Accepted Manuscript

Identification problem of transition models for repeated measurement data with nonignorable missing values

Kosuke Morikawa, Yutaka Kano

PII: S0047-259X(17)30131-8
DOI: https://doi.org/10.1016/j.jmva.2017.12.007
Reference: YJMVA 4316

To appear in: *Journal of Multivariate Analysis*

Received date : 6 March 2017

Please cite this article as: K. Morikawa, Y. Kano, Identification problem of transition models for repeated measurement data with nonignorable missing values, *Journal of Multivariate Analysis* (2017), https://doi.org/10.1016/j.jmva.2017.12.007

Identification problem of transition models for repeated measurement data with nonignorable missing values

Kosuke Morikawa[a,*], Yutaka Kano[a]

[a]Graduate School of Engineering Science, Osaka University, Osaka 560-8531, Japan

Abstract

In this paper, we consider a transition model on a response variable to describe repeated measurement data and we provide sufficient conditions to check model identifiability when analyzing data with nonignorable missing values. The sufficient conditions can give us intuitive model characteristics to achieve identifiability. In addition to the model assumptions on the response variable, a parametric model of the missing-data mechanism is often assumed. In this article, we consider identifiability in two situations: (i) both the response variable distribution and the missing-data mechanism are parametric; (ii) one of them is nonparametric, i.e., the global model is semiparametric. Useful identifiable models are proposed on the basis of these conditions. We also present an application to data of a comparative trial of two dosages of depot medroxyprogesterone acetate.

Keywords: Drop-out, identifiability, incomplete data, not missing at random, repeated measurement data, transition model.

1. Introduction

In clinical studies, researchers generally hope to obtain complete data but this does not always happen. In fact, subjects are typically told that they can drop out anytime they want, in accordance with human participant protection protocols. There are thus many drop-outs in some experiments. From a statistical perspective, missing values complicate the data analysis, because ignoring the missing-data mechanism can lead to inappropriate inference.

For example, Machin et al. [7] report results of a comparative trial of two dosages of depot medroxyprogesterone acetate (DMPA, 100 mg and 150 mg) where over 40% of subjects are missing at the endpoint. In the DMPA trial test, 1151 female subjects were divided into two dosages randomly and took DMPA in every quarter, over one year. They reported the results of DMPA as binary data: if a subject experienced amenorrhea, i.e., absence of a menstrual period in a woman of reproductive age, it was coded as 1; otherwise, it was coded 0. Whether a subject experienced amenorrhea or not was based on her menstrual diary. Each subject thus generated a sequence according to whether or not she experienced amenorrhea in the successive reference periods. The number of women with each sequence is shown in Table 1, where "×" means missingness. For example, "0 1 × ×" means amenorrhea is absent in first period, but present the next period, and the data were missing for the third and fourth periods. These data have been analyzed by several authors using various approaches; see, e.g., [2, 8, 16]. We will analyze the same data with a new approach.

In the analysis of repeated measure data, serial correlations among response variables in $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iT})$ may not be ignored and any statistical model for $\boldsymbol{Y}_i$ has to take the correlations into account. Serial correlation is typically incorporated into the model either via a conditional or a marginal approach. Conditional models describe the serial correlation by modeling Y_t, which is the response at time t, given not only covariates X but also $Y_1, \ldots, Y_{t-1}$, which are responses recorded earlier than time t. The approach is intuitive and simple, and the serial correlations are obtained easily from the conditional model [12].

*Corresponding author.
Email addresses: morikawa@sigmath.es.osaka-u.ac.jp (Kosuke Morikawa), kano@sigmath.es.osaka-u.ac.jp (Yutaka Kano)