Bayesian Clustering with Regression

Peter Mueller*, Fernando Quintana, and Gary Rosner

M.D. Anderson Cancer Center, Pontifica Univ Catolica de Chile

E-Mail: pm@wotan.mdacc.tmc.edu

Abstract: We propose a model for covariate-dependent clustering, i.e., we develop a probability model for random partitions that is indexed by covariates. The motivating application is inference for a clinical trial. As part of the desired inference we wish to define clusters of patients. Defining a prior probability model for cluster memberships should include a regression on patient baseline covariates. We build on product partition models (PPM). We define an extension of the PPM to include the desired regression. This is achieved by including in the cohesion function a new factor that increases the probability of experimental units with similar covariates to be included in the same cluster.

   We discuss implementations suitable for continuous, categorical, count and ordinal covariates.

1