The Behavior of the Likelihood Ratio Test For Testing Missingness.

N. Hens[1], M. Aerts[1], G. Molenberghs[1], H. Thijs[1]

[1] Center for Statistics, Limburgs Universitair Centrum, Universitaire Campus, B-3590 Diepenbeek, Belgium
Presenting author: Niel Hens

Abstract: To asses the sensitivity of conclusions to model choices in the context of selection models for non-random dropout, one can oppose the different missing mechanisms to each other; e.g. by the likelihood ratio tests. The finite sample behavior of the null distribution and the power of the likelihood ratio test is studied under a variety of missingness mechanisms.

Keywords: Missing Data, Sensitivity Analysis, Likelihood Ratio Test, Missing Mechanisms.

1 Introduction

In a longitudinal setting, units are measured on several occasions. It is not unusual that a sequence is not fully observed, due to intermediate missingness and dropout. In the context of maximum likelihood inference, Rubin (1976) classified missing data into three types, namely missing completely at random, missing at random and missing not at random. Diggle and Kenward (1994) use a selection model to represent such a process. A selection model consists of two parts: a measurement part and a missingness process part. Such a model relies on strong and untestable assumptions. Not only the distributional assumptions can be misspecified but also the presence of missing data can have a large impact. In classical theory, the asymptotic distribution of the likelihood ratio test is a chisquare distribution with degrees of freedom equal to the difference in number of parameters. Careful considerations have to be made when using this result to test for missing not at random as shown by Rotnitzky et al (2000). We will first provide a motivating example from Rotnitzky et al (2000), then we will introduce selection models. In a simulation study, we will illustrate the finite sample behavior of the likelihood ratio test and we will conclude with some current research topics.

2 A Motivating Example

The following example is used in Rotnitzky et al (2000). Let $Y_1, \cdots, Y_n$ be a sample of n observations from a normal distribution with mean β and variance σ^2. Suppose there is missingness in this sample which is possibly related to the outcome itself. Let us denote this conditional probability by

$$P_c(y; \alpha_0, \alpha_1) = e^{H(\alpha_0 + \alpha_1 (y-\beta)/\sigma)}$$

where α_0 and α_1 are unknown parameters and $H(.)$ is a known function assumed to have its first three derivatives at α_0 non-zero. Interest goes out to test whether $\alpha_1 = 0$ which corresponds to missing completely at random. We thus consider two random variables (R, Y) where R is a binary indicator, which is 1 if Y is observed and 0 otherwise. The contribution of one individual to the loglikelihood is thus

$$\begin{aligned} & r[-\log\sigma - (y-\beta)^2/(2\sigma^2) + H\{\alpha_0 + \alpha_1(y-\beta)/\sigma\}] \\ + \quad & (1-r)[\log E\{1 - P_c(y; \alpha_0, \alpha_1)\}] \end{aligned}$$

For n individuals the loglikelihood $L_n(\beta, \sigma, \alpha_0, \alpha_1)$ is the sum of n such terms., If we have a look at the score vector at the null point $\beta, \sigma, \alpha_0, \alpha_1 = 0$ we obtain the following equations.

$$\begin{aligned} & r(y-\beta)/\sigma^2 \\ & r(-\sigma^2 + (y-\beta)^2)/\sigma^3 \\ & rH'(\alpha_0) - (1-r)\frac{H'(\alpha_0)e^{H(\alpha_0)}}{1 - e^{H(\alpha_0)}} \\ & rH'(\alpha_0)(y-\beta)/\sigma \end{aligned}$$

We can see that this score vector is degenerate at this particular parameter point. Equivalently, the information matrix calculated from expected second derivatives is singular at this parameter point.
Rotnitzky et al (2000) show that likelihood-based inference with a singular information matrix can have some consequences with respect to the distribution of the likelihood ratio test. Depending on the nature of the model either the asymptotic distribution can be a mixture of χ^2-distributions or the convergence rate is very slowly. Due to these demerits the application of the asymptotic distribution has to be considered with care. We will illustrate this behavior in the context of selection models by simulations.

3 Selection Models

Let us assume that for subject i, $i = 1, \cdots, N$, a sequence of responses Y_{ij} is measured at several occasions $j = 1, 2, \ldots, J$. Let R_{ij} be a missingness indicator and assume that y_{i1} is always observed. Then $r_{ij} = 0$ if y_{ij} is missing and $r_{ij} = 1$ if y_{ij} is observed. The measurement part of the model of Diggle and Kenward (1994) is given by

$$\mathbf{Y_i} = (Y_{i1}, \ldots, Y_{iJ}) \sim N(X_i\beta, \Sigma_i), \quad i = 1, \ldots, N,$$

where β is a vector of fixed effects, X_i is a matrix containing covariate values and Σ_i is a covariance matrix. The missingness process is described by

$$\text{logit}[Pr(R_{ij} = 1|y_{i,j-1}, y_{ij})] = \psi_0 + \psi_1 y_{i,j-1} + \psi_2 y_{ij},$$

where $Pr(R_{ij} = 1|y_{i,j-1}, y_{ij})$ is the probability for the i^{th} subject to drop out at time j. If ψ_2 differs from zero, the missingness process is non-random. Let us denote

$$g(\mathbf{h_{id}}, y_{id}) = Pr(R_{id} = 1|y_{i,d-1}, y_{id})$$

with d the time of dropout and $\mathbf{h_{id}} = (y_{i1}, \ldots, y_{i,d-1})$ the history of y_{id}, which we now restrict to depend on the previous measurement only. The total loglikelihood has the form

$$\ell = \sum_{i=1}^{N} (r_i \ell_i^c + (1 - r_i)\ell_i^i),$$

with ℓ_i^i the contribution for an incompleter

$$\ell_i^i = \ln f(\mathbf{h_{id}}) + \sum_{j=2}^{d_i - 1} \ln[1 - g(\mathbf{h_{ij}}, y_{ij})] + \ln \int f(y_{id}|\mathbf{h_{id}}) g(\mathbf{h_{id}}, y_{id})\, dy_{id}$$

and ℓ_i^c the contribution for a completer

$$\ell_i^c = \ln f(\mathbf{y_i}) + \sum_{j=2}^{J} \ln[1 - g(\mathbf{h_{ij}}, y_{ij})].$$

The likelihood ratio test statistic for testing MNAR versus MAR is then given by

$$G = -2[\ell_{MNAR} - \ell_{MAR}].$$

Due to the difference in only one parameter, the distribution of this statistic can be misleadingly expected to be $\chi^2(1)$. Based on this statistic Kenward (1998) and Molenberghs et al (2001) rejected the null hypothesis of missing at random on a value of 5.11, which corresponds to a P-value of 0.02 for their data example (Mastitis in dairy cattle). They compared this result with the Wald test (P-value of 0.002) and concluded that the asymptotic approximations are not very accurate. Rotnitzky et al. (2000) state that the regular assumptions of the likelihood ratio test statistic do not hold in this case due to the singular information matrix. In the next paragraph, we will illustrate the behavior of the likelihood ratio test statistic for the different missingness parameters in a simple setting.

4 Simulations

For this small simulation study 400 similar datasets were generated in 4 different settings. Each dataset consists of 200 subjects, each with two measurements generated from a bivariate normal distribution. Consider the following bivariate normal distribution, based on a compound symmetry covariance matrix:

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 4 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 2 \\ 2 & 4 \end{pmatrix}\right]. \tag{1}$$

The dropout process was generated according to the following model

$$\text{logit}[P(R_i = 1|Y_{i1}, Y_{i2})] = -2 + \psi_1 Y_{i1} + \psi_2 Y_{i2} \tag{2}$$

where ψ_1 and ψ_2 were chosen according to four different settings. In setting 1, the null hypothesis is $\psi_1 = 0$, given that $\psi_2 = 0$, while in setting 2 the null hypothesis is $\psi_1 = 0$, given that $\psi_2 \neq 0$. Setting 3 considers a test for $\psi_2 = 0$, given that $\psi_1 = 0$ and finally in setting 4 $\psi_2 = 0$ is tested, given that $\psi_1 \neq 0$. In the next table an overview of the different simulation settings is given.

	Data under H_0 with	
	$\psi_2 = 0$	$\psi_2 \neq 0$
$H_0 : \psi_1 = 0$	Setting 1	Setting 2
	$\psi_1 = 0$	$\psi_1 \neq 0$
$H_0 : \psi_2 = 0$	Setting 3	Setting 4

Figure 1 shows plots of the simulated null-distributions together with approximating χ^2-distribution.

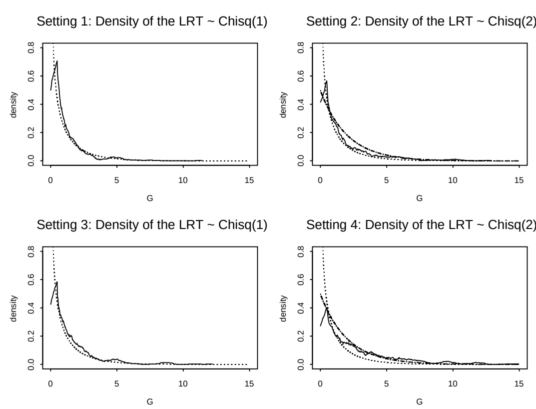

FIGURE 1. Density plots (dots) of the different settings with approximating χ^2-distribution (full line).

5 Discussion and Further Research

From the literature and the simulation settings, it is clear that the likelihood ratio test for testing missing not at random does not fulfill the regular assumptions. The use of classical asymptotic results might clearly lead to false results. A study of the theoretical asymptotical distribution and a power simulation study are topics of current research.

References

Diggle, P.J. and Kenward, M.G. (1994) Informative dropout in longitudinal data analysis (with discussion). *Applied Statistics*, **43**, 49–93.

Kenward, M.G. (1998) Selection models for repeated measurements with nonrandom dropout: an illustration of sensitivity. *Statistics in Medicine*, **17**, 2723-2732.

Little, R.J.A. (1976) Inference about means from incomplete multivariate data. Biometrika, **63**, 593–604.

Molenberghs, G., Verbeke, G., Thijs, T., Lesaffre, E. and Kenward, M.G. (2001) Influence analysis to assess sensitivity of the dropout process. *Computational Statistics and Data Analysis*, **37**, 93–113.

Rotnitzky A., Cox D.R., Bottai M., Robins J. (2000) Likelihood-based inference with singular information matrix. *Bernouilli*, **6**(2), 243–284.