



Statistical inference for familial disease models assuming exchangeability



Dale Bowman*

Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152, United States
The Pooled Resource Open-Access ALS Clinical Trials Consortium, United States

ARTICLE INFO

Article history:

Received 18 January 2016

Received in revised form 10 August 2016

Accepted 11 August 2016

Available online 20 August 2016

Keywords:

Exchangeable

Exchangeable binary

Exchangeable multinomial

ABSTRACT

Binary data often occur in epidemiology studies clustered by families. When an assumption of exchangeability of response is reasonable, the exchangeable multinomial distribution may be used to model binary and multinomial responses. Conditional probabilities of a new family member being diagnosed given family members previously diagnosed are derived. Applications are made to real familial disease data on IPF and ALS.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

In many epidemiologic studies there is interest in the way diseased cases occur within families. It is apparent that the disease status of family members will generally not be independent due to genetic and environmental factors common to family members. When only the number of affected per family is reported, information on individual covariates that could also affect response cannot be used in the subsequent analysis and it may be reasonable to assume that responses of family members are exchangeable (Yu and Zelterman, 2002). The lack of independence in exchangeable binary responses creates an obstacle when selecting an appropriate model. Yu and Zelterman (2008) argue that the best approach for representing sums of exchangeable binary responses may be to use generalized estimating equations (GEE's) (Lipsitz et al., 1991), as other models did not demonstrate a great reduction in deviance over the GEE's. Often, however, interest is in obtaining quantities from a probability model for use in assessing risk or in the development of predictive models. When this is the case, GEE's provide little help. Yu and Zelterman (2002) argue that exact probability methods should be used when modeling sums of exchangeable binary variables, as often the data is sparse and approximate methods may yield unreliable results.

In this paper, the exchangeable binary model of Bowman and George (1995) is used as a probability model for familial disease. This probability model is used to construct the conditional probability that a previously undiagnosed family member has positive disease status given the disease status of the other family members (that have been diagnosed). This model is extended to a situation where more than two disease classifications are possible. For example in a study of diabetes, it is possible to classify family members as having Type 1 diabetes, Type 2 diabetes, and no diabetes. Interest may then be in determining the conditional probability of an undiagnosed family member being diagnosed with Type 1 diabetes given the disease status of the remaining family members.

This paper is arranged as follows: Section 2 describes the exchangeable model with k possible disease responses. In the third section, maximum likelihood estimation of the parameters of the exchangeable multinomial distribution is discussed. Section 4 describes an application to data on interstitial pulmonary fibrosis (IPF) and to neurological data on amyotrophic

* Correspondence to: Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152, United States.
E-mail address: ddb Bowman@memphis.edu.

lateral sclerosis (ALS) and a simulation study is conducted to examine the properties of the estimates. Results are discussed in Section 5.

2. Model

Let random variables X_1, X_2, \dots, X_n be the disease response types of n family members. Each family member can have one of k different disease responses O_1, \dots, O_k . Assume that the disease status of i of the family members has been assessed and of the i members, there are R_j type O_j responses for $j = 1, \dots, k$ so that $\sum_{j=1}^k R_j = i$. Interest is in the probability of each response type for an unassessed family member given the responses of the assessed members, that is probabilities of the form

$$P(X_{i+1} = O_\ell | R_j = r_j \text{ type } O_j \text{ responses } j = 1, \dots, k) \tag{2.1}$$

for $\ell = 1, \dots, k$. In the simplest case where $k = 2$ the two disease response types could indicate presence or absence of the disease. More complex disease response types can be developed for $k > 2$.

2.1. Exchangeable assumption

Assume that the disease responses of family members are exchangeable in the sense that $P(X_{\pi_1} = x_1, \dots, X_{\pi_n} = x_n) = P(X_1 = x_1, \dots, X_n = x_n)$ for all permutations (π_1, \dots, π_n) of $(1, \dots, n)$. Define marginal probabilities $\tau_{r_1, r_2, \dots, r_{k-1}}$ as

$$P(X_1 = \dots = X_{r_1} = O_1, \dots, X_{r_1+r_2+\dots+r_{k-2}+1} = \dots = X_{r_1+r_2+\dots+r_{k-1}} = O_{k-1}) \tag{2.2}$$

where $\tau_{0, \dots, 0} = 1$. If we let r_j be the number of type O_j responses in the i assessed family members for $j = 1, \dots, k$ so that $\sum_{j=1}^k r_j = i$ then the joint probability of r_j type O_j responses for $j = 1, \dots, k$ is found as

$$\frac{n!}{\prod_{i=1}^k r_i!} \sum_{t=0}^{r_k} (-1)^t \binom{r_k}{t} \sum_{\ell_1+\dots+\ell_{k-1}=t} \binom{t}{\ell_1, \dots, \ell_{k-1}} \tau_{r_1+\ell_1, \dots, r_{k-1}+\ell_{k-1}} \tag{2.3}$$

see [George et al. \(2016\)](#). Using (2.3), the probability the $(i + 1)$ st assessed family member is of type O_j given the responses of the previously assessed family members is

$$\frac{\sum_{t=0}^{r_k} (-1)^t \binom{r_k}{t} \sum_{\ell_1+\dots+\ell_{k-1}=t} \binom{t}{\ell_1, \dots, \ell_{k-1}} \tau_{r_1+\ell_1, \dots, r_j+1+\ell_j, \dots, r_{k-1}+\ell_{k-1}}}{\sum_{t=0}^{r_k} (-1)^t \binom{r_k}{t} \sum_{\ell_1+\dots+\ell_{k-1}=t} \binom{t}{\ell_1, \dots, \ell_{k-1}} \tau_{r_1+\ell_1, \dots, r_{k-1}+\ell_{k-1}}} \tag{2.4}$$

From this formulation, the probability of an unassessed family member having a particular disease response type given the status of family members is seen to depend on the number of previously assessed family members and the number of family members having each response type.

2.2. A model with two disease response types

Let $X_i = 1$ if the i th family member has a disease and 0 otherwise. Then assuming exchangeability, X_1, \dots, X_n form a sequence of exchangeable binary random variables. Suppose that i of the family members have been assessed for the disease. Let $r = \sum_{j=1}^i X_j$, the number of assessed family members with positive disease status. Based only on an assumption of exchangeability of responses [Bowman and George \(1995\)](#) show that the probability $R = r$ is given by

$$P(R = r) = \binom{i}{r} \sum_{j=0}^{i-r} (-1)^j \binom{i-r}{j} \lambda_{r+j}, \tag{2.5}$$

where marginal response probabilities λ_r are equal to $P(X_1 = 1, \dots, X_r = 1)$. Then the desired probabilities of the $(i + 1)$ st family member having positive disease status ($X_{i+1} = 1$) or negative disease status ($X_{i+1} = 0$) conditioned on the number of family members with positive disease status are given by

$$f_{X_{i+1} | \mathbf{X}}(x_{i+1} | \mathbf{x}) = \begin{cases} \frac{\sum_{k=0}^{i-r} (-1)^k \binom{i-r}{k} \lambda_{r+1+k}}{\sum_{k=0}^{i-r} (-1)^k \binom{i-r}{k} \lambda_{r+k}} & x_{i+1} = 1 \\ \frac{\sum_{k=0}^{i+1-r} (-1)^k \binom{i+1-r}{k} \lambda_{r+k}}{\sum_{k=0}^{i-r} (-1)^k \binom{i-r}{k} \lambda_{r+k}} & x_{i+1} = 0. \end{cases} \tag{2.6}$$

Download English Version:

<https://daneshyari.com/en/article/1151480>

Download Persian Version:

<https://daneshyari.com/article/1151480>

[Daneshyari.com](https://daneshyari.com)