ABSTRACT

In many structured stochastic models, and specifically in genetic analysis, the statistics of primary importance for estimation and testing are functions of unobservable latent variables. Computation of likelihoods requires (often infeasible) summation over these latent variables, and the distributions of resulting likelihood-based statistics are unknown. Recently, the notions of fuzzy p-values and confidence levels have been introduced into the statistics literature as a way to describe the uncertainty inherent in a randomized test. In latent variable problems, our uncertainty is that inherent in the distribution of latent variables, given observed data. In this context, the natural definition of a fuzzy p-value is the distribution, given observed data, of that function of latent variables that would be the p-value were the latent variables observed.

These ideas are applied in the context of likelihood-based inference of genetic linkage of DNA affecting a trait to that controlling a set of DNA markers. We show how fuzzy p-values can summarize both the strength of evidence for linkage and the uncertainty about that evidence. The approach also provides a simple solution to the problem of providing a global significance level for the multiple dependent tests performed in testing for linkage. Realizations from the fuzzy p-value distribution may be obtained efficiently with only two sets of Monte Carlo realizations, one from the unconditional distribution of latent inheritance patterns, and the other conditional on observed marker data. No resimulation of data random variables is required, and the procedure is thus more robust to any assumed trait or marker model than many other current approaches.