

Bayesian Sample size Determination for Longitudinal Studies with Continuous Response using Marginal Models

Taban Baghfalaki;

Department of Statistics, Faculty of Mathematical Sciences,
Tarbiat Modares University, Tehran, Iran.

Received: 26 Aug 2015

Revised: 17 Sep 2017

Extended Abstract

Paper pages (1-20)

Introduction

Longitudinal study designs are common in a lot of scientific researches, especially in medical, social and economic sciences. The reason is that longitudinal studies allow researchers to measure changes of each individual over time and often have higher statistical power than cross-sectional studies. Choosing an appropriate sample size is a crucial step in a successful study. A study with insufficient sample size may have a small statistical power to detect significant effects and may lead to incorrect answers to many important research questions. On the other hand, a study with more than what a sample size should be wastes the resources and budget. In longitudinal studies, because of the complexity of the design of experiment, including the selection of the number of individuals and the number of repeated measurements, the sample size determination is less studied. This paper uses a simulation-based method to determine the sample size from a Bayesian perspective. For this purpose several Bayesian criteria of sample size determination for a longitudinal study using marginal model are used. Most of the methods of determining the sample size are based on creation of one hypothesis. In this paper, in addition to using this method, we also present a method to determine the sample size for multiple hypothesis testing. Using several examples the proposed Bayesian methods are discussed.

Material and method

Based on the Bayesian sample size determination approach proposed by Wang and Gelfand (2002), which is a simulation-based approach, two sets of priors are considered. The first set is called the “fitting” or “analysis” priors which are used for analyzing data. The other set of priors is called “design” or “sampling” sets which drawing upon expertise, we may speculate upon a variety of informative scenarios regarding the unknown parameters and capture each with a suitable sampling prior. In the first step of the simulation-based approach, one generates the parameters from the design priors then a sample from the generated parameters is obtained. Note that the frequentist/classical viewpoint of the sample size determination approach require a point estimate of the variance and the smallest meaningful difference (Diggle et al., 2002). This information is elicited based on pilot data or the opinion of experts. The use of design priors in the proposed approach of Wang and Gelfand (2002) replaces this part of the classical sample size determination approach. After generating data using the design priors in this stage, they are analyzed by the fitting priors. In this paper, non-informative priors are used as fitting priors, as we assume there is no prior information or any expert opinion to be used to construct a prior

distribution. If there is any prior information to be used to elicit an appropriate prior distribution, one may use it in the step of fitting prior and for sure, if the specification of the prior is correct, will achieve a better result. Also, in this paper, four criteria are used for determining sample size in longitudinal studies: Bayesian power criterion (BPC), average length criterion (ALC), average posterior variance criterion (APVC) and average converge criterion (ACC).

The simulation-based approach of Wang and Gelfand (2002) can be summarized as follows:

1. Specify the value of the effect of an important regression coefficient which is the interest parameter and specify the design and fitting priors.
2. For each sample size, the following steps are repeated M times:
 - i. Generate values of the unknown parameters from their design priors.
 - ii. Simulate the values of covariates from continuous or discrete distributions and the response variable from its distribution.
 - iii. Analyze the generated data set of step (ii) using the fitting priors.
 - iv. Calculate BPC.
3. Fit a curve or surface through Bayesian power values and find an adequate sample size for a desired power using interpolation. In this paper, the curve is fitted using a polynomial regression.

In step 2(iv), one can calculate ACC, APVC, or ALC and determine the sample size based on these criteria.

References

1. Diggle, P., Heagerty, P., Liang, K. and Zeger, S. (2002). *Analysis of Longitudinal Data*. Oxford:Oxford University Press.
2. Wang, F., and Gelfand, A. E. (2002). A simulation-based approach to Bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*, 193-208.

Keywords: Bayesian analysis, longitudinal data, marginal model, posterior distribution, Bayesian power criterion

*Corresponding author: t.baghfalaki@modares.ac.ir