A SIMULATION STUDY OF A MINIMUM DISTANCE ESTIMATOR FOR FINITE MIXTURES UNDER CENSORING

René Ferland and Sorana Froda
Université du Québec à Montréal

Jean Lavigne
MDS Pharma Services

In this paper we introduce a minimum distance estimator of the weights in a finite mixture when the data are censored. Our estimator is a natural extension of the estimator of Choi and Bulgren (1968). In order to check the robustness of this minimum distance estimator we perform a simulation study.

1. Introduction

In this paper we consider data $Y_1, \ldots, Y_n$ from the following finite mixtures model

$$(1.1) \qquad H(t) = \sum_{j=1}^{k} \pi_j F_j(t) = \sum_{j=1}^{k} \pi_j F(t, \theta_j), \quad \pi_j \geq 0, \sum_{j=1}^{k} \pi_j = 1,$$

where k is known, and the parameters θ_j, $j = 1, \ldots, k$ are either known or unknown. The data Y_i, $i = 1, \ldots, n$ are times (to relapse or recovery, e.g.) and therefore it makes sense to assume that $F_j(t)$ is some typical distribution in survival analysis, e.g. exponential, Weibull or Gompertz. In medical studies, many illnesses are actually mixtures of two or more conditions. The rate of survival (relapse, failure) can be different for each of these conditions, but the conditions or cause of death may be hard to identify. In a cohort, one may have both categorized and uncategorized data (patients e.g.): the uncategorized data form a mixture of unknown proportions or weights, weights to be estimated, while the categorized ones can offer initial estimates of the parameter values, θ_j, $j = 1, \ldots, k$. McLachlan and Basford (1988, Chapter 4) describe various practical settings where one has some knowledge of the components F_j in (1.1) and is interested in estimating the proportions π_j. In particular, they quote Choi (1979) who proposes mixture models in the case of two mutually exclusive causes of failure (competing risks).

In the medical (lifetime) context it is typical for the data to be censored, i.e., often one observes (T_i, δ_i) where, for each $i = 1, \ldots, n$, $T_i = \min(Y_i, C_i)$, with C_i a censoring time independent of Y_i, and δ_i an indicator variable such that

$$\delta_i = \begin{cases} 1 & \text{if } Y_i \leq C_i \\ 0 & \text{if } Y_i > C_i. \end{cases}$$

Keywords and phrases: finite mixture; minimum distance estimator; censored data; robustness.