ESTIMATING A DISTRIBUTION FUNCTION BASED ON MINIMA-NOMINATION SAMPLING

BY MARTIN T. WELLS AND RAM C. TIWARI

Cornell University and University of North Carolina

The nonparametric maximum likelihood estimator of a distribution function based on a maxima-nomination sample has been derived recently by Boyles and Samaniego (1986). In this article we study minima-nominations for the case of censored data.

1. Introduction. Let $X_{i1}, \ldots, X_{iK_i}$, $i = 1, \ldots, n$ be independent identically distributed (i.i.d.) random variables (r.v.'s) having a common continuous distribution function F with support $(0,\infty)$. Denote the vector $(X_{i1}, \ldots, X_{iK_i})$ by $\mathbf{X}_i$, $i = 1, \ldots, n$. Define the map $\Pi_i : \mathbb{R}^{K_i} \to \mathbb{R}$ such that Π_i maps $\mathbf{X}_i$ into a particular element in $\mathbf{X}_i$, say X_i $(i = 1, \ldots, n)$. We shall call X_i the nominee of $\mathbf{X}_i$ and the collection $\{X_i : i = 1, \ldots, n\}$ is called the nomination sample. The case when $\Pi_i(\mathbf{X}_i) = \max_{1 \leq j \leq K_i} X_{ij}$ has been studied by Willemain (1980) and Boyles and Samaniego (1986). Another important case is where $\Pi_i(\mathbf{X}_i) = \min_{1 \leq j \leq K_i} X_{ij}$; that is, when the nominee of $\mathbf{X}_i$ is the minimum. As an example of such a data generating process suppose that a factory has n identical machines; the i^{th} machine having K_i components $(i = 1, \ldots, n)$. Suppose also that each machine is set up as a series system of i.i.d. components with common d.f. F. Let $\mathbf{X}_i$ be the life lengths of the components in the i^{th} machine. As soon as the first component fails the entire machine fails, these first failure times for the entire factory are $(X_1, \ldots, X_n)$, the nomination sample. A reliability engineer may be interested in inference about the components of the machines, that is about F, rather than the machines itself.

Another example of such a data generating process is the following. Suppose a consumer has a known number of options from which he/she has to make a single decision. The wise consumer will usually choose the option that costs the least and hence the nominee will be the option of minimal cost. Although the distribution of all option costs is unknown, one would like to be able to draw some inference about this distribution from the nomination sample.

In this note we consider the estimation of the distribution function with a nomination sample in the presence of random censoring. This estimator is derived

AMS 1980 subject classifications. Primary 62G05; secondary 60G15, 60G55.

Key words and phrases. Martingale limit theorems, nomination sampling, nonparametric maximum likelihood, censored data, Kaplan-Meier estimator.