Bayesian Analysis (2008) **3**, Number 2, pp. 269–296

Bayesian Sample Size Determination for Binomial Proportions

Cyr E. M'Lan*, Lawrence Joseph† and David B. Wolfson‡

Abstract. This paper presents several new results on Bayesian sample size determination for estimating binomial proportions, and provides a comprehensive comparative overview of the subject. We investigate the binomial sample size problem using generalized versions of the Average Length and Average Coverage Criteria, the Median Length and Median Coverage Criteria, as well as the Worst Outcome Criterion and its modified version. We compare sample sizes derived from highest posterior density and equal-tailed credible intervals. In some cases, we derive, for the first time, closed form sample size formulae, and where this is not possible, we describe various numerical approaches. These range in complexity from Monte Carlo simulations to more sophisticated curve fitting techniques, third order analytic approximations, and exact, but more computationally-intensive, methods. We compare the accuracy and efficiency of the different computational methods for each of the criteria and make recommendations about which methods are preferred. Finally, we consider, again for the first time, issues surrounding prior robustness on the choice of sample size. Examples are given throughout the text.

Keywords: Bayesian design, binomial distribution, Monte Carlo numerical approximation, robustness, sample size determination.

1 Introduction

Sample size determination for accurate estimation of a binomial parameter is arguably the most common design situation faced by statisticians. For example, consider designing a study to estimate the prevalence of osteoporosis in women aged 80 and older. Suppose that a previous study (Kmetic et al 2002) has provided an estimate of 42%, with 95% credible interval of (28%, 56%), but it is now desired to estimate the accuracy to within a total interval width of 5% (that is, ± 2.5%). What should the sample size be?

A large literature exists for this problem, including both frequentist (Lachin 1981, Lemeshow et al. 1990) and Bayesian approaches (Adcock 1988, Pham-Gia and Turkkan 1992, Joseph, Wolfson and du Berger 1995, Adcock 1997). From a frequentist viewpoint, the most common methods are based on confidence interval formulae derived from normal approximations to the binomial distribution, and require a point estimate of the binomial parameter as input into the sample size formula. There are at least three

*Department of Statistics, University of Connecticut, Storrs, CT, mailto:mlan@stat.uconn.edu

†Division of Clinical Epidemiology, McGill University Health Centre, Montreal, Quebec, Canada,mailto:lawrence.joseph@mcgill.ca

‡Department of Mathematics and Statistics, McGill University, Montreal, Quebec, Canada, mailto:wolfson@math.mcgill.ca

 DOI:10.1214/08-BA310

drawbacks to these methods: First, the binomial parameter will almost never be known with high accuracy at the planning stage of the experiment, since if it were accurately known the experiment would not need to be carried out. Therefore, one of the most crucial inputs cannot be accurately known, casting doubt on the sample size estimated. Second, the normal approximation is well-known to be inaccurate for small sample sizes, and even for large sample sizes when the binomial parameter is near 0 or 1 (Brown, Cai and DasGupta 2001). Finally, the sample size formulae are typically for confidence intervals rather than tolerance intervals, so that there is no probability statement about how often the desired 95% coverage will be achieved once the data are collected (but see Satten and Kupper 1990 for a notable exception).

Bayesian sample size methods use prior information about the binomial parameter rather than a point estimate, and fully account for the uncertainty in the predicted data, thus offering an attractive alternative to the frequentist formulae. Bayesian methods can avoid relying on normal approximations by using exact highest posterior density (HPD) intervals rather than less efficient equal-tailed intervals. However, using Bayesian methods is not without challenges, both numerical and in choosing from among the many Bayesian criteria that have appeared in the literature.

This paper presents several new results on Bayesian sample size determination for estimating binomial proportions, and provides a comprehensive comparative overview of the subject. Joseph et al. (1995) considered three HPD interval based criteria for Bayesian binomial sample size determination. We extend this early work considerably by applying a total of six criteria, generalizing two of those considered by Joseph et al. (1995), and considering three additional criteria. These criteria are reviewed in Section 2. In Section 3 we present four different methods to compute the various criterion functions, while Section 4 applies these methods to computing the sample sizes themselves, including providing sample sizes for the osteoporosis study introduced above. For each criterion, we compare sample sizes derived from highest posterior density to those from equal-tailed credible intervals. For the first time, closed form sample size formulae are derived for some cases. In other situations, we describe various numerical approaches, ranging in complexity from Monte Carlo simulations to more sophisticated curve fitting techniques, third order analytic approximations, and exact, but more computationally-intensive, methods. Within each criterion, we compare the accuracy and efficiency of the different computational techniques, and, case by case, make recommendations about which methods are preferred. In Section 5, again for the first time, we discuss robustness of the choice of sample size, both by considering additional criteria, and by defining new classes of prior distributions. We provide some concluding remarks in the final section.

2 Bayesian sample size criteria for binomial parameters

Let p be the binomial parameter to be estimated based on a sample size of n. For the rest of this paper, following Joseph et al. (1995), we assume the following prior-likelihood model: $p \sim \mathbf{Be}(a, b)$, $a, b > 0$, and $\mathbf{x}_n | p \sim \mathbf{Bin}(n, p)$, $n \geq 2$, where $\mathbf{Be}(a, b)$ indicates a beta distribution with parameters a and b, and $\mathbf{Bin}(n, p)$ represents the binomial distribution, with parameters n and p. As a result, the marginal predictive distribution of $\mathbf{x}_n$ is Beta-Binomial with

closed form solution described in kruschke

$$\boxed{p_{X_n}(\mathbf{x}_n|n,a,b) = \binom{n}{\mathbf{x}_n}\frac{\mathbf{B}(a+\mathbf{x}_n, n+b-\mathbf{x}_n)}{\mathbf{B}(a,b)}, \qquad \mathbf{x}_n = 0,1,\cdots,n,} \tag{1}$$

where $\mathbf{B}(a,b)$ indicates the beta function with parameters a and b. For a given sample data point $\mathbf{x}_n$, the posterior distribution of p, $\pi(p|\,\mathbf{x}_n,n,a,b)$, is $\mathbf{Be}(a+\mathbf{x}_n, n+b-\mathbf{x}_n)$. Let $\mathtt{HPD}_L(\mathbf{x}_n,n,a,b,l) = (u,v),\ u<v$, be the corresponding highest posterior density ($\mathtt{HPD}$) interval for p of given length l and let $\mathtt{HPD}_C(\mathbf{x}_n,n,a,b,1-\alpha)$ be an $\mathtt{HPD}$ interval for θ of given posterior coverage $1-\alpha$. Define $l^\star_{1-\alpha}(\mathbf{x}_n|\,n,a,b) = \int_{\mathtt{HPD}_C(\mathbf{x}_n,n,a,b,1-\alpha)} dp$ and $\alpha^\star_l(\mathbf{x}_n|\,n,a,b) = \int_{\mathtt{HPD}_L(\mathbf{x}_n,n,a,b,l)} \pi(p|\,\mathbf{x}_n,n,a,b)\,dp$ to be the actual length and the actual posterior coverage of an $\mathtt{HPD}$ interval of nominal coverage $1-\alpha$ and of nominal length l, respectively.

In this paper we consider the following six Bayesian sample size criteria: the average length criterion of order k, $\mathbf{ALC}_k$, the average coverage criterion of order k, $\mathbf{ACC}_k$, where k is an integer, the worst outcome criterion, **WOC**, the modified worst outcome criterion, **MWOC**, the median length criterion, **MLC**, and the median coverage criterion **MCC**. The $\mathbf{ALC}_k$, $\mathbf{ACC}_k$, **MCC**, and **MLC** were recently proposed by M'lan, Joseph and Wolfson (2006). The $\mathbf{ALC}_k$ and $\mathbf{ACC}_k$ are natural extensions of earlier criteria, **ALC** and **ACC** discussed by Joseph et al. (1995), while the **WOC** and **MWOC** were discussed by Joseph and Bélisle (1997). These criteria are now briefly reviewed in our context of binomial sample size calculations.

2.1 The k-th average length criterion, ALC_k

The $\mathbf{ALC}_k$ seeks the minimum n such that

$$\left(\sum_{\mathbf{x}_n=0}^{n}\left\{l^\star_{1-\alpha}(\mathbf{x}_n|\,n,a,b)\right\}^k p_{X_n}(\mathbf{x}_n|\,n,a,b)\right)^{1/k} \ \le\ l, \tag{2}$$

where p_{X_n} is given by (1). Thus, the $\mathbf{ALC}_k$, fixes the posterior coverage of $\mathtt{HPD}$ intervals to be $1-\alpha$, and finds the smallest n that provides a k-th mean length of at most l. When $k=1$, the $\mathbf{ALC}_k$ reduces to the **ALC**, and the $\mathbf{ALC}_2$ is asymptotically equivalent to the PGT $-\,(ii)$ criterion of Pham-Gia and Turkkan (1992), based on the marginal mean posterior variance. We show in Appendix A that $\mathbf{ALC}_{k=\infty}$ corresponds to the **WOC**.

2.2 The k-th average coverage criterion, ACC_k

The $\mathbf{ACC}_k$ finds the minimum sample size n such that

$$\left(\sum_{\mathbf{x}_n=0}^{n}\left\{\alpha^\star_l(\mathbf{x}_n|\,n,a,b)\right\}^k p_{X_n}(\mathbf{x}_n|\,n,a,b)\right)^{1/k} \ \ge\ 1-\alpha\,. \tag{3}$$

In contrast to the $\mathbf{ALC}_k$, the $\mathbf{ACC}_k$ fixes the length of the `HPD` interval to be l and determines the smallest n that provides a k-th average posterior coverage of at least $1-\alpha$. When $k=1$, the $\mathbf{ACC}_k$ reduces to the **ACC**.

2.3 The worst outcome criteria, WOC and MWOC

Stricter than either the $\mathbf{ALC}_k$ or the $\mathbf{ACC}_k$, the **WOC** and its modified version, the **MWOC** (Joseph et al. 1997) guarantee the desired posterior coverage and `HPD` length either over all anticipated data sets or over a subset, $\mathcal{S}_n$ of possible data sets, respectively. For the **MWOC**, $\mathcal{S}_n$ could be a $100(1-\gamma)\%$ credible region of the marginal predictive distribution, p_{X_n}. Fixing the length at l, the **MWOC** seeks the minimum n such that

$$\inf_{\mathbf{x}_n \in \mathcal{S}_n} \alpha_l^{\star}(\mathbf{x}_n |\, n, a, b) \;\geq\; 1-\alpha\,, \tag{4}$$

while the **WOC** simply sets $\mathcal{S}_n = \big\{0, 1, 2, \cdots, n\big\}$.

It is easy to see that the **MWOC** is also equivalent to minimizing

$$\sup_{\mathbf{x}_n \in \mathcal{S}_n} l_{1-\alpha}^{\star}(\mathbf{x}_n |\, n, a, b) \;\leq\; l \tag{5}$$

since both equations (4) and (5) lead to choosing the minimum value of the set $\mathfrak{N}$ of sample sizes,

$$\mathfrak{N} = \left\{ n: \sup_{\mathbf{x}_n \in \mathcal{S}_n} l_{1-\alpha}^{\star}(\mathbf{x}_n |\, n, a, b) \leq l, \quad \inf_{\mathbf{x}_n \in \mathcal{S}_n} \alpha_l^{\star}(\mathbf{x}_n |\, n, a, b) \geq 1-\alpha \right\}. \tag{6}$$

Hence the **WOC** and **MWOC** are simultaneously "fixed length" and "fixed coverage" criteria. When $\mathcal{S}_n = \big\{0, 1, 2, \cdots, n\big\}$, we can refer to (4) as the **WCOC** (worst outcome criterion defined in terms of coverage) and (5) as the **WLOC** (worst outcome criterion defined in terms of length). For computational reasons (see section 3), we prefer formulation (5).

2.4 The median coverage criterion, MCC, and median length criterion, MLC

The median coverage criterion, **MCC**, seeks the smallest n such that

$$\mathrm{med}_{0 \leq \mathbf{x}_n \leq n}\; \alpha_l^{\star}(\mathbf{x}_n |\, n, a, b) \;\geq\; 1-\alpha, \tag{7}$$

while the median length criterion, **MLC**, seeks the smallest n such that

$$\mathrm{med}_{0 \leq \mathbf{x}_n \leq n}\; l_{1-\alpha}^{\star}(\mathbf{x}_n |\, n, a, b) \;\leq\; l. \tag{8}$$

Although the above six criteria are defined in terms of `HPD` intervals which minimize the sample sizes given $1-\alpha$ and l, we also consider easier-to-compute equal-tailed intervals.

In the determination of sample sizes based on the credible interval criteria introduced here, there are two major practical issues: (i) Computation of the criterion function, and (ii) Determination of the optimal sample size using the criterion function. In the next two sections, we discuss various approaches to each of these two steps. We begin the two-stage procedure for sample size computations with a presentation of several approaches to the computation of the various criterion functions.

3 Computation of Bayesian criterion functions and HPD intervals for p

3.1 Exact `HPD` interval computation

Our discussion in this section focusses on computation of exact `HPD` intervals for p. Although numerical, the method is "exact" in the sense that one can specify any decimal accuracy that is desired. Without loss of generality, we discuss cases where $a + x_n > 1$ and $b + n - \mathbf{x}_n > 1$. All other cases lead to posterior densities which are monotonic or flat, for which computation of `HPD` intervals is straightforward.

When the length, l, of the `HPD` interval is fixed in advance, Corollary B.2 of Appendix B shows that any `HPD` interval $(p_1(\mathbf{x}_n), p_2(\mathbf{x}_n))$ for estimating p satisfies

$$\begin{aligned}\operatorname{logit}\bigl(p_1(\mathbf{x}_n)\bigr) \quad &= \quad \log\left(\exp\left\{\frac{(a-1+\mathbf{x}_n)}{n+a+b-2}\,\varpi\right\} - 1\right) \;- \\ &\qquad \log\left(\exp\left(\varpi\right) \,-\, \exp\left\{\frac{(a-1+\mathbf{x}_n)}{n+a+b-2}\,\varpi\right\}\right), \qquad (9)\end{aligned}$$

and $\operatorname{logit}\bigl(p_2(\mathbf{x}_n)\bigr) = \operatorname{logit}\bigl(p_1(\mathbf{x}_n)\bigr) + \varpi$ for some $\varpi > 0$, where $\operatorname{logit}(p) = \log\left(\frac{p}{1-p}\right)$. To emphasize the dependence of the interval $(p_1(\mathbf{x}_n), p_2(\mathbf{x}_n))$ on ϖ, we write $(p_1(\mathbf{x}_n,\varpi), p_2(\mathbf{x}_n,\varpi))$. To determine the optimal $\varpi^\star$, one can find the zero of the function $G_p(\varpi) = \bigl[p_2(\mathbf{x}_n,\varpi) - p_1(\mathbf{x}_n,\varpi) - l\bigr]^2$ via the Newton-Raphson method. Exploitation of the form in (9) leads to a considerable gain in efficiency over the generic `HPD` interval calculations previously used (Joseph et al, 1995; Hashemi et al., 1997). Let $\alpha_l^\star(\mathbf{x}_n|\, n, a, b)$ be the coverage of the `HPD` interval of length l corresponding to $\mathbf{x}_n$. We may repeat these `HPD` coverage calculations for each $\mathbf{x}_n = 0, 1, \cdots, n$, anticipating their use in calculating the various criterion functions.

When the coverage, $1-\alpha$, of the `HPD` interval is fixed in advanced, one only needs to find the zero of the function $G_p(\varpi) = \Bigl[F_p\bigl(p_2(\mathbf{x}_n,\varpi)\bigr) - F_p\bigl(p_1(\mathbf{x}_n,\varpi)\bigr) - (1-\alpha)\Bigr]^2$ to determine the value of $\varpi^\star$ which guarantees a coverage of $1-\alpha$. Let $l_{1-\alpha}^\star(\mathbf{x}_n|\, n, a, b)$ be the length of the `HPD` interval of fixed coverage $1-\alpha$, $(p_1(\mathbf{x}_n,\varpi^\star), p_2(\mathbf{x}_n,\varpi^\star))$. Again, we will typically need to repeat these calculations of `HPD` lengths for $\mathbf{x}_n = 0, 1, \cdots, n$.

Once the HPD intervals have been determined for each $\mathbf{x}_n = 0, 1, \cdots, n$, the criterion function is obtained as some functional of these HPD intervals with respect to the marginal distribution, $p_{X_n}(\mathbf{x}_n|n, a, b)$ in (1). For example, we will need to find the average, or the median, or the worst possible outcome, depending on the criterion used. More precisely, for the $\mathbf{ALC}_k$ and the $\mathbf{ACC}_k$, one must calculate, respectively,

$$\begin{aligned}
\mathrm{alc}_k(n,a,b) &= \Big(\sum_{\mathbf{x}_n=0}^{n} \left\{ l^{\star}_{1-\alpha}(\mathbf{x}_n|\, n,a,b) \right\}^k p_{X_n}(\mathbf{x}_n|\, n,a,b) \Big)^{1/k}, \text{ and} \\
\mathrm{acc}_k(n,a,b) &= \Big(\sum_{\mathbf{x}_n=0}^{n} \left\{ \alpha^{\star}_{l}(\mathbf{x}_n|\, n,a,b) \right\}^k p_{X_n}(\mathbf{x}_n|\, n,a,b) \Big)^{1/k}.
\end{aligned}$$

While one has to compute the HPD intervals for each $\mathbf{x}_n = 0, \cdots, n$, symmetry in the HPD intervals reduces the set to $\mathbf{x}_n = 0, \cdots, (n+b-a)/2$ or $\mathbf{x}_n = (n+b-a)/2, \cdots, n$, the choice depending on which set has more points. A similar symmetrical property holds for the marginal distribution $p_{X_n}(\mathbf{x}_n|\, n, a, b)$. For n large, this amounts to saving approximately half of the computational load.

3.2 Third order approximations to the length and coverage of HPD intervals

Rather than computing exact HPD intervals, one can consider first and third order approximations to credible intervals. First order approximations do not distinguish between HPD and equal-tailed intervals, leading to larger sample sizes. It is therefore worthwhile to investigate higher order approximations, as discussed by Welch and Peers (1963), Peers (1968), and Severini (1991), and Mukerjee and Dey (1993).

Define $l^{\mathtt{HPD}}_{1-\alpha}(\mathbf{x}_n|\, n, a, b)$ and $l^{\mathtt{EQ}}_{1-\alpha}(\mathbf{x}_n|\, n, a, b)$ to be the third order approximations of the length of HPD and equal tailed intervals, respectively. Let $z = z_{\alpha/2}$ be the upper $\alpha/2$ point of the standard normal distribution and let $N = n + a + b$ be the "extended" sample size, that is, the sum of the sample size and prior parameters. A modification of the third order approximation in Peers (1968) leads to the following result:

$$\begin{aligned}
l^{\mathtt{HPD}}_{1-\alpha}(\mathbf{x}_n|\, n,a,b) &= \frac{2}{N\sqrt{v_1(\mathbf{x}_n)}} \Big\{ z - (z^3+3z)\frac{v_2(\mathbf{x}_n)-1}{4N} + z\frac{v_2(\mathbf{x}_n)}{2N} + \\
&\qquad 5(z^3+3z)\frac{v_2(\mathbf{x}_n)-2}{18N} - z\frac{v_2(\mathbf{x}_n)-2}{N} \Big\}, \qquad (10)
\end{aligned}$$

where $v_1(\mathbf{x}_n) = \dfrac{1}{\mathbf{x}_n + a} + \dfrac{1}{n+b-\mathbf{x}_n}$, and $v_2(\mathbf{x}_n) = \dfrac{n+b-\mathbf{x}_n}{\mathbf{x}_n+a} + \dfrac{\mathbf{x}_n+a}{n+b-\mathbf{x}_n}$, for $x_n = 0, 1, \cdots, n$. The details are laid out in appendix C.

A third order estimate of the coverage of an HPD interval of length l, $\alpha^{\mathtt{HPD}}_{l}(\mathbf{x}_n|\, n, a, b)$, can then be obtained by first solving the third degree polynomial equation,

$l^{\mathtt{HPD}}_{1-\alpha}(\mathbf{x}_n|\, n, a, b) = l$, in z and recovering the coverage by setting $\alpha^{\mathtt{HPD}}_l(\mathbf{x}_n|\, n, a, b) = 2\Phi^{-1}(z) - 1$. The third order approximate lengths $l^{\mathtt{HPD}}_{1-\alpha}(\mathbf{x}_n|\, n, a, b)$ and the coverages $\alpha^{\mathtt{HPD}}_l(\mathbf{x}_n|\, n, a, b)$ are used in place of $l^{\star}_{1-\alpha}(\mathbf{x}_n|\, n, a, b)$ and $\alpha^{\star}_l(\mathbf{x}_n|\, n, a, b)$, respectively, in the expressions for the criterion functions.

A third order approximation for equal-tailed intervals is given by

$$l^{\mathtt{EQ}}_{1-\alpha}(\mathbf{x}_n|\, n, a, b) \quad = \quad l^{\mathtt{HPD}}_{1-\alpha}(\mathbf{x}_n|\, n, a, b) \;+\; 4z\,\frac{v_2(\mathbf{x}_n) - 2}{9N^2\sqrt{v_1(\mathbf{x}_n)}}\,, \tag{11}$$

and hence similar methods apply to this case.

3.3 HPD interval computation via Monte Carlo methods

General Monte Carlo techniques for approximating HPD intervals for a given coverage are presented by Tanner (1993), Hyndman (1996), and Chen and Shao (1999). The technique by Chen and Shao (1999) is the most efficient, carrying a computational load similar to equal-tailed intervals. The method applies when the posterior distribution is unimodal, the case here. For our beta prior-binomial likelihood model, the algorithm can be summarized as follows: Simulate M independent random values, $p_1, \cdots, p_M$, from $\mathbf{Be}(a + \mathbf{x}_n, n + b - \mathbf{x}_n)$. Consider the set of all Monte Carlo credible intervals of coverage $1 - \alpha$, $\big(p_{(j)},\, p_{(j+[(1-\alpha)M])}\big)$ for $j = 1, \cdots, M - [(1-\alpha)M]$, and choose the interval with the minimum length as an estimate of the HPD interval of fixed coverage $1 - \alpha$. As a result, $l^{\star}_{1-\alpha}(\mathbf{x}_n|\, n, a, b)$ is estimated by

$$\widehat{l}^{\star}_{1-\alpha}(\mathbf{x}_n|\, n, a, b) = \min_{1\le j\le M-[(1-\alpha)M]} \Big(p_{(j+[(1-\alpha)M])} - p_{(j)}\Big).$$

In the same spirit, when the length of an HPD interval is fixed in advance, M'lan et al. (2006) proposed estimating the HPD coverage, $\alpha^{\star}_l(\mathbf{x}_n|\, n, a, b)$, by

$$\widehat{\alpha}^{\star}_l(\mathbf{x}_n|\, n, a, b) = \max_{1\le j\le M} \frac{\#\big\{1 \le k \le M:\; p_j \le p_k \le p_j + l\big\}}{M}.$$

Here, one considers all Monte Carlo credible intervals $\big(p_{(j)},\, p_{(j)} + l\big)$, $j = 1, \cdots, n$, of length l and chooses the interval with the largest coverage.

To estimate the criterion functions, $\mathrm{alc}_k(n, a, b)$ and $\mathrm{acc}_k(n, a, b)$ for a given n, first generate m observations $(p_i, \mathbf{x}^i_n)$: generate $p_i \sim \mathbf{Be}(a, b)$, and then simulate $\mathbf{x}^i_n \sim \mathbf{Bin}(n,\, p_i)$ for $i = 1, \cdots, m$. For each $\mathbf{x}^i_n$ compute $\widehat{l}^{\star}_{1-\alpha}(\mathbf{x}^i_n|\, n, a, b)$ and $\widehat{\alpha}^{\star}_l(\mathbf{x}^i_n|\, n, a, b)$ as described above. The following approximations to the criterion functions are then available:

$$\widehat{\mathrm{alc}}_k(n, a, b) \quad = \quad \left(\frac{1}{m}\sum_{i=1}^{m}\big[\widehat{l}^{\star}_{1-\alpha}(\mathbf{x}^i_n|\, n, a, b)\big]^k\right)^{1/k} \quad \text{and} \tag{12}$$

$$\widehat{\mathrm{acc}}_k(n, a, b) \quad = \quad \left(\frac{1}{m}\sum_{i=1}^{m}\big[\widehat{\alpha}^{\star}_l(\mathbf{x}^i_n|\, n, a, b)\big]^k\right)^{1/k}. \tag{13}$$

Applications of these Monte Carlo methods in different settings can be found in Joseph et al. (1997), Wang and Gelfand (2002), and M'lan et al. (2006). Similar algorithms can be constructed for the **WOC**, **MWOC**, **MLC** and **MCC**.

4 Determination of the optimal sample size

In sections 4.1, 4.2, and 4.4, we show how each of the approaches in sections 3.1, 3.2, and 3.3, respectively, provide different methods of sample size determination. In section 4.3 we provide computationally fast and flexible sample size formulae for $\mathbf{ALC}_k$, **WOC**, and **MLC**. We then follow with general guidelines about when to use each method. Mathematical details are given in Appendix D.

4.1 Use of the exact expressions to the criterion functions

Except for very small n, all criterion functions discussed in Section 2 are strictly monotonic in n, ensuring a unique solution to our sample size problem. This, together with the asymptotic results in appendix D suggests that a variety of approaches can be used to compute the sample size. A bisectional search seems sufficient for the $\mathbf{ACC}_k$. For the length criteria, **ALC**, **WLOC**, **MWOC**, and **MLC**, we recommend using plots that are ostensibly linearly related to their sample sizes to determine the sample size. This suggestion relies mainly on the following observations:

- Graph (a) of Figure 1 indicates that there is a linear relationship between $\dfrac{1}{\text{alc}^2(n,a,b)}$ and n.
- Similar linear relationships seems to hold for $\dfrac{1}{\text{wloc}^2(n,a,b)}$ and $\dfrac{1}{\text{mlc}^2(n,a,b)}$, as illustrated in graph (b).

These observations suggest determining sample sizes as follows:

1. Estimate the linear function that passes through the points, $\big(n, 1/(\text{length criterion function})^2\big)$, by the method of least squares.
2. Equate this regression function to $\dfrac{1}{l^2}$ and solve for n.

Similar algorithms can be constructed for the **MLC**, **WLOC**, and **MWOC**.

One clear advantage of this technique over a bisectional search is that the estimated linear regression function is reusable to determine sample sizes for different values of l. In addition, the linear regression line displayed in graph (a) of Figure 1 suggests that the **ALC** sample size based on the prior parameter $(3,1)$ is always smaller than that of the pair $(3,2)$ for any length l. Graph (b) of Figure 1 suggests that $n_{\text{alc}} < n_{\text{mloc}} < n_{\text{woc}}$

for $(a, b) = (1, 1)$ irrespective of l. We have also observed that the slope of the line for the criteria **WLOC** and **MLC** is independent of (a, b), so again any estimated line for a given (a, b) can be used for other pairs of prior parameters and for any length. These observations are supported by Propositions D.3 and D.4 of Appendix D. Another disadvantage of a bisectional search is that if Monte Carlo estimates are used, the monotonicity in n of the criterion functions is not preserved.

4.2 Use of the third order approximations to the criterion functions

Instead of using "exact" expressions, one could use the third order approximations given by (10) to approximate the criterion function. While there can be some loss of accuracy, there can be considerable gains in running times. For example, for large **ALC** or **ACC** sample sizes, say $n = 10,000$, third order approximations can take as little as one second, compared to 30 minutes for an exact calculation. While 30 minutes may not be prohibitive, the run times considerably lengthen if prior robustness is a concern (see Section 5), and there are times when a quick approximation if preferable, for example during "live" consultations about study design.

Tables 1, 2, and 3 provide comparisons between the sample sizes obtained using the third order approximations and exact calculations for $(a, b) = (1, 1)$, $(2, 2)$, $(3, 3)$, $(4, 4)$, $(4, 1)$, $(4, 2)$, and $(4, 3)$. Because the sample size problem is symmetric in a and b, the sample sizes for $(a, b) = (1, 4), (2, 4)$, and $(3, 4)$ are equal to the sample sizes for $(a, b) = (4, 1), (4, 2)$, and $(4, 3)$, respectively. Tables 1 and 2 show that all methods perform well when the prior distribution for p is concentrated away from the endpoints 0 and 1. Table 3, however, indicates that the accuracy of the third order approximations is higher than that of the first order approximations, especially for values of p near 0 or 1, where posterior densities for p are highly asymmetric. Unlike the third order approximations, the first order approximations do not distinguish between `HPD` and equal-tailed intervals. Table 3 also indicates that the accuracy of the third order approximation is particular remarkable when $\min(a, b) \geq 2$ and when $n > a + b$; that is when the sample size dominates the prior information, $a + b$. A limited simulation study has confirmed this behavior. We have also carried out some preliminary testing of our third order approximations for non-integer values of a and b, finding that the sample sizes they provide remain accurate. For example, for a Jeffreys' prior with $(a, b) = (1/2, 1/2)$, exact computation and our third order approximation all lead to $n_{\text{alc}} = 151, 617, 265, 1071$ and $n_{\text{acc}} = 226, 910, 452, 1817$ for $(1 - \alpha, l) = (.95, .1), (.95, .05), (.99, .1), (.99, .05)$, respectively.

4.3 Sample size formulae

Propositions D.3 and D.4 suggest the following approximate $\mathbf{ALC}_k$, **WOC**, and **MLC** sample size formulae for p:

Theorem 4.1. *For the* $\mathbf{ALC}_k$*, an approximate sample size formula is:*

$$n_p = 4\frac{z^2_{\alpha/2}}{l^2}\left(\frac{\mathbf{B}(a+k/2,b+k/2)}{\mathbf{B}(a,b)}\right)^{2/k} - a - b, \qquad a>1,\ b>1. \tag{14}$$

Proof. The goal is to approximately solve the equation $\mathrm{alc}_k(n,a,b) = l$. Equation (23) in Proposition D.3 in appendix D suggests that $\mathrm{alc}_k(n,a,b) \approx \dfrac{2\,z_{\alpha/2}\;c^k_p(a,b)}{\sqrt{n+a+b}}$, where $c^k_p(a,b) = \{\mathbf{B}(a+k/2,b+k/2)/\mathbf{B}(a,b)\}^{1/k}$. Equate $\dfrac{2\,z_{\alpha/2}\;c^k_p(a,b)}{\sqrt{n+a+b}}$ to l and solve for n. This completes the proof. $\square$

Theorem 4.2. *For the* **WOC***, an approximate sample size formula is:*

$$n_p \quad = \quad \frac{z^2_{\alpha/2}}{l^2} - a - b\,, \qquad a,b>1. \tag{15}$$

Proof. Equation (24) in Proposition D.3 in the appendix D suggests that $\mathrm{wloc}(n,a,b) \approx \dfrac{z_{\alpha/2}}{\sqrt{n+a+b}}$. Equate $z_{\alpha/2}/\sqrt{n+a+b}$ to l and solve for n. This completes the proof. $\square$

Theorem 4.3. *For the* **MLC***, an approximate sample size formula is:*

$$n_p \quad = \quad \frac{3}{4}\frac{z^2_{\alpha/2}}{l^2} - \frac{1}{3}(a+b+3) \qquad a,b>1. \tag{16}$$

Proof. Proposition D.4 in the appendix D suggests that $\mathrm{mlc}(n,a,b) \approx 2z_{\alpha/2}\sqrt{\dfrac{(N+a+b)(3N-a-b)}{16N^2(N+1)}}$. Solve the approximate equation $\dfrac{(N+a+b)(3N-a-b)}{16N^2(N+1)} = \dfrac{l^2}{4\,z^2_{\alpha/2}}$ for n. An expansion of the solution as a Taylor series yields $n_p = \dfrac{3}{4}\dfrac{z^2_{\alpha/2}}{l^2} - 1 - \dfrac{1}{3}(a+b) + o\left(\dfrac{l^2}{z^2_{\alpha/2}}\right)$, which completes the proof. $\square$

These closed form sample size formulae allow direct comparisons between the different criteria and display how the choice of prior parameters a and b affect the sample sizes. In addition, these sample size formula can be compared to those arising from a frequentist approach. For example, we have $n_p + a + b = 4\dfrac{z^2_{\alpha/2}}{l^2}\dfrac{a\,b}{(a+b)(a+b+1)}$ for the $\mathbf{ALC}_2$ and $n_p + a + b = \dfrac{z^2_{\alpha/2}}{l^2}$ for the **WOC**, while frequentist sample sizes corresponding to setting $p = \dfrac{a}{a+b}$ and $p = \dfrac{1}{2}$ are $n_p = 4\dfrac{z^2_{\alpha/2}}{l^2}\dfrac{a\,b}{(a+b)^2}$ and $n_p = \dfrac{z^2_{\alpha/2}}{l^2}$, respectively.

We next discuss how fitting a line through Monte Carlo estimates of the criterion function values against n can be used to determine sample sizes.

4.4 Sample size determination via curve-fitting of Monte Carlo estimates

The approach described in this section depends on the Monte Carlo estimates of the criterion functions derived in section 3.3. To reduce Monte Carlo errors, Müller and Parmigiani (1995) and Müller (1999) advocate fitting local regression curves to Monte Carlo estimates of functions of interest in Bayesian optimal design. Applying this method to sample size calculations based on the **ACC** for example, we first plot the pairs $\big(n_i, \widehat{\alpha}_l^\star(\mathbf{x}_{n_i} | n_i, a, b)\big)$ for various n_i, randomly generated from an appropriate interval which includes the optimal sample size. We then fit a smooth curve to these points. The equation of the local smoothing curve, $f(n)$, describes the relation between the $\widehat{\mathrm{acc}}_k(n_i, a, b)$ and n_i, and solving the equation $f(n) = l$ deterministically in n provides an estimate of the sample size. We directly fit a local curve through the pairs $(n_i, \widehat{\mathrm{acc}}_k(n_i, a, b))$ for a grid of appropriately chosen points n_i, and use the information contained in predicted values from the curve to determine the sample size n, as illustrated in graph (b) of Figure 2. The Monte Carlo estimates $\widehat{\mathrm{acc}}_k(n_i, a, b)$ are defined in equation (13). A similar algorithm yields sample sizes for the median coverage criterion **MCC**.

For the $\mathbf{ALC}_k$ we make use of the linear relationship between $\dfrac{1}{\mathrm{alc}_k^2(n, a, b)}$ and n that was described in Section 4.1. We then fit a linear regression of the form $\dfrac{1}{\widehat{\mathrm{alc}}_k^{\,2}(n_i, a, b)} = e_1 + e_2 n_i$, estimate $\hat{e}_1$ and $\hat{e}_2$ through the method of least squares, and solve the equation $\hat{e}_1 + \hat{e}_2 n = \dfrac{1}{l^2}$ in n to determine the sample size. The Monte Carlo estimates $\widehat{\mathrm{alc}}_k(n_i, a, b)$ defined in equation (12) are plotted in graph (a) of Figure 2. Similar algorithms can be constructed for the other length-based criteria, **WOC**, **MWOC**, and **MLC**.

The construction of an efficient grid for the $\mathbf{ALC}_k$, **MLOC** and **WOC** relies mainly on an initial guess of the sample size, $\tilde{n}$, for example using a third order approximation and/or our sample size formula. To proceed, construct, an interval centered at $\tilde{n}$, say, of length $\max(200, \tilde{n}/10)$, and generate random uniform integer points, n_i, within this interval. Similar methods can be constructed for the $\mathbf{ACC}_k$ and **MCOC**.

4.5 General guidelines for sample size determination

Having defined a variety of computational methods for our many sample size criteria, in practice one needs to decide which methods are preferable. Our experience suggests the following:

1. For the $\mathbf{ALC}_k$, **WOC** and the **MLC**, use the approximate sample size formulae

in equations (14), (15), (16) when $a, b \geq 1$ are not too far apart. If a, b are far apart (for example when $a/b > 50$) and $\min(a, b) \geq 2$, use the third order approximation. This sample size should be recalculated, however, if it is smaller than $a+b$, via the exact computations as described in Section 4.1, or the regression-based Monte Carlo approach of Section 4.4. When $\min(a, b) < 1$, again use the exact computations or the regression-based Monte Carlo approach. For smaller sample sizes the exact method works well, but for running time issues, for sizes larger than 2000, the Monte Carlo approach is preferable, typically running in a few seconds, regardless of the sample size. We do not recommend the use of exact calculations for **WOC** and **MWOC**, as these criteria often lead to large sample sizes. The sample sizes provided by Monte Carlo approaches are random, so a repetition of the calculation can lead to different sample sizes. The computing time depends on the choice of m and M (see section 3.3), not so much on n. We suggest that at least five Monte Carlo sample sizes be calculated to provide an idea of the variability associated with the Monte Carlo algorithm. The values $m = M = 2000$ are often satisfactory, but these can be increased to, say, 5000, if the variance is too large.

2. For the $\mathbf{ACC}_k$ and **MCC**, use the third order approximation when $a, b \geq 1$ are not too far apart. If a, b are far apart, or if the sample size is smaller than $a+b$, use an exact computation or a Monte Carlo curve fitting approach. When $\min(a, b) < 1$, again use exact computations or the regression-based Monte Carlo approach.

The sample size formula and the third order approach are also easier to program compared to the exact and Monte Carlo approaches. The more accurate exact method is much harder to program, depending on many tuning parameters that need to be chosen appropriately to obtain convergence of the Newton-Raphson algorithm. If there is a hyperdistribution on a set of prior inputs (see section 5), we again recommend Monte Carlo techniques.

Table 1 provides several examples of Bayesian sample sizes for the **ALC**, **WOC** and **MLC**, allowing comparisons between the sizes given by the different criteria, and comparisons across computational methods. Table 2 provides similar results for the **ACC** and **MCC**. As expected, the **WOC** provides the largest sample sizes, with no consistent ordering seen for the other criteria. For symmetric prior distributions, all methods seem to lead to very similar sample sizes. As shown in Table 3, however, when skewed prior distributions are used, there can be substantial differences in sample size estimates from exact and approximate methods, and within the approximate methods, between the first and third order approximations.

4.6 Example

We now return to the example introduced in Section 1, and apply our methods to calculating sample sizes for accurate estimation of the prevalence of osteoporosis in elderly women.

Table 4 contains the sample sizes using a beta(20.5, 28.25) prior density, $1-\alpha = 0.95$, and $l = 0.05$. The results suggest sample sizes of either 1133, 1420, or 1487, depending on the criterion chosen. Of course, if one can afford the costs involved in recruiting 1487 subjects, the WOC sample size guarantees the desired width and coverage regardless of the data set that eventually arises, and so is the gold standard. In this case, there is only a small reduction in sample size to 1420 if one decides that attaining the desired width and coverage on average is sufficient, so one may prefer the WOC sample size. This occurs because the prior density concentrates not far from $p = 0.5$, the probability associated with the highest variance for a binomial distribution. In other examples with p concentrated away from 0.5, a larger drop may suggest the reverse decision. The **MCC** and **MLC** are substantially smaller, suggesting that 1133 will result in the desired width and coverage half the time, but that half the data sets will result in lengths and/or coverages that do not meet the target. The final sample size can be chosen based on the above considerations.

5 Robust Bayesian Sample Size Determination

Bayesian sample size calculations take the uncertainty inherent in the estimation of p into account. Yet these calculations still depend on the prior inputs, (a, b), both to generate the predictive distribution and to form the posterior distribution for p. While there is no universal notion of robustness, various ideas have been presented in the Bayesian literature. We present two such ideas and apply them to the sample size calculation problem. These methods are not "fully Bayesian", since more than one prior distribution is considered, but they are still useful in practice, in assessing the effect of prior choice on the sample size.

The first idea is to expand the range of prior distributions being considered. This can be done in different ways:

(i) Replace the single conjugate prior distribution, $\mathbf{Be}(a, b)$, by a class of conjugate prior distributions, $\Gamma = \{\mathbf{Be}(a, b),\ a, b \in \mathrm{T}\}$, where T is a subset of $\mathbb{R}^+$, and study how the sample sizes vary across Γ. In a spirit similar to the **WOC**, one can then use the largest sample size since it guarantees that the sample size criterion holds for all the prior distributions in Γ.

(ii) Enlarge the class of prior distributions by imposing a hyper-distribution, $\pi(a, b)$, $(a, b) \in \mathrm{T}$ on the prior parameters. Here the single sample size that is reported is robust in the sense that it considers more heterogeneous p's.

(iii) Select a new family of prior distribution that includes the Beta distributions as a special or limiting case. One example of such family is the three-parameter generalized Beta distribution, denoted $p \sim GB3(a, b, \lambda)$. See Chen and Novick (1984) for more details.

The second idea is to replace the sample size criterion with a more robust criterion, in the spirit of Adcock (1997, eq. 4.9). Suppose $\pi(a, b)$, $(a, b) \in \mathrm{T}$, is a hyper-distribution on the prior parameters (a, b). Replace, for example, the **ACC** by a criterion that finds the minimum sample size n such that

$$\int_{\mathrm{T}} \left\{ \sum_{\mathbf{x}_n=0}^{n} \alpha_l^\star(\mathbf{x}_n | \, n, a, b) \, p_{X_n}(\mathbf{x}_n | \, n, a, b) \right\} \pi(a, b) \, da \, db \; \geq \; 1 - \alpha \,, \tag{17}$$

where $\alpha_l^\star(\mathbf{x}_n | \, n, a, b)$ is the posterior coverage of an HPD interval of length l for p given the data $\mathbf{x}_n$ and a, b. Here we average both with respect to the predictive marginal distribution of $\mathbf{x}_n$ and π. Similar criteria could be defined for the **ALC**, **WOC**, and **MLC**.

In practice, sample size calculations via curve-fitting to Monte Carlo estimates is the best choice for case (iii) and for the revised criteria employing hyperpriors. The third order approximations are a good choice for cases (i) and (ii). Exact computations are case-specific, and, perhaps, too slow to be of practical use for assessing robustness, because the criteria functions need to be calculated many times for a single sample size. Below we discuss how some of the above robustness ideas can be implemented in practice.

Example 1: Suppose one is willing to assume a symmetric Beta distribution, which might be the case when one expects p to be near 0.5, but one is a priori uncertain as to how near to 0.5. Under this scenario, one could select the prior family $\Gamma = \Big\{ \mathbf{Be}(a,\, a),\ a > 0 \Big\}$. Graph a) of Figure 3 displays the **ACC**, **ALC**, and $\mathbf{ALC}_2$ third order approximations to the sample sizes for $1 \leq a \leq 50$ when $1 - \alpha = .95$ and $l = 0.1$. The maximum sample sizes over Γ for the **ACC**, **ALC**, and $\mathbf{ALC}_2$ are $n = 346, 345$, and 346, respectively. With $p = 0.5$, the corresponding frequentist sample size is 385, larger than all the reported Bayesian sample sizes. More generally, one could believe that p is near some π_0, and use the class $\Gamma = \Big\{ \mathbf{Be}(a+1,\, \frac{a}{\pi_0} - a + 1),\ a > 0 \Big\}$; these priors all having a mode at π_0.

Example 2: Another way to create a class of prior distributions is to set the amount of prior information, $a + b$, to a constant n_0, the number of prior observations to which the prior information is equivalent. Suppose, for example, one decides to set $n_0 = 6$ and to consider only the integer pairs (a, b) such that $a \leq b$ in order to reinforce the idea that $p \leq .5$ is more likely than $p > .5$. In this case, one would restrict attention to $(a, b) = (1, 5), (2, 4)$, and $(3, 3)$. **ALC** third order approximations to the sample sizes corresponding to $(a, b) = (1, 5), (2, 4)$, and $(3, 3)$ for $1 - \alpha = .95$ and $l = .1$, are 151, 277, and 319, while those corresponding to the **ACC** are 199, 297, and 327 respectively. In general, any pair (a, b) with $b = n_0 - a$ and $0 < a \leq n_0/2$ would be appropriate, leading to $\Gamma = \Big\{ \mathbf{Be}\big(n_0\pi,\, n_0(1-\pi)\big),\ 0 < \pi \leq 0.5 \Big\}$. Graph b) of Figure 3 displays the **ACC**, **ALC**, and $\mathbf{ALC}_2$ sample sizes for $1 \leq a \leq 3$ when $n_0 = 6$, $l = 0.1$ and $1 - \alpha = .95$. The maximum sample sizes for the **ACC**, **ALC**, and $\mathbf{ALC}_2$ are 327, 319, and 323, respectively.

Example 3: Suppose one fixes the amount of prior information to some $n_0 > 2$ as in example 2. It is well known in the Bayesian literature that one can gain robustness by adding a hierarchical level to the prior distribution. For example, one could consider $p|\pi \sim \mathbf{Be}(n_0\pi,\, n_0(1-\pi))$ and $\pi \sim \mathbf{Be}(\delta, \gamma; \pi_l, \pi_u)$ with δ, γ, π_l and π_u known quantities and where $0 \le \pi_l < \pi_u \le 1$. The notation $\pi \sim \mathbf{Be}(\delta, \gamma; \pi_l, \pi_u)$ represents a Beta distribution with support on $[\pi_l, \pi_u]$, $\pi_l < \pi_u$. This prior model is known as an "Imprecise Beta Model". In this context, one could use the robust average coverage criterion in (17) or its average length counterpart of order k, for example,

$$\left(\int_{\mathcal{X}_n} \left\{ l^{\star}_{1-\alpha}(\mathbf{x}_n |\, n, n_0\pi, n_0(1-\pi)) \right\}^k p_{X_n,\pi}(\mathbf{x}_n, \pi |\, n)\, d\mathbf{x}_n \right)^{1/k} \le\ l, \tag{18}$$

where $l^{\star}_{1-\alpha}(\mathbf{x}_n |\, n, n_0\pi, n_0(1-\pi))$ is the length of an HPD interval of coverage $1-\alpha$ for p given π and $\mathbf{x}_n$. An approximate sample size formula for (18) is

$$\begin{aligned} n_p \;=\; & 4\frac{z^2_{\alpha/2}}{l^2} \left(\int_{\pi_l}^{\pi_u} \frac{\mathbf{B}(n_0\pi + k/2, n_0(1-\pi) + k/2)}{\mathbf{B}(n_0\pi, n_0(1-\pi))} \times \right. \\ & \left. \qquad \frac{(\pi - \pi_l)^{\delta-1}(\pi_u - \pi)^{\gamma-1}}{\mathbf{B}(\delta,\gamma)(\pi_u - \pi_l)^{\delta+\gamma-1}}\, d\pi \right)^{2/k} - n_0, \end{aligned} \tag{19}$$

when $n_0\pi_l \ge 1$ and $n_0(1-\pi_u) \ge 1$. For example, let $\delta = 3$, $\gamma = 5$, $\pi_l = 0.0$, $\pi_u = 1.0$, $n_0 = 6$, $1-\alpha = 0.95$, and $l = 0.1$. The curve-fitting approach based on $m = M = 2000$ leads to sample sizes of 246, 262 and 277 for the $\mathbf{ALC}, \mathbf{ALC}_2$ and $\mathbf{ACC}$, respectively. Equation (19) suggests sample sizes of 252 and 269, respectively, for the $\mathbf{ALC}$ and $\mathbf{ALC}_2$. These two values were used to generate the grid of appropriately chosen n's for our curve-fitting approach.

6 Conclusion

While the binomial sample size problem has been investigated in the past, we provide a unified treatment, which has previously not been available. We examine several new analytic and computational methods specifically tailored to this problem, compare them for efficiency and accuracy, and make recommendations as to which method is best for each situation. We discuss a third order approximation that is simple to implement and highly accurate when $n > a + b$. For the first time, we present accurate closed form formulae for several criteria. We point out that there is a linear relationship between sample sizes and the square-inverse-of-length based criterion functions. We use this linear relationship in the context of curve-fitting to Monte Carlo sample size estimates. Although these linear relations have not been previously discussed in the context of Bayesian sample sizes, they are typical of frequentist sample size calculations. Software implementing all of the methods discussed in this paper is available from the first author.

Clearly, deciding which criterion function to use for sample size determination will be a question of personal taste or, perhaps, depend on the particular situation. For example, if it is important to accommodate a possible, although unlikely, catastrophic data set, then one might use one of the "worst outcome" criteria. In practice, the user can compute sample sizes across a range of different criteria, and based on the information provided by all calculations, reach a compromise as to the size of the sample needed and the acceptability of the criterion.

Appendix

A Proof that $\mathbf{ALC}_\infty = \mathbf{WOC}$

Proposition A.1. *Let $n(k, 1-\alpha, l)$ denote the sample sizes under the $\mathbf{ALC}_k$. Then*

$$n(k+1, 1-\alpha, l) \quad \geq \quad n(k, 1-\alpha, l)\,.$$

Proof. The proof of this proposition is entirely based on the natural ordering of the L_k-norm. Without loss of generality, assume that $n = n(k+1, 1-\alpha, l) < \infty$. Then

$$\sum_{\mathbf{x}_n=0}^{n} l^\star_{1-\alpha}(\mathbf{x}_n|\, n, a, b)\big\}^{k+1}\, p_{X_n}(\mathbf{x}_n|\, n, a, b) \;\leq\; l^{k+1}.$$

This implies also that n satisfies

$$\sum_{\mathbf{x}_n=0}^{n} \big\{l^\star_{1-\alpha}(\mathbf{x}_n|\, n, a, b)\big\}^{k}\, p_{X_n}(\mathbf{x}_n|\, n, a, b) \;\leq\; l^{k}, \tag{20}$$

because the L_k-norm increases monotonically as k increases. Therefore $n(k+1, l, 1-\alpha)$ is larger than the smallest bound of all n satisfying equation (20); that is, $n(k, l, 1-\alpha)$. $\square$

Hence the sequence of sample sizes $n(k, 1-\alpha, l)$ is increasing as k increases.

Proposition A.2. *Assume $\mathcal{X}_n$ is a discrete set and that the sequence of sample sizes $n(k, 1-\alpha, l)$ is bounded. Then* $\mathbf{WOC} = \mathbf{ALC}_\infty$.

Proof. Let $w_1, w_2, \cdots, w_m$ and $a_1, a_2, \cdots, a_m$ be sequences of m non-negative real numbers with $\sum_{i=1}^m w_i = 1$ and $\sup a_i < \infty$. Then,

$$\lim_{k\longrightarrow\infty} \left(\sum_{i=1}^{m} w_i a_i^k\right)^{1/k} = \sup a_i\,. \tag{21}$$

For let j be the index such that $a_j = \sup a_i$. Then, we have $w_j^{1/k} a_j \leq \left(\sum_{i=1}^{m} w_i a_i^k\right)^{1/k} \leq a_j$ and (21) follows as k tends to ∞.

Since the sequence of sample sizes $n(k, 1-\alpha, l)$ is bounded it must converge, which in turn implies that the criterion $\mathbf{ALC}_\infty$ is well-defined. A straightforward application of equation (21) leads to

$$\sup_{0 \le \mathbf{x}_n \le n} l^\star_{1-\alpha}(\mathbf{x}_n | \, n, a, b) \;=\; \lim_{k \longrightarrow \infty} \left(\sum_{\mathbf{x}_n = 0}^{n} \left\{ l^\star_{1-\alpha}(\mathbf{x}_n | \, n, a, b) \right\}^k p_{X_n}(\mathbf{x}_n | \, n, a, b) \right)^{1/k},$$

for any given n, and the proof is complete. □

B Method for the exact computation of HPD intervals for p

Proposition B.1. *Let ϕ be a random variable with density $f_\phi(\omega) \propto \dfrac{\exp(\alpha\,\omega)}{(1+\exp(\omega))^{\alpha+\beta}}$, $-\infty < \omega < \infty$. The* `HPD` *interval of given length l is the interval $(\phi_s, \phi_s + l)$ where*

$$\phi_s = \log\Big(\exp\big(\tfrac{\alpha}{\alpha+\beta}\, l\big) - 1\Big) - \log\Big(\exp(l) - \exp\big(\tfrac{\alpha}{\alpha+\beta}\, l\big)\Big).$$

Proof. It is not difficult to prove that the posterior density of ϕ is continuous and strongly unimodal (thus unimodal) on $\mathbb{R}$, and $\lim_{|\omega| \to \infty} f_\phi(\omega) = 0$. Under such conditions, the `HPD` region of size l is an interval of length l which must satisfy $f_\phi(\phi_s) = f_\phi(\phi_s + l)$. As a result, ϕ_s satisfies $\dfrac{\exp(\phi_s + l) + 1}{1 + \exp(\phi_s)} = \exp\big(\tfrac{\alpha}{\alpha+\beta}\, l\big)$. Hence, $\phi_s = \log\Big(\exp\big(\tfrac{\alpha}{\alpha+\beta}\, l\big) - 1\Big) - \log\Big(\exp(l) - \exp\big(\tfrac{\alpha}{\alpha+\beta}\, l\big)\Big)$. □

Corollary B.2. *Let $p \sim \mathbf{Be}(\alpha, \beta)$ with $\alpha, \beta > 1$. Let (p_1, p_2) be an* `HPD` *interval. There exists $\varpi > 0$ such that $logit(p_1) = \log\Big(\exp\big(\tfrac{\alpha-1}{\alpha+\beta-2}\,\varpi\big) - 1\Big) - \log\Big(\exp(\varpi) - \exp\big(\tfrac{\alpha-1}{\alpha+\beta-2}\,\varpi\big)\Big)$ and $logit(p_2) = logit(p_1) + \varpi$, where $logit(p) = \log\left(\dfrac{p}{1-p}\right)$.*

Proof. When $\alpha, \beta > 1$, f_p is continuous and unimodal on the interval $(0, 1)$, and $\lim_{p \to 0} f_p(\omega) = 0 = \lim_{p \to 1} f_p(\omega)$. Under these conditions, every `HPD` region is an interval that satisfies $f_p(p_1) = f_p(p_2)$ or equivalently $g_p\big(\text{logit}(p_1)\big) = g_p\big(\text{logit}(p_2)\big)$ with $g_p(\omega) = f_p\left(\dfrac{\exp(\omega)}{1+\exp(\omega)}\right) = \dfrac{\exp\{(\alpha-1)\,\omega\}}{(1+\exp(\omega))^{\alpha+\beta-2}}$. Now apply Proposition B.1. □

C Third order approximations to the length of HPD intervals when a and b are integers

Although one might contemplate using third order approximations to `HPD` and equal-tailed intervals as in Peers (1968), these results alone are not sufficient because they are undefined for $\mathbf{x}_n = 0$ and $\mathbf{x}_n = n$. This phenomenon is well-known in the frequentist literature where a popular practice is to replace $\mathbf{x}_n$ and n by $\mathbf{x}_n + \kappa/2$ and $n + \kappa$, respectively, to avoid problems at these endpoints. The values $\kappa = 1, 2, 4$ are used most often in practice (Brown, Cai and DasGupta, 2001). We instead employ a modification of the results in Peers (1968). We first observe that the family of posterior distributions for p given $\mathbf{x}_n \in \{0, 1, \cdots, n\}$ under our model, Model 1, $\mathbf{x}_n | \, p \sim \mathbf{Bin}(n, p)$ and $p \sim \mathbf{Be}(a, b)$ for a, b positive integers, is a subset of the family of posterior distribution for p given $\mathbf{y}_n \in \{0, 1, \cdots, n + a + b\}$, obtained under Model 2, $\mathbf{y}_n | \, p \sim \mathbf{Bin}(n + a + b, p)$ and $p \sim \mathbf{Be}(0, 0)$. We therefore use the third order approximations in Peers (1968) under Model 2 for $\mathbf{y}_n = a + 0, a + 1, \cdots, a + n$, to approximate intervals under Model 1 for $\mathbf{x}_n = 0, 1, \cdots, n$, respectively.

Define $l^{\mathtt{HPD}}_{1-\alpha}(\mathbf{x}_n | \, n, a, b)$ and $l^{\mathtt{EQ}}_{1-\alpha}(\mathbf{x}_n | \, n, a, b)$ to be the third order approximations of the lengths of HPD and equal tailed intervals, respectively. Let $z = z_{\alpha/2}$ be the upper $\alpha/2$ point of the standard normal distribution and let $N = n + a + b$ be the "extended" sample size. We have:

$$\begin{aligned} l^{\mathtt{HPD}}_{1-\alpha}(\mathbf{x}_n | \, n, a, b) &= \frac{2}{N\sqrt{v_1(\mathbf{x}_n)}} \left\{ z - (z^3 + 3z)\frac{v_2(\mathbf{x}_n) - 1}{4N} + z\frac{v_2(\mathbf{x}_n)}{2N} + \right. \\ &\qquad \left. 5(z^3 + 3z)\frac{v_2(\mathbf{x}_n) - 2}{18N} - z\frac{v_2(\mathbf{x}_n) - 2}{N} \right\}, \\ l^{\mathtt{EQ}}_{1-\alpha}(\mathbf{x}_n | \, n, a, b) &= l^{\mathtt{HPD}}_{1-\alpha}(\mathbf{x}_n | \, n, a, b) + 4z\frac{v_2(\mathbf{x}_n) - 2}{9N^2\sqrt{v_1(\mathbf{x}_n)}}. \end{aligned}$$

where $v_1(\mathbf{x}_n) = \dfrac{1}{\mathbf{x}_n + a} + \dfrac{1}{n + b - \mathbf{x}_n}$, and $v_2(\mathbf{x}_n) = \dfrac{n + b - \mathbf{x}_n}{\mathbf{x}_n + a} + \dfrac{\mathbf{x}_n + a}{n + b - \mathbf{x}_n}$, for $x_n = 0, 1, \cdots, n$.

D Limiting results for the ACC, ALC$_k$, MLC and WOC

D.1 Preliminary results

The lemma below proved in Billingsley (1995, p.338 and 340), is reproduced here for convenience.

Lemma D.1. *Let X be a random variable and $\{X_n\}_{n\geq 1}$ be a sequence of random variables such that $X_n \Longrightarrow^d X$ (convergence in distribution).*

a) *Assume furthermore that the X_n are uniformly integrable. Then X is integrable and $\mathbf{E}(X_n) \longrightarrow \mathbf{E}(X)$.*

b) *If $|X_n|$ is uniformly bounded, then the X_n are uniformly integrable.*

c) *Let h be a Borel function and $(h_n)_{n\geq 1}$ a sequence of Borel functions. Denote D*

the set of x for which $h_n(x_n) \longrightarrow h(x)$ fails for some sequence $x_n \longrightarrow x$. Suppose that $D \subset \mathbb{R}$ and $\Pr(X \in D) = 0$. Then $h_n(X_n) \Longrightarrow^d h(X)$.

In section D.2, we provide four asymptotic results and we use these results to derive asymptotic expressions for the criterion functions.

D.2 Asymptotic limits for the criterion functions

The asymptotic expressions of Propositions D.2, D.3 and D.4 below also provide reassurance that our sample size criteria are well-defined. We start with the **ACC**, followed by the $\mathbf{ALC}_k$, **WOC**, and **MLC**.

Proposition D.2. *When $a, b > 1$,*

$$\lim_{n \longrightarrow \infty} acc(n, a, b) = 1\,. \tag{22}$$

Proof. Let $\mathbf{I}_{(u,v)}(\omega)$ be the indicator function for the set (u, v). Let $f(\mathbf{x}_n|p)$ be the binomial probability mass function $\mathbf{x}_n|p \sim \mathbf{Bin}(n, p)$, and let $f(p)$ be the density of a Beta random variable with parameters a and b. We have

$$\begin{aligned} \mathrm{acc}(n, a, b) &= \sum_{\mathbf{x}_n=0}^{n} \left(\int_{p(\mathbf{x}_n,l)}^{p(\mathbf{x}_n,l)+l} \pi(p|\,\mathbf{x}_n, n, a, b)\, dp \right) p_{X_n}(\mathbf{x}_n|n, a, b)\ , \\ &= \int_0^1 \sum_{\mathbf{x}_n=0}^{n} \mathbf{I}_{\left(p(\mathbf{x}_n,l),\, p(\mathbf{x}_n,l)+l\right)}(p)\, f(\mathbf{x}_n|p)\, f(p) dp\ , \end{aligned}$$

where $\big(p(\mathbf{x}_n, l),\, p(\mathbf{x}_n, l) + l\big)$ is the `HPD` interval of length l. Given p, the binomial series, $\sum_{\mathbf{x}_n=0}^{n} \mathbf{I}_{\left(p(\mathbf{x}_n,l),\, p(\mathbf{x}_n,l)+l\right)}(p)\, f(\mathbf{x}_n|p)$, is uniformly bounded by 1, so that the Lebesgue dominated convergence theorem implies

$$\lim_{n \longrightarrow \infty} \mathrm{acc}(n, a, b) = \int_0^1 \lim_{n \longrightarrow \infty} \sum_{\mathbf{x}_n=0}^{n} \mathbf{I}_{\left(p(\mathbf{x}_n,l),\, p(\mathbf{x}_n,l)+l\right)}(p)\, f(\mathbf{x}_n|p)\ f(p) dp\ ,$$

if $\lim_{n \longrightarrow \infty} \sum_{\mathbf{x}_n=0}^{n} \mathbf{I}_{\left(p(\mathbf{x}_n,l),\, p(\mathbf{x}_n,l)+l\right)}(p)\, f(\mathbf{x}_n|p)$ exists almost everywhere. Since the posterior distribution is unimodal and takes the same value at the endpoints when $a, b > 1$, $p(\mathbf{x}_n, l)$ is the unique solution of the equation $\log f_p(\psi + l|\,\mathbf{x}_n, n) = \log f_p(\psi|\,\mathbf{x}_n, n)$, or equivalently, $\dfrac{\log(1-\psi) - \log(1-\psi-l)}{\log(\psi+l) - \log(\psi)} = \dfrac{a-1+\mathbf{x}_n}{b-1+n-\mathbf{x}_n} = \zeta(\psi)$, say. The function ζ has a continuous inverse function, ζ^{-1} and we have $p(\mathbf{x}_n, l) = \zeta^{-1}\left(\dfrac{a-1+\mathbf{x}_n}{b-1+n-\mathbf{x}_n}\right)$. Thus $\lim_{n \longrightarrow \infty} p(\mathbf{x}_n, l) = \zeta^{-1}(\frac{p}{1-p})$ with probability 1. Hence, a straightforward application of Lemma D.1 implies

$$\lim_{n \longrightarrow \infty} \mathrm{acc}(n, a, b) = \int_0^1 \mathbf{I}_{\left(\zeta^{-1}(p/(1-p)),\, \zeta^{-1}(p/(1-p))+l\right)}(p)\ \ f(p) dp\ .$$

Now note that $(\zeta^{-1}(p), \zeta^{-1}(p)+l)$ is the unique HPD interval of length l of the posterior distribution $g(\psi) \propto \psi^p(1-\psi)^{1-p}$ which has mode at p. Hence

$$\mathbf{I}_{\left(\zeta^{-1}(p/(1-p)),\, \zeta^{-1}(p/(1-p))+l\right)}(p) = 1\,.$$

This completes the proof. □

Since the lengths of HPD intervals for p have no closed form expression, we use the usual first order approximations to HPD intervals to obtain limiting expressions for the $\mathbf{ALC}_k$, **WOC**, and **MLC**. Let $l_p(\mathbf{x}_n|n,a,b) = 2z_{\alpha/2}\sqrt{\mathbf{Var}(p|\,\mathbf{x}_n,n,a,b)}$ be the length of this approximate interval, where $\mathbf{Var}(p|\,\mathbf{x}_n,n,a,b) = \dfrac{(a+\mathbf{x}_n)(n+b-\mathbf{x}_n)}{(n+a+b)^2(n+a+b+1)}$ and $z_{\alpha/2}$ is the upper $\alpha/2$ point of the standard normal distribution.

Proposition D.3. *For $a, b > 1$, we have*

$$\lim_{n\longrightarrow\infty} \sqrt{N}\left\{\frac{\mathbf{E}\big[l_p^k(X_n,n,a,b)\big]}{2^k z_{\alpha/2}^k}\right\}^{1/k} = \left\{\frac{\mathbf{B}(a+k/2,b+k/2)}{\mathbf{B}(a,b)}\right\}^{1/k}. \tag{23}$$

In particular,

$$\lim_{n\longrightarrow\infty} \sqrt{N}\left[\lim_{k\longrightarrow\infty}\left\{\frac{\mathbf{E}\big[l_p^k(X_n,n,a,b)\big]}{2^k z_{\alpha/2}^k}\right\}^{1/k}\right] = \frac{1}{2}\,. \tag{24}$$

Proof. Set $Y_n = \dfrac{X_n}{n}$, and let $\mathcal{F}_\backslash = \left\{\prime, \frac{\infty}{\backslash}, \frac{\in}{\backslash}, \cdots, \frac{\backslash-\infty}{\backslash}, \infty\right\}$ be the set of points where the probability mass function of Y_n is positive. Application of Theorem B.1 in M'lan, Joseph, and Wolfson (2006) leads to $Y_n \Longrightarrow^d p$, where $p \sim \mathbf{Be}(a,b)$. Setting $h_n(y) = \dfrac{(a+ny)^{k/2}(n+b-ny)^{k/2}}{N^{k/2}(N+1)^{k/2}}$ and $h(y) = y^{k/2}(1-y)^{k/2} = \lim_n h_n(y)$, $y \in [0,1]$, Lemma D.1 implies that $h_n(Y_n) \Longrightarrow^d h(p)$. In addition, the sequence of functions $h_n(x)$ is uniformly bounded by 1. These results together lead to

$$\begin{aligned}
N^{k/2}\frac{\mathbf{E}\big[l_p^k(X_n,n,a,b)\big]}{2^k z_{\alpha/2}^k} &= \sum_{\mathbf{x}_n\in\mathcal{X}_n} \frac{(a+\mathbf{x}_n)^{k/2}(n+b-\mathbf{x}_n)^{k/2}}{N^{k/2}(N+1)^{k/2}}\, p_{X_n}(\mathbf{x}_n|n,a,b)\,,\\
&= \sum_{y_n\in\mathcal{F}_\backslash} h_n(\mathbf{y}_n)\, p_{Y_n}(\mathbf{y}_n|n,a,b) = \mathbf{E}\big(h_n(Y_n)\big)\,,\\
&\longrightarrow \int_0^1 h(y)\frac{y^{a-1}(1-y)^{b-1}}{\mathbf{B}(a,b)}\,dy = \frac{\mathbf{B}(a+k/2,b+k/2)}{\mathbf{B}(a,b)}\,,
\end{aligned}$$

which completes the proof of the first result.

The second result, $\lim_{k\longrightarrow\infty}\left\{\dfrac{\mathbf{B}(a+k/2,b+k/2)}{\mathbf{B}(a,b)}\right\}^{1/k} = \dfrac{1}{2}$ follows from Stirling's formula. □

Proposition D.4. *Assume that $a, b > 1$. We have*

$$\sqrt{N}\frac{\mathrm{med}_{\mathbf{x}_n \in \mathcal{X}_n}\Big[l_p(\mathbf{x}_n, n, a, b)\Big]}{2\, z_{\alpha/2}} \quad \approx \quad \sqrt{\frac{(N+a+b)(3N-a-b)}{16N(N+1)}} \tag{25}$$

and converges to $\dfrac{\sqrt{3}}{4}$ *as* $n \longrightarrow \infty$*, where the median is over the set* $\{0, 1, \cdots, n\}$.

Proof. The median of the posterior variances for p for $\mathbf{x}_n = 0, 1, \cdots, n$, are attained approximately at $x^\star = \dfrac{n + 2(b-a)}{4}$. Here we have

$$\mathbf{Var}(p|\, x^\star, n, a, b) = \frac{(N+a+b)(3N-a-b)}{16N^2(N+1)}\,,$$

which completes the proof. □

Equations (23), (24) and (25) demonstrate that the convergence rates to zero of the criterion functions $\mathrm{alc}_k(n, a, b)$, $\mathrm{wloc}(n, a, b)$ and $\mathrm{mlc}(n, a, b)$ are of the order $\dfrac{1}{\sqrt{n}}$ when $a, b > 1$. These results also imply that, asymptotically, $\mathrm{alc}_k(n, a, b) \approx \dfrac{2\, z_{\alpha/2}\, c_p^k(a, b)}{\sqrt{n+a+b}}$, $\mathrm{wloc}(n, a, b) \approx \dfrac{z_{\alpha/2}}{\sqrt{n+a+b}}$ and $\mathrm{mlc}(n, a, b) \approx 2\, z_{\alpha/2}\sqrt{\dfrac{(N+a+b)(3N-a-b)}{16N^2(N+1)}}$, where $c_p^k(a, b) = \Big\{\mathbf{B}(a + k/2, b + k/2) \,/\, \mathbf{B}(a, b)\Big\}^{1/k}$. The above asymptotic expressions for $\mathrm{alc}_k(n, a, b)$, $\mathrm{wloc}(n, a, b)$ and $\mathrm{mlc}(n, a, b)$ are exploited in Section 4.3 to derive approximate sample size formulae.

References

Adcock, C. J. (1988). "A Bayesian Approach to Calculating Sample Sizes." *The Statistician: Journal of the Institute of Statisticians*, 37: 433–439.

— (1997). "Sample Size Determination: A Review." *The Statistician: Journal of the Institute of Statisticians*, 46: 261–283.

Billingsley, P. (1995). *Probability and Measure*. John Wiley & Sons.

Brown, L. D., Cai, T. T., and DasGupta, A. (2001). "Interval Estimation for a Binomial Proportion (Pkg: P101-133)." *Statistical Science*, 16(2): 101–133.

Chen, J. J. and Novick, M. R. (1984). "Bayesian Analysis for Binomial Models with Generalized Beta Prior Distributions." *Journal of Educational Statistics*, 9: 163–175.

Chen, M.-H. and Shao, Q.-M. (1999). "Monte Carlo Estimation of Bayesian Credible and HPD Intervals." *Journal of Computational and Graphical Statistics*, 8: 69–92.

Hashemi, L., Nandram, B., and Goldberg, R. (1997). "Bayesian Analysis for a Single 2×2 Table." *Statistics in Medicine*, 16: 1311–1328.

Hyndman, R. J. (1996). "Computing and Graphing Highest Density Regions." *The American Statistician*, 50: 120–126.

Joseph, L., Belisle, P., and Bélisle, P. (1997a). "Bayesian Sample Size Determination for Normal Means and Differences between Normal Means." *The Statistician: Journal of the Institute of Statisticians*, 46: 209–226.

Joseph, L., Belisle, P., Du Berger, R., and Bélisle, P. (1997b). "Bayesian and Mixed Bayesian/likelihood Criteria for Sample Size Determination." *Statistics in Medicine*, 16: 769–781.

Joseph, L., Wolfson, D. B., and Berger, R. d. (1995). "Sample Size Calculations for Binomial Proportions Via Highest Posterior Density Intervals." *The Statistician: Journal of the Institute of Statisticians*, 44: 143–154.

Lachin, J. M. (1981). "Introduction to Sample Size Determination and Power Analysis for Clinical Trials." *Controlled Clinical Trials*, 2: 93–113.

M'Lan, C. E., Joseph, L., and Wolfson, D. B. (2006). "Bayesian Sample Size Determination for Case-control Studies." *Journal of the American Statistical Association*, 101(474): 760–772.

Mukerjee, R. and Dey, D. K. (1993). "Frequentist Validity of Posterior Quantiles in the Presence of a Nuisance Parameter: Higher Order Asymptotics." *Biometrika*, 80: 499–505.

Müller, P. (1998). "Simulation Based Optimal Design." *Bayesian Statistics 6*, 90: 1322–1330.

Müller, P. and Parmigiani, G. (1995). "Optimal Design Via Curve Fitting of Monte Carlo Experiments." *Journal of the American Statistical Association*, 90: 1322–1330.

Peers, H. W. (1968). "Confidence Properties of Bayesian Interval Estimates." *Journal of the Royal Statistical Society, Series B: Methodological*, 30: 535–544.

Pham-Gia, T. and Turkkan, N. (1992). "Sample Size Determination in Bayesian Analysis (Disc: P399-404)." *The Statistician: Journal of the Institute of Statisticians*, 41: 389–397.

Satten, G. A. and Kupper, L. L. (1990). "Sample Size Requirements for Interval Estimation of the Odds Ratio." *American Journal of Epidemiology*, 131: 177–184.

Severini, T. A. (1991). "On the Relationship between Bayesian and Non-Bayesian Interval Estimates." *Journal of the Royal Statistical Society, Series B, Methodological*, 53: 611–618.

Tanner, M. A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer-Verlag Inc.

Welch, B. L. and Peers, H. W. (1963). "On Formulae for Confidence Points Based on Integrals of Weighted Likelihoods." *Statistical Methods in Medical Research*, 4: 311–337.

Table 1: **Table of ALC, WOC, and MLC sample sizes using the indicated methods with** $(a, b) = (1, 1), (2, 2), (3, 3), (4, 4), (4, 1), (4, 2)$**, and** $(4, 3)$**.**

(a, b)	coverage $1 - \alpha$	length l	**ALC**			**WOC**			**MLC**		
			Exact HPD	**First order formula**	**Third order** HPD	**Exact** HPD	**First order formula**	**Third order** HPD	**Exact** HPD	**First order formula**	**Third order** HPD
$(1, 1)$	.95	.10	234	235	234	381	383	381	285	287	285
		.05	945	946	945	1534	1535	1534	1149	1151	1149
$(2, 2)$	.95	.10	295	296	295	379	381	379	285	286	285
		.05	1195	1196	1195	1532	1533	1532	1149	1150	1149
$(3, 3)$	.95	.10	319	320	319	377	379	377	284	286	284
		.05	1295	1296	1295	1530	1531	1530	1149	1150	1149
$(4, 4)$	.95	.10	331	332	331	375	377	375	282	285	282
		.05	1348	1349	1348	1528	1529	1528	1147	1149	1147
$(4, 1)$	.95	.10	174	177	174	378	380	378	284	286	284
		.05	718	721	718	1531	1532	1531	1148	1150	1148
$(4, 2)$	.95	.10	277	278	277	377	379	377	284	286	284
		.05	1127	1128	1127	1530	1531	1530	1149	1150	1149
$(4, 3)$	.95	.10	318	321	318	376	378	376	283	285	283
		.05	1294	1295	1294	1529	1530	1529	1148	1149	1148

Table 2: **Table of ACC and MCC sample sizes using the indicated methods with** $(a, b) = (1, 1), (2, 2), (3, 3), (4, 4), (4, 1), (4, 2)$**, and** $(4, 3)$**.**

(a, b)	coverage $1 - \alpha$	length l	**ACC**		**MCC**	
			Exact HPD	**Third order** HPD	**Exact** HPD	**Third order** HPD
$(1, 1)$	.95	.10	274	274	285	285
		.05	1105	1105	1149	1149
$(2, 2)$	.95	.10	311	311	285	285
		.05	1259	1259	1149	1149
$(3, 3)$	.95	.10	327	327	284	284
		.05	1329	1329	1149	1149
$(4, 4)$	.95	.10	336	336	282	282
		.05	1368	1368	1147	1147
$(4, 1)$	.95	.10	223	223	284	284
		.05	913	913	1148	1148
$(4, 2)$	.95	.10	297	297	284	284
		.05	1207	1207	1149	1149
$(4, 3)$	.95	.10	326	326	283	283
		.05	1328	1328	1148	1148

Table 3: **Table of sample sizes for estimating p for various choices (a, b), leading to skewed prior distributions.**

$1-\alpha = .99$ and $l = .01$										
$(\mathbf{a}, \mathbf{b})$	**ALC**					**ACC**				
	Exact		**First order**	**Third order**		**Exact**		**Third order**		
	HPD	equal	formula	HPD	equal	HPD	equal	HPD	equal	
$(1, 49)$	3997	4035	4015	3995	4034	7881	7889	7881	7889	
$(1, 99)$	1930	1998	1959	1926	1996	3957	3976	3957	3976	
$(1, 149)$	1196	1285	1229	1188	1281	2540	2570	2540	2570	
$(1, 199)$	804	908	836	792	903	1781	1822	1780	1821	
$(1, 249)$	551	665	580	538	660	1287	1341	1285	1340	
$(1, 299)$	370	489	392	357	483	924	995	920	993	
$(1, 349)$	230	350	244	219	345	633	725	624	721	
$(1, 399)$	115	235	120	108	231	375	501	357	493	
$(1, 449)$	17	136	16	16	133	100	303	80	286	

$1-\alpha = .95$ and $l = .01$									
$(\mathbf{a}, \mathbf{b})$	**ALC**					**ACC**			
	Exact		**First order**	**Third order**		**Exact**		**Third order**	
	HPD	equal	formula	HPD	equal	HPD	equal	HPD	equal
$(2, 48)$	5125	5139	5137	5125	5139	6270	6278	6270	6278
$(2, 98)$	2531	2559	2555	2531	2559	3142	3159	3142	3158
$(2, 148)$	1598	1640	1634	1598	1640	2006	2032	2006	2032
$(2, 198)$	1096	1152	1143	1096	1151	1394	1430	1394	1430
$(2, 248)$	769	838	827	768	837	995	1043	995	1042
$(2, 298)$	530	612	599	529	611	702	763	701	762
$(2, 348)$	342	437	421	340	434	468	545	465	543
$(2, 398)$	187	293	276	182	290	268	366	261	363
$(2, 448)$	53	171	151	45	166	82	213	69	208

Table 4: **Sample sizes for an osteoporosis prevalence study. The parameter inputs were a beta(20.5, 28.25) prior density, $1-\alpha = 0.95$, and $l = 0.05$.**

Criterion	Sample Size
$\mathbf{ACC}_1$	1420
$\mathbf{ACC}_2$	1420
$\mathbf{ALC}_1$	1418
$\mathbf{ALC}_2$	1419
MLC	1133
MCC	1133
WOC	1487

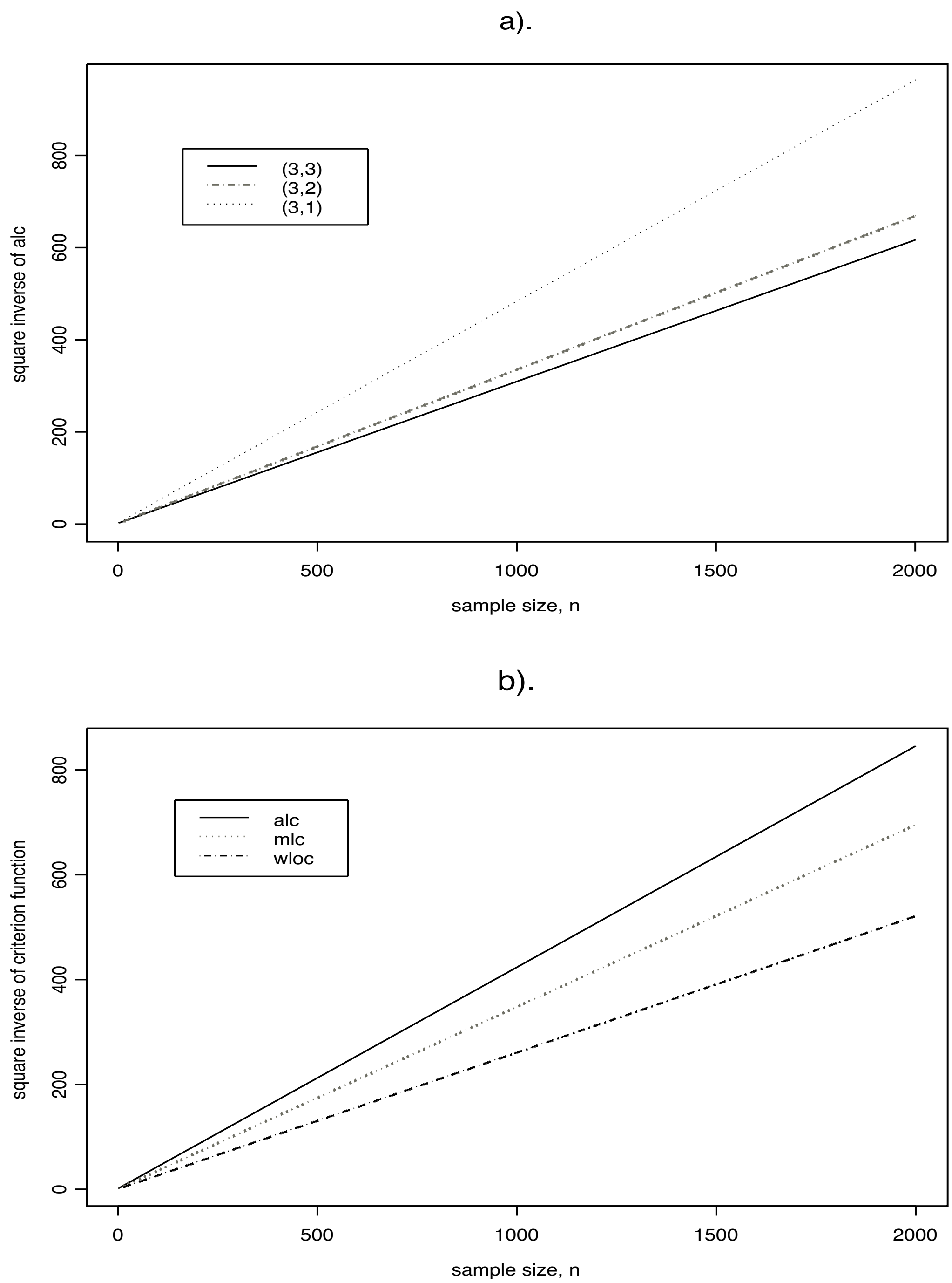


Figure 1: **a)** Graph of $\frac{1}{\mathrm{alc}^2(n,a,b)}$ as a function of n, with $(a,b) = (3,3), (3,2), (3,1)$ and $1-\alpha = 0.95$. **b)** Graph of $\frac{1}{\mathrm{alc}^2(n,1,1)}$, $\frac{1}{\mathrm{mlc}^2(n,1,1)}$, and $\frac{1}{\mathrm{wloc}^2(n,1,1)}$ as a function of n, when $1-\alpha = 0.95$.

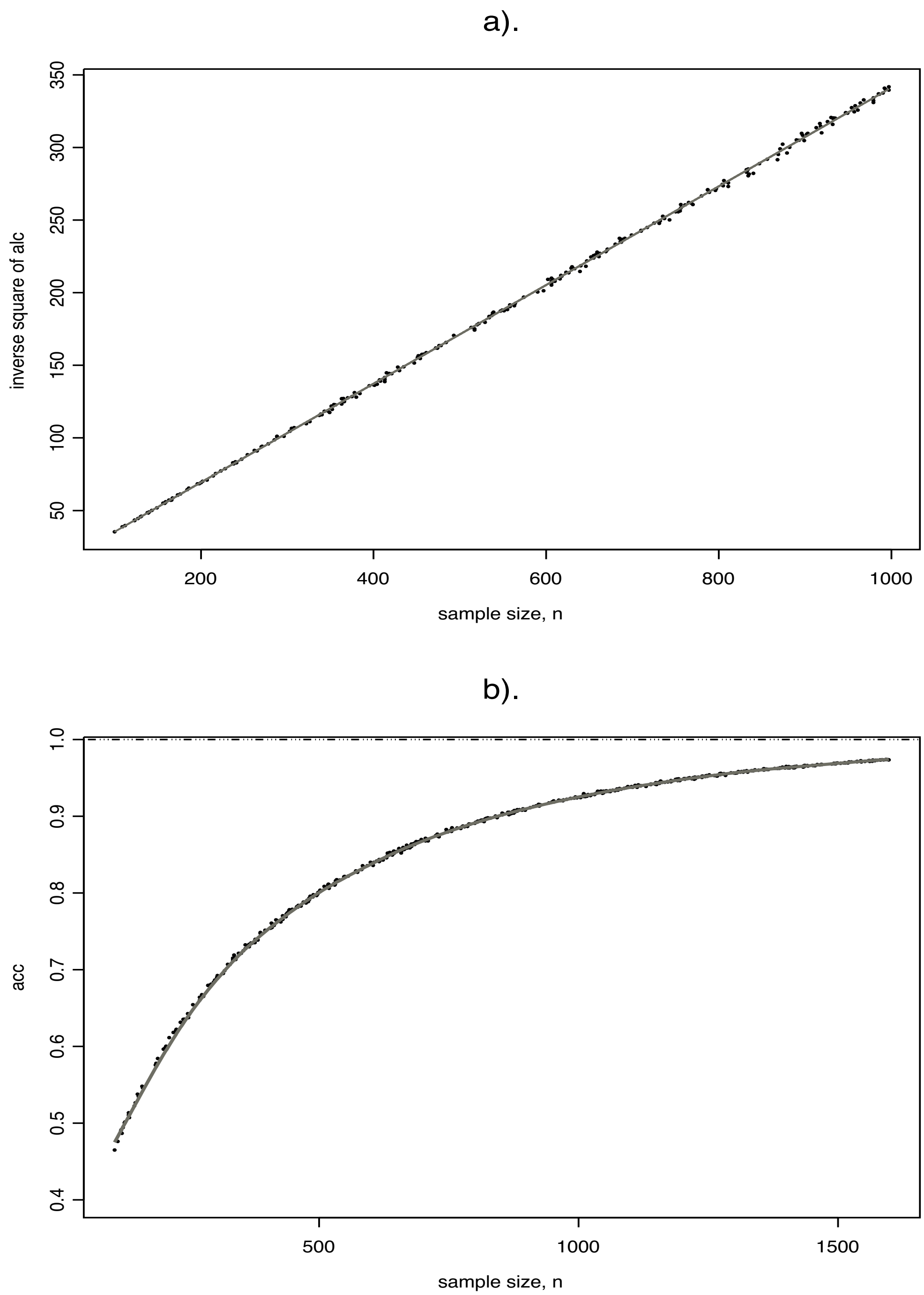
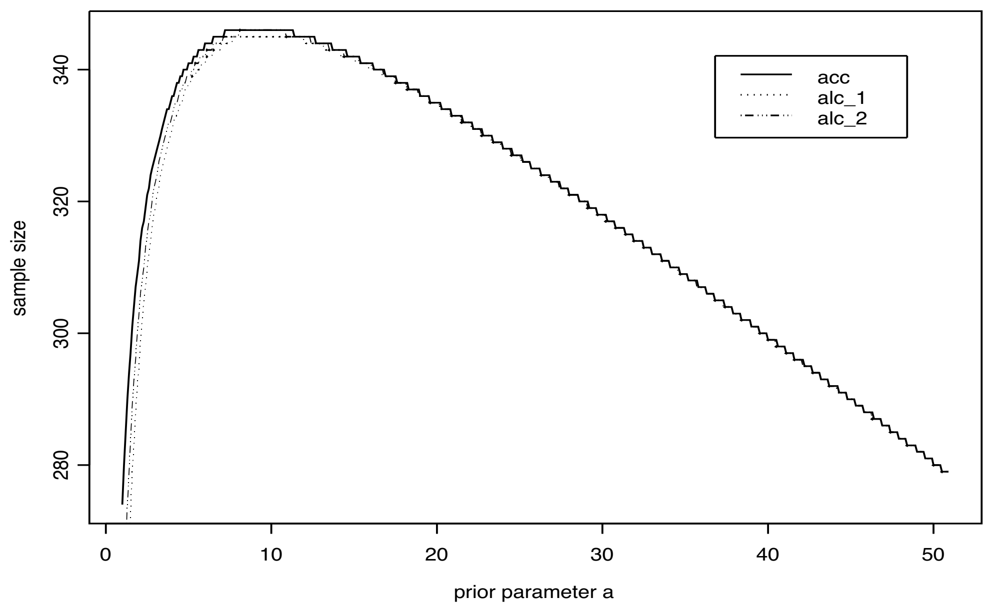


Figure 2: **a)** Graph of 500 Monte Carlo pairs $\left(n_i,\ \dfrac{1}{\widehat{\mathrm{alc}}_k^{\,2}(n_i,2,2)}\right)$ using $m = M = 1000$. **b)** Graph of 500 Monte Carlo pairs $(n_i,\ \widehat{\mathrm{acc}}_k(n_i,2,2))$ using $m = M = 1000$.

a).



b).

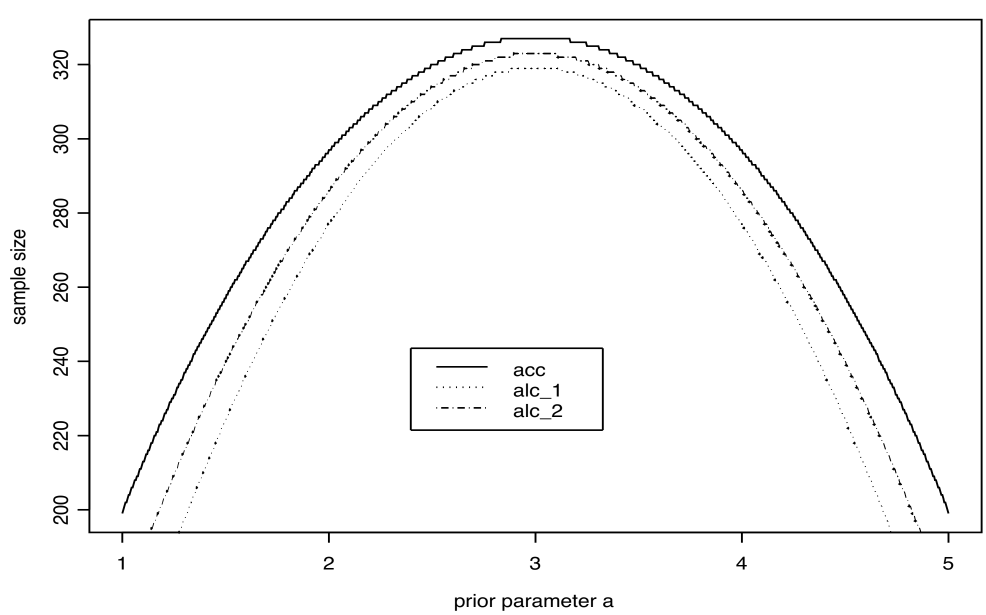


Figure 3: **a)** Graph of the $\mathbf{ACC}, \mathbf{ALC}$, and $\mathbf{ALC}_2$ sample sizes as a function of a when considering a symmetric prior, $\mathbf{Be}(a, a)$ for $a = 1, \cdots, 50$. $1 - \alpha = .95$ and $l = 0.1$. **b)** Graph of the $\mathbf{ACC}, \mathbf{ALC}$, and $\mathbf{ALC}_2$ sample sizes as a function of a when considering a $\mathbf{Be}(a, n_0 - a)$ prior distribution, with $a = 1, \cdots, 5$, $n_0 = 6$, $1 - \alpha = .95$, and $l = 0.1$.