Computational Statistics and Data Analysis 117 (2018) 19–31

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Simultaneous estimation based on empirical likelihood and general maximum likelihood estimation

Junyong Park

Department of Mathematics and Statistics, University of Maryland Baltimore County, MD 21250, USA

ARTICLE INFO

Article history:
Received 2 January 2017
Received in revised form 1 August 2017
Accepted 2 August 2017
Available online 9 August 2017

Keywords:
Empirical likelihood
General maximum likelihood
Estimation of mean vector

ABSTRACT

One typical problem in simultaneous estimation of mean values is estimating means of normal distributions, however when normality or any other distribution is not specified, more robust estimation procedures are demanded. A new estimation procedure is proposed based on empirical likelihood which does not request any specific distributional assumption. The new idea is based on incorporating empirical likelihood with general maximum likelihood estimation. One well-known nonparametric estimator, the linear empirical Bayes estimator, can be interpreted as an estimator based on empirical likelihood under some framework and it is shown that the proposed procedure can improve the linear empirical Bayes estimator. Numerical studies are presented to compare the proposed estimator with some existing estimators. The proposed estimator is applied to the problem of estimating mean values corresponding to high valued observations. Simulations and real data example of gene expression are provided.

1. Introduction

When there is $(X_1, \ldots, X_n)$ where $X_i \sim f(x|\theta_i)$, simultaneous estimation of parameter vector $(\theta_1, \ldots, \theta_n)$ has been considered by many researchers through the compound decision theory. θ_is are deterministic sequences and one main interest is to obtain an estimate of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)$, say $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)$, with small risk

$$R(\boldsymbol{\theta}) = \sum_{i=1}^{k} E(\hat{\theta}_i - \theta_i)^2 \tag{1}$$

based on an observational vector $(X_1, \ldots, X_n)$. This is known as a compound decision problem.

In the framework of empirical Bayes estimation, (X_i, θ_i)'s are independent and identical samples where $X_i|\theta_i \sim f(x|\theta_i)$ and $\theta_i \sim G(\theta)$ for an unknown G. For any estimator δ, the Bayes rule (say $\delta^*(X)$) is an estimator minimizing $E(\delta(X) - \theta)^2$ which is obtained from $\delta^*(x) = \arg\min_\delta E[(\delta(X) - \theta)^2 | X = x]$. The compound decision problem can be considered as a specific case of the empirical Bayes estimation if $G(\theta) = \frac{1}{n}\sum_{i=1}^{n} I(\theta_i \le \theta)$ is an empirical distribution of $(\theta_1, \ldots, \theta_n)$.

There are typical examples such as estimating mean vector for normal distributions and that of Poisson distributions. In estimation of mean vector, there are numerous literatures on the mean vector estimation when the distribution of x given θ, $f(x|\theta)$, has the form of parametric distribution. More specifically, $X_i \sim f(x|\theta_i)$ for $1 \le i \le n$ and estimate $(\theta_1, \ldots, \theta_n)$ based on $(X_1, \ldots, X_n)$ under square loss function. Under the normality of X_i given θ_i, Johnstone and Silverman (2004) used the empirical Bayes estimator for specific prior distributions on θ_is, Brown and Greenshtein (2009) used the idea of smoothing in the nonparametric empirical Bayes estimator, Jiang and Zhang (2009) showed some optimality of estimator for the

E-mail address: junpark@umbc.edu.

http://dx.doi.org/10.1016/j.csda.2017.08.003

Bayes estimator based on general maximum likelihood estimator (GMLE) and Park (2014) used the shrinkage estimator from conditional maximum likelihood estimation under normality. All these methods assume that data are generated from $f(x|\theta) = \phi(x-\theta)$ where ϕ is the density function of a standard normal distribution. Additionally, there are some nonparametric estimators which do not request any specific form of $f(x|\theta)$ such as hard or soft shrinkage estimators in Donoho and Johnstone (1994, 1995). Furthermore, Robbins (1983, 1985) proposed the linear empirical Bayes (LEB) estimator which depends on some moment conditions. Under the situation that the conditional variance is constant, the LEB estimator is the variant of the James–Stein's estimator in James and Stein (1961). All these nonparametric approaches do not request any specific form of $f(x|\theta)$, however their performances are very sensitive to the configuration of $(\theta_1, \ldots, \theta_n)$ in compound decision problem and G in the empirical Bayes problem. For example, the hard and soft shrinkage estimators are efficient when there are some structural assumption on mean vector such as sparsity. Under the Bayesian framework, the LEB estimator is derived when the prior G is assumed to be the normal distribution. This may lead to poor performances of the LEB estimator when the true G has more than one mode. This phenomenon will be demonstrated later through simulations.

All existing methods either assume $f(x|\theta)$ is known or suffer from the sensitivity to the structure of θs (or G). To avoid these, we propose to use the empirical likelihood (EL) estimation which does not request the specific form of $f(x|\theta)$ and use the General maximum likelihood estimation (GMLE) to estimate G. In the context of the empirical likelihood estimation, it is commonly assumed that X_i's are i.i.d. samples from some unknown distribution and one main interest is to estimate a common mean, say θ_0, and then test the hypothesis of $H_0 : \theta_0 = 0$. Our problem is different from the classical empirical likelihood estimation problem in the sense that X_i's are independent, but not identical since θ_is can be different in our setting.

Throughout this paper, we propose a new method which assume that none of $f(x|\theta)$ and G are known. Our proposed estimator is based on the quadratic optimization with moment constraints. We shall show that the linear empirical Bayes (LEB) estimator is understood under our framework, however the LEB estimator is actually obtained without considering all given constraints. We present theoretical relationship between our estimator and the LEB estimator and show advantage of our proposed estimator over existing estimators through numerical studies.

This paper is organized as follows. In Section 2, we provide a brief review on the estimation of mean vector estimation and propose a new approach to estimate mean vector based on empirical likelihood and general maximum likelihood estimation. We also provide some relationship between our proposed procedure and some existing estimator. We present numerical studies for comparison in Section 3. In Section 4, we apply our idea to estimate the mean values corresponding to high valued observations and Section 5 shows a real example of gene expression data. Some concluding remarks are presented in Section 6.

2. Estimator based on empirical likelihood and general maximum likelihood

Suppose we have $(X_1, \ldots, X_n)$ and $X_i \sim f(x|\theta_i)$ for a given θ_i where $\theta_i \sim G(\cdot)$. Here, both $f(x|\theta)$ and G are unknown in our context while most of estimation problems are based on known $f(x|\theta)$ except nonparametric estimators such as the LEB estimator and hard/soft estimators. Our goal is to estimate $\theta_i = E(X_i|\theta_i)$ with only moment conditions such as $Var(X_i|\theta_i) = \sigma^2$ for some known σ^2 instead of assumption of known $f(x|\theta)$. Without loss of generality, we can assume $\sigma = 1$. Since we assume that $f(x|\theta)$ is unknown, our procedure will cover a broad class of $f(x|\theta)$. Typically, θ is the location parameter leading to $f(x|\theta) = q(x-\theta)$ for some density q which is usually generated from an additive structure $X_i = \theta_i + \epsilon_i$ for $\epsilon_i \sim q(\epsilon)$. However, the additive structure is not required in our estimation problem.

We assume that both $f(x|\theta)$ and G are unknown, so our goal is to develop an estimator which does not require specific form of $f(x|\theta)$ and G. We propose an estimator which is more robust than estimators using specific form of $f(x|\theta)$ or G.

To accomplish our goal to develop a new type of estimator, we approximate the optimal Bayes estimator which will be defined below. The optimal Bayes rule requests information on $f(x|\theta)$ and G, however $f(x|\theta)$ and G are assumed to be unknown in our problem, so our main idea is to approximate $f(x|\theta)$ based on the empirical likelihood and approximate G based on general maximum likelihood estimation (GMLE).

First, we briefly review the Bayes estimator under square loss function. Under the square loss function, the Bayes estimator denoted by $t_G(x) = \text{argmin}_\delta E((\delta-\theta)^2|x)$ is

$$t_G(x) = E(\theta|x) = \frac{\int \theta f(x|\theta)dG(\theta)}{\int f(x|\theta)dG(\theta)}. \tag{2}$$

In particular, when $f(x|\theta)$ is either normal or Poisson distribution, the Bayes estimator depends only on the marginal distribution of x, $f(x) = \int f(x|\theta)dG(\theta)$. See Brown and Greenshtein (2009) for normal distribution and Park (2012) and Brown et al. (2013) for Poisson distribution. Ideally, when both $f(x|\theta)$ and G are known, the Bayes estimator obtains the optimal risk, say $R^*(G) = E(\theta - t_G(X))^2$. For any estimator $t(x)$, we actually have $E(\theta - t(X))^2 = R^*(G) + E(t_G(X) - t(X))^2$. Therefore, if either $f(x|\theta)$ or G is misspecified, $t(x)$ may be different from $t_G(x)$ which can lead to $E(t_G(X) - t(X))^2 > 0$. In most of the literatures on empirical Bayes estimation, it is assumed that $f(x|\theta)$ is known and only G is unknown. On the other hand, some nonparametric estimators do not request any information on $f(x|\theta)$ and G, however their performances are very sensitive to the structure of G. To avoid two difficulties from assuming $f(x|\theta)$ or the structure of G, we propose an estimator based on both the empirical likelihood (EL) estimation and GMLE.

In the following subsections, we introduce our proposed estimator based on EL estimation and GMLE. We also demonstrate that the linear empirical likelihood (LEB) estimator in Robbins (1983) can be understood under our framework and present some relationship between the LEB estimator and our proposed estimator.