Communications
in
Mathematical
Research
25(2)(2009), 123–136

Semi-empiricial Likelihood Confidence Intervals for the Differences of Two Populations Based on Fractional Imputation*

Bai Yun-xia[1], Qin Yong-song[2], Wang Li-rong[3] and Li Ling[2]
(*1. Pharmacy Department, Baotou Medical College, Baotou, Neimenggu, 014040*)
(*2. College of Mathematical Sciences, Guangxi Normal University, Guilin, Guangxi, 541004*)
(*3. Loudi Technician College, Loudi, Hunan, 417000*)

Communicated by Guo Jian-hua

Abstract: Suppose that there are two populations x and y with missing data on both of them, where x has a distribution function $F(\cdot)$ which is unknown and y has a distribution function $G_\theta(\cdot)$ with a probability density function $g_\theta(\cdot)$ with known form depending on some unknown parameter θ. Fractional imputation is used to fill in missing data. The asymptotic distributions of the semi-empirical likelihood ration statistic are obtained under some mild conditions. Then, empirical likelihood confidence intervals on the differences of x and y are constructed.

Key words: empirical likelihood, confidence intervals, fractional imputation, missing data

2000 MR subject classification: 62G05, 62E20

Document code: A

Article ID: 1674-5647(2009)02-0123-14

1 Introduction

Missing datas are common in opinion polls, market research surveys, medical studies and other scientific experiments. In this situation, the usual inference procedure cannot be applied directly. A common method for handling incomplete data is to impute a value for each missing variables and then apply usual statistical methods to the "complete data" as if they were true observations. Missing data analysis covers a variety of problems that are often seen in practical applications (see [1]). Owen[2]–[5] first put forward the technique of empirical likelihood in nonparametric statistics. Recently, Wang and Rao[6]–[7] use a empirical likelihood method to construct confidence intervals for the response means in linear

*Received date: Dec. 13, 2007.
Foundation item: The NSF (10661003) of China, SRF for ROCS, SEM ([2004]527), the NSF (0728092) of Guangxi, and Innovation Project of Guangxi Graduate Education ([2006]40).

and nonparametric models. In this paper, we focus on constructing confidence intervals for various differences of two populations x and y which have distribution functions $F(\cdot)$ and $G_\theta(\cdot)$, where $F(\cdot)$ is unknown and $G_\theta(\cdot)$ with probability density function $g_\theta(\cdot)$ is of known form depending on some unknown parameter θ. Let Δ denote the differences of x and y such as the differences of the means and the distribution functions of two populations. This model is often seen in practical applications. For example, doctors intend to use a new medicine A to cure a specified illness. B is also used to treat the disease. B is known very well and A is less known by doctors. Many clinic experiments should be done to obtain the data of curative effects between A and B. We are interested in comparing noticeable differences between A and B. We take B and A as x and y, respectively. Thus, we want to know the differences Δ of x and y, such as

$$\Delta = Ex - Ey, \qquad \Delta = P(x \le y), \qquad \Delta = F(x_0) - G_{\theta_0}(x_0) \ \ (x_0 \text{ is fixed})$$

and so on. In this paper, we suppose the following information is available

$$E_F\omega(x, \theta, \Delta) = 0, \tag{1.1}$$

where ω is a function of known form.

In this paper, we use the information (1.1) to construct empirical likelihood confidence intervals on Δ. Thus, we can test the hypotheses on the differences Δ of x and y. We suppose that the hypothesis is

$$\mathrm{H}_0: \quad \Delta = \Delta_0 \text{ for some known } \Delta_0.$$

If we wish to test the hypothesis that there is no noticeable differences between x and y, we let $\Delta_0 = 0$. If Δ_0 is in the above interval, we accept the hypothesis; otherwise, the hypothesis should be rejected.

Consider the following simple random samples of incomplete data associated with populations (x, δ_x) and (y, δ_y),

$$(x_i,\ \delta_{x_i}), \quad i = 1, \cdots, m, \qquad (y_j,\ \delta_{y_j}), \quad j = 1, \cdots, n,$$

where

$$\delta_{x_i} = \begin{cases} 0, & \text{if } x_i \text{ is missing;} \\ 1, & \text{else,} \end{cases}$$

$$\delta_{y_j} = \begin{cases} 0, & \text{if } y_j \text{ is missing;} \\ 1, & \text{else.} \end{cases}$$

Suppose that x and y are missing completely at random (MCAR)

$$P(\delta_{x_i} = 1|x) = P(\delta_{x_i} = 1) = P_1(\text{constant}),$$
$$P(\delta_{y_j} = 1|y) = P(\delta_{y_j} = 1) = P_2(\text{constant}).$$

We also assume that (x, δ_{x_i}) and (y, δ_{y_j}) are independent. Let

$$r_x = \sum_{i=1}^m \delta_{x_i}, \qquad m_x = m - r_x, \qquad r_y = \sum_{i=1}^n \delta_{y_j}, \qquad m_y = n - r_y,$$

$$s_{r_x} = \{i : \delta_{x_i} = 1, i = 1, \cdots, m\}, \qquad s_{m_x} = \{i : \delta_{x_i} = 0, i = 1, \cdots, m\},$$

$$s_{r_j} = \{j : \delta_{y_j} = 1, j = 1, \cdots, n\}, \qquad s_{m_y} = \{j : \delta_{y_j} = 0, j = 1, \cdots, n\}.$$