Assessing accuracy of statistical inferences by resamplings with applications in the reliability theory

Yuri K. Belyaev
Dept of Mathematics and Mathematical Statistics
Umeå University, SE-901 87 Umeå
Dept of Forest Economics
Swedish University of Agricultural Sciences, SE-901 83 Umeå
Sweden
yuri.belyaev@math.umu.se

Abstract

Commonly the analysis of statistical data includes three steps. In the first step after a pilot analysis of a given statistical data an appropriate statistical model is selected. In the second step statistical inferences, e.g. point estimation, classification, etc., are realized. On the third step the accuracies of the obtained statistical inferences are evaluated. The main aim of this communication is illustration how the proper resampling methods can be used in evaluation accuracies of the realized statistical inferences. Three examples with different statistical inferences will be represented.

In the first example we consider statistical data containing intervals related to exact values of random variables (r.v.s), e.g. life times (l.t.s) of tested elements. Rather complex structured data one can obtain in analysis of so-called willingness-to-pay (WTP) problems. Suppose that an improvement of elements' reliability is planned. Then the appropriate increasing of price, that consumers will be ready to pay for it, is essentially to know for future development of production the elements with better reliability. Necessary information can be collected by a contingent valuation (CV) where each randomly sampled respondent should freely select acceptable to him interval containing appropriate future price of the elements with improved reliability characteristics. The intervals containing selected prices are dependent on the unobserved values of selected prices. We call these intervals as self-selected. The ends of the self-selected intervals can be essentially rounded. The related to such type of data statistical model has been suggested by Belyaev and Kriström (2009) with application to improvement of environment. Parametric and non-parametric ML-estimators for WTP-distribution of the unobserved exact WTP-values can be obtained and their accuracy can be consistently estimated by the resampling methods.

In the second example we suppose that for each element values of one or several auxiliary variables are registered before life testing. The auxiliary variables can be used to identify non-tested elements with short future l.t.s. Special nearest neighbors (NN-) classifiers can be applied. The most important their characteristics are the cross-classification (CC-) probabilities. The cross-validated (CV-) estimators of the CC-probabilities can be found by using the training sets with the auxiliary variables. The distribution of deviations of the CV-estimators from CC-probabilities can be evaluated by using resampling from blocks of the explanatory variables.

In the third example we consider approach to selection linear heteroscedastic model which regression function is a multinomial of several explanatory variables. By using resampling it is possible consistently recognize that a term in the multinomial should be excluded or should not be excluded as the size of data is growing.