

Estimation following Adaptively Randomised Clinical Trials

Coad, Steve

Queen Mary, University of London, School of Mathematical Sciences

Mile End Road

London E1 4NS, United Kingdom

E-mail: d.s.coad@qmul.ac.uk

Introduction

Suppose that two treatments are being compared in a clinical trial. As the trial progresses, one of the treatments may look more promising and it would be desirable to allocate a higher proportion of patients to this treatment. In such cases, response-adaptive randomisation is used. The simplest such rules may be represented as urn models, in which balls of different types are added to or removed from the urn according to the previous assignments and responses. Having carried out the trial, we will want to estimate parameters of interest. Although the usual estimators will be approximately unbiased for trials with moderate to large numbers of patients, their biases may be appreciable for small to moderate-sized trials and the corresponding confidence intervals may also have coverage probabilities far from the nominal values.

Consider the following two-parameter model for the patient responses: y_{11}, y_{12}, \dots are random variables with density $f_1(y; \theta, \eta)$ and y_{21}, y_{22}, \dots are random variables with density $f_2(y; \theta, \eta)$, where θ is the parameter of interest, η is a nuisance parameter and $(\theta, \eta) \in \Omega$. The model is adaptive in the sense that patients are assigned to treatments using response-adaptive randomisation, so that the probability of assigning patient $k + 1$ to treatment 1 is a function of the previous k assignments and responses. The aim of this work is to obtain an approximately pivotal quantity for θ , from which corrected confidence intervals may be constructed which have coverage probabilities close to the nominal values for trials with a small number of patients. The results will then be applied to two examples.

The signed root transformation

It is well known that the likelihood function is not affected by the adaptive nature of the model; see, for example, Berger and Wolpert (1984). Thus, the log-likelihood function based on y_{11}, \dots, y_{1,n_1} and y_{21}, \dots, y_{2,n_2} is

$$\ell(\theta, \eta) = \sum_{k=1}^{n_1} \log\{f_1(y_{1k}; \theta, \eta)\} + \sum_{k=1}^{n_2} \log\{f_2(y_{2k}; \theta, \eta)\}.$$

Write $\hat{\theta}$ and $\hat{\eta}$ for the maximum likelihood estimators, and $\hat{\eta}_\theta$ for the restricted maximum likelihood estimator of η when θ is fixed. Let

$$\ell_{ij}(\theta, \eta) = \frac{\partial^{i+j}}{\partial \theta^i \partial \eta^j} \ell(\theta, \eta),$$

and suppose that

$$\frac{1}{n} \ell(\theta, \eta) \rightarrow \kappa(\theta, \eta) \quad \text{and} \quad \frac{1}{n} \ell_{ij}(\theta, \eta) \rightarrow \kappa_{ij}(\theta, \eta)$$

in $P_{\theta, \eta}$ -probability for all $(\theta, \eta) \in \Omega$ as $n \rightarrow \infty$, where κ_{ij} is not necessarily the partial derivative of κ .

Let

$$\Lambda_\theta = \ell(\hat{\theta}, \hat{\eta}) - \ell(\theta, \hat{\eta}_\theta)$$

and

$$Z_n = Z_n(\theta) = \sqrt{2\Lambda_\theta} \text{ sign}(\theta - \hat{\theta}).$$

Then Z_n is the first component of the bivariate signed root transformation (e.g. Bickel and Ghosh, 1990), which is asymptotically standard normal as $n \rightarrow \infty$ under modest conditions. Thus, Z_n may be treated as a first approximation to a pivotal quantity. The aim of this work is to find data-dependent quantities $\hat{\mu}_n$ and $\hat{\sigma}_n$ such that

$$Z_n^* = \frac{Z_n - n^{-\frac{1}{2}}\hat{\mu}_n}{\hat{\sigma}_n}$$

is asymptotically standard normal to third order in the very weak sense of Woodroffe (1986).

Correction terms

Let $N_i(n)$ denote the number of patients on treatment i after n assignments for $i = 1, 2$ and suppose that

$$\frac{N_i(n)}{n} \rightarrow \rho_i(\theta, \eta)$$

in $P_{\theta, \eta}$ -probability for almost every $(\theta, \eta) \in \Omega$ as $n \rightarrow \infty$, where ρ_i is a continuous function on Ω . To define the correction terms, write

$$\kappa^{20}(\theta, \eta) = \kappa_{20}(\theta, \eta) - \frac{\kappa_{11}(\theta, \eta)^2}{\kappa_{02}(\theta, \eta)}$$

and

$$\kappa^{30}(\theta, \eta) = \kappa_{30}(\theta, \eta) - 3\frac{\kappa_{21}(\theta, \eta)\kappa_{11}(\theta, \eta)}{\kappa_{02}(\theta, \eta)} + 3\frac{\kappa_{11}(\theta, \eta)^2\kappa_{12}(\theta, \eta)}{\kappa_{02}(\theta, \eta)^2} - \frac{\kappa_{11}(\theta, \eta)^3\kappa_{03}(\theta, \eta)}{\kappa_{02}(\theta, \eta)^3}.$$

Note that $-\kappa^{20}(\theta, \eta)$ is the average observed Fisher information for θ per observation.

Now, it may be shown that

$$E_{\theta, \eta}(Z_n) \simeq \frac{1}{\sqrt{n}}\mu(\theta, \eta)$$

in the very weak sense, where

$$\begin{aligned} \mu(\theta, \eta) = & \frac{1}{\sqrt{-\kappa^{20}(\theta, \eta)}} \left[-\frac{1}{3} \frac{\kappa^{30}(\theta, \eta)}{\kappa^{20}(\theta, \eta)} - \frac{1}{2} \frac{\kappa_{12}(\theta, \eta)}{\kappa_{02}(\theta, \eta)} + \frac{1}{2} \frac{\frac{\partial}{\partial \theta} \kappa^{20}(\theta, \eta)}{\kappa^{20}(\theta, \eta)} + \frac{\frac{\partial}{\partial \eta} \kappa_{11}(\theta, \eta)}{\kappa_{02}(\theta, \eta)} \right. \\ & \left. - \frac{1}{2} \frac{\kappa_{11}(\theta, \eta)}{\kappa_{02}(\theta, \eta)} \frac{\frac{\partial}{\partial \eta} \kappa^{20}(\theta, \eta)}{\kappa^{20}(\theta, \eta)} - \frac{1}{2} \frac{\kappa_{11}(\theta, \eta) \{2 \frac{\partial}{\partial \eta} \kappa_{02}(\theta, \eta) - \kappa_{03}(\theta, \eta)\}}{\kappa_{02}(\theta, \eta)^2} \right]. \end{aligned}$$

This can be estimated by $\hat{\mu}_n = \mu(\hat{\theta}, \hat{\eta})$. We also obtain the very weak approximation

$$E_{\theta, \eta} \left\{ \left(Z_n - \frac{\hat{\mu}_n}{\sqrt{n}} \right)^2 \right\} \simeq 1 + \frac{m(\theta, \eta)}{n} = \sigma_n^2(\theta, \eta),$$

say, where $m(\theta, \eta)$ has a very complicated form. Let $\hat{\sigma}_n^2 = \sigma_n^2(\hat{\theta}, \hat{\eta})$.

Given a desired confidence level $0 < \gamma < 1$, let

$$\mathcal{I}_n = \{ \theta : |Z_n^*(\theta)| \leq \Phi^{-1}(\frac{1+\gamma}{2}) \},$$

where Φ denotes the standard normal distribution function. Then

$$P_{\theta, \eta}(\theta \in \mathcal{I}_n) = \gamma + o(\frac{1}{n})$$

as $n \rightarrow \infty$, in the very weak sense. Thus, \mathcal{I}_n is an approximate 100 $\gamma\%$ confidence interval for θ .

Remark 1. For normal models, $\kappa_{ij} = 0$ for $i + j \geq 3$, and the above formulae for $\mu(\theta, \eta)$ and $m(\theta, \eta)$ simplify dramatically.

Remark 2. When n is replaced by a family of stopping times $t = t_a$ depending on a parameter $a \geq 1$, more general formulae for $\mu(\theta, \eta)$ and $m(\theta, \eta)$ are obtained which involve the limit of a/t_a and its first derivatives.

Examples

Example 1. Normal model. Suppose that y_{11}, y_{12}, \dots are normal random variables with mean μ and unit variance and that y_{21}, y_{22}, \dots are normal random variables with mean ν and unit variance. Let $\theta = (\mu - \nu)/2$ and $\eta = (\mu + \nu)/2$. Then the log-likelihood function is

$$\ell(\theta, \eta) = -\frac{1}{2} \left[\sum_{k=1}^{n_1} \{y_{1k} - (\theta + \eta)\}^2 + \sum_{k=1}^{n_2} \{y_{2k} - (\eta - \theta)\}^2 \right],$$

where $\theta, \eta \in \mathbf{R}$.

Consider the response-adaptive rule studied by Hayre and Gittins (1981) that randomises patients to treatment 1 with probability $w(\hat{\theta})/\{1 + w(\hat{\theta})\}$, where

$$w(\theta) = \begin{cases} \sqrt{1 + 10\theta}, & \theta > 0, \\ \frac{1}{\sqrt{1 + 10|\theta|}}, & \theta \leq 0. \end{cases}$$

Then we know that

$$\rho_1(\theta, \eta) = \frac{w(\theta)}{1 + w(\theta)}$$

and

$$\rho_2(\theta, \eta) = \frac{1}{1 + w(\theta)}.$$

It follows easily that

$$\kappa_{20}(\theta, \eta) = \kappa_{02}(\theta, \eta) = -1, \quad \kappa_{11}(\theta, \eta) = -\rho_1(\theta, \eta) + \rho_2(\theta, \eta)$$

and $\kappa_{ij}(\theta, \eta) = 0$ for $i + j \geq 3$. So we have

$$\kappa^{20}(\theta, \eta) = -4\rho_1(\theta, \eta)\rho_2(\theta, \eta).$$

It may then be shown that

$$\mu(\theta, \eta) = \frac{1}{4} \frac{\{1 - 2\rho_1(\theta, \eta)\}}{\{\rho_1(\theta, \eta)\rho_2(\theta, \eta)\}^{\frac{3}{2}}} \frac{\partial \rho_1(\theta, \eta)}{\partial \theta}$$

and

$$\sigma_n^2(\theta, \eta) = 1 + \frac{\mu(\theta, \eta)^2}{n}.$$

Example 2. Binary model. Suppose that y_{11}, y_{12}, \dots are Bernoulli random variables with parameter p_1 and that y_{21}, y_{22}, \dots are Bernoulli random variables with parameter p_2 . Let $\theta = \sqrt{p_1 q_2 / (p_2 q_1)}$ and $\eta = \sqrt{p_1 p_2 / (q_1 q_2)}$, where $q_i = 1 - p_i$ for $i = 1, 2$. Then the log-likelihood function is

$$\ell(\theta, \eta) = \sum_{k=1}^{n_1} y_{1k} \log(\theta\eta) + \sum_{k=1}^{n_2} y_{2k} \log\left(\frac{\eta}{\theta}\right) - n_1 \log(1 + \theta\eta) - n_2 \log\left(1 + \frac{\eta}{\theta}\right),$$

where $\theta, \eta > 0$.

Consider the drop-the-loser rule introduced by Ivanova (2003) that randomises patients to the treatments using an urn model. The initial urn composition is one ball of each treatment type and an immigration ball. When the immigration ball is drawn, one ball of each treatment type is added. When one of the treatment type balls is drawn and there is a success, the ball is returned; otherwise, the ball is not returned. Then we know that

$$\rho_1(\theta, \eta) = \frac{\theta(1 + \theta\eta)}{2\theta + \eta + \theta^2\eta}$$

and

$$\rho_2(\theta, \eta) = \frac{\theta + \eta}{2\theta + \eta + \theta^2\eta}.$$

It follows that

$$\kappa_{20}(\theta, \eta) = -\frac{\eta(\theta^2 + 1 + 2\theta\eta)}{\theta(1 + \theta\eta)(\theta + \eta)(2\theta + \eta + \theta^2\eta)}, \quad \kappa_{02}(\theta, \eta) = \frac{\theta^2}{\eta^2}\kappa_{20}(\theta, \eta)$$

and

$$\kappa_{11}(\theta, \eta) = \frac{1 - \theta^2}{(1 + \theta\eta)(\theta + \eta)(2\theta + \eta + \theta^2\eta)}.$$

So we have an expression for $\kappa^{20}(\theta, \eta)$. Using the above results and further calculations, we can obtain a formula for $\mu(\theta, \eta)$, which does not simplify. The expression for $m(\theta, \eta)$ has a complicated form and has not yet been calculated.

Simulation results

The simulation results are based on 10,000 replications. In order to assess the accuracy of the approximations, the coverage probabilities are reported for both the first-order pivot, Z_n , and the corrected pivot, Z_n^* .

Monte Carlo results for Example 1 are presented in Tables 1 and 2 when $n = 25$ and $n = 50$, respectively. Note that one patient was allocated to each treatment initially, as in Woodrooffe and Coad (2002). Results are also included for the expected sample size on the better treatment, $E(N_1)$. We took $\nu = 0$, so that $\theta = \eta$. The results show that the use of Z_n always yields coverage probabilities within two standard errors of the nominal values, even when θ is large, and that there is no noticeable improvement when Z_n^* is used.

Table 1. Monte Carlo estimates of expected sample sizes and coverage probabilities for the normal model when $n = 25$

θ	$E(N_1)$	Z_n		Z_n^*	
		95%	90%	95%	90%
0.5	17.0	0.952	0.901	0.952	0.900
0.75	18.0	0.953	0.897	0.952	0.894
1.0	18.6	0.952	0.903	0.950	0.902
1.25	19.0	0.952	0.903	0.952	0.903
1.5	19.3	0.952	0.906	0.952	0.906
2.0	19.8	0.953	0.903	0.952	0.903

Table 2. Monte Carlo estimates of expected sample sizes and coverage probabilities for the normal model when $n = 50$

θ	$E(N_1)$	Z_n		Z_n^*	
		95%	90%	95%	90%
0.5	34.9	0.951	0.902	0.949	0.900
0.75	36.6	0.950	0.903	0.949	0.903
1.0	37.8	0.952	0.903	0.952	0.904
1.25	38.7	0.952	0.902	0.952	0.902
1.5	39.4	0.954	0.903	0.954	0.902
2.0	40.4	0.953	0.904	0.953	0.904

Monte Carlo results for Example 2 are presented in Tables 3 and 4 when $n = 25$ and $n = 50$, respectively. Note that one patient was allocated to each treatment initially. The estimates of θ and η were also modified by adding 0.5 to the numbers of successes and failures. Results are again included for $E(N_1)$. This time, the results show that the use of Z_n always yields coverage probabilities less than the nominal values, especially if the difference between p_1 and p_2 is large. However, when $Z_n - \hat{\mu}_n/\sqrt{n}$ is used, most of the coverage probabilities are within two standard errors of the nominal values.

Table 3. Monte Carlo estimates of expected sample sizes and coverage probabilities for the binary model when $n = 25$

p_1	p_2	$E(N_1)$	Z_n		$Z_n - \hat{\mu}_n/\sqrt{n}$	
			95%	90%	95%	90%
0.7	0.5	13.9	0.932	0.880	0.951	0.902
0.5	0.3	13.7	0.927	0.866	0.943	0.897
0.4	0.2	13.7	0.940	0.862	0.961	0.898
0.55	0.5	12.8	0.942	0.893	0.955	0.908
0.7	0.3	15.1	0.933	0.863	0.961	0.896
0.8	0.5	14.7	0.933	0.870	0.958	0.909

Table 4. Monte Carlo estimates of expected sample sizes and coverage probabilities for the binary model when $n = 50$

p_1	p_2	$E(N_1)$	Z_n		$Z_n - \hat{\mu}_n/\sqrt{n}$	
			95%	90%	95%	90%
0.7	0.5	28.6	0.943	0.887	0.949	0.899
0.5	0.3	28.1	0.941	0.887	0.952	0.899
0.4	0.2	28.0	0.944	0.893	0.955	0.901
0.55	0.5	25.9	0.942	0.886	0.949	0.895
0.7	0.3	31.6	0.934	0.881	0.947	0.896
0.8	0.5	30.5	0.937	0.878	0.950	0.892

Extension to higher dimensions

Now suppose that there are $p \geq 2$ nuisance parameters, so that $\eta = (\eta_1, \dots, \eta_p)^T$. Then we can extend the approach which has been developed for a single nuisance parameter to higher dimensions. However, such a generalisation is considerably more complicated, both algebraically and analytically.

To appreciate the nature of the calculations involved, consider the correction term $\mu(\theta, \eta)$. Although this will have the same form as that for the single nuisance parameter case, the formulae for

$\kappa^{20}(\theta, \eta)$ and $\kappa^{30}(\theta, \eta)$ are now much more complicated. For example, $\kappa_{11}(\theta, \eta)$ becomes a p -vector and $\kappa_{02}(\theta, \eta)$ a $p \times p$ matrix.

Discussion

We have shown how to construct corrected confidence intervals for the parameter of interest for data from an adaptively randomised clinical trial when there is a nuisance parameter. The accuracy of the approximations has been assessed by simulation for two examples.

The focus here has been interval estimation and the biases of the usual estimators have not been studied. For the binary case, Coad and Ivanova (2001) show that the biases are of order $1/n$ and approximate them for several urn designs.

A similar approach can be applied when a stopping time is incorporated. For some related work, see Coad and Woodroffe (1997) for the censored survival data case and Coad and Govindarajulu (2000) for the binary case.

REFERENCES

- Berger, J.O. and Wolpert, R.L. (1984). *The Likelihood Principle*. Hayward, California: Institute of Mathematical Statistics.
- Bickel, P.J. and Ghosh, J.K. (1990). A decomposition of the likelihood ratio statistic and the Bartlett correction - a Bayesian argument. *Ann. Statist.* **18**, 1070-1090.
- Coad, D.S. and Govindarajulu, Z. (2000). Corrected confidence intervals following a sequential adaptive clinical trial with binary responses. *J. Statist. Plann. Inf.* **91**, 53-64.
- Coad, D.S. and Ivanova, A. (2001). Bias calculations for adaptive urn designs. *Sequential Anal.* **20**, 91-116.
- Coad, D.S. and Woodroffe, M.B. (1997). Approximate confidence intervals after a sequential clinical trial comparing two exponential survival curves with censoring. *J. Statist. Plann. Inf.* **63**79-96.
- Hayre, L.S. and Gittins, J.C. (1981). Sequential selection of the larger of two normal means. *J. Amer. Statist. Assoc.* **76**, 696-700.
- Ivanova, A. (2003). A play-the-winner type urn design with reduced variability. *Metrika* **58**, 1-13.
- Woodroffe, M. (1986). Very weak expansions for sequential confidence levels. *Ann. Statist.* **14**, 1049-1067.
- Woodroffe, M. and Coad, D.S. (2002). Corrected confidence sets for sequentially designed experiments: Examples. *Sequential Anal.* **21**, 191-218.

RÉSUMÉ

Supposons que deux traitements soient, dans le cadre d'une étude clinique, comparés avec modification du ratio de randomisation en fonction des réponses observées. A la fin de l'étude, nous nous intéressons à l'estimation des paramètres d'intérêt. Bien que les estimateurs usuels soient approximativement sans biais pour un moyen ou grand nombre de patient, leurs biais pourraient être appréciables pour les essais de petite ou moyenne taille et les intervalles de confiance correspondants pourraient également avoir des probabilités de couverture loin de la valeur nominale. Un modèle adaptatif à deux paramètres est étudié dans lequel il y a un paramètre d'intérêt et un paramètre de nuisance. Les intervalles de confiance corrigés, basés sur la transformation racine signé, sont construits pour le paramètre d'intérêt. Les intervalles ont des probabilités de couverture proche de la valeur nominale pour des essais de petite taille. La précision des approximations est évaluée par simulation pour les deux exemples.