Journal of Statistical Planning and Inference 138 (2008) 3915 – 3926

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

On Bayesian estimators in multistage binomial designs

Pierre Bunouf*, Bruno Lecoutre

ERIS, Laboratoire de Mathématiques Raphaël Salem, CNRS-Université de Rouen, Avenue de l'Université, BP 12, 76801 Saint-Etienne-du-Rouvray, France

ARTICLE INFO

Article history:
Received 29 December 2006
Received in revised form
12 September 2007
Accepted 14 February 2008
Available online 28 March 2008

Keywords:
Sequential analysis
Fisher information
Beta-*J* distribution
Haldane prior
Uniform prior
Frequentist properties

ABSTRACT

A new class of Bayesian estimators for a proportion in multistage binomial designs is considered. Priors belong to the beta-*J* distribution family, which is derived from the Fisher information associated with the design. The transposition of the beta parameters of the Haldane and the uniform priors in fixed binomial experiments into the beta-*J* distribution yields bias-corrected versions of these priors in multistage designs. We show that the estimator of the posterior mean based on the corrected Haldane prior and the estimator of the posterior mode based on the corrected uniform prior have good frequentist properties. An easy-to-use approximation of the estimator of the posterior mode is provided. The new Bayesian estimators are compared to Whitehead's and the uniformly minimum variance estimators through several multistage designs. Last, the bias of the estimator of the posterior mode is derived for a particular case.

1. Introduction

Many publications relate the bias of the maximum likelihood estimator (MLE) in multistage designs (see e.g. Jennison and Turnbull, 2000, Chapter 8). Whitehead (1986) suggested subtracting to the MLE an estimate of its bias. Emerson and Fleming (1990) derived an estimator by conditioning the (unbiased) MLE from the first-stage data on the stopping stage and the outcome variable. Later, Jung and Kim (2004) showed that this bivariate statistics is sufficient and complete in multistage designs involving discrete binomial observations. They gave an explicit form of this estimator, which is the uniformly minimum variance unbiased estimator (UMVUE) by application of Rao–Blackwell theorem.

The frequentist approach proposes non-unique solutions based on orderings in the observations space. Emerson and Fleming (1990) compared the characteristics of the median unbiased estimator for the normal mean based on the MLE ordering and the stage-wise ordering, in which results corresponding to earlier termination are more extreme than those which terminate later (Armitage, 1958).

On the other side, most of Bayesian statisticians are still reluctant to transgress the stopping rule principle (i.e. a valid inference is based on the likelihood only) in spite of explicit attempts to incorporate the stopping rule into non-subjective priors (see e.g. Bernardo and Smith, 1994; Ye, 1993). Nevertheless, the conditioning on the design d is fully justified in the Bayesian approach. According to de Cristofaro (2004), Bayes' formula must make explicit reference to d and the preexperimental evidence e_0. Denoting i the statistical information, it becomes

$$p(\theta|i,e_0,d) \propto p(\theta|e_0,d)p(i|\theta,e_0,d). \tag{1}$$

Applying Jeffreys' rule to the likelihood of binomial multistage design, Bunouf and Lecoutre (2006) derived a corrected version of the Jeffreys prior and formalized the beta-*J* distribution family. The corrected Jeffreys prior was used for interval estimation in

* Corresponding author.
E-mail addresses: pierre.bunouf@gmail.com (P. Bunouf), bruno.lecoutre@univ-rouen.fr (B. Lecoutre).

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.02.014

Bunouf and Lecoutre (2007). The idea followed in this article is to derive parameterizations of the beta-J priors leading to optimal Bayesian estimators.

The corrected versions of the Haldane and the uniform priors are introduced in Section 2. We also define the estimator of the posterior mean based on the corrected Haldane prior and the estimator of the posterior mode based on the corrected uniform prior. An approximation of the estimator of the posterior mode for two-stage designs is derived in Section 3. In Section 4, we extend some comparisons between MLE, UMVUE and Whitehead's estimator, reported in Chang et al. (1989), to the new Bayesian estimators. The characteristics of the estimators in other designs to assess the probability of rare events are discussed. In Section 5, the bias of the estimator of the posterior mode is derived for the two-stage designs allowing an early stopping if at least one success is observed at stage 1.

2. Corrected Haldane and uniform priors

Let $d_{\mathrm{Bin}^{\otimes K}}$ denote a K-stage binomial design ($K \geqslant 2$) involving successive binomial trials X_k of fixed sizes n_k ($k = 1, \ldots, K$). After each stage, the available data are analyzed and a decision whether to continue or to stop the experiment is made. We denote by $Y_k = \sum_{i=1}^{k} X_i$ the number of successes and by $v_k = \sum_{i=1}^{k} n_i$ the sample size accrued until the kth analysis. The variable M is the number of analyses actually performed. Let J_k ($k = 1, \ldots, K-1$) be the continuation region from stage k to $k+1$ for Y_k, M is then the first k such that $Y_k \notin J_k$. The stopping rule is determined by the probability $P_\theta(M \geqslant k) = P_\theta(Y_1 \in J_1, Y_2 \in J_2, \ldots, Y_{k-1} \in J_{k-1})$, which is the sum of the probabilities

$$p(x_1, \ldots, x_{k-1} | d_{\mathrm{Bin}^{\otimes K}}, \theta) = \binom{n_1}{x_1} \cdots \binom{n_{k-1}}{x_{k-1}} \theta^{y_{k-1}} (1-\theta)^{v_{k-1} - y_{k-1}}$$

for the sequences $(x_1, \ldots, x_{k-1})$ on the $k-1$ dimension restriction

$$\mathscr{R}_{(k)} = \{(x_1, \ldots, x_i) : y_i \in J_i ; i = 1, \ldots, k-1\}. \tag{2}$$

Denoting h a function of $x_1, \ldots, x_K$ independent of θ, the log-likelihood is

$$\begin{aligned} \mathrm{Log}(L(\theta; m, x_1, \ldots, x_m)) = h(x_1, \ldots, x_K) &+ [x_1 \mathrm{Log}(\theta) + (n_1 - x_1)\mathrm{Log}(1-\theta)] \\ &+ 1_{m \geqslant 2}[x_2 \mathrm{Log}(\theta) + (n_2 - x_2)\mathrm{Log}(1-\theta)] \\ &+ \cdots + 1_{m=K}[x_K \mathrm{Log}(\theta) + (n_K - x_K)\mathrm{Log}(1-\theta)]. \end{aligned} \tag{3}$$

Define $r_k = n_k / n_1$ ($k = 2, \ldots, K$), the Fisher information conditional on $d_{\mathrm{Bin}^{\otimes K}}$ is given by

$$\begin{aligned} I(\theta | d_{\mathrm{Bin}^{\otimes K}}) &= -E_\theta \left[\frac{\partial^2}{\theta^2} \mathrm{Log}(L(\theta; m, x_1, \ldots, x_m)) \right] \\ &= \frac{n_1}{\theta(1-\theta)} [1 + r_2 P_\theta(M \geqslant 2) + \cdots + r_K P_\theta(M = K)]. \end{aligned} \tag{4}$$

The beta-J distribution family is conjugate to the likelihood in (3). Its density is based on the components in θ of (4), the three positive scalars a, b, c, and the design parameters r_k and J_k, such that

$$\mathrm{Be}^J(a, b, c; r_2, \ldots, r_K; J_1, \ldots, J_{K-1}) \propto \theta^{a-1}(1-\theta)^{b-1}[1 + r_2 P_\theta(M \geqslant 2) + \cdots + r_K P_\theta(M = K)]^c. \tag{5}$$

If $c = 0$, (5) reduces to the beta distribution $\mathrm{Be}(a, b)$. When using beta priors in multistage designs, the stopping rule is ignored as a and b characterize the preexperimental information on θ. Conversely, the beta-J priors, assuming $c > 0$, are design-dependent and allow a correction for the bias. We show this property in K-stage Bernoulli design, where the experiment stops early if one success is observed. The stopping rule is defined by $P_\theta(M \geqslant k) = (1-\theta)^{k-1}$. If $c > 0$ in (5), the higher the value of K is, the more the weight on the low values of θ is. The correction of the prior is proportional in K to the positive bias induced by the stopping rule on the MLE (i.e. $\hat{\theta}^{\mathrm{ML}} = 1/M$).

The property of bias correction of the beta-J priors is examined in the field of point estimation. In binomial experiments ($K = 1$), the estimator of the posterior mean based on the Haldane prior $\vartheta | H \sim \mathrm{Be}(0, 0)$ and the estimator of the posterior mode based on the uniform prior $\vartheta | U \sim \mathrm{Be}(1, 1)$ both coincide with the unbiased MLE (i.e. $\hat{\theta}^{\mathrm{ML}} = Y_M / v_M$). Let $I(\theta)$ be the Fisher information associated with binomial experiments of size n, the Haldane prior is proportional to $I(\theta)$, i.e.

$$\vartheta | H \sim \mathrm{Be}(0, 0) \propto I(\theta) = \frac{n}{\theta(1-\theta)}.$$

Extending this relationship to multistage binomial designs, we define the *corrected Haldane prior*, which is proportional to the conditional Fisher information (4), i.e.

$$\vartheta | d_{\mathrm{Bin}^{\otimes K}}, H \sim \mathrm{Be}^J(0, 0, 1; r_2, \ldots, r_K; J_1, \ldots, J_{K-1}) \propto I(\theta | d_{\mathrm{Bin}^{\otimes K}}). \tag{6}$$

Table 1
Range of the bias of MLE, MEAN and MODE for the two-stage designs with $n_1 = n_2 = 5$ and $s_{1,u} = 1, 2, 3, 4, 5$. Values of θ where the bias is minimum ($\theta_{\min}$), null (θ_0) and maximum ($\theta_{\max}$). There can be one or two values of θ_0

	$s_{1,u}=1$	$s_{1,u}=2$	$s_{1,u}=3$	$s_{1,u}=4$	$s_{1,u}=5$
MLE bias	$[0, 0.033]$	$[0, 0.044]$	$[0, 0.047]$	$[0, 0.044]$	$[0, 0.033]$
$\theta_{\max}$	0.167	0.333	0.500	0.667	0.833
MEAN bias	$[-0.012, 0.014]$	$[-0.013, 0.018]$	$[-0.010, 0.021]$	$[-0.003, 0.022]$	$[-0.001, 0.025]$
$\theta_{\min}, \theta_{\max}$	0.559, 0.117	0.737, 0.288	0.862, 0.466	0.955, 0.655	0.316, 0.853
θ_0	0.313	0.051 & 0.524	0.154 & 0.716	0.285 & 0.900	0.438
MODE bias	$[-0.013, 0.007]$	$[-0.015, 0.010]$	$[-0.016, 0.013]$	$[-0.009, 0.014]$	$[-0.0004, 0.023]$
$\theta_{\min}, \theta_{\max}$	0.487, 0.091	0.680, 0.255	0.831, 0.430	0.928, 0.615	0.294, 0.850
θ_0	0.225	0.058 & 0.434	0.147 & 0.630	0.260 & 0.819	0.381

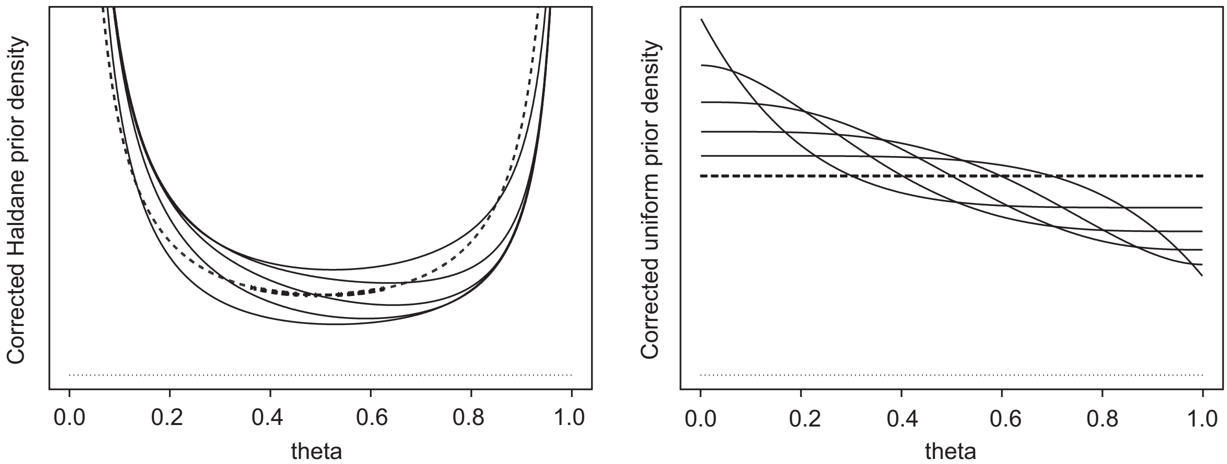

Fig. 1. Densities of the corrected (—) and the uncorrected (- - -) Haldane prior (left) and uniform prior (right) for the two-stage designs with $n_1 = n_2 = 5$ and $s_{1,u} = 1, 2, 3, 4, 5$.

Eq. (6) implicitly determines the value $c = 1$, which is the extent to which the Haldane prior is corrected by the stopping rule. This choice of the parameter c may also be used for the corrected version of the uniform prior (i.e. *corrected uniform prior*), which is defined as

$$\vartheta | d_{\mathrm{Bin}^{\otimes K}}, U \sim \mathrm{Be}^{J}(1, 1, 1; r_2, \ldots, r_K; J_1, \ldots, J_{K-1}). \tag{7}$$

Although prior distributions with c not equal to 1 are considered elsewhere (see e.g. Bunouf and Lecoutre, 2006 where $c = \frac{1}{2}$), in this paper we consider only the case $c = 1$. The link between estimators and beta prior parameters allowing unbiasedness in binomial experiments is extended to multistage designs. So, we denote by MEAN the estimator of the posterior mean based on the corrected Haldane prior and by MODE the estimator of the posterior mode based on the corrected uniform prior.

The bias of these two Bayesian estimators is first examined in simple two-stage designs. Let $s_{1,u}$ be the upper stopping boundary such that, if $X_1 \geqslant s_{1,u}$, the experiment stops at stage 1 based on the sample size $n_1 = 5$. Otherwise, the experiment continues to stage 2 involving an additional sample of size $n_2 = 5$.

For any value of $s_{1,u}$ in $\{1, \ldots, 5\}$, the bias of MLE is positive. Table 1 gives the range of bias of MEAN, MODE and MLE, and the values of θ where the bias is minimum, null and maximum. The bias curves of the three estimators are also displayed for $s_{1,u} = 1, 2$ in Fig. 2, Section 4.1.

This first exploration shows a strong decrease of the magnitude of the bias of the Bayesian estimators in contrast to MLE. Fig. 1 displays the density curves of the corrected and the uncorrected versions of the Haldane prior used for MEAN and the uniform prior used for MODE. Compared with the uncorrected versions, the corrected priors overweight the densities for $\theta < s_{1,u}/n_1$, allowing the correction for the bias.

3. Approximation of the posterior mode based on the corrected uniform prior

In this section, we derive an easy-to-use approximation of the posterior mode based on the corrected uniform prior for two-stage designs. The related estimator will be noted MODE^A. Its use in K-stage designs when $K \geqslant 3$ is discussed in Section 4.

We consider the general two-stage design $d_{\mathrm{Bin}^{\otimes 2}}$ with sample sizes (n_1, n_2). The experiment stops at stage 1 if $Y_1 \leqslant s_{1,l}$ or if $Y_1 \geqslant s_{1,u}$. Let $\mathrm{Bin}(n, \theta; x)$ be the probability function for the binomial outcome x and $C\mathrm{Bin}(n, \theta; J)$ be the cumulative probability function for the binomial outcome interval J. Based on $J_1 = [s_{1,l} + 1, s_{1,u} - 1]$, the corrected uniform prior (7) becomes

$$\vartheta | d_{\mathrm{Bin}\otimes 2}, U \sim \mathrm{Be}^J(1,1,1;r_2;J_1) \propto (1 + r_2\mathrm{CBin}(n_1,\theta;J_1)).$$

The density mode of the posterior distribution $\mathrm{Be}^J(y_m+1, v_m-y_m+1, 1; r_2; J_1)$ is the solution in θ of

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\mathrm{Log}(p(\theta | d_{\mathrm{Bin}\otimes 2}, U, (m, y_m))) = 0. \tag{8}$$

Denoting $\beta(,)$ the beta function, $\mathrm{CBin}(n_1,\theta;[0,s_{1,u}-1])$ can be expressed using the incomplete beta function $I_\theta(,,)$ (Abramowitz and Stegun, 1964), such that

$$\begin{aligned}\mathrm{CBin}(n_1,\theta;[0,s_{1,u}-1]) &= \mathrm{CBin}(n_1,1-\theta;[s_{1,u},n_1])\\ &= \Pr(\mathrm{Be}(n_1-s_{1,u}+1,s_{1,u})<1-\theta)\\ &= \frac{1}{\beta(n_1-s_{1,u}+1,s_{1,u})}\int_0^{1-\theta}\xi^{n_1-s_{1,u}}(1-\xi)^{s_{1,u}-1}\,\mathrm{d}\xi\\ &= I_{1-\theta}(n_1-s_{1,u}+1,s_{1,u}).\end{aligned}$$

Similar expression for $\mathrm{CBin}(n_1,\theta;[0,s_{1,l}])$ leads to a new form of (8), i.e.

$$\begin{aligned}&\frac{y_m}{\theta}-\frac{v_m-y_m}{1-\theta}+\frac{\mathrm{d}}{\mathrm{d}\theta}[\mathrm{Log}(1+r_2\mathrm{CBin}(n_1,\theta;[s_{1,l}+1,s_{1,u}-1]))]\\ &\quad=\frac{y_m-v_m\theta}{\theta(1-\theta)}+\frac{f'(\theta)}{f(\theta)}=0,\end{aligned} \tag{9}$$

where $f(\theta)=1+r_2I_{1-\theta}(n_1-s_{1,u}+1,s_{1,u})-r_2I_{1-\theta}(n_1-s_{1,l},s_{1,l}+1)$.

The integral form of $I_{1-\theta}$ eases the derivation of $f(\theta)$ in θ:

$$\begin{aligned}f'(\theta)&=\frac{-r_2}{\beta(n_1-s_{1,u}+1,s_{1,u})}\theta^{s_{1,u}-1}(1-\theta)^{n_1-s_{1,u}}+\frac{r_2}{\beta(n_1-s_{1,l},s_{1,l}+1)}\theta^{s_{1,l}}(1-\theta)^{n_1-s_{1,l}-1}\\ &=\frac{r_2}{1-\theta}[(n_1-s_{1,u}+1)\mathrm{Bin}(n_1,\theta;s_{1,u}-1)-(n_1-s_{1,l})\mathrm{Bin}(n_1,\theta;s_{1,l})].\end{aligned}$$

Then, (9) can be reexpressed as

$$\frac{\theta}{y_m-v_m\theta}=\frac{1}{r_2}\frac{1+r_2\mathrm{CBin}(n_1,\theta;[s_{1,l}+1,s_{1,u}-1])}{(n_1-s_{1,u}+1)\mathrm{Bin}(n_1,\theta;s_{1,u}-1)-(n_1-s_{1,l})\mathrm{Bin}(n_1,\theta;s_{1,l})}. \tag{10}$$

The resolution of (10) uses the characteristics of the holomorphic functions in θ defined by the two terms of the equality. The calculation (see Appendix A) yields the estimator MODE^A, defined as

$$\hat{\theta}^{\mathrm{MODE}^A}=\frac{Y_m t}{1+v_m t}, \tag{11}$$

where

$$t=\frac{1}{r_2}\frac{1+r_2\mathrm{CBin}(n_1,Y_m/v_m;[s_{1,l}+1,s_{1,u}-1])}{(n_1-s_{1,u}+1)\mathrm{Bin}(n_1,Y_m/v_m;s_{1,u}-1)-(n_1-s_{1,l})\mathrm{Bin}(n_1,Y_m/v_m;s_{1,l})}.$$

In one-sided designs based on either $s_{1,l}$ or $s_{1,u}$, only one term $\mathrm{Bin}(n_1, Y_m/v_m; s_{1,l})$ or $\mathrm{Bin}(n_1,Y_m/v_m;s_{1,u}-1)$, respectively, is kept in the denominator of t.

4. Comparison with other estimators

The new Bayesian estimators MEAN, MODE and MODE^A are compared to MLE, Whitehead's estimator (WHI) and UMVUE. Let $B_{d_{\mathrm{Bin}\otimes K},\hat{\theta}}$ be the bias function of an estimator $\hat{\theta}$ in the design $d_{\mathrm{Bin}\otimes K}$, i.e.

$$B_{d_{\mathrm{Bin}\otimes K},\hat{\theta}}(\theta)=E_{d_{\mathrm{Bin}\otimes K},\theta}(\hat{\theta})-\theta,$$

WHI is obtained by resolving iteratively the equality $\tilde{\theta} = \hat{\theta}^{\mathrm{ML}} - B_{d_{\mathrm{Bin}\otimes K},\hat{\theta}^{\mathrm{ML}}}(\tilde{\theta})$. Jung and Kim (2004) derived the useful explicit form of UMVUE, which is

$$\hat{\theta}^{\mathrm{UMVU}} = \frac{\sum_{\mathscr{R}_{(m)}} \binom{n_1-1}{x_1-1}\binom{n_2}{x_2}\cdots\binom{n_m}{x_m}}{\sum_{\mathscr{R}_{(m)}} \binom{n_1}{x_1}\binom{n_2}{x_2}\cdots\binom{n_m}{x_m}},$$

where $\mathscr{R}_{(m)}$ is the $m-1$ dimension restriction as defined in (2).

The investigated frequentist characteristics are the bias and the efficiency relative to MLE, which is defined as

$$RE_{d_{\mathrm{Bin}\otimes K},\hat{\theta}}(\theta) = \frac{\mathrm{MSE}_{d_{\mathrm{Bin}\otimes K},\hat{\theta}^{\mathrm{ML}}}(\theta)}{\mathrm{MSE}_{d_{\mathrm{Bin}\otimes K},\hat{\theta}}(\theta)},$$

where $\mathrm{MSE}_{d_{\mathrm{Bin}\otimes K},\hat{\theta}}(\theta) = E_{d_{\mathrm{Bin}\otimes K},\theta}(\hat{\theta}-\theta)^2$ is the mean square error. We also discuss the interpretation of estimates with respect to design considerations.

4.1. One-sided two-stage designs

In Table 1 (Section 2), we compared the biases of MLE, MEAN and MODE. In this section, the comparison is extended to UMVUE, WHI and MODE^A in the one-sided two-stage designs with $n_1 = n_2 = 5$ and $s_{1,u} = 1, 2$. Fig. 2 displays the curves of bias and efficiency relative to MLE of the estimators.

As shown in Table 2, estimates UMVUE at stage 1 correspond to MLE, and the correction for unbiasedness only holds on the estimates at stage 2. Fig. 3 displays the curves of the correction relative to MLE of the estimates UMVUE, WHI, MEAN, MODE, MODE^A in function of the estimates MLE. For all the estimators, the strongest correction occurs for $\hat{\theta}^{\mathrm{ML}} = s_{1,u}/n_1$. Unlike the Bayesian estimators where the correction is stronger at stage 1 and UMVUE where there is no correction at stage 1, WHI depends on MLE only.

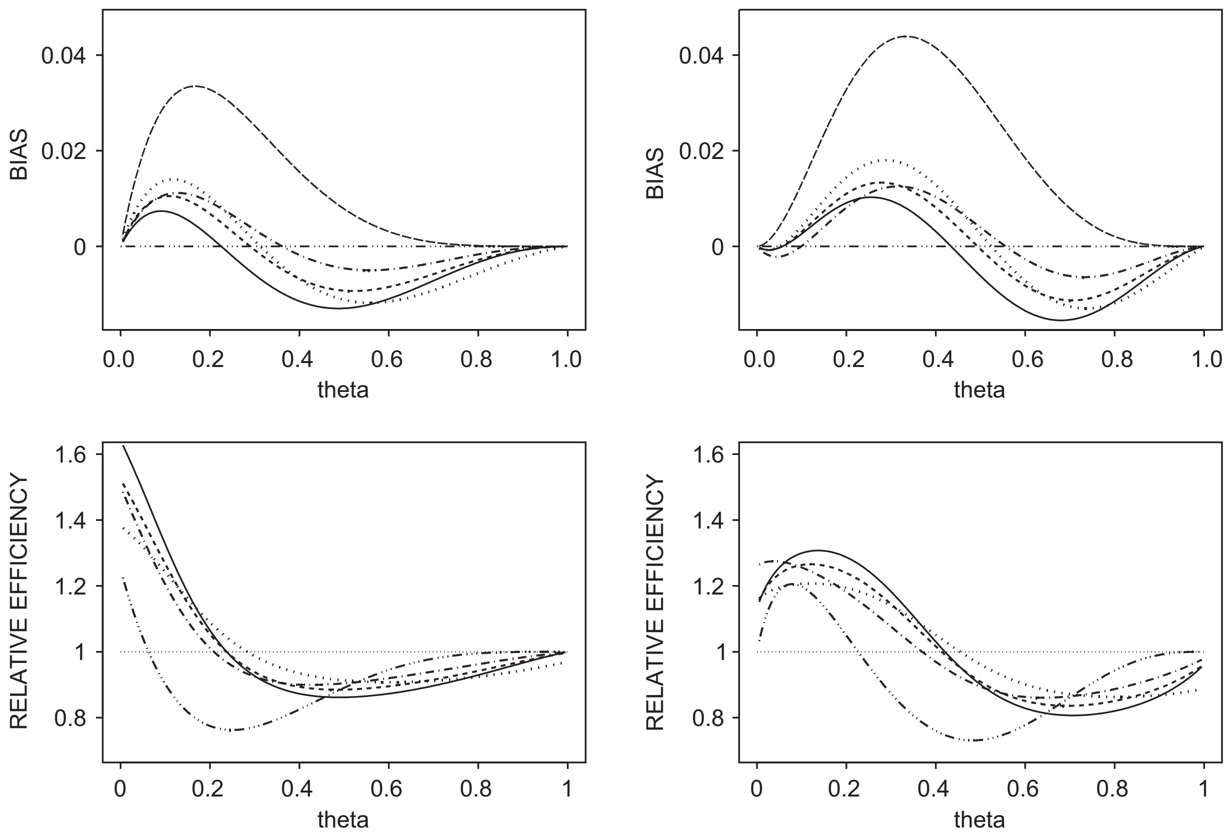

Fig. 2. Bias of MLE (– –), UMVUE (– · · · · –), WHI (– · –), MEAN (…), MODE (—) and MODE^A (- - -) for the two-stage designs with $n_1 = n_2 = 5$ and $s_{1,u} = 1$ (left-top) and $s_{1,u} = 2$ (right-top). Efficiency relative to MLE with $s_{1,u} = 1$ (left-bottom) and $s_{1,u} = 2$ (right-bottom).

Table 2
Estimates MLE, UMVUE, WHI, MEAN, MODE and MODE^A for the two-stage designs with $n_1 = n_2 = 5$ and $s_{1,u} = 1, 2$

(M, Y_m)	MLE	UMVUE	WHI	MEAN	MODE	MODE^A
$s_{1,u} = 1$						
(1, 1)	0.200	0.200	0.167	0.169	0.153	0.160
(1, 2)	0.400	0.400	0.383	0.371	0.366	0.373
(1, 3)	0.600	0.600	0.597	0.586	0.593	0.594
(1, 4)	0.800	0.800	0.800	0.797	0.800	0.800
(1, 5)	1	1	1	1	1	1
(2, 0)	0	0	0	0	0	0
(2, 1)	0.100	0	0.075	0.087	0.084	0.084
(2, 2)	0.200	0	0.166	0.181	0.176	0.178
(2, 3)	0.300	0	0.272	0.281	0.277	0.280
(2, 4)	0.400	0	0.383	0.385	0.384	0.386
(2, 5)	0.500	0	0.492	0.490	0.492	0.492
$s_{1,u} = 2$						
(1, 2)	0.400	0.400	0.356	0.359	0.339	0.346
(1, 3)	0.600	0.600	0.579	0.567	0.558	0.568
(1, 4)	0.800	0.800	0.798	0.788	0.796	0.796
(1, 5)	1	1	1	1	1	1
(2, 0)	0	0	0	0	0	0
(2, 1)	0.100	0.100	0.089	0.093	0.094	0.094
(2, 2)	0.200	0.143	0.172	0.185	0.183	0.183
(2, 3)	0.300	0.167	0.259	0.278	0.274	0.274
(2, 4)	0.400	0.182	0.356	0.376	0.369	0.371
(2, 5)	0.500	0.192	0.464	0.477	0.472	0.475
(2, 6)	0.600	0.200	0.579	0.583	0.581	0.583

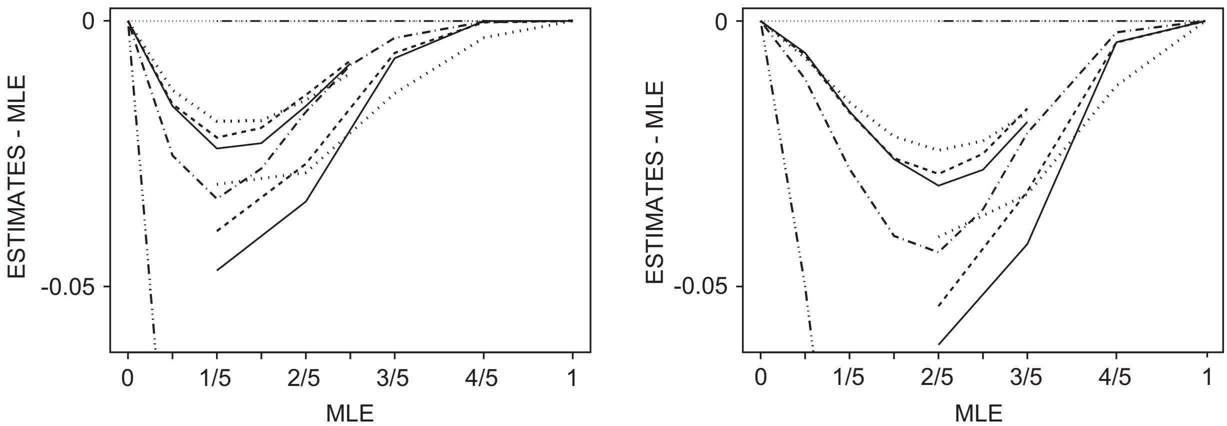

Fig. 3. Correction relative to MLE of the estimates UMVUE (– · · · · –), WHI (– · –), MEAN (...), MODE (—), MODE^A (- - -) in function of the estimates MLE for the two-stage designs with $n_1 = n_2 = 5$ and $s_{1,u} = 1$ (left) and $s_{1,u} = 2$ (right). Except for WHI, there are two separate curves by estimator for the estimates at stages 1 and 2.

4.2. *Chang three-stage designs*

The second comparison refers to the three-stage designs for frequentist testing procedures given in Chang et al. (1987). These designs were developed for cancer clinical trials to minimize the average sample size, and were already used in Chang et al. (1989) to compare MLE, UMVUE and WHI. For the set of hypotheses $H_0 : \theta \leqslant \theta_0$ versus $H_1 : \theta \geqslant \theta_1$, H_0 is accepted if $Y_k \leqslant s_{k,l}$ and rejected if $Y_k \geqslant s_{k,u}$ $(k = 1, 2, 3)$. The experiment stops early if a decision can be made at stage 1 or stage 2.

In 'Chang I' design, the tested values are $(\theta_0, \theta_1) = (0.05, 0.20)$. The design is based on the sample sizes $(n_1, n_2, n_3) = (15, 15, 10)$ and the boundaries $(s_{1,l}, s_{2,l}, s_{3,l}; s_{1,u}, s_{2,u}, s_{3,u}) = (0, 1, 4; 4, 5, 5)$. In 'Chang II' design, the parameters are $(n_1, n_2, n_3) = (20, 15, 15)$ and $(s_{1,l}, s_{2,l}, s_{3,l}; s_{1,u}, s_{2,u}, s_{3,u}) = (3, 8, 15; 9, 12, 16)$ to test the values $(\theta_0, \theta_1) = (0.20, 0.40)$. Both Chang designs ensure a 90% power for a nominal α-risk set to 0.05.

Fig. 4 shows the curves of bias and efficiency relative to MLE. Compared to MLE, the advantage of the Bayesian estimators and WHI is substantial for $\theta \in [s_{1,l}/n_1, s_{1,u}/n_1]$. This interval of θ is relevant in the experimental context as it contains both the values (θ_0, θ_1) involved in the hypotheses and the continuation regions for Y_k/v_k $(k = 1, 2)$, which are $J'_k = [(s_{k,l} + 1)/v_k, (s_{k,u} - 1)/v_k]$. In compensation for its unbiasedness, UMVUE exhibits less advantageous characteristics in terms of relative efficiency.

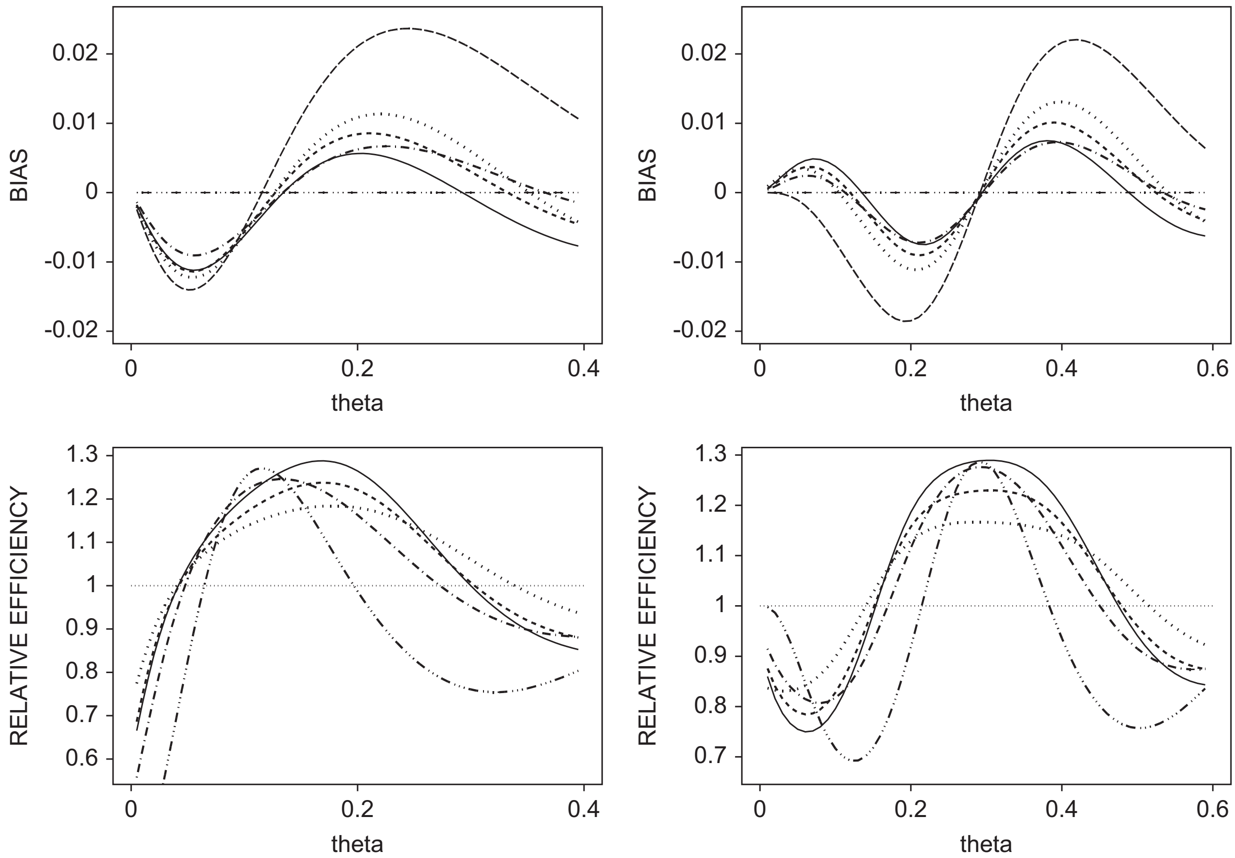

Fig. 4. Bias of MLE (– –), UMVUE (– · · · –), WHI (– · –), MEAN (…), MODE (—), MODE^A (- - -) for Chang I design (top-left) and Chang II design (top-right). Efficiency relative to MLE for Chang I design (bottom-left) and Chang II design (bottom-right).

Combining the bias and the relative efficiency, MODE offers particularly attractive characteristics among the design-dependent estimators in Chang designs. Although the approximation MODE^A takes into account the stopping rule for stage 1 only, the frequentist characteristics relative to MODE are almost preserved. As a general rule, the use of MODE^A in K-stage designs ($K \geqslant 3$) may be advocated when the stopping rule for stages k ($k \geqslant 2$) has a moderate influence and agrees with the stopping rule for stage 1. This condition is achieved when $r_{k+1} \leqslant r_k$ ($k = 2, \ldots, K-1$) and $J'_{k+1} \subset J'_k$ ($k = 1, \ldots, K-2$).

More generally, in hypothesis testing context, attention must be paid on the design specificities. Chang designs favor a stop at stage 1 if a decision on the hypotheses H_0 or H_1 can be made early. In some other designs, the stopping rule for stage 1 is based on a decision in favor of one hypothesis, whereas the stopping rule for the other interim stages is based on a decision in favor of the other hypothesis. In this case, the approximation MODE^A should not be used in place of MODE.

4.3. Planning experiments to assess the probability of rare events

An experiment based on K successive samples of size $n_k = 5$ ($k = 1, \ldots, K$) is planned to assess the probability of a rare event, says a life threatening adverse event within a clinical trial, with the expectation that $\theta < 0.1$. The experiment stops as soon as at least one event is observed (i.e. $s_{k,u} = 1$, $k = 1, \ldots, K-1$). The value of K which optimizes the characteristics of the estimators is sought.

Table 3 provides the estimates if one event is observed at stage $M = 1, 2, 3, 5, 10$ for $K = 3, 5, 10$. MEAN and MODE are the estimators which are the most influenced by K. If $M = 1$, estimates MEAN vary from 0.152 ($K = 3$) to 0.112 ($K = 10$), and estimates MODE vary from 0.127 ($K = 3$) to 0.062 ($K = 10$). Conversely, the variation of the estimates WHI from 0.157 ($K = 3$) to 0.152 ($K = 10$) is weak, and the estimates UMVUE are the sample proportion 0.200. If $M \geqslant 2$, estimates UMVUE are null and irrelevant in such experimental situation.

Fig. 5 shows the curves of bias and efficiency relative to MLE for $K = 5, 10$ and $\theta \in [0, 0.1]$. The frequentist properties of MODE are remarkable, with a small bias magnitude and a strongly increasing relative efficiency as θ gets closer to zero. Conversely, the poor characteristics of the approximation MODE^A are due to the lack of agreement between the stopping rules for stage 1 and stages k ($k = 2, \ldots, K-1$), as $J'_k \not\subset J'_1$.

Fig. 5 also reveals the high sensitivity to K of MEAN and MODE in terms of frequentist characteristics. When planning the design, K should not be over-estimated to avoid an inadequate correction of the estimates by the stopping rule. Conversely,

Table 3
Estimates MLE, UMVUE, WHI, MEAN, MODE, MODE^A for $Y_m = 1$ at stage $M = 1, 2, 3, 5, 10$ in the K-stage designs with $n_k = 5$ $(k = 1, \ldots, K)$ and $s_{k,u} = 1$ $(k = 1, \ldots, K-1)$ for $K = 3, 5, 10$

		MLE	UMVUE	WHI	MEAN	MODE	MODE^A
$(M, Y_m) = (1, 1)$	$K = 3$	0.200	0.200	0.157	0.152	0.127	0.160
	$K = 5$	0.200	0.200	0.152	0.133	0.096	0.160
	$K = 10$	0.200	0.200	0.152	0.112	0.062	0.160
$(M, Y_m) = (2, 1)$	$K = 3$	0.100	0	0.066	0.079	0.073	0.084
	$K = 5$	0.100	0	0.058	0.068	0.058	0.084
	$K = 10$	0.100	0	0.054	0.056	0.041	0.084
$(M, Y_m) = (3, 1)$	$K = 3$	0.070	0	0.041	0.055	0.052	0.059
	$K = 5$	0.070	0	0.035	0.048	0.044	0.059
	$K = 10$	0.070	0	0.031	0.040	0.032	0.059
$(M, Y_m) = (5, 1)$	$K = 5$	0.040	0	0.020	0.031	0.030	0.037
	$K = 10$	0.040	0	0.016	0.026	0.023	0.037
$(M, Y_m) = (10, 1)$	$K = 10$	0.020	0	0.007	0.015	0.014	0.019

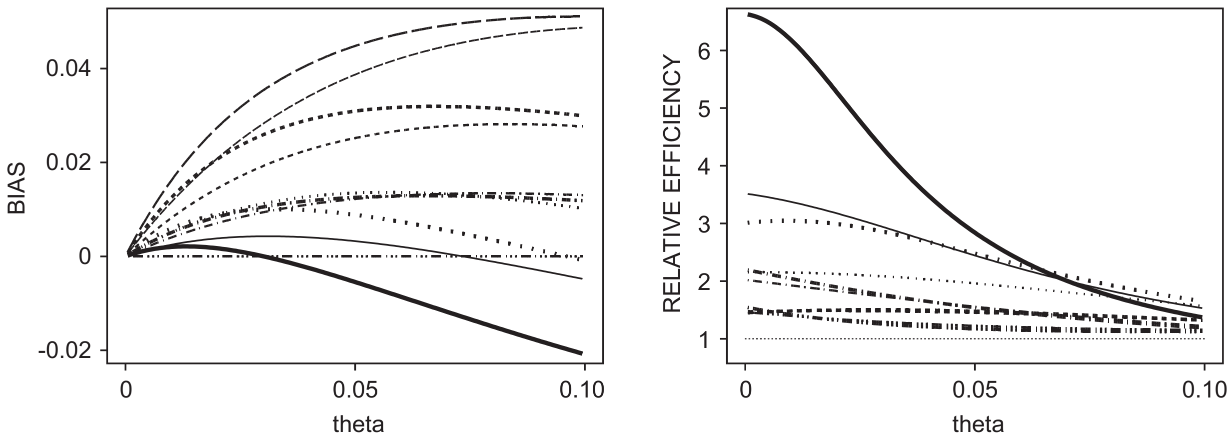

Fig. 5. Bias of MLE (– –), UMVUE (– · · · –), WHI (– · –), MEAN (…), MODE (—) and MODE^A (- - -) (left), and efficiency relative to MLE (right) in the K-stage designs with $n_k = 5$ $(k = 1, \ldots, K)$ and $s_{k,u} = 1$ $(k = 1, \ldots, K-1)$ for $K = 5, 10$ (thinner and thicker lines, respectively).

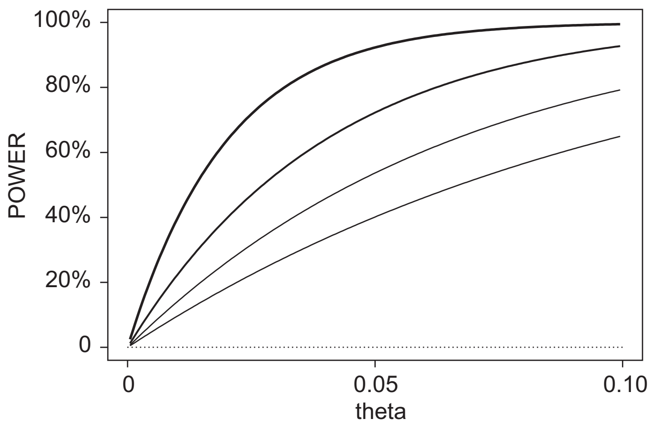

Fig. 6. Power curves of the test rejecting H_0 if at least one event is observed in the K-stage designs with $n_k = 5$ $(k = 1, \ldots, K)$ and $s_{k,u} = 1$ $(k = 1, \ldots, K-1)$ for $K = 2, 3, 5, 10$ (from thinnest to thickest line).

the design should allow a reasonable probability of observing at least one event in the experiment, which is

$$\text{prob} = P_\theta(\text{at least one event is observed}) = 1 - (1-\theta)^{5K}. \tag{12}$$

A choice of K allowing prob in (12) to be between 0.75 and 0.95 for plausible expectations of θ is a sound strategy. Such values of prob are guaranteed by $K = 5$ for $\theta \in [0.054, 0.112]$, and by $K = 10$ for $\theta \in [0.027, 0.058]$. The frequentist characteristics of MEAN and MODE are optimal under these conditions.

Further considerations on the design can be based on the frequentist test for the set of hypotheses $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ ($\theta_0 < \theta_1$), where H_0 is rejected if at least one event is observed in the experiment, otherwise H_0 is accepted. Whatever the ordering in the observations space, the power function is prob in (12). Fig. 6 displays the power curves for $K = 2, 3, 5, 10$. One adjusts K such that the type 1 and the type 2 errors ensure the nominal frequentist risks α and β, respectively.

5. Bias of the estimator of the corrected mode

In this section, we derive an approximation of the bias of MODE in function of the bias of MLE for the two-stage designs with $n_1 = n_2$ and $s_{1,u} = 1$ (i.e. the experiment stops at stage 1 if at least one event is observed). Such two-stage designs are used when the outcome is a critical event, such as a severe adverse event in clinical trial.

The bias of MLE is first derived. The expectation of an estimator $\hat{\theta}$ is

$$E_{d_{\mathrm{Bin}\otimes 2},\theta}(\hat{\theta}) = \sum_{i=1}^{n_1} \binom{n_1}{i} \theta^i (1-\theta)^{n_1-i} \underbrace{[\hat{\theta}_{(1,i)} + \hat{\theta}_{(2,i)}(1-\theta)^{n_1}]}_{D(\hat{\theta};i,\theta)}. \tag{13}$$

The derivation of $D(\hat{\theta}; i, \theta)$ in (13) for MLE yields

$$D(\hat{\theta}^{\mathrm{ML}}; i, \theta) = \frac{i}{n_1} + (1-\theta)^{n_1} \frac{i}{2n_1} = \frac{i}{n_1}\left(1 + \frac{(1-\theta)^{n_1}}{2}\right),$$

which results to the bias of MLE, i.e.

$$B_{d_{\mathrm{Bin}\otimes 2},\hat{\theta}^{\mathrm{ML}}}(\theta) = \sum_{i=1}^{n_1} \binom{n_1}{i} \theta^i (1-\theta)^{n_1-i} \frac{i}{n_1} \frac{(1-\theta)^{n_1}}{2} = \frac{1}{2}\theta(1-\theta)^{n_1}.$$

The calculation of the bias of MODE involves a new quadratic approximation of MODE, noted $\mathrm{MODE}^{\mathrm{Q}}$, which is

$$\hat{\theta}^{\mathrm{MODE}^{\mathrm{Q}}} = \frac{Y_m}{v_m}\left(1 - \frac{n_1}{v_m}\left(1 - \frac{Y_m}{v_m}\right)^{n_1-0.5} + \frac{n_1}{v_m}\left(1 - \frac{Y_m}{v_m}\right)^{2n_1-1}\right). \tag{14}$$

Details on the derivation of $\mathrm{MODE}^{\mathrm{Q}}$ are given in Appendix B.1. The expectation of $\mathrm{MODE}^{\mathrm{Q}}$ is calculated in Appendix B.2. It yields the approximation of the bias of MODE in function of the bias of MLE and $\mu_q(\theta)$, which is the q-order non-centered moment of the distribution $\mathrm{Bin}(n_1, \theta)$, such that

$$\begin{aligned} B^{\mathrm{App}}_{d_{\mathrm{Bin}\otimes 2},\hat{\theta}^{\mathrm{MODE}}}(\theta) = B_{d_{\mathrm{Bin}\otimes 2},\hat{\theta}^{\mathrm{ML}}}(\theta) + \left(1 + \frac{(1-\theta)^{n_1}}{4}\right) \\ \times \left[\frac{\mu_{2n_1-1}(1-\theta)}{n_1^{2n-1}} - \frac{\mu_{2n_1}(1-\theta)}{n_1^{2n_1}} - \frac{\mu_{n_1-0.5}(1-\theta)}{n_1^{n_1-0.5}} + \frac{\mu_{n_1+0.5}(1-\theta)}{n_1^{n_1+0.5}}\right]. \end{aligned} \tag{15}$$

The curves of the bias of MODE, $\mathrm{MODE}^{\mathrm{Q}}$ and $B^{\mathrm{App}}_{d_{\mathrm{Bin}\otimes 2},\hat{\theta}^{\mathrm{MODE}}}$ are displayed for the two-stage design with $n_1 = n_2 = 5$ and $s_{1,u} = 1$ in Fig. 7.

6. Concluding remarks

The integration of the stopping rule in the prior results in a justified Bayesian answer to the issue of point estimation in multistage binomial designs. The beta-J distribution family allows formalizing the design-corrected versions of the Haldane and the uniform priors. We showed that the estimator of the posterior mean based on the corrected Haldane prior and the estimator of the posterior mode based on the corrected uniform prior have good frequentist properties in terms of bias and efficiency relative to MLE, and coherent interpretation. An easy-to-use approximation of the estimator of the posterior mode was derived for two-stage designs. The conditions of its use in K-stage designs ($K \geqslant 3$) were established. We explored several designs where the new Bayesian estimators compare favorably with MLE, UMVUE and Whitehead's estimator.

Acknowledgments

The authors are indebted to the two anonymous referees and the associate editor for their helpful comments.

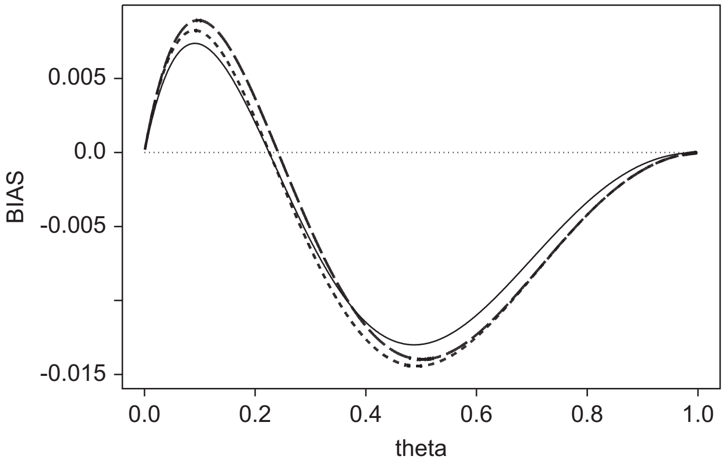

Fig. 7. Bias of MODE (—) and MODE^Q (- - -), and approximated bias of MODE (– –) for the two-stage design with $n_1 = n_2 = 5$ and $s_{1,u} = 1$.

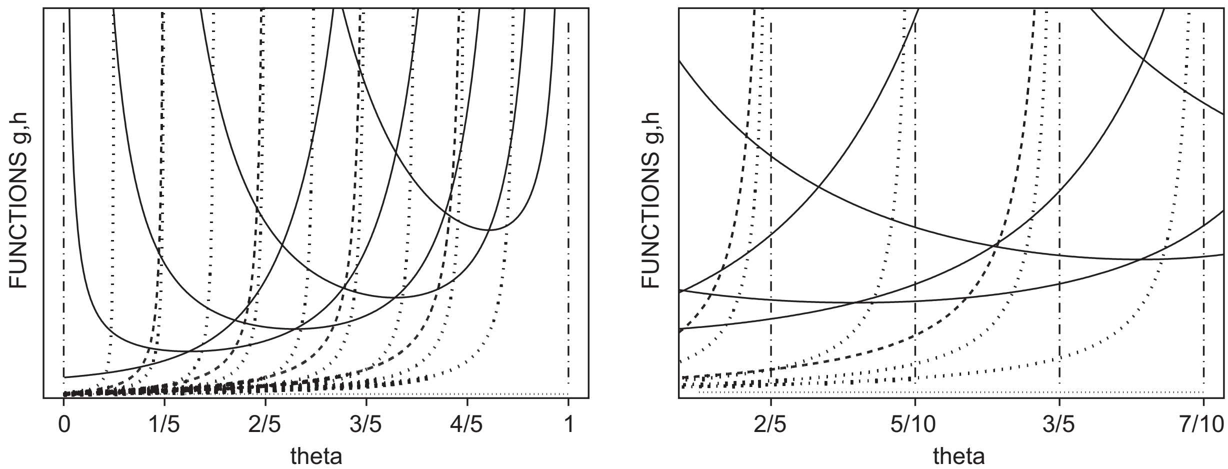

Fig. A1. Holomorphic functions $g(\theta)$ for $x_1 = 1, 2, 3, 4$ (- - -) and $y_2 = 1, \ldots, 9$ (…), and, $h(\theta)$ (—) for $s_{1,u} = 1, \ldots, 5$ in the one-sided two-stage designs with $n_1 = n_2 = 5$.

Appendix A. Resolution of Eq. (10)

The observations space for the design $d_{\mathrm{Bin}\otimes 2}$ is $\mathscr{S} = \bigcup_{k=1}^{2} \mathscr{S}_k$ where $\mathscr{S}_1 = \{(1, x_1) : x_1 \leqslant s_{1,l}$ and/or $x_1 \geqslant s_{1,u}\}$ and $\mathscr{S}_2 = \{(2, y_2) : s_{1,l} + 1 \leqslant y_2 \leqslant n_2 + s_{1,l} + 1$ and/or $s_{1,u} - 1 \leqslant y_2 \leqslant n_2 + s_{1,u} - 1\}$. If $y_2 = 0$ and $y_m = v_m$ $(m = 1, 2)$, the solutions of (10) are 0 and 1, respectively. For any other observation, the solution is the value in θ of the intersection of the curves of the two holomorphic functions

$$g(\theta) = \frac{\theta}{y_m - v_m\theta},$$
$$h(\theta) = \frac{1}{r_2} \frac{1 + r_2 \mathrm{CBin}(n_1, \theta; [s_{1,l} + 1, s_{1,u} - 1])}{(n_1 - s_{1,u} + 1)\mathrm{Bin}(n_1, \theta; s_{1,u} - 1) - (n_1 - s_{1,l})\mathrm{Bin}(n_1, \theta; s_{1,l})}. \tag{A.1}$$

For example, consider the one-sided two-stage designs with $n_1 = n_2 = 5$ and $s_{1,u}$ in $\{1, \ldots, 5\}$. The left graphics in Fig. A1 shows the curves of $g(\theta)$ for $x_1 = 1, 2, 3, 4$ at stage 1 and $y_2 = 1, \ldots, 9$ at stage 2, and, the curves of $h(\theta)$ for $s_{1,u} = 1, \ldots, 5$.

For any (m, y_m), the curve of $h(\theta)$ crosses the vertical asymptote of $g(\theta)$ at the point $(y_m/v_m, h(y_m/v_m))$ (see enlargement in the right graphics). Then, the value of θ of the horizontal projection of this point on the curve of $g(\theta)$ approaches the θ-axis value of the intersection of the curves of $g(\theta)$ and $h(\theta)$. This value of θ is the solution of the equation $g(\theta) = h(y_m/v_m)$, i.e.

$$\theta = \frac{y_m h(y_m/v_m)}{1 + v_m h(y_m/v_m)}.$$

In one-sided designs with an upper stopping boundary $s_{1,u}$ (resp. with a lower stopping boundary $s_{1,l}$), the function h is positive (resp. negative) yielding a negative (resp. positive) correction relative to the sample proportion y_m/v_m. In two-sided designs with both $s_{1,l}$ and $s_{1,u}$, there are two asymptotes at the values of θ nullifying the denominator of (A.1). The correction relative to y_m/v_m is then negative or positive depending on the value of (m, y_m).

Appendix B. Details of calculations for Section 5

B.1. Derivation of the quadratic approximation MODE^Q

In this appendix, we derive the quadratic approximation MODE^Q for the two-stage designs with $n_1 = n_2$ and $s_{1,u} = 1$. From (11), the approximation MODE^A is

$$\hat{\theta}^{\text{MODE}^A} = \frac{Y_m}{v_m + n_1 \dfrac{1}{1 + (v_m/(v_m - Y_m))^{n_1}}}. \tag{B.1}$$

As mentioned in Table 2, the approximation MODE^A results in a slight overestimation of MODE. The simple expression (B.1) allows applying a correction on the denominator to refine the approximation, such as

$$\hat{\theta}^{\text{MODE}^A} \simeq \frac{Y_m}{v_m + n_1 \dfrac{1}{1 + (v_m/(v_m - Y_m))^{n_1 - 0.5}}} = \frac{Y_m}{v_m}\left(1 + \frac{n_1}{v_m}\frac{1}{1 + \dfrac{1}{(1 - Y_m/v_m)^{n_1 - 0.5}}}\right)^{-1}.$$

Noting that the component $(1 - Y_m/v_m)^{n_1 - 0.5}$ is small for reasonable high values of n_1, a first Taylor development in θ leads to

$$\hat{\theta}^{\text{MODE}^A} \simeq \frac{Y_m}{v_m}\left(1 - \frac{n_1}{v_m}\frac{(1 - Y_m/v_m)^{n_1 - 0.5}}{1 + (1 - Y_m/v_m)^{n_1 - 0.5}}\right).$$

Using the same argument, a second Taylor development yields the quadratic approximation MODE^Q in (14).

B.2. Derivation of the expectation of MODE^Q

The calculation of $D(\hat{\theta}; i, \theta)$ in (13) for MODE^Q leads to

$$\begin{aligned} D(\hat{\theta}^{\text{MODE}^Q}; i, \theta) &= \hat{\theta}^{\text{MODE}^Q}_{(1,i)} + (1 - \theta)^{n_1}\hat{\theta}^{\text{MODE}^Q}_{(2,i)} \\ &= \frac{i}{n_1}\left[1 - \left(1 - \frac{i}{n_1}\right)^{n_1 - 0.5} + \left(1 - \frac{i}{n_1}\right)^{2n_1 - 1} + \frac{1}{2}(1 - \theta)^{n_1} \right. \\ &\quad \left. - \frac{(1 - \theta)^{n_1}}{4}\left(1 - \frac{i}{2n_1}\right)^{n_1 - 0.5} + \frac{(1 - \theta)^{n_1}}{4}\left(1 - \frac{i}{2n_1}\right)^{2n_1 - 1}\right]. \end{aligned} \tag{B.2}$$

The approximation of the two terms $(1 - i/2n_1)$ by $(1 - i/n_1)$ in (B.2) allows a factorization, which is

$$\begin{aligned} D(\hat{\theta}^{\text{MODE}^Q}; i, \theta) &\simeq \frac{i}{n_1} + \frac{i}{2n_1}(1 - \theta)^{n_1} \\ &\quad + \left(1 + \frac{(1 - \theta)^{n_1}}{4}\right)\left(\frac{i}{n_1}\left(1 - \frac{i}{n_1}\right)^{2n_1 - 1} - \frac{i}{n_1}\left(1 - \frac{i}{n_1}\right)^{n_1 - 0.5}\right). \end{aligned} \tag{B.3}$$

Eq. (B.3) is now used in (13) to approach the expectation of MODE^Q. Using the variable $X \sim \text{Bin}(n_1, \theta)$, one gets

$$\begin{aligned} E_{d_{\text{Bin}\otimes 2},\theta}(\hat{\theta}^{\text{MODE}^Q}) &\simeq \theta + \frac{\theta}{2}(1 - \theta)^{n_1} + \left(1 + \frac{(1 - \theta)^{n_1}}{4}\right) \\ &\quad \times \left(\frac{1}{n_1^{2n_1}}E_\theta[X(n_1 - X)^{2n_1 - 1}] - \frac{1}{n_1^{n_1 + 0.5}}E_\theta[X(n_1 - X)^{n_1 - 0.5}]\right) \\ &= \theta + \frac{\theta}{2}(1 - \theta)^{n_1} + \left(1 + \frac{(1 - \theta)^{n_1}}{4}\right) \\ &\quad \times \left(\frac{1}{n_1^{2n_1}}E_{1-\theta}[(n_1 - X)X^{2n_1 - 1}] - \frac{1}{n_1^{n_1 + 0.5}}E_{1-\theta}[(n_1 - X)X^{n_1 - 0.5}]\right). \end{aligned}$$

References

Abramowitz, M., Stegun, I.A., 1964. Handbook of Mathematical Functions. Dover, New York.
Armitage, P., 1958. Numerical studies in the sequential estimation of a binomial parameter. Biometrika 45, 1–15.
Bernardo, J.M., Smith, A.F.M., 1994. Bayesian Theory. Wiley, New York.
Bunouf, P., Lecoutre, B., 2006. Bayesian priors in sequential binomial design. C.R. Acad. Sci. Paris, Ser. I 343, 339–344.
Bunouf, P., Lecoutre, B., 2007. On Bayesian confidence intervals in multistage designs. Technical Report, University of Rouen.
Chang, M.N., Therneau, T.M., Wieand, H.S., Cha, S.S., 1987. Designs for group sequential phase II clinical trials. Biometrics 43, 865–874.
Chang, M.N., Wieand, H.S., Chang, V.T., 1989. The bias of the sample proportion following a group sequential phase II clinical trial. Statist. Med. 8, 563–570.
de Cristofaro, R., 2004. On the foundations of likelihood principle. J. Statist. Plann. Inference 126, 401–411.
Emerson, S.S., Fleming, T.R., 1990. Parameter estimation following group sequential hypothesis testing. Biometrika 77, 875–892.
Jennison, C., Turnbull, B.W., 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall, London.
Jung, S., Kim, K.M., 2004. On the estimation of the binomial probability in multistage clinical trials. Statist. Med. 23, 881–896.
Whitehead, J., 1986. On the bias of maximum likelihood estimation following a sequential test. Biometrika 73, 573–581.
Ye, K., 1993. Reference priors when the stopping rule depends on the parameter of interest. J. Amer. Statist. Assoc. 88, 360–363.