

LINEAR INTERVAL REGRESSION FOR CENSORED RESPONSE VARIABLE

PAVEL KALUZNY¹, LEON BOBROWSKI¹,
CORNELIA ENĂCHESCU²

The method is discussed for multiple linear regression model for situations, where values of the dependent (response) variable y can only be given in the form of interval bounds $[y_j^-, y_j^+]$. Such situations can arise from measurement errors, statistical survey studies and, notably, from censoring frequently encountered in survival analysis. The method allows for the intervals to be specified individually for each data point; the interval bounds may coincide for intervals of zero length, as in usual point-wise regression, and need not be finite. Model estimation is based on minimization of a convex piecewise linear (*CPL*) penalty function, and does not make any specific assumptions as to the probability distribution of the data. The *CPL* penalty function gives robustness to the model and may be extended to enforce sparse regression coefficients for problems with large number of covariates. We describe the problem of interval regression, present the solution method and the associated optimization problem, and demonstrate practical results on simulated data and publicly available example of survival data.

Keywords: interval regression, censored data, convex piecewise linear function, CPL function.

¹ Nalecz Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland.

² Academia Romana, Institutul de Statistica Matematica si Matematica Aplicata, Cal. 13 Septembrie, nr. 13, Bucuresti, Romania