DISCRETE MULTIVARIATE ANALYSIS USING LOG—LINEAR MODELS

Robert J. Flowers*
Instituto de Ingenieria, UJAT.

ABSTRACT

A procedure is defined for obtaining maximum likelihood estimates for the parameters of a family of log—linear models which includes the general log—linear model, logit models, multinomial logit models, and association models as well as models for which maximum likelihood estimates have not been obtained. The desired model is defined by means of transformation matrices much as is done for the Grizzle—Starmer—Koch procedure. A major difference is that the observations are assumed to follow a Poisson distribution.

Key words: General log-linear models, logit models, association models, Newton—Raphson algorithm, Poisson distribution.

Resumen

Se define un procedimiento para obtener estimadores de máxima-verosimilitud para los parametros de una familia de modelos loglineales los que incluye el modelo loglineal general, modelos de logit, modelos de logit multinomial, modelos de asociación, y modelos para los que todavia no se hayan obtenido estimadores de máxima-verosimilitud. El modelo deseado se define por medio de matrices de transformación análogo al procedimiento de Grizzle Starmer-Koch. Una diferencia importante es que se supone que las observaciones siguen una distribución de Poisson.

Palabras clave: Modelo loglineal general, modelo de logit, modelo de asociación, algoritmo de Newton-Raphson, distribución de Poisson.

Introduction

A rather flexible methodology was developed by Grizzle, Starmer, and Koch (1969) for performing regression analyses of categorical data using either linear or log - linear models. One of the weaknesses of their procedure, which is commonly referred to as the GSK procedure, is that in the presence of zero cells the sample estimate of the variance is zero. This problem is usually handled by adding 1/2 to the zero cells or to all cells. Grizzle, Starmer, and Koch suggest adding $1/\gamma$ to each cell where γ is the total number of cells. These choices are somewhat arbitrary. A more suitable solution is to add 1/2 to each cell only to obtain initial estimates of the variance-covariance matrix. Cell estimates may be obtained using the weighted least squares estimates from which new estimates of the variance-covariance matrix are obtained. The variance-covariance matrix may be iteratively corrected until convergence is obtained. It will be shown here that by the use of a Newton-Raphson algorithm that a broad class of log-linear models can be defined for which maximum likelihood estimates may be obtained. The form of these models will be altered slightly from that of Grizzle, Starmer, and Koch. First, the transformations will be made on the cell frequencies rather than on the cell proportions. Second, the data will be assumed to follow a Poisson distribution rather than a multinomial distribution.

* The work for this article was done while the author was employed by Texas Transportation Institute, Texas A&M University.