COX PROPORTIONAL HAZARDS ANALYSIS FOR HAPLOTYPE-BASED ASSOCIATION ANALYSIS: THE STOCHASTIC-EM ALGORITHM

Author's name(s): Tregouet David-Alexandre, Tiret Laurence.
Affiliation(s): INSERM U525, 91 boulevard de l'Hôpital, 75634 Paris, France.
Email: tregouet@chups.jussieu.fr; tiret@chups.jussieu.fr;
Phone: 33 1 40 77 96 93; Fax: 33 1 40 77 97 28

Topic area: Statistical methods for association studies in human genetics
Key words: survival data, haplotypes

Abstract

It is now widely recognized that haplotype information inferred from genotypes can be of great interest to better characterize the role of a candidate gene in the etiology of a complex trait in the context of association studies. Haplotype-based analysis may help in differentiating the true effect of a polymorphism from what is due to its linkage disequilibrium with other variant(s). Haplotypes may serve as better markers for unknown functional variants than single polymorphisms. Lastly, they may define functional units whose effects cannot be predicted from what is known of the individual effect of each variant. Several works have recently advocated the simultaneous estimation of haplotypes frequencies and haplotypes effects in order to get better efficiency in parameter estimation (Niu et al., 2002; Tregouet et al., 2002; Epstein and Satten, 2003; Lake et al., 2003; Stram et al., 2003; Tanck et al., 2003; Tregouet et al., 2004). To our knowledge, most of the available models can deal with binary and/or quantitative phenotype analysis but none has yet discussed the application of haplotype-based association analysis to a survival outcome. We here described how the recently proposed Stochastic-EM (SEM) (Tregouet et al., 2004) algorithm can be extended to detect haplotypes effects in censored data analysis using a standard Cox proportional hazards formulation. This algorithm is a stochastic variant of the EM algorithm where the E-step is replaced by a stochastic step that consists in filling any ambiguous haplotypic pair by a single draw from its conditional distribution on the observed data and then to run a M step on the reconstructed pseudo-completed data sample. We showed how the partial likelihood formulation of the Cox model (Cox 1972) can be used at both steps of the SEM algorithm.
Application of this model to real data analysis will be given and will illustrate the interest of this algorithm to test for the association between haplotypes and a survival outcome in prospective studies.

References


