DETECTION OF GENE COPY NUMBER CHANGES IN CGH MICROARRAYS USING A SPATIALLY CORRELATED MIXTURE MODEL

P. Broët

Université Paris-XI, INSERM U472, Villejuif, France

† E-mail: broet@vjf.inserm.fr

Comparative genomic hybridization (CGH) microarray is an emerging tool in bioclinical research that allows to identify genomic alterations. In oncology, where carcinogenesis is associated with complex chromosomic alterations, CGH arrays can be used for detailed analysis of genomic changes in copy number (in terms of gains or loss of genetic information) in the tumor sample. There are broad similarities between CGH array studies and differential gene expression studies as difference in signal intensities are used in both cases to identify features where changes are occurring. There are nevertheless clear differences that render less appropriate the use of the standard statistical techniques proposed for investigating differential gene expression changes. In contrast to transcriptome-oriented studies where dependencies between gene measurement intensities are mostly related to complex biological pathway interactions, dependencies between gene sequence changes for CGH microarray studies are related to chromosomic location. Here, we present a Bayesian spatially structured mixture model where we link the mixture weights to Markov random fields. From this model, posterior probabilities of belonging to each of the considered states (modal copy, deletion, amplification) are estimated for each genomic sequence and used to classify them and estimate the false discovery rate. Using simulated realistic data sets, we show the good performance of our model. We also present results on cancer related datasets.