated to any one of aforementioned restrictions but eventually it can be explained by an order restriction which is unimodality. The most immediate application of this situation is as follows. The old and new treatments are applied to each of two groups, A and B, of patients whose survival functions are $1-F(t)$ and $1-G(t)$, for $t>0$, respectively. Let two distribution functions F and G cross at a point t_0, called crossing point, i.e.,

$$1-F(t) > 1-G(t) \text{ for } t < t_0 \text{ and } 1-F(t) < 1-G(t) \text{ for } t > t_0$$

then the patients in group A have lower chance of survival before time t_0 but higher chance after t_0.

2. Duality and Estimation

Let $\mathcal{A}$ be the set of functions on $X = \{1, 2, \ldots, k\}$ which are isotonic with respect to the partial order $x_1 \geq \cdots \geq x_{i_0-1} \geq x_{i_0} \leq x_{i_0+1} \leq \cdots \leq x_k$ for fixed $1 < i_0 < k$. The set $\mathcal{A}$ is a polyhedral cone and hence is finitely generated. The dual $\mathcal{A}^{*w}$, of $\mathcal{A}$ is the set of all functions making an obtuse angle with each of the generators of $\mathcal{A}$. In this case,

$$\begin{aligned}\mathcal{A}^{*w} = \{h \;\; : \;\; & \sum_{j=1}^{i} h(x_j)w(x_j) \leq 0, \text{ for } i = 1, \ldots, i_0 - 1, \\ & \sum_{j=1}^{i} h(x_j)w(x_j) \geq 0, \text{ for } i = i_0, \ldots, k, \;\; \sum_{j=1}^{k} h(x_j)w(x_j) = 0\}.\end{aligned}$$

Without loss of generality we assume that both random samples are observed at $s_1, s_2, \ldots, s_k$, where $-\infty = s_0 < s_1 < s_2 < \cdots < s_k \leq \infty$. Let d_{1i} (d_{2i}) be the number of observations from F (G) distribution at s_i. Now we need to find F and G, which maximize the following function under the restriction.

$$\prod_{i=1}^{k} \{F(s_i) - F(s_{i-1})\}^{d_{1i}} \{G(s_i) - G(s_{i-1})\}^{d_{2i}}.$$

Suppose we have a crossing point in $(s_{i_0-1}, s_{i_0},]$. Since F and G are both nondecreasing and arbitrary, we only need to find F and G which satisfy

$$\begin{aligned} \sum_{j=1}^{i}[F(s_i) - F(s_{i-1})] &\leq \sum_{j=1}^{i}[G(s_i) - G(s_{i-1})] \text{ for } i = 1, \ldots, i_0 - 1, \\ \sum_{j=1}^{i}[F(s_i) - F(s_{i-1})] &\geq \sum_{j=1}^{i}[G(s_i) - G(s_{i-1})] \text{ for } i = i_0, \ldots, k. \end{aligned}$$

This estimation problem is quite similar to the ordinary stochastic ordering problem which is studied extensively by several researchers including Robertson and Wright(1981), Dykstra(1982), Wang(1996) and many others. For details the readers refer to Barlow and Brunk (1972) or Robertson *et. al.*(1988).

3. Consistency

It can be shown that the estimator is strongly consistent. More precise reslut can be obtained using law of the iterated logarithm. See Kiefer(1961) and Robertson and Wright (1974) for relevant works.

REFERENCES

Barlow, R.E., and Brunk, H.D. (1972). The isotonic regression problem and its dual, *J. Amer. Statist. Assoc.*, **67**, 140-7.

Dykstra, R.L. (1986). Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *J. Amer. Statist. Assoc.*, **77**, 621-8.

Kiefer, J. (1961). On large deviations of the empiric distribution function of vector random variables and a law of the iterated logarithm. *Pacific J. Math.* **11**, 649-60.

Robertson, T., and Wright, F.T. (1974). On the maximum likelihood estimation of stochastically ordered random variables. *Ann. Statist.*, **2**, 528-34.

Robertson, T., and Wright, F.T. (1981). Likelihood ratio tests for and against stochstic ordering between multinomial populations. *Ann. Statist.*, **10**, 1234-45.

Robertson. T., Wright, F.T., and Dykstra, R.L. (1988). *Order restricted statistical inference. Wiley.*

Wang, Y. (1996). A likelihood ratio test against stochastic ordering in several populations. *J. Amer. Statist. Assoc.*, **91**, 1676-83.