ESTIMATING EVOLUTIONARY PATHWAYS AND GENETIC PROGRESSION SCORES WITH RTREEMIX

Jörg Rahnenführer[1], Jasmina Bogojeska[2], Adrian Alexa[2], André Altmann[2], Thomas Lengauer[2]

[1]Fakultät Statistik, Technische Universität Dortmund, Vogelpothsweg 87, 44227 Dortmund, Germany

[2]Max Planck Institute for Informatics, Stuhlsatzenhausweg 85, 66123 Saarbrücken, Germany

Email: rahnenfuehrer@statistik.uni-dortmund.de

In genetics, many evolutionary pathways can be modeled on the molecular level by the ordered accumulation of permanent changes. We have developed the class of mixture models of mutagenetic trees (Beerenwinkel et al., 2005a) that provides a suitable statistical framework for describing these processes. These models have been successfully applied to describe disease progression in cancer and in HIV. In cancer, progression is modeled by the accumulation of lesions in tumor cells such as chromosomal losses or gains (Ketter et al., 2007). In HIV, the accumulation of drug resistance-associated mutations in the genome is known to be associated with disease progression. Mutations in the genome of the dominant strain in the infecting virus population arise when a patient receives a specific medication.

From such evolutionary models, genetic progression scores can be derived that assign measures for the disease state to single patients (Rahnenführer et al., 2005). Progression of a single patient along such a model is typically correlated with increasingly poor prognosis. In the cancer application, we showed that higher genetic progression scores are significantly associated with shorter expected survival times in glioblastoma patients (Rahnenführer et al., 2005) and times until recurrence in meningioma patients (Ketter et al., 2007).

We present applications in this framework as well as the easy-to-use and compute-efficient R package *Rtreemix* for estimating such mixtures of evolutionary models from cross-sectional data. *Rtreemix* builds up on efficient C/C++ code provided in the Mtreemix package (Beerenwinkel et al., 2005b) for estimating mixture models. It contains additional new functions for estimating genetic progression scores with corresponding bootstrap confidence intervals for estimated model parameters. Furthermore, the stability of the estimated evolutionary mixture models can be analyzed (Bogojeska et al., 2008).

Beerenwinkel, N., Rahnenführer, J., Däumer, M., Hoffmann, D., Kaiser, R., Selbig, J., Lengauer, T. (2005a) Learning multiple evolutionary pathways from cross-sectional data, Journal of Computational Biology, 12(6), 584-598.

Beerenwinkel, N., Rahnenführer, J., Kaiser, R., Hoffmann, D., Selbig, J., Lengauer, T. (2005b) Mtreemix: a software package for learning and using mixture models of mutagenetic trees, Bioinformatics, 21(9), 2106-2107.

Bogojeska, J., Rahnenführer, J., Lengauer, T. (2008) Stability analysis of mixtures of mutagenetic trees, BMC Bioinformatics, 9(1): 165.

Rahnenführer, J., Beerenwinkel, N., Schulz, W.A., Hartmann, C., Deimling, A.V., Wullich, B., Lengauer, T. (2005) Estimating cancer survival and clinical outcome based on genetic tumor progression scores, Bioinformatics, 21(10), 2438-2446.

Ketter, R., Urbschat, S., Henn, W., Feiden, W., Beerenwinkel, N., Lengauer, T., Steudel, W.-I., Zang, K.D., Rahnenführer, J. (2007) Application of oncogenetic trees mixtures as a biostatistical model of the clonal cytogenetic evolution of meningiomas, International Journal of Cancer, 121(7), 1473-1480.