



Bayesian non-linear regression models with skew-elliptical errors: Applications to the classification of longitudinal profiles

Rolando De la Cruz

Departamento de Salud Pública, Escuela de Medicina, Pontificia Universidad Católica de Chile, Marcoleta 434, Casilla 114D, Santiago, Chile

ARTICLE INFO

Article history:

Received 28 November 2007

Received in revised form 13 June 2008

Accepted 12 August 2008

Available online 28 August 2008

ABSTRACT

Typically, the fundamental assumption in non-linear regression models is the normality of the errors. Even though this model offers great flexibility for modeling these effects, it suffers from the same lack of robustness against departures from distributional assumptions as other statistical models based on the Gaussian distribution. It is of practical interest, therefore, to study non-linear models which are less sensitive to departures from normality, as well as related assumptions. Thus the current methods proposed for linear regression models need to be extended to non-linear regression models. This paper discusses non-linear regression models for longitudinal data with errors that follow a skew-elliptical distribution. Additionally, we discuss Bayesian statistical methods for the classification of observations into two or more groups based on skew-models for non-linear longitudinal profiles. Parameter estimation for a discriminant model that classifies individuals into distinct predefined groups or populations uses appropriate posterior simulation schemes. The methods are illustrated with data from a study involving 173 pregnant women. The main objective in this study is to predict normal versus abnormal pregnancy outcomes from beta human chorionic gonadotropin data available at early stages of pregnancy.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

In assisted reproduction programs, once a pregnancy has been achieved, the power of predicting its evolution is important for the medical team as well as for the patient. The determination of pregnanediol or of placental protein 14 levels has been postulated as a predictive factor; however, the determination of the human chorionic gonadotropin is, seemingly, the most adequate parameter for the detection of early developmental alterations during pregnancy.

Human chorionic gonadotropin (HCG) is a peptide hormone produced in pregnancy which is made by the embryo soon after conception and, later, by the syncytiotrophoblast (part of the placenta). Its role is to prevent the disintegration of the corpus luteum of the ovary and thereby maintain progesterone production that is critical for pregnancy in humans. HCG may have additional functions, for instance, it is thought that it affects the immune tolerance of the pregnancy. Early pregnancy testing is generally based on the detection or measurement of HCG.

In obstetrics, it is well known that HCG is one of the clinical variables which shows dramatic changes in women during pregnancy. It has also been established that values of the HCG are different in women who have normal pregnancies with terminal deliveries from that in women who have miscarriages or other types of adverse pregnancy outcomes.

In a period of two years, 173 consecutive pregnancies at the private clinic in vitro fertilization unit were analyzed by determining the sub-unit beta human chorionic gonadotropin (β -HCG) concentrations between two and 12 weeks after oocyte collection. Consequently, pregnancy outcomes were divided into two groups: normal and abnormal. The women

E-mail address: rolando@med.puc.cl.

were classified as having undergone normal pregnancies if they had a normal delivery, or abnormal pregnancies if they had any complication resulting in a non-terminal delivery and loss of the fetus. Levels of the β -HCG for these 173 women range from 3 to 120,000 mIU/ml (milli International Units per milli-liter). Exploratory plots (not shown) show a high frequency in the interval 0–20,000 mIU/ml, for this reason usually this data are modeled in the log scale, to obtain normality and a more reasonable range scale. However, initial exploratory plots (not shown) confirm that in both groups the left tail of the underlying distribution descends more slowly than the right tail. Thus the skew-regression models will estimate and test for the skewness in the data more formally.

It is well known that skewness and heavy tails are often present in the many datasets. The standard assumption of the multivariate normality for the error term in multivariate regression in many situations does not agree with the real data. The most common approach adopted to solve this disagreement is transformation of the variable. However, there are many problems with the arbitrary choice of the transformation, especially under multivariate data. Thus in the cases where the assumption of normality is not tenable, more flexible models can be adopted to accommodate skewness and heavy tails. The recent literature has a tendency to propose a more flexible multivariate distribution for the error that can be more realistic (see, for example, Azzalini and Capitanio (1999), DiCiccio and Monti (2004) and Genton (2004)). In this direction, Sahu et al. (2003) propose a new class of distributions by introducing skewness in multivariate elliptically symmetric distributions. The class contains many standard families including the multivariate skew-normal and the skew- t distributions. They give practical applications in linear regression models.

Also, other authors have studied regression models with skew distributions. Fernández and Steel (1998, 2000) treat Bayesian modeling of fat tails and skewness and issues of existence of the posterior (moments) in linear regression. Branco and Dey (2002) considered a linear regression model under a skewed heavy tailed error distribution. Ferreira and Steel (2007b) introduce a novel method for the generation of a class of multivariate skewed distributions with applications in regression analysis, and applied this class to firm size distributions in Ferreira and Steel (2004), while Ferreira and Steel (2007a) explores the practical comparison of various ways of modeling multivariate skewness through posterior odds, using prior matching. A more general constructive representation of classes of skewed distributions is provided in Branco and Dey (2001) and Ferreira and Steel (2006). For a detailed account of skew models see Genton (2004) and the references therein.

It is well known that discriminant analysis is an important tool in statistics, but until recently it could not be applied to longitudinal data, and in the other hand, longitudinal studies are increasingly more often in medical research. The works developed in this area assume normal distributions in the models. From this point of view, we believe that the research to develop statistical tools with nonstandard assumptions to analyze this type of data is a significant contribution to the field. Recently, some authors have studied discriminant analysis with longitudinal data. These works considered linear and non-linear models to describe the longitudinal profiles in each group (see Tomasko et al. (1999), Brown et al. (2000), Marshall and Barón (2000), Brant et al. (2003), Wernecke et al. (2004), De la Cruz-Mesía and Quintana (2007) and De la Cruz-Mesía et al. (2007)). Extensions to the case of multiple responses have been considered by Marshall et al. (2008).

In this paper we develop non-linear regression models for longitudinal data with skew-elliptical distributions. We confine ourselves here to the multivariate skew-elliptical distributions, as presented in Sahu et al. (2003). From this family, we studied the multivariate skew-normal and skew- t distributions. This last distribution is flexible enough to accommodate skewness and heavy tails. Although, there are others who study skewness of the normal and Student- t distributions in the literature (see Azzalini and Dalla-Valle (1996), Fernández and Steel (1998, 1999), Azzalini and Capitanio (1999, 2003) and Jones and Faddy (2003), among others). The advantage of this distribution is that it is constructed to derive an easy implementation of the Bayesian inference method. The principal application of these models will be in the context of discriminant analysis with biomedical data, that is, the classification of subjects into one of two or more groups based on longitudinal markers using skew-elliptical non-linear regression models.

In our example the main objective is to explore a classification technique for predicting the outcome of pregnancy on the basis of longitudinal β -HCG measurements. Our approach is fully Bayesian and provides the posterior (or predictive) probability of the outcome of pregnancy in women based on the longitudinal marker. Medical team can then make decisions on the basis of these probabilities.

The paper is organized as follows: Section 2 contains brief descriptions of the definitions of the elliptical and skew-elliptical distributions. In Section 3, we introduce the skew-elliptical non-linear regression model, and an appropriate posterior simulation scheme based on the Gibbs sampling algorithm is described. Section 4 describes the statistical methodology for the classification of longitudinal data using skew-elliptical non-linear regression models. Section 5 investigates the use of these flexible distributions, applying them to simulated data. Section 6 illustrates the proposed longitudinal method using data from Santiago, Chile on the β -HCG measured in women with normal and abnormal pregnancy outcomes. Finally, Section 7 discusses the results.

2. Skew-elliptical distributions

Before presenting the skew-elliptical non-linear regression model, it is useful to recall the definition of the elliptical and skew-elliptical distributions, and define the skew distributions which will be the focus of our attention in this paper.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات