Evidence evaluation for discrete data

Colin Aitken*,

The School of Mathematics and Maxwell Institute,

The University of Edinburgh, Edinburgh, UK

c.g.g.aitken@ed.ac.uk,

Erica Gold,

Department of Language and Linguistic Science,

The University of York, York, UK

erica.gold@york.ac.uk

Abstract

Methods for the evaluation of evidence in the form of measurements by way of the likelihood ratio are becoming more widespread. There are few methods for the evaluation of evidence in the form of discrete data (counts) by this way. Two suggestions of models where the responses are discrete data are described. The models are basic models, the exact situations for which will rarely occur in practice. However, they illustrate issues that need to be considered in the analysis of discrete data and provide a foundation on which other models may be built. Two levels of variation are assumed. The evidence is in the form of counts obtained in the investigation of a crime. There is evidence associated with the criminal and evidence associated with a suspect, who may or may not be the criminal. There is variation in these counts within subjects (the criminal and the suspect). There is also variation between subjects which is needed to help develop the likelihood ratio; models for this variation take the form of prior distributions on the parameters of the models used for the within-subject variation. The likelihood ratio compares the probability of the evidence if the criminal and the suspect are the same person with the probability of the

evidence if they are different people. The first suggestion assumes the responses (counts) within subjects are independent and follow a Poisson model. The expectation of the Poisson distribution has a gamma distribution between subjects. The second suggestion applies to situations when there is an underlying ordering in the responses with autocorrelation between adjacent responses. The responses are binary and a bivariate binomial model is considered. There are three parameters for the bivariate binomial model and beta distributions are assumed for their prior distributions. Examples of the performances of the models are illustrated with simulated data in the context of a problem in forensic phonetics. Performances are assessed through use of the likelihood ratio and its behaviour in association with changes in the responses associated with the criminal and the suspect. Prior parameter values can be interpreted in terms of personal beliefs about the underlying phenomena which the models attempt to represent. Variation in the results with variation in the choice of prior parameter values is investigated. There is discussion of the problems particular to the evaluation of evidence for discrete data, with suggestions for further work.