A NEW METHOD FOR THE DETECTION OF BREAKPOINTS AND GENE COPY NUMBER CHANGES IN ARRAY CGH DATA

E. Budinská[1], E. Gelnarová[1], M. G. Schimek[2,1]

[1]*Centre of Biostatistics and Analyses, Masaryk University, Brno, Czech Republic*
[2]*Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Graz, Austria*

Email: *michael.schimek@meduni-graz.at*

Array CGH experiments have become a powerful technique for analyzing changes in DNA, by comparing a DNA under investigation with a control DNA. These microarray experiments produce a huge amount of data and special statistical techniques for detecting changed regions are necessary. We introduce a new, rather simple method for the detection of breakpoints and gene copy number changes in array CGH data. The method uses the quantile smoothing approach proposed by Eilers & Menezes (Bioinformatics, 21, 1146–1153, 2005) as an important step of the data pre-processing. The new method detects breakpoints based on the assumption of rank order dependence of copy number changes and the jump character of these changes in the log ratios. The method is sequential and is based on monitoring changes in the variability of the distribution of log ratios using a moving window of fixed width. The variability of distribution of log ratios is estimated in each window by a modified version of the median absolute deviation. The idea of the method is that for those windows which cover breakpoints, the high variability is characteristic. When the variability of a window exceeds the estimated critical level, the breakpoint is detected. The critical level is derived as a quantile of the empirical distribution of variability of the dataset. The performance of the method is demonstrated using simulated and publicly available data sets.