Journal of Statistical Planning and Inference 143 (2013) 593–602

Contents lists available at SciVerse ScienceDirect

Journal of Statistical Planning and Inference

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

On the efficiency of nonparametric variance estimation in sequential dose-finding

Chih-Chi Hu, Ying Kuen Cheung*

Department of Biostatistics, Columbia University, 722 West 168th Street, New York, NY 10032, USA

ARTICLE INFO

Article history:
Received 1 November 2010
Received in revised form
22 February 2012
Accepted 11 August 2012
Available online 2 September 2012

Keywords:
Homoscedasticity
Least squares estimate
Phase I trials
Quantile estimation
Stochastic approximation

ABSTRACT

Dose-finding in clinical studies is typically formulated as a quantile estimation problem, for which a correct specification of the variance function of the outcomes is important. This is especially true for sequential study where the variance assumption directly involves in the generation of the design points and hence sensitivity analysis may not be performed after the data are collected. In this light, there is a strong reason for avoiding parametric assumptions on the variance function, although this may incur efficiency loss. In this paper, we investigate how much information one may retrieve by making additional parametric assumptions on the variance in the context of a sequential least squares recursion. By asymptotic comparison, we demonstrate that assuming homoscedasticity achieves only a modest efficiency gain when compared to nonparametric variance estimation: when homoscedasticity in truth holds, the latter is at worst 88% as efficient as the former in the limiting case, and often achieves well over 90% efficiency for most practical situations. Extensive simulation studies concur with this observation under a wide range of scenarios.

1. Introduction

We consider quantile estimation in the context of dose-finding study where patients are tested in successive groups of size m. Precisely, let X_i denote the dose given to the patients in the ith group, and Y_{ij} denote a continuous biomarker from the jth patient in the group. A response is said to occur if the outcome Y_{ij} exceeds a threshold t_0. The objective is to estimate the dose θ such that $\pi(\theta)=p$ for some pre-specified p, where $\pi(x) \coloneqq \mathrm{pr}(Y_{ij}>t_0|X_i=x)$. This clinical setting is not uncommon, and there is also a wide range of applications in other areas such as reliability testing and bioassay. However, quantile estimation based on continuous data has received relatively little attention in the literature. In practice, this problem is often dealt with by using sequential methods based on the dichotomised data $V_{ij} \coloneqq I(Y_{ij}>t_0)$, where $I(A)$ is indicator of the event A, such as the logit-MLE (Wu, 1985) or the continual reassessment method (O'Quigley et al., 1990). These methods, using the binary data to estimate θ, provide general solutions without imposing strong assumptions on the characteristics of Y_{ij}. On the other hand, this approach can result in substantial information loss due to dichotomisation. Cheung (2010) demonstrates that, with group size $m=3$ and normal data, the asymptotic efficiency of an optimal logit-MLE using the dichotomised data V_{ij} is at most 80% of a corresponding Robbins and Monro (1951) procedure using the continuous data Y_{ij}; and the efficiency loss becomes more substantial with a larger m or a more extreme target p. Having said this, we acknowledge that there is an ongoing need for designs and models for clinical situations with truly binary outcomes that are not results of dichotomisation of continuous outcomes. This paper, however, focuses on the relative

* Corresponding author.
E-mail address: yc632@columbia.edu (Y.K. Cheung).

0378-3758/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jspi.2012.08.014

efficiency of a least squares recursion using the continuous data under various assumptions on Y_{ij}. Generally, we consider the regression model

$$Y_{ij} = M(X_i) + \sigma(X_i)Z_{ij}, \tag{1}$$

where the noise Z_{ij} is standard normal. Among the earliest proposals to address this problem, Eichhorn and Zacks (1973) study sequential search procedures for θ under the assumptions that the mean function $M(x)$ is linear in x and the standard deviation is known and is constant, i.e., $\sigma(x) = \sigma$. Recently, Cheung and Elkind (2010) describe a novel application of the stochastic approximation method that leaves both $M(x)$ and $\sigma(x)$ unspecified subject to the constraint that θ is uniquely defined, and propose to estimate $\sigma(x)$ nonparametrically. These two sets of assumptions represent two extreme approaches, and raise the question whether there is a reasonable middle ground. Specifically, this paper focuses on the estimation of the standard deviation function, and investigates how much efficiency may be retrieved by imposing stronger assumptions on $\sigma(x)$ than that in Cheung and Elkind (2010) while keeping the mean $M(x)$ unspecified. Our investigation will be conducted in the context of a sequential least squares recursion described in Section 2. Section 3 derives the asymptotic distribution of an proposed estimator for θ. Section 4 reviews Wu's (1985) logit-MLE as a comparison method of the least squares recursion. Efficiency comparison is given in Section 5, and concluding remarks in Section 6. Technical details are put in Appendix.

2. Least squares recursion

Under model (1), Cheung and Elkind (2010) show that solving $\pi(\theta) = p$ is equivalent to solving $f(\theta) = t_0$, where $f(x) \coloneqq M(x) + z_p\sigma(x)$ and z_p is the upper pth percentile of standard normal. For brevity in discussion, we may assume here that the objective function f is continuous and strictly increasing so that the solution θ exists uniquely. An important class of models that satisfies this assumption is models with increasing mean $M(x)$ and constant coefficient of variation across doses. Conditions 1–3, below make precise statements of the assumptions that are much less restrictive.

Now, pretend that $f(x) = t_0 + b(x-\theta)$ for some $b > 0$, and suppose also that we can observe an asymptotically unbiased variable $U_{i,n}$ of $f(X_i)$ for group i. A least squares estimate $\hat{\theta}_n$ of θ based on the first n groups of observations can be obtained by solving

$$\frac{1}{n}\sum_{i=1}^{n}[U_{i,n} - \{t_0 + b(X_i - \hat{\theta}_n)\}] = 0. \tag{2}$$

Then we may set the next dose

$$X_{n+1} = \hat{\theta}_n. \tag{3}$$

The least squares recursion formed by (2) and (3) in essence is identical to the adaptive design proposed by Lai and Robbins (1979). A subtle difference is that the unbiased variable $U_{i,n}$ is chosen based on the assumption about the variance function $\sigma(x)$.

*Case*1 (known variance): When $\sigma(x)$ is completely known, a natural choice is to define $U_{i,n} = \overline{Y}_i + z_p\sigma(X_i)$, where $\overline{Y}_i = m^{-1}\sum_{j=1}^{m} Y_{ij}$ is the average of the measurements in group i.

*Case*2 (heteroscedasticity): When $\sigma(x)$ is unknown and unspecified, we may define $U_{i,n} = \overline{Y}_i + z_p\lambda_m^{1/2}s_i$, where s_i^2 is the sample variance of the measurements in group i,

$$\lambda_m = \frac{(m-1)\Gamma^2\{(m-1)/2\}}{2\Gamma^2(m/2)} \tag{4}$$

and $\Gamma(\cdot)$ is the gamma function. Note that the form of λ_m in (4) ensures $E(\lambda_m^{1/2}s_i) = \sigma(X_i)$ so that $U_{i,n}$ is unbiased for $f(X_i)$.

Under both Cases 1 and 2, the observed variable $U_{i,n}$ is unbiased for $f(X_i)$, and $U_{i,n}$ and $U_{j,n}$ are mutually independent for $i \neq j$. Therefore, using the same techniques as in Lai and Robbins (1979), we can then verify that the least squares recursion formed by (2) and (3) is identical to the nonparametric Robbins–Monro procedure under these two cases: $X_{n+1} = X_n - (nb)^{-1}(U_{n,n} - t_0)$, where $b > 0$ is the same as the assumed slope used in the least squares estimation (2). Hence, the standard convergence results of stochastic approximation apply so that $X_n \to \theta$ with probability one; for example, see Sack (1958). In addition, if $b < 2f'(\theta)$, the distribution of $\sqrt{n}(X_n - \theta)$ will converge weakly to a mean zero normal with variance equal to $\alpha_1\sigma^2(\theta)$ under Case 1 and $\alpha_1\alpha_2\sigma^2(\theta)$ under Case 2, where $\alpha_1 = [mb\{2f'(\theta) - b\}]^{-1}$ and $\alpha_2 = 1 + mz_p^2(\lambda_m - 1)$. In other words, the asymptotic relative efficiency due to the knowledge of $\sigma(x)$ is equal to α_2. To illustrate the magnitude, the efficiency $\alpha_2 = 2.87, 2.35, 2.17$ for $m = 2, 3, 4$ and $p = 0.10$. The efficiency gain is quite substantial, and is not surprising because Cases 1 and 2 in a sense represent two extremities of assumptions.

Case 3 (homoscedasticity): When $\sigma(x)$ is identical to an unknown constant σ for all x, we may choose $U_{i,n} = \overline{Y}_i + z_p\hat{\sigma}_n$ where $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^{n} s_i^2$.