Computational Statistics and Data Analysis 90 (2015) 47–60

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Robust and efficient estimation of effective dose

Rohana J. Karunamuni [a,*], Qingguo Tang [b], Bangxin Zhao [c]

[a] Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada T6G 2G1
[b] School of Economics and Management, Nanjing University of Science and Technology, Nanjing 210094, China
[c] Department of Statistical and Actuarial Sciences, Western University, London, Ontario, Canada N6A 5B7

ARTICLE INFO

Article history:
Received 24 April 2014
Received in revised form 9 March 2015
Accepted 6 April 2015
Available online 27 April 2015

Keywords:
Dose–response curve
Effective dose
Maximum likelihood
Weighted least squares
Minimum distance methods

ABSTRACT

In dose–response studies, experimenters are often interested in estimating the effective dose ED_p, the dose at which the probability of response is p, $0 < p < 1$. For instance, in pharmacology studies one is typically interested in estimating $ED_{0.5}$, whereas in toxicology studies the main interest is ED_p for smaller values of p. In this context, methods based on parametric, semiparametric, and nonparametric models have been developed. Traditional estimators based on parametric models are generally efficient but are not robust to model misspecification. On the other hand, nonparametric estimators are robust to model misspecification but are less efficient. Semiparametric methods are a compromise. Two new parametric methods are presented in this paper for estimating ED_p using minimum-distance techniques. It is shown that the proposed estimators are efficient under the model and simultaneously have some desirable robustness properties. The asymptotic properties such as consistency and asymptotic normality are studied. Small-sample and robustness properties of the proposed estimators are examined and are compared with traditional estimators using Monte Carlo studies. Two real-data examples are also analyzed.

1. Introduction

Dose–response is a relationship that describes the changes occurring in an organism when it is exposed to different levels (doses) of a chemical compound or drug (an agent). Dose–response experiments are routinely conducted in pharmacology and toxicology, and the main interest is to study the relationship between the exposure levels of the agents and the responses obtained. In a standard dose–response experiment, the study subjects are randomized into several subgroups. The outcome of interest is usually measured at several increasing dose levels, denoted as x_j ($j = 1, 2, \ldots, K$, i.e., K different increasing dosages). In each subgroup, the number of individuals who show a response is observed. At a given dose x, one typically assumes that the response Y is a Bernoulli random variable with probability of "success" being $\eta(x)$, i.e., $\Pr(Y = 1|x) = \eta(x)$. The statistical problem concerns the estimation of "effective dose" levels, defined as $ED_p = \eta^{-1}(p)$ with $0 < p < 1$, where $\eta^{-1}(.)$ is the inverse function of $\eta(.)$. Note that ED_p can be interpreted as the dose at which the probability of response is p. For example, if $p = 0.5$, then $ED_{0.5}$ is the dose that produces a desired effect in half of the test population. Pharmacology studies typically focus on estimating $ED_{0.5}$, whereas in toxicology studies the main interest is estimating ED_p for smaller values of p (Yuan and Yin, 2011). The importance of estimating extreme percentage points, such as $ED_{0.90}$ and $ED_{0.95}$, is also well-known (Wetherill, 1963).

* Corresponding author.
E-mail address: R.J.Karunamuni@ualberta.ca (R.J. Karunamuni).

http://dx.doi.org/10.1016/j.csda.2015.04.001
0167-9473/

The most commonly used method is to assume a parametric model for the dose–response curve: $\eta_\theta(x) = F(\mathbf{z}^T\theta)$, where $\theta = (\alpha, \beta)^T$ are unknown parameters, $\mathbf{z}(x) = (1, x)^T$, and F is some known distribution function, also known as the *link function*. Commonly used links include the probit and logit functions (Berkson, 1944; McCullagh and Nelder, 1989; Morgan, 1992). Then the effective dose level ED_p is given by

$$ED_p = \frac{F^{-1}(p) - \alpha}{\beta}. \tag{1.1}$$

A "plug-in" estimator of ED_p can be obtained by replacing the parameter θ by an estimator. For this purpose, traditional methods such as the maximum likelihood and weighted least squares approaches can easily be employed, and the resulting estimators are efficient under the model (i.e., when the link function is correctly specified). However, these estimators can be highly unstable if the assumed link function is not completely correct, and they are generally not robust if the data are slightly contaminated. In general, a minor instability in the model can have severe consequences on the accuracy of ED_p based on the above estimators; see Morgan (1992) for a comprehensive account of the parametric estimation methods for dose–response curves.

Nonparametric methods, on the other hand, are highly robust because they do not rely on a particular form of the link function. Many nonparametric methods have been proposed in the literature (Schmoyer, 1984; Müller and Schmitt, 1988; Kelly and Rice, 1990; Mukhopadhyay, 2000; Dette et al., 2005; Park and Park, 2006; Bornkamp and Ickstadt, 2009; Dette and Scheder, 2010). Nonparametric methods are flexible, and the shape of the dose–response curve is mainly determined by the data. In general, nonparametric estimates are consistent under widely applicable regularity conditions. However, compared with parametric models, nonparametric methods are less efficient, and it is difficult to extrapolate the dose–response curve beyond the range of the observed dose levels. To retain the advantages of parametric and nonparametric approaches, Yuan and Yin (2011) proposed a promising semiparametric method that takes a weighted average of the two methods, where the weight is chosen by minimizing the mean integrated squared error; see also Nottingham and Birch (2000). Some disadvantages of both nonparametric and semiparametric methods are (a) the corresponding estimators of ED_p do not have a closed form, and (b) they might not be uniquely defined when the estimated dose–response curve is not monotone. For these and other reasons, some practitioners prefer parametric methods. However, the lack of robustness of such methods to link misspecification and the presence of outliers are of great concern in practice. In this paper we address these issues and propose possible alternative approaches.

We present two new parametric methods for estimating ED_p using minimum-distance techniques. These methods have a degree of automatic robustness to model misspecification (Donoho and Liu, 1988). The first method is based on the *Hellinger distance* approach (Beran, 1977; Simpson, 1987), while the second method is based on the *symmetric chi-squared* distance (Lindsay, 1994, 2004). In fact, there is a near-equivalence relationship between the Hellinger distance and the symmetric chi-squared distance. Thus, estimators based on these two approaches appear to have similar asymptotic properties. We estimate the parameter θ using these two approaches and then construct plug-in estimators of ED_p based on formula (1.1). We show that the proposed estimators of ED_p achieve efficiency under the model and simultaneously have desirable robustness properties such as robustness to link misspecification. We use Monte Carlo studies to compare the finite-sample performance with that of the traditional estimators. Excellent efficiency properties and automatic robustness make the proposed estimators appealing in practical applications.

Robustness of estimation in the present context has been addressed by Hamilton (1979), who described and compared various robust methods for estimation of $ED_{0.50}$. Huang (2001) investigated, largely by simulation, the effects of carrying out a logistic analysis when the dose–response curve is incorrectly specified. The object of interest is a confidence interval on $ED_{0.50}$. Neuhaus (1999) discussed binary regression when the response variable is possibly misclassified.

An alternative approach to attain robustness to link misspecification is through the choice of design, which is an approach that has been investigated both theoretically and by simulation by various researchers (Huang, 2002; Biedermann et al., 2006; Woods et al., 2006; Adewale and Wiens, 2009; Adewale and Xu, 2009; Dette et al., 2008; Li and Wiens, 2011).

The remainder of this paper is organized as follows. Section 2 develops the traditional and proposed minimum-distance estimators of ED_p. Section 3 discusses the asymptotic properties of the proposed estimators, and Section 4 presents their finite-sample behavior. Two real data applications are given in Section 5, and Section 6 presents a discussion. The proofs of the main results are deferred to Appendix.

2. Parametric estimation of effective dose

2.1. Traditional estimators

In a dose–response experiment, the usual experimental procedure is to apply a dose x_j ($j = 1, 2, \ldots, K$, i.e., K different dosages) to each of n_j individuals and to record the number of responses (e.g., the number of cures). Assume that the number of responses observed is m_j for dose x_j, $j = 1, 2, \ldots, K$. Then the ratio $\pi_{j,N} = m_j/n_j$ is a sufficient statistic for the probability of response at dose x_j : $\Pr(Y = 1 \mid x_j)$, $j = 1, 2, \ldots, K$, where the random variable Y is as defined in the Introduction and $N = \sum_{j=1}^{K} n_j$, $K \geq 2$. Assume that $\Pr(Y = 1|x) = F(\mathbf{z}^T\theta)$, where F is some known distribution function. Note that we are dealing with N independent Bernoulli random variables, but not all are identically distributed; for a trial at dose x_j, the