Log-symmetric Regression Models: A semiparametric approach

Luis Hernando Vanegas[1,a], *Gilberto A. Paula*[2]

[1]hvanegasp@gmail.com, Departamento de Estatística, Universidad Nacional de Colombia
[a] Instituto de Matemática e Estatística, Universidade de São Paulo
[2] giapaula@ime.usp.br, Instituto de Matemática e Estatística, Universidade de São Paulo

Accelerated failure time models or log-location-scale models with a specified error distribution and non-informative right-censored observations have received a lot of attention in recent years. The accelerated failure time model is appealing as it allows to specify a multiplicative effect of covariates acting on the quantiles (of any order) of the failure time distribution, which enables a straightforward parameter interpretation. Although these regression models are very interesting, they have limitations, for instance, usually just one parameter of the failure time distribution is modeled, and they do not admit the presence of nonparametric effects in their systematic component. Therefore, in this work, a very flexible accelerated failure time model is proposed, where the location and scale parameters of the log-lifetime distribution are modeled by using semi-parametric functions of explanatory variables, and whose nonparametric components are approximated by natural cubic splines or P-splines. The flexibility provided by the systematic component under this model lies in its capacity to relate the distribution of the lifetime (or failure time) with a set of covariates using a sum of arbitrary functions, whose functional forms are estimated from the data. Obviously, if the log-scale parameter is specified to be constant, this approach retains the direct link between the multiplicative effect of covariates and the quantiles (of any order) of the failure time distribution. In addition, if the location and scale parameters of the log-lifetime distribution are specified to be affected by covariates, the regression parameters can be interpreted by taking into account their multiplicative effect acting on the median and the skewness (or the relative dispersion) of the failure time distribution. Furthermore, the random component of the model is described by a very flexible class of probability distributions (i.e., the log-symmetric class), which in turn induces a wide range of shapes for the failure or hazard rate function. Particular cases of this approach include models based on the Birnbaum-Saunders and Birnbaum-Saunders-t distributions, which have been extensively studied in the context of failure times under the fatigue or cumulated damage assumption (see, e.g., Barros et al. (2008); Paula et al. (2012)). In addition, some of the log-symmetric distributions exhibit heavier tails than those of the log-normal one, which allows to estimate the model parameters in a robust manner under the presence of extreme or outlying observations. The `R` package `ssym` is developed. This package implements the statistical methodologies addressed in this work.

Keywords: skewness, asymmetric responses, maximum penalized likelihood estimator.

Bibliography

Barros M. and Paula G. A. and Leiva V. (2008) A new class of survival regression models with heavy-tailed errors: robustness and diagnostics. Lifetime Data Analysis, 14, 316-332.