Journal of Statistical Planning and Inference 167 (2015) 90–101

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Nearly second order three-stage design for estimating a product of several Bernoulli proportions

Zohra Benkamra [a], Mekki Terbeche [b], Mounir Tlemcani [a,*]

[a] Department of Mathematics, L.A.A.R, University of Science and Technology Mohamed Boudiaf, USTO-MB, Oran, Algeria

[b] Department of Mathematics, University of Oran Es-Senia, Algeria

ARTICLE INFO

Article history:
Received 12 July 2014
Received in revised form 15 February 2015
Accepted 21 May 2015
Available online 29 May 2015

MSC:
62L10
62L12

Keywords:
Sequential allocation
Second order
Lower bound
Asymptotic optimality
Strong law of large numbers
Rate of convergence
Two-stage
Three-stage
Sampling
Product of Bernoulli proportions

ABSTRACT

We give a second order lower bound for the variance incurred by a three-stage procedure for estimating a product of means by allocation from independent Bernoulli populations. The asymptotic analysis is derived from the trivial case of two proportions which allows one to construct an exact policy with an exact lower bound. We extend the result to the case of several proportions and rigorously prove the nearly second order asymptotic optimality of the proposed scheme using the rate of convergence in the strong law of large numbers which is upper limited by the central limit theorem. The results are validated *via* Monte-Carlo simulations and are very promising for the exact second order optimality.

1. Introduction

Assume that for $n \geq 2$, independent random variables $X_1, \ldots, X_n$ are observable from Bernoulli populations $\mathcal{P}_1, \ldots, \mathcal{P}_n$ with probabilities of success $p_1, \ldots, p_n$ respectively. In non linear estimation, cf. Page (1987); Shapiro (1985); Hardwick and Stout (1996), the problem of estimating the product $p = p_1 p_2 \ldots p_n$ has many applications in reliability engineering and risk, cf. Rekab (1992a); Benkamra et al. (2012). Because designers are generally risk averse, a crucial objective of such problems is the reduction of the variance incurred by the sampling method leading to a discrete optimization problem under the constraint of a total sample size T fixed. In the case of a Bayesian framework, this objective turns to minimize the Bayes risk associated with squared error loss plus a cost per failure in the case of Bernoulli proportions, cf. Woodroofe and Hardwick (1990); Hardwick and Stout (1996); Benkamra et al. (2013). Such problems can be solved by dynamic programming but this technique becomes costly and intractable in the case of several parameters. Assuming T is large enough, sequential

* Corresponding author. Tel.: +213 773759981.
E-mail address: mounir.tlemcani@gmail.com (M. Tlemcani).

http://dx.doi.org/10.1016/j.jspi.2015.05.006
0378-3758/

allocation procedures are good alternatives to derive asymptotically optimal partitions $M_1 + \cdots + M_n = T$ in the sense that the produced variance reaches a lower bound when T goes to infinity and M_i is the sample size taken to estimate p_i from observations in the population $\mathcal{P}_i$ for $i = 1, \ldots, n$. In the case of two proportions ($n = 2$), optimal procedures can be obtained and solved analytically when the coefficients of variation of the associated Bernoulli populations are known, cf. Page (1990). Unfortunately, these coefficients are unknown in practice since they depend themselves on the unknown proportions $p_1, \ldots, p_n$. Hence, adaptive sampling or allocation can be made sequentially, based on accruing data. It is well known that fully sequential designs, in which one updates information after each observation, are the most powerful. However, they are rarely used in practice because of the difficulties with their analysis and implementation. The most common issue of these drawbacks is the restriction of the sequential form to few stages experiment, in which allocation can be made in group for the current stage, based on accruing data from the previous one, cf. Hardwick and Stout (1995, 2002). In the literature, while two-stage and three-stage designs have received extensive analysis leading typically in linear estimation to first and second order asymptotic optimality respectively, cf. Ghosh et al. (1997); Rekab and Tahir (2004) and the references therein, there does not appear, to our knowledge, to be work in which second order asymptotic optimality is rigorously analyzed in the case of non linear estimation. The focus in this work is on second order asymptotic optimality of 2- and 3-stage designs for estimating a product of several Bernoulli proportions under the constraint of a total and fixed sample size T large enough. Wu (2013) has obtained a second order lower bound for the Bayes risk in a similar problem in software reliability with only two proportions. Both fully sequential and multi-stage designs were considered but the convergence results were roughly presented by Monte-Carlo simulation in order to validate second order optimality with respect to the total sample size.

The purposes of this work are to provide rigorous analysis of asymptotic optimality of both 2- and 3-stage designs when estimating a product of several Bernoulli proportions in a fixed framework. The analysis is based on the trivial case of two proportions which has the advantage that the underlined optimization problem can be solved analytically so that exact allocation rules with exact lower bound can be obtained. In Section 2, we present the setting of the optimization problem of the sample variance as a function of allocation under the constraint of a total sample size fixed T. In Section 3, we give analytical optimality conditions in a particular case of two proportions, as far as the exact lower bound can be obtained and an asymptotic expansion as $T \to \infty$ can be done for both the optimality conditions and the optimal value of the scaled variance. This will be the key point to derive, in Section 4, first and second order lower bounds for the scaled variance. We then show how the rate of convergence of any asymptotic allocation scheme can be controlled by the rate of convergence in the strong law of large numbers. In Section 5, we discuss the limitation of the rate of convergence in first stage designs when the initial sample size is chosen as a function of T according to Rekab (1992b), and we deduce the weakness of two-stage designs to be second order in Section 6. This leads us, in Section 7, to propose a three-stage design which is nearly second order asymptotically optimal with respect to the second order lower bound obtained in Section 3 and we validate our convergence results by some experiments *via* Monte-Carlo simulation in Section 8. In Section 9, we give final remarks and conclusion.

Throughout this paper, we use standard notations for asymptotic comparison in probability as follows: $f = o(g)$, $f = \Theta(g)$, and $f = o_{\mathbf{P}}(g)$; respectively, as $T \to \infty$, means that f is dominated by g asymptotically with probability one, f is dominated and subjected to g asymptotically with probability one, and $f/g \to 0$ in probability; respectively, as $T \to \infty$.

2. The sample variance as a function of allocation

Let M_i be the sample size allocated to estimate the proportion p_i, $i = 1, \ldots, n$. The maximum likelihood estimator (M.L.E) for the product $p = p_1 p_2 \ldots p_n$ is

$$\hat{p} = \hat{p_1}\hat{p_2} \ldots \hat{p_n},$$

where $\hat{p_i}$ is the sample proportion,

$$\hat{p_i} = \frac{S_{iM_i}}{M_i},$$

and S_{iM_i} counts the number of success in M_i Bernoulli trials in the population $\mathcal{P}_i$.

Assuming independence within and across populations, and using the fact that $\hat{p_i}$ is an unbiased estimator of p_i, the sample variance of $\hat{p}$ can be written as

$$\mathbb{V}ar(\hat{p}) = p^2 \left(\prod_{i=1,n} \left(1 + \frac{c_i^2}{M_i} \right) - 1 \right), \quad (1)$$

where c_i is the coefficient of variation of the Bernoulli population $\mathcal{P}_i$, i.e.,

$$c_i = \sqrt{\frac{1}{p_i} - 1}.$$