Statistical inference for the semi-parametric proportional reversed hazard model for left-censored and zero-inflated data[1]

Christian Paroissin [1] & Magdalena Pereda Vivo [2]

[1] *Universite de Pau et des Pays de l'Adour, E2S UPPA, CNRS, LMAP, Pau, France et christian.paroissin@univ-pau.fr*

[2] *Universite de Pau et des Pays de l'Adour, E2S UPPA, CNRS, LMAP, Pau, France et mpvivo@univ-pau.fr*

In many areas (like toxicology, ecotoxicology, chemistry, geosciences and more generally in environmental sciences, as few examples), studies are based on data obtained by some analytical methods. However, with such approaches, one may have only partial information. For instance, when dealing with concentration measurements with an analytical method, one will observe an exact measurement only if it is larger than a certain threshold, called limit of quantification (LOQ): in other words, one has only the information that the concentration lies between zero and this limit. Such a situation is called left-censoring.

Censoring is well-known phenomena in statistics. It appears, for instance, in biostatistics. When dealing with lifetime data, some duration may not be observed exactly since the event occurs later than a certain time point. A typical case is when one performs a medical study over a given period: all the lifetimes longer than this period then get censored. Such a situation is known as right censoring, and it has been investigated rather extensively in the literature. Left censoring has been less studied by statisticians.

Besides, in some situations there could be the absence of the substance under consideration, and this will lead to true zeros. In statistical terms, we talk about data with zero excess.

In this work, we aim to analyse data subject to left censoring with inflation of zeros. It is not possible to distinguish a zero value from a positive left-censored observation. In the literature, there are two articles which propose a mixture model for this type of data, but they assumed a parametric distribution for the strict positive values. Here, we consider a semi-parametric regression model, more precisely the proportional reversed hazard model, for the positive part. We can find in Grouwels (2015) a semi-parametric regression model but they propose a Cox model for the positive part. We find more natural to use the reversed hazard function instead of the hazard function when dealing with left-censoring. For that, we propose a semi-parametric mixture model for dealing with left-censored zero-inflated data and analyse the influence of the covariates on the variables. Furthermore, we provide an EM algorithm to estimate the different parameters and we study the asymptotic properties of the estimators. This methodology has been applied to simulated data.

[1] *This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement No 945416.*