The Statistician (1997)
46, *No.* 2, *pp.* 209–226

Bayesian sample size determination for normal means and differences between normal means

By LAWRENCE JOSEPH†

Montreal General Hospital and McGill University, Montreal, Canada

and PATRICK BÉLISLE

Montreal General Hospital, Canada

[Received June 1995. Revised October 1996]

SUMMARY

Several criteria for Bayesian sample size determination based on lengths and coverages of posterior credible intervals have recently appeared in the literature. Some but not all of these have been applied to estimating sample sizes for normal distributions. In this paper, these criteria are applied to find sample sizes for single normal means and differences between two normal means, both when the variances are known and when they are unknown. Fully Bayesian approaches as well as mixed Bayesian–likelihood approaches are considered. In the case of the difference between two normal means, situations with equal and unequal variances of the two distributions are considered. In addition to the rule that assumes equally sized groups, optimal solutions are determined which allow the sizes drawn from the two populations to differ while minimizing their sum. Exact closed form solutions are available for many of the situations, whereas numerical algorithms are described for others.

Keywords: Bayesian approach; Credible interval; Normal distribution; Optimal design; Predictive distribution; Sample size

1. Introduction

Sample size determination is an important component of the planning of studies in many fields. In accord with its stature in statistics, the normal distribution plays a central role in estimating sample size requirements (Desu and Raghavarao, 1990; Lipsey, 1990; Lemeshow *et al.*, 1990). Standard frequentist sample size formulae for the normal distribution such as

$$n \geqslant \frac{4\sigma^2 Z^2_{1-\alpha/2}}{l^2} \tag{1}$$

guarantee that a $100(1-\alpha)\%$ confidence interval will be of total length l, provided that the variance is *a priori* known to be exactly equal to σ^2, where $Z_{1-\alpha/2}$ is the $(1-\alpha/2)$-percentile of the standard normal distribution. There are three serious limitations to this formula. The first is that before the experiment is carried out σ will almost never be known with very high precision, but the sample size suggested by inequality (1) is directly proportional to the square of σ. The second is that, regardless of the value of σ used in inequality (1), final inferences are calculated on the basis of the observed data, which are of course not known at the planning stage. The third is that prior information may be available about the mean, and neglecting this can lead to a larger sample size than is necessary.

In this paper, Bayesian approaches to sample size determination are discussed which

†*Address for correspondence*: Division of Clinical Epidemiology, Department of Medicine, Montreal General Hospital, 1650 Cedar Avenue, Montreal, Quebec, H3G 1A4, Canada.
E-mail: joseph@binky.ri.mgh.mcgill.ca

 0039–0526/97/46209

address all three of these problems. Rather than a single value, a prior distribution is assigned to σ which reflects the pre-experimental uncertainty about its value. This prior distribution is then used to create a predictive distribution for the data, which places a weight on each data set that may occur. The prior information about σ is updated with the information in the data to form posterior inferences through Bayes theorem, often leading to sample sizes that are different from those which would occur under non-Bayesian formulations.

Each potential data vector of length n leads to a different interval length for fixed coverage, or conversely to a different coverage for a fixed interval length. Averaging these lengths or coverages over the predictive distribution for the data leads to two different criteria for Bayesian sample size determination. A third conservative criterion is obtained by considering the set $\mathscr{S}$ that consists of, for example, 95% or 99% of the most likely data values according to the predictive distribution, and simultaneously ensuring a sufficiently small interval length and sufficiently large coverage for all data vectors in $\mathscr{S}$. The above three general criteria, first compared in the context of binomial sampling in Joseph *et al.* (1995a), are reviewed in Section 2.

Many researchers have previously considered Bayesian approaches to sample size determination. For a single normal mean, Adcock (1988) developed closed form formulae for the cases with known and unknown variance by averaging the coverage of fixed length posterior credible sets over the predictive distribution of the data. We review Adcock's formulae in Section 3 below, and we use similar techniques to derive closed form formulae for the case of average lengths of fixed coverage posterior credible sets, as well as 'worst case' criteria. For a comparison of two or more normal means, DasGupta and Vidakovic (1997) have investigated a Bayesian approach to sample size estimation in analysis of variance.

Bayesian sample size estimation for a single binomial parameter has received considerable attention: Adcock (1987, 1992, 1995), Pham-Gia and Turkkan (1992), Pham-Gia (1995) and Joseph *et al.* (1995a, b). The case of a difference between two binomial parameters has been considered by Joseph *et al.* (1996).

Other work related to Bayesian sample size determination includes Adcock (1993), Berger (1985), Goldstein (1981), Gould (1993), Hutton and Owens (1993) and Spiegelhalter and Freedman (1986).

Sections 3, 4 and 5 apply the three criteria defined in Section 2 to the cases with a single normal mean (for variance known and unknown), the difference between two normal means with equal variances and the difference between two normal means with unequal variances respectively. In all cases, the fully Bayesian approach and a mixed Bayesian–likelihood approach are considered. The former utilizes the prior information for both the construction of the predictive distribution of the data and for posterior inferences, whereas the latter substitutes a normalized likelihood for posterior inferences while retaining the prior information for deriving the predictive distribution of the data. The mixed Bayesian–likelihood approach is intended for investigators who will use non-informative prior distributions or non-Bayesian techniques when reporting final inferences, but who recognize the utility of prior information for planning. In Sections 4 and 5, equal-sized groups as well as optimal allocation of the total sample size are considered. Section 6 contains several examples, while a summary discussion is found in Section 7. A computer program in the S-PLUS language that calculates sample sizes for all situations discussed in this paper is available by sending the electronic mail message 'send samplesize-mean from S' to `statlib@lib.stat.cmu.edu`.

2. Criteria for Bayesian sample size determination

Let $\theta \in \Theta$ be the parameter of interest, $f(\theta)$ the prior distribution, $x = (x_1, \ldots, x_n)$ the data of sample size n, $\mathscr{X}$ the data space, $f(x)$ the predictive marginal distribution of the data and $f(\theta|x)$ the posterior distribution for θ given data x. Then

$$f(x) = \int_{\Theta} f(x|\theta)\, f(\theta)\, \mathrm{d}\theta \tag{2}$$

and

$$f(\theta|x) = f(x|\theta)\, f(\theta)/f(x).$$

When using inequality (1), it suffices to specify l, α and a value for σ. In a Bayesian approach, however, specifying l, α and a prior distribution for θ is insufficient, since a decision is required regarding the data, which are of course unknown at the planning stage. Consideration of the predictive distribution of the data given by expression (2) leads to the following three criteria.

2.1. *Average coverage criterion*

For a fixed posterior interval length l, we can determine the sample size by finding the smallest n such that the equation

$$\int_{\mathscr{X}} \left\{ \int_{a}^{a+l} f(\theta|x, n)\, \mathrm{d}\theta \right\} f(x)\, \mathrm{d}x \geqslant 1 - \alpha \tag{3}$$

is satisfied. This average coverage criterion (ACC) ensures that the mean coverage of posterior credible intervals of length l, weighted by $f(x)$, is at least $1 - \alpha$.

Adcock (1988) first proposed the use of an ACC in the context of estimating normal means, where the interval $(a,\ a + l)$ was chosen to be a symmetric tolerance region around the mean. Joseph *et al.* (1995a) proposed that the interval $(a,\ a + l)$ be chosen to be a highest posterior density interval for asymmetric posterior distributions. Although in general these can lead to different sample sizes, for normal posterior densities they are equivalent, since highest posterior density regions are simply symmetric intervals around the mean.

2.2. *Average length criterion*

For a fixed posterior credible interval coverage of $1 - \alpha$, we can also determine the sample size by finding the smallest n such that

$$\int_{\mathscr{X}} l'(x, n)\, f(x)\, \mathrm{d}x \leqslant l, \tag{4}$$

where $l'(x, n)$ is the length of the $100(1 - \alpha)\%$ posterior credible interval for data x, determined by solving

$$\int_{a}^{a+l'(x,n)} f(\theta|x, n)\, \mathrm{d}\theta = 1 - \alpha$$

for $l'(x, n)$ for each value of $x \in \mathscr{X}$. As before, a can be chosen to give highest posterior density intervals or symmetric intervals, which coincide for symmetric unimodal densities.

This average length criterion (ALC) ensures that the mean length of $100(1 - \alpha)\%$ posterior credible intervals weighted by $f(x)$ is at most l. It does not appear to have been previously applied to the normal distribution. Since most researchers will report intervals of fixed coverage (usually 95%) regardless of their length, it can be argued that the ALC is more conventional than the ACC.

2.3. *Worst outcome criterion*

Cautious investigators may not be satisfied with the 'average' assurances provided by the ACC and the ALC criteria. Therefore, a conservative sample size can also be determined by the smallest n satisfying the equation

$$\inf_{x\in\mathscr{S}}\left\{\int_a^{a+l(x,n)} f(\theta|x,n)\,\mathrm{d}\theta\right\} \geqslant 1-\alpha, \tag{5}$$

where $\mathscr{S}$ is a suitably chosen subset of the data space $\mathscr{X}$. For example, this worst outcome criterion (WOC) ensures that, if $\mathscr{S}$ consists of the most likely 95% of the possible $x \in \mathscr{X}$, then there is 95% assurance that the length of the $100(1-\alpha)\%$ posterior credible interval will be at most l.

3. Sample sizes for single normal mean

Throughout this section we assume that the data vector $X = (X_1, X_2, \ldots, X_n)$ consists of exchangeable components from a normal distribution where μ is the unknown normal mean and λ is the precision of the data, defined as $\lambda = 1/\sigma^2$, where σ^2 is the variance. We also assume a normal–gamma conjugate prior distribution for (μ, λ), so that *a priori*

$$\lambda \sim \text{gamma}(\nu, \beta),$$

$$\mu|\lambda \sim N(\mu_0, n_0\lambda).$$

Sections 3.1 and 3.2 treat the cases of known and unknown precision from a fully Bayesian perspective respectively, whereas Section 3.3 considers mixed Bayesian–likelihood approaches. The standard Bayesian distributional results used in this paper can be found in Appendix A of Bernardo and Smith (1994).

3.1. *Sample sizes for single normal mean with known precision*

If the precision is known *a priori* to be equal to λ, then it can be easily shown that *a posteriori*

$$\mu|x \sim N(\mu_n, \lambda_n),$$

where

$$\lambda_n = (n+n_0)\lambda,$$

$$\mu_n = \frac{n_0\mu_0 + n\bar{x}}{n_0+n},$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Since the posterior precision depends only on n and does not vary with the particular observed data vector x, all three criteria (ACC, ALC and WOC) lead to the same solution, which is also equivalent to that given by Adcock (1988):

$$n \geqslant \frac{4Z_{1-\alpha/2}^2}{\lambda l^2} - n_0. \tag{6}$$

If a non-informative prior is used such that $n_0 = 0$, then inequality (6) reduces to the classical formulation (1).

3.2. *Sample sizes for single normal mean with unknown precision*

If the precision is unknown, then the marginal posterior distribution of μ is given by

$$\mu|x \sim \sqrt{\left\{\frac{\beta_n}{(n+n_0)(\nu+n/2)}\right\}}\, t_{2\nu+n} + \mu_n,$$

where

$$\mu_n = \frac{n_0\mu_0 + n\bar{x}}{n+n_0},$$

$$\beta_n = \beta + \frac{1}{2}ns^2 + \frac{1}{2}\frac{nn_0}{n+n_0}(\bar{x}-\mu_0)^2,$$

$$ns^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

and $t_{2\nu+n}$ represents a t-distribution with $2\nu+n$ degrees of freedom. Since the posterior precision varies with the data x, different criteria will lead to different sample sizes.

3.2.1. *Average coverage criterion.* Adcock (1988) has shown that the ACC sample size is given by

$$n = \frac{4\beta}{\nu l^2}\, t^2_{2\nu;1-\alpha/2} - n_0. \tag{7}$$

Since ν/β is the prior mean for precision λ, the ACC sample size for unknown precision is similar to that for known precision, in that we only need to substitute the prior mean precision for λ in inequality (6) and exchange the normal quantile Z with a quantile from a $t_{2\nu}$-distribution. Since the degrees of freedom of the t-distribution do not increase with the sample size, equation (7) can lead to sample sizes that are substantially different from those from inequalities (1) and (6) even when n is large. See Section 6 for examples of this.

3.2.2. *Average length criterion.* When estimating a single normal mean with unknown precision λ with a gamma(ν, β) prior distribution on λ, the ALC (4) is satisfied (see Appendix A) when n is sufficiently large that

$$2t_{n+2\nu,1-\alpha/2}\sqrt{\left\{\frac{2\beta}{(n+2\nu)(n+n_0)}\right\}}\,\frac{\Gamma\left(\frac{n+2\nu}{2}\right)\Gamma\left(\frac{2\nu-1}{2}\right)}{\Gamma\left(\frac{n+2\nu-1}{2}\right)\Gamma(\nu)} \leqslant l. \tag{8}$$

Although it does not appear feasible to solve inequality (8) explicitly for n, the left-hand side is straightforward to calculate given ν, β, n_0, α and n. Therefore, the exact minimum n satisfying inequality (8) can be found by a bisectional search algorithm.

3.2.3. *Worst outcome criterion.* Let $\mathscr{S}$ be the subset of $\mathscr{X}$ such that

$$\int_{\mathscr{S}} f(x)\,\mathrm{d}x = 1 - w,$$

where $f(x)$ is given by expression (2) and $f(x) \geqslant f(y)$ for all $x \in \mathscr{S}$ and $y \notin \mathscr{S}$. Thus $\mathscr{S}$ is a highest posterior density region according to the predictive distribution (2). Then it can be shown (see Appendix B) that, when estimating a single normal mean with unknown precision λ with a gamma(ν, β) prior distribution on λ, the WOC (5) is satisfied when n is sufficiently large that

$$\frac{l^2(n+2\nu)(n+n_0)}{8\beta\{1+(n/2\nu)F_{n,2\nu;1-w}\}} \geqslant t^2_{n+2\nu;1-\alpha/2}, \tag{9}$$

where $F_{n,2\nu;1-w}$ denotes the $100(1-w)$-percentile of an F-distribution with n and 2ν degrees of freedom. As in the previous section, the smallest n satisfying inequality (9) can be found by a bisectional search. If $\mathscr{S} = \mathscr{X}$, then the sample size is not defined, since $F_{n,2\nu;1-w} \to \infty$ as $w \to 0$, so that inequality (9) cannot be satisfied for any n.

3.3. *Mixed Bayesian–likelihood approaches for a single normal mean*

Mixed Bayesian–likelihood approaches (Joseph *et al.*, 1996) use the prior distribution to derive the predictive distribution of the data but assume that one will use only the likelihood function for final inferences. These are intended to satisfy investigators who recognize that prior information is important for planning but prefer to base final inferences only on the data. For example, they can be used by investigators who plan to report 95% confidence intervals. The fully Bayesian closed form formulae (7)–(9) do not apply to the mixed Bayesian–likelihood approach (see Appendix C). However, a simulation algorithm can be devised to approximate the required sample size. For the ACC, for example, we seek the minimum n such that

$$\int_{\mathscr{X}} \left\{ \int_a^{a+l} f(\mu|x, n)\,\mathrm{d}\mu \right\} f(x)\,\mathrm{d}x \geqslant 1 - \alpha \tag{10}$$

holds, where $f(x)$ is the predictive distribution of the data x given a gamma(ν, β) prior distribution on λ, $f(\mu|x, n)$ is the posterior distribution derived from the non-informative prior distribution $f(\mu, \lambda) \propto \lambda^{-1}$ and $(a, a+l)$ forms a symmetric or, equivalently, highest density posterior credible interval for μ. Therefore,

$$(a, a+l) = (\bar{x} - l/2, \bar{x} + l/2),$$

$$\mu|x \sim \bar{x} + \sqrt{\left\{\frac{ns^2}{n(n-1)}\right\}}\, t_{n-1},$$

$$\int_{\bar{x}-l/2}^{\bar{x}+l/2} f(\mu|x)\,\mathrm{d}\mu = 2p_t\left(\frac{l}{2}\sqrt{\left\{\frac{n(n-1)}{ns^2}\right\}}; n-1\right),$$

where $p_t(c; d)$ is the area between 0 and c under a t-density with d degrees of freedom. Since the area depends on the data only through ns^2, the following algorithm finds the approximate sample size.

(a) Select an initial estimate of the sample size n.
(b) Generate m values of the random variable ns^2. Since

$$ns^2 \sim \text{gamma}\left(\frac{n-1}{2}, \frac{\lambda}{2}\right),$$

and $\lambda \sim \text{gamma}(\nu, \beta)$,

$$ns^2 \sim \text{gamma} - \text{gamma}\left(\nu, 2\beta, \frac{n-1}{2}\right),$$

i.e. ns^2 follows a gamma–gamma distribution. A random variable x follows a gamma–gamma distribution if

$$f(x|\nu, \beta, n) = \frac{\Gamma(\nu + n)\beta^\nu}{\Gamma(\nu)\,\Gamma(n)} \frac{x^{n-1}}{(\beta + x)^{\nu+n}}$$

for $x > 0$, $\nu > 0$, $\beta > 0$ and $n > 0$. See Bernardo and Smith (1994), p. 430.

(c) For each of the m values of ns_i^2, $i = 1, 2, \ldots, m$, in step (b), calculate

$$\text{coverage}(ns_i^2) = 2p_t\left(\frac{l}{2}\sqrt{\left\{\frac{n(n-1)}{ns_i^2}\right\}}; n-1\right).$$

(d) Then

$$\frac{1}{m}\sum_{i=1}^{m} \text{coverage}(ns_i^2)$$

approximates the average coverage that forms the left-hand side of inequality (10).

Repeating steps (b)–(d) for values of n selected by a bisectional search procedure will lead to an approximately correct sample size. The accuracy of this estimate increases with increasing m.

Similarly, the WOC sample size can be approximated by ensuring that

$$2p_t\left(\frac{l}{2}\sqrt{\left\{\frac{n(n-1)}{ns_i^2}\right\}}; n-1\right) \geqslant 1-\alpha$$

for 95% or 99% (say) of the ns_i^2 values for a given value of n.

For the ALC, further simplifications (see Appendix C) obviate the need to simulate data. This is because it can be shown that the average length of the likelihood-based posterior credible region for μ of level $1-\alpha$ is given by

$$\frac{2}{\sqrt{\{n(n-1)\}}} E\{\sqrt{(ns^2)}\} t_{n-1;1-\alpha/2},$$

and that

$$E\{\sqrt{(ns^2)}\} = \sqrt{(2\beta)} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\Gamma(\nu-\frac{1}{2})}{\Gamma(\nu)}.$$

Therefore, the average length given a gamma(ν, β) prior distribution on the precision λ for a sample size n is given by

$$2t_{n-1;1-\alpha/2}\sqrt{\left\{\frac{2\beta}{n(n-1)}\right\}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\Gamma(\nu-\frac{1}{2})}{\Gamma(\nu)}. \tag{11}$$

Finding the minimum n such that expression (11) is less than the desired length l provides the exact ALC sample size under a mixed Bayesian–likelihood approach.

4. Sample sizes for difference between two normal means—common precision

In this section we consider independent random vectors $X_1 = (X_{11}, X_{21}, \ldots, X_{n_1 1})$ and $X_2 = (X_{12}, X_{22}, \ldots, X_{n_2 2})$ such that $X_{ij} \sim N(\mu_j, \lambda)$, $i = 1, 2, \ldots, n_j$, $j = 1, 2$. We again assume normal–gamma prior conjugate prior densities, so that

$$\lambda \sim \text{gamma}(\nu, \beta)$$

and

$$\mu_j|\lambda \sim N(\mu_{0j}, n_{0j}\lambda), \qquad j = 1, 2.$$

We seek the sample sizes n_1 and n_2 required to estimate $\theta = \mu_1 - \mu_2$ under the ACC, ALC and WOC criteria. Taking a fully Bayesian approach, Sections 4.1 and 4.2 consider the cases of known and unknown precision respectively, whereas Section 4.3 discusses the mixed Bayesian–likelihood approach. Throughout, we consider both equal sample sizes, when $n_1 = n_2$, as well as optimal allocation, where the sum $n_1 + n_2$ is minimized allowing $n_1 \neq n_2$. Ideally, for optimal allocation, the minimum value of $n_1 + n_2$ should be selected that fulfils the relevant criterion. This can be difficult, however, because of the large number of possible combinations of n_1 and n_2. Therefore, we have chosen to find the combination of n_1 and n_2 that minimizes the posterior variance. For practical purposes, there will be little if any difference between these two strategies.

4.1. *Known common precision*

If the common precision λ is exactly known, then after observing data vectors x_1 and x_2, of sizes n_1 and n_2 respectively, the posterior distribution of θ is

$$\theta|x_1, x_2 \sim N\left\{\mu_{n_2 2} - \mu_{n_1 1}, \frac{\lambda(n_{01} + n_1)(n_{02} + n_2)}{n_1 + n_2 + n_{01} + n_{02}}\right\},$$

where

$$\mu_{n_j j} = \frac{n_{0j}\mu_{0j} + n_j \bar{x}_j}{n_{0j} + n_j}, \qquad j = 1, 2.$$

As was the case in Section 3.1, the posterior variance depends on the data only through the sample sizes n_1 and n_2, so that the ACC, ALC and WOC sample sizes coincide. For $n_1 = n_2$, as shown in Appendix D, the sample size is given by

$$n_1 = n_2 \geqslant \frac{-B + \sqrt{(B^2 - 4AC)}}{2A}, \tag{12}$$

where $A = \lambda^2$,

$$B = \lambda^2(n_{01} + n_{02}) - 8\lambda Z^2_{1-\alpha/2}/l^2$$

and

$$C = n_{01} n_{02} \lambda^2 - \frac{4(n_{01} + n_{02})\lambda Z^2_{1-\alpha/2}}{l^2}.$$

If $B^2 - 4AC \leqslant 0$, then the prior information is sufficient, and no sampling is required.

Optimal allocation minimizes the posterior variance, which is minimized when $n_1 + n_{01} = n_2 + n_{02}$. Therefore the minimum sample size for n_1 satisfies

$$n_1 \geqslant \frac{8}{\lambda l^2} Z^2_{1-\alpha/2} - n_{01}, \tag{13}$$

from which $n_2 = n_1 + n_{01} - n_{02}$ can be calculated. If $n_{01} = n_{02}$, then the sample size given by inequality (12) reduces to that given by inequality (13).

4.2. *Common but unknown precision*

In the case of common but unknown precision, if the prior information on the unknown parameters μ_1, μ_2 and λ is from the normal–gamma family as given above, then the posterior density of $\theta = \mu_1 - \mu_2$ is

$$\theta|x_1, x_2 \sim A + \sqrt{\left(\frac{B}{2CD}\right)}\, t_{2C},$$

where

$$A = E(\theta|x_1, x_2) = \frac{n_2\bar{x}_2 + n_{02}\mu_{02}}{n_2 + n_{02}} - \frac{n_1\bar{x}_1 + n_{01}\mu_{01}}{n_1 + n_{01}},$$

$$B = 2\beta + n_1 s_1^2 + n_2 s_2^2 + \frac{n_1 n_{01}}{n_1 + n_{01}}(\bar{x}_1 - \mu_{01})^2 + \frac{n_2 n_{02}}{n_2 + n_{02}}(\bar{x}_2 - \mu_{02})^2,$$

$$C = \frac{n_1 + n_2}{2} + \nu$$

and

$$D = \frac{(n_1 + n_{01})(n_2 + n_{02})}{n_1 + n_{01} + n_2 + n_{02}}.$$

The posterior variance depends on the particular data vectors x_1 and x_2, so that different criteria lead to different sample sizes.

4.2.1. *Average coverage criterion.* As shown in Appendix E, the ACC is satisfied when

$$\frac{(n_1 + n_{01})(n_2 + n_{02})}{n_1 + n_2 + n_{01} + n_{02}} \geqslant \frac{4\beta}{\nu l^2} t^2_{2\nu;1-\alpha/2} \tag{14}$$

holds. This equation can be solved explicitly if $n = n_1 = n_2$, so that the sample size is given by

$$n \geqslant \frac{-B + \surd(B^2 - 4AC)}{2A}, \tag{15}$$

where

$$A = \frac{\nu l^2}{4},$$

$$B = \frac{\nu l^2}{4}(n_{01} + n_{02}) - 2\beta t^2_{2\nu,1-\alpha/2}$$

and

$$C = \frac{n_{01} n_{02} \nu l^2}{4} - \beta t^2_{2\nu;1-\alpha/2}(n_{01} + n_{02}).$$

A reasonable criterion for optimal sample size selection for unequal n_1 and n_2 is to minimize the expected posterior variance. Since

$$\bar{x}_i \sim \mu_{0i} + \sqrt{\left\{\frac{\beta_i(n_i + n_{0i})}{\nu_i n_i n_{0i}}\right\}} t_{2\nu_i}, \qquad i = 1, 2,$$

and

$$n_i s_i^2 \sim \text{gamma} - \text{gamma}\left(\nu_i, 2\beta_i, \frac{n_i - 1}{2}\right), \qquad i = 1, 2,$$

it can be shown that the expected variance of θ given the data is given by

$$\frac{1}{(n_1 + n_{01})(n_2 + n_{02})} \frac{n_1 + n_2 + n_{01} + n_{02}}{n_1 + n_2 + 2\nu - 2} \left\{2\beta + \frac{(n_1 + n_2 - 2)\,\Gamma(\nu - 1)}{\Gamma(\nu)} + \frac{2\beta}{\nu - 1}\right\}.$$

This implies that the minimum variance for fixed $n_1 + n_2$ occurs when

$$\frac{1}{(n_1 + n_{01})(n_2 + n_{02})}$$

is minimized, i.e. when

$$n_1 + n_{01} = n_2 + n_{02}. \tag{16}$$

Solving inequality (14) with this constraint provides the ACC optimal sample size.

4.2.2. *Average length criterion.* In Appendix F, it is shown that the ALC for the difference between two normal means with unknown but equal precision is satisfied if

$$2t_{n_1+n_2+2\nu;1-\alpha/2}\sqrt{\left\{\frac{2\beta(n_1+n_{01}+n_2+n_{02})}{(n_1+n_2+2\nu)(n_1+n_{01})(n_2+n_{02})}\right\}}\frac{\Gamma\left(\frac{n_1+n_2+2\nu}{2}\right)\Gamma\left(\frac{2\nu-1}{2}\right)}{\Gamma\left(\frac{n_1+n_2+2\nu-1}{2}\right)\Gamma\left(\frac{2\nu}{2}\right)} \leqslant l. \tag{17}$$

The sample size can then be found for either equal or unequal n_1 and n_2 by a bisectional search. For $n_1 \neq n_2$, constraint (16) applies.

4.2.3. *Worst outcome criterion.* The WOC sample sizes n_1 and n_2 for the difference between two normal means with equal but unknown precision are such that

$$\frac{l^2}{8\beta}\frac{(n_1+n_{01})(n_2+n_{02})}{n_1+n_{01}+n_2+n_{02}}\frac{n_1+n_2+2\nu}{1+\{(n_1+n_2)/2\nu\}F_{n_1+n_2,2\nu;1-w}} \geqslant t^2_{n_1+n_2+2\nu;1-\alpha/2}.$$

Again, a bisectional search can be used to find the sample sizes required for cases with equal and unequal sizes in each group.

4.3. *Mixed Bayesian–likelihood approach*

As in Section 3.3, the exact results of Sections 4.2.1–4.2.3 do not apply to the mixed Bayesian–likelihood approach, and therefore an approximate sample size can be found via simulations from the preposterior density of the data. Thus samples from the sufficient statistic $(\bar{x}_1, \bar{x}_2, n_1 s_1^2, n_2 s_2^2)$ are required. Although samples can be straightforwardly generated by first drawing samples from the prior distribution of (μ_1, μ_2, λ) and then from $(\bar{x}_1, \bar{x}_2, n_1 s_1^2, n_2 s_2^2)$ given (μ_1, μ_2, λ), it is more efficient in terms of computing time to generate $(\bar{x}_1, \bar{x}_2, n_1 s_1^2, n_2 s_2^2)$ directly. Given a random sample of $(\bar{x}_1, \bar{x}_2, n_1 s_1^2, n_2 s_2^2)$ vectors, the ACC, ALC and WOC sample sizes can be approximated by a bisectional search. The integrals on the left-hand sides of equations (3) and (4) are replaced by average values over the simulated data vectors, and we replace the exact infimum in inequality (5) by ensuring that the appropriate proportion of data points in the simulation satisfies the criterion. Appendix G provides an efficient method to simulate the required random samples. Unlike the fully Bayesian approach, optimal solutions are not of interest here, since, if prior information is not utilized in the calculations of posterior variances, then the optimal solution has $n_1 = n_2$.

5. Sample sizes for difference between two normal means—unequal precisions

In this section we consider independent random vectors $X_1 = (X_{11}, X_{21}, \ldots, X_{n_1 1})$ and $X_2 = (X_{12}, X_{22}, \ldots, X_{n_2 2})$ such that $X_{ij} \sim N(\mu_j, \lambda_j)$, $i = 1, 2, \ldots, n_j$, $j = 1, 2$. Once again we assume normal–gamma prior conjugate prior densities, so that

$$\lambda_j \sim \text{gamma}(\nu_j, \beta_j), \qquad j = 1, 2,$$

and

$$\mu_j|\lambda \sim N(\mu_{0j}, n_{0j}\lambda), \qquad j = 1, 2.$$

5.1. *Known precisions*

If the precisions λ_1 and λ_2 are exactly known, then after observing data vectors x_1 and x_2, of sizes n_1 and n_2 respectively, the posterior distribution of θ is

$$\theta|x_1, x_2 \sim N\left(\mu_{n_2 2} - \mu_{n_1 1}, \frac{\lambda_{n_1 1}\lambda_{n_2 2}}{\lambda_{n_1 1} + \lambda_{n_2 2}}\right),$$

where $\lambda_{n_i i} = \lambda_i(n_{0i} + n_i)$ and

$$\mu_{n_i i} = \frac{\lambda_i(n_{0i}\mu_{0i} + n_i\bar{x}_i)}{\lambda_{n_i i}} = \frac{n_{0i}\mu_{0i} + n_i\bar{x}_i}{n_{0i} + n_i}.$$

A calculation analogous to that in Appendix D shows that, if $n_1 = n_2$, then the sample size which simultaneously satisfies the ACC, ALC and WOC is given by

$$n \geqslant \frac{-B + \sqrt{(B^2 - 4AC)}}{2A}, \tag{18}$$

where $A = \lambda_1\lambda_2$,

$$B = \lambda_1 n_{02}\lambda_2 + \lambda_2 n_{01}\lambda_1 - \frac{4Z^2_{1-\alpha/2}}{l^2}(\lambda_1 + \lambda_2),$$

$$C = n_{01}\lambda_1 n_{02}\lambda_2 - \frac{4Z^2_{1-\alpha/2}}{l^2}(n_{01}\lambda_1 + n_{02}\lambda_2).$$

For $n_1 \neq n_2$, the optimal sample size is obtained when

$$n_1 + n_{01} \geqslant \frac{4}{l^2} Z^2_{1-\alpha/2}\left\{\frac{1}{\sqrt{(\lambda_1\lambda_2)}} + \frac{1}{\lambda_1}\right\}$$

and

$$n_2 + n_{02} = \sqrt{\left(\frac{\lambda_1}{\lambda_2}\right)}(n_1 + n_{01}).$$

5.2. *Unknown precisions*

The analytical results of Section 4 no longer apply when $\lambda_1 \neq \lambda_2$. For large n_1 and n_2, the posterior distribution of $\theta = \mu_1 - \mu_2$ can be approximated by

$$\theta|x \approx N(\mu_{n1} - \mu_{n2}, \lambda^*),$$

where

$$\mu_{nj} = \frac{n_{0j}\mu_{0j} + n_j\bar{x}_j}{n_j + n_{0j}}, \qquad j = 1, 2,$$

$$\lambda^* = \left\{\frac{2\beta_{n_1 1}}{(n_1 + n_{01})(2\nu_1 + n_1 - 2)} + \frac{2\beta_{n_2 2}}{(n_2 + n_{02})(2\nu_2 + n_2 - 2)}\right\}^{-1}.$$

For smaller values of n_1 and n_2, a simulation similar to that of Section 4.3 is suggested. This case is simpler than the case of common precision, since $(\bar{x}_1, n_1 s_1^2)$ is independent of $(\bar{x}_2, n_2 s_2^2)$. However, since the exact posterior distribution of $\theta = \mu_1 - \mu_2$ does not have a standard form, the required interval lengths and coverages must be obtained numerically. In most applications, the normal approximation should provide adequate sample sizes. For the mixed Bayesian–likelihood approach, the above methods can be used, setting $n_{0j} = \beta_j = 0$ and $\nu_j = -\frac{1}{2}$, $j = 1, 2$.

Since

$$\bar{x}_j \sim \mu_{0j} + \sqrt{\left\{\frac{\beta_j(n_j + n_{0j})}{\nu_j n_j n_{0j}}\right\}} t_{2\nu_j}$$

and

$$n_j s_j^2 \sim \text{gamma} - \text{gamma}\left(\nu_j, 2\beta_j, \frac{n_j - 1}{2}\right), \qquad j = 1, 2,$$

the expected posterior variance can be shown to be

$$E\{V(\theta|x_1, x_2)\} = \sum_{j=1}^{2} \frac{\beta_j}{(n_j + n_{0j})(\nu_j - 1)},$$

which is minimized if and only if

$$n_2 + n_{02} = \sqrt{\left(\frac{\beta_2}{\beta_1}\frac{\nu_1 - 1}{\nu_2 - 1}\right)}(n_1 + n_{01}). \tag{19}$$

Thus optimal sample sizes that minimize the posterior variance can be calculated as above, but subject to constraint (19).

6. Examples

In practice, the prior information will be different in every problem, so it is impossible to provide exhaustive tables. Therefore, Table 1 presents a variety of examples that illustrate the relationships between the various criteria discussed for the case of estimating sample size requirements for a single normal mean.

Examples 1–3 of Table 1 show that the Bayesian approach can provide larger sample sizes than the frequentist approach, even though prior information is incorporated in the final inferences. The same examples also illustrate that the sample size provided by the ALC tends to be smaller than that of the ACC when $1 - \alpha$ is near 1 and l is not near 0. This is because

TABLE 1
Sample sizes for estimating a single normal mean†

Example	ν	β	n_0	l	$1 - \alpha$	*Freq*	*ACC*	*ALC*	*WOC (95%)*
1	2	2	10	0.5	0.99	107	330	160	589
						MBL →	345	171	604
2	2	2	10	0.2	0.95	385	761	595	2152
						MBL →	771	606	2174
3	2	2	10	0.2	0.80	165	226	248	914
						MBL →	237	259	916
4	2	2	10	0.2	0.50	46	45	61	245
						MBL →	55	72	252
5	100	100	100	0.2	0.95	385	289	288	344
						MBL →	392	389	473
6	100	100	10	0.2	0.95	385	379	378	436
						MBL →	393	389	473

†Freq is the sample size computed using inequality (1), and the ACC, ALC and WOC sample sizes are calculated using expressions (7), (8) and (9) respectively. The mixed Bayesian–likelihood (MBL) sample sizes are calculated from the results of Section 4.3. When exact formulae were not available, sufficient samples were simulated to reduce the Monte Carlo error to less than 1% of the total sample size.

coverage probabilities are bounded above by 1, so that maintaining the required average coverage becomes more difficult as $1-\alpha$ becomes larger. Similarly, since l is bounded below by 0, maintaining an average length of l becomes more difficult as l approaches 0, leading to the larger sizes for the ALC compared with the ACC in example 4.

Example 5 shows that, with a large amount of prior information on both λ and μ, the Bayesian approach leads to smaller sample sizes than the frequentist approach, but, as expected, the mixed Bayesian likelihood approach provides similar sample sizes to the frequentist approach. With large amounts of prior information on λ but not on μ, similar sample sizes are provided by all criteria, as suggested by example 6, with the WOC criteria somewhat higher than the rest.

Similar situations can be constructed for the difference between two normal means. As just one example, let $\nu_1 = \nu_2 = \beta_1 = \beta_2 = 10$, $n_{01} = n_{02} = 20$, $l = 0.2$ and $1-\alpha = 0.95$. Then the standard frequentist formula given by

$$n_1 = n_2 = \frac{4(\sigma_1^2 + \sigma_2^2) Z_{1-\alpha/2}^2}{l^2}$$

suggests 769 as a sample size, whereas the ACC, ALC and WOC (95%) sample sizes are given by 844, 823 and 1222 respectively. The mixed Bayesian–likelihood sample sizes are 864, 842 and 1249, for the ACC, ALC and WOC (95%) respectively. Since the prior parameters are identical for the two means, the optimal solution is identical with that for equal sizes. However, if the prior parameters are changed such that more is known *a priori* about μ_1 than μ_2, with $\nu_1 = \beta_1 = 18$, $\nu_2 = \beta_2 = 2$, $n_{01} = 18$, $n_{02} = 2$, $l = 0.2$ and $1-\alpha = 0.95$, then the ACC, ALC and WOC (95%) sample sizes are given by $n_1 = n_2 = 1149$, 1044 and 2554 respectively, whereas the optimal sizes (n_1, n_2) are (937, 1311), (864, 1210) and (1954, 2708) respectively, leading to economies of approximately 2%, 1% and 9% respectively. Larger economies will be realized when the amount of prior information is more unbalanced between the two populations. The standard frequentist solution remains at $n_1 = n_2 = 769$, which may be inadequate in this case.

7. Discussion

In this paper we have developed criteria for Bayesian sample size determination that address the limitations of frequentist calculations. However, the solutions are non-unique, not only because the choice of prior parameters may not be obvious, but also because the different criteria can lead to substantially different sample sizes. The choice between a criterion that takes an average over the predictive distribution of the unknown data and a worst case criterion would seem to depend on the degree of risk that we are willing to take in any given experiment. The choice between the ACC and the ALC appears somewhat arbitrary, although the convention to report 95% intervals regardless of the data would seem to favour the ALC. The consideration of all criteria should lead to a more informed choice.

In all cases, computations can be straightforwardly and quickly performed either exactly or via simulations. When necessary, excellent starting values for the bisectional search can be found by using approximations to the exact formulae and Stirling's approximation to the gamma function. The techniques in this paper can be extended to non-conjugate prior distributions by appropriate simulation algorithms. Recent progress in Bayesian computing algorithms may be helpful in this regard.

Acknowledgements

The authors wish to thank Marie-Pierre Aoun for help in preparing the manuscript. This work is supported in part by the Natural Sciences and Engineering Research Council of

Canada. L. Joseph is a research scholar supported by the Fonds de la Recherche en Santé du Québec.

Appendix A: derivation of average length for single normal mean

Following a procedure similar to that in Adcock (1988), and under the conditions of Section 3, the length $l(x)$ of the highest posterior density interval of probability coverage $1-\alpha$ is such that

$$P\left\{|\mu-\mu_n| \leqslant \frac{l(x)}{2}\bigg|x\right\}=1-\alpha,$$

where

$$\mu_n=\frac{n_0\mu_0+n\bar{x}}{n_0+n}.$$

Let

$$U_1'=(n+n_0)(\mu-\mu_n)^2,$$
$$U_2'=ns^2$$

and

$$U_3'=\left(\frac{1}{n}+\frac{1}{n_0}\right)^{-1}(\bar{x}-\mu_0)^2.$$

Then, conditional on σ^2,

$$U_i=\frac{U_i'}{2\beta}\sim\frac{1}{2\beta}\sigma^2\chi^2_{\nu_i}, \qquad i=1,2,3,$$

where $\nu_1=\nu_3=1$, $\nu_2=n-1$ and U_1, U_2 and U_3 are independent. The length $l(x)$ is such that

$$P\left\{\frac{n+n_0}{2\beta}(\mu-\mu_n)^2 \leqslant \frac{n+n_0}{2\beta}\frac{l^2(x)}{4}\bigg|x\right\}=1-\alpha.$$

Since the length $l(x)$ is a function only of the statistic U_2+U_3,

$$P\left\{U_1 \leqslant \frac{n+n_0}{2\beta}\frac{l^2(U_2,U_3)}{4}\bigg|U_2,U_3\right\}=1-\alpha$$

$$\Leftrightarrow P\left\{\frac{U_1}{1+U_2+U_3} \leqslant \frac{1}{1+U_2+U_3}\frac{n+n_0}{2\beta}\frac{l^2(U_2,U_3)}{4}\bigg|U_2,U_3\right\}=1-\alpha$$

$$\Leftrightarrow P\left\{F_{1,n+2\nu} \leqslant \frac{n+2\nu}{1+U_2+U_3}\frac{n+n_0}{2\beta}\frac{l^2(U_2+U_3)}{4}\right\}=1-\alpha,$$

since, conditional on U_2 and U_3,

$$\frac{U_1}{1+U_2+U_3}\sim\frac{1}{n+2\nu}F_{1,n+2\nu}.$$

Then

$$\frac{n+2\nu}{1+U_2+U_3}\frac{n+n_0}{2\beta}\frac{l^2(U_2+U_3)}{4}=F_{1,n+2\nu;1-\alpha}=t^2_{n+2\nu;1-\alpha/2},$$

and the average length of the highest posterior density region of probability coverage $1-\alpha$ is

$$E\{l(U_2+U_3)\}=2t_{n+2\nu,1-\alpha/2}\sqrt{\left\{\frac{2\beta}{(n+2\nu)(n+n_0)}\right\}}E\{\sqrt{(1+U_2+U_3)}\}$$

$$= 2t_{n+2\nu,1-\alpha/2}\sqrt{\left\{\frac{2\beta}{(n+2\nu)(n+n_0)}\right\}}\frac{\Gamma\left(\frac{n+2\nu}{2}\right)\Gamma\left(\frac{2\nu-1}{2}\right)}{\Gamma\left(\frac{n+2\nu-1}{2}\right)\Gamma\left(\frac{2\nu}{2}\right)},$$

from which inequality (8) follows. Note that, if a random variable $(n/m)Z \sim F_{m,n}$, then

$$E\{\sqrt{(1+Z)}\} = \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m+n-1}{2}\right)}\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)}.$$

Appendix B: derivation of worst outcome criterion sample size for single normal mean

As above, the probability coverage of the highest posterior density region of length l centred at the posterior mean μ_n can be given as

$$P\left(|\mu-\mu_n| \leqslant \frac{l}{2}\middle|x\right) = P\left(F_{1,n+2\nu} \leqslant \frac{n+2\nu}{1+U_2+U_3}\frac{n+n_0}{2\beta}\frac{l^2}{4}\middle|U_2+U_3\right),$$

which is a decreasing function of U_2+U_3. Since U_2+U_3 is marginally distributed as $(n/2\nu)F_{n,2\nu}$, the probability coverage is greater than $1-\alpha$ with probability $1-w$ if the sample size n is such that

$$P\left(F_{1,n+2\nu} \leqslant \frac{n+2\nu}{1+(n/2\nu)F_{n,2\nu,1-w}}\frac{n+n_0}{2\beta}\frac{l^2}{4}\right) \geqslant 1-\alpha,$$

which directly leads to inequality (9).

Appendix C: derivation of average length for mixed Bayesian–likelihood approach

In the mixed Bayesian–likelihood approach, the prior information about σ^2 is not used for inference (U_1) but is used for the preposterior distribution of the data (U_2 and U_3), invalidating the exact results of Appendixes A and B. Therefore, approximate sample sizes are derived via simulation of the posterior densities.

Since $\bar{x}$ and ns^2 are independent given μ and λ, their joint preposterior distribution is

$$f(\bar{x}, ns^2) = \iint f(\bar{x}, ns^2|\mu,\lambda)\, f(\mu,\lambda)\, \mathrm{d}\mu\, \mathrm{d}\lambda$$

$$= \iint f(\bar{x}|\mu,\lambda)\, f(ns^2|\mu,\lambda)\, f(\mu|\lambda)\, f(\lambda)\, \mathrm{d}\mu\, \mathrm{d}\lambda$$

$$\propto \iint \lambda^{1/2} \exp\{-\tfrac{1}{2}n\lambda(\mu-\bar{x})^2\}\lambda^{(n-1)/2}(ns^2)^{(n-3)/2}\exp\left(-\frac{\lambda}{2}ns^2\right)\lambda^{1/2}$$

$$\times \exp\{-\tfrac{1}{2}n_0\lambda(\mu-\mu_0)^2\}\lambda^{\nu-1}\exp(-\beta\lambda)\, \mathrm{d}\mu\, \mathrm{d}\lambda.$$

Therefore,

$$f(ns^2|\bar{x}) \propto (ns^2)^{(n-3)/2}\iint \lambda^{(n-1)/2+\nu}\exp\left[-\frac{\lambda}{2}\{n(\mu-\bar{x})^2+n_0(\mu-\mu_0)^2+ns^2+2\beta\}\right]\mathrm{d}\mu\, \mathrm{d}\lambda,$$

which after some straightforward algebraic manipulations can be shown to be

$$ns^2|\bar{x} \sim \text{gamma}-\text{gamma}\left\{\frac{2\nu+1}{2}, \frac{nn_0(\bar{x}-\mu_0)^2}{n+n_0}+2\beta, \frac{n-1}{2}\right\}.$$

The average length of the highest posterior density interval of coverage $1-\alpha$ is

$$\frac{2}{\sqrt{\{n(n-1)\}}} E\{\sqrt{(ns^2)}\} t_{n-1;1-\alpha/2}.$$

For further calculations, it is necessary to note that if $Y \sim \text{gamma}(a, b)$ then

$$E(Y^t) = \frac{\Gamma(a+t)}{\Gamma(a) b^t}.$$

Thus

$$\begin{aligned} E\{\sqrt{(ns^2)}\} &= E[E\{\sqrt{(ns^2)}|\lambda\}] \\ &= E\left\{\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\sqrt{2}}{\sqrt{\lambda}}\right\} \\ &= \sqrt{(2\beta)} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\Gamma(\nu - \frac{1}{2})}{\Gamma(\nu)}, \end{aligned}$$

so that the ALC is satisfied if

$$2 t_{n-1;1-\alpha/2} \sqrt{\left\{\frac{2\beta}{n(n-1)}\right\}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \frac{\Gamma(\nu - \frac{1}{2})}{\Gamma(\nu)} \leqslant l,$$

as stated in expression (11).

Appendix D: derivation of Bayesian sample size for difference between two normal means when variances are known

The posterior distribution of $\theta = \mu_2 - \mu_1$ is

$$\theta | x_1, x_2 \sim N\left(\mu_{n_2 2} - \mu_{n_1 1}, \frac{\lambda_{n_1 1} \lambda_{n_2 2}}{\lambda_{n_1 1} + \lambda_{n_2 2}}\right),$$

where $\lambda_{n_i i} = \lambda_i(n_{0i} + n_i)$ and $\mu_{n_i i} = (n_{0i}\mu_{0i} + n_i \bar{x}_i)/(n_{0i} + n_i)$, for $i = 1, 2$. The posterior variance depends only on n and is not a function of the data (x_1, x_2). Therefore, to find the ACC, ALC or WOC sample sizes, the equation

$$\sqrt{\left(\frac{\lambda_{n_1 1} \lambda_{n_2 2}}{\lambda_{n_1 1} + \lambda_{n_2 2}}\right)} \frac{l}{2} \geqslant Z_{1-\alpha/2}$$

must be satisfied. Straightforward algebra leads to the condition

$$\lambda_1 \lambda_2 n^2 + \left\{\lambda_1 n_{02} \lambda_2 + \lambda_2 n_{01} \lambda_1 - \frac{4 Z_{1-\alpha/2}^2}{l^2} (\lambda_1 + \lambda_2)\right\} n + n_{01} \lambda_1 n_{02} \lambda_2 - \frac{4 Z_{1-\alpha/2}^2}{l^2} (n_{01} \lambda_1 + n_{02} \lambda_2) \geqslant 0,$$

from which inequalities (12) and (18) follow.

Appendix E: derivation of exact average coverage criterion solution for difference between two normal means with equal but unknown precision

In this appendix the results of Adcock (1988) are extended to the case of a difference between two normal means. Let

$$U_1' = \frac{(n_1 + n_{01})(n_2 + n_{02})}{n_1 + n_{01} + n_2 + n_{02}} \{\theta - E(\theta|x)\}^2,$$

$$U_2' = n_1 s_1^2,$$

$$U_3' = n_2 s_2^2,$$

$$U_4' = \frac{n_1 n_{01}}{n_1 + n_{01}} (\bar{x}_1 - \mu_{01})^2$$

and

$$U_5' = \frac{n_2 n_{02}}{n_2 + n_{02}} (\bar{x}_2 - \mu_{02})^2.$$

Then, conditional on σ^2, $U_i' \sim \sigma^2 \chi^2_{\nu_i}$, $i = 1, \ldots, 5$, where $\nu_1 = \nu_4 = \nu_5 = 1$, $\nu_2 = n_1 - 1$ and $\nu_3 = n_2 - 1$. Next, define $U_i = U_i'/2\beta$, $i = 1, \ldots, 5$. Then, conditional on σ^2, $U_1, \ldots, U_5$ are independently distributed with

$$U_i \sim \frac{\sigma^2}{2\beta} \chi^2_{\nu_i}, \qquad i = 1, \ldots, 5,$$

where the ν_i are as given above.

The ACC will be satisfied if

$$E\left(P\left[\{\theta - E(\theta|x_1, x_2)\}^2 \leqslant \frac{l^2}{4} \middle| \bar{x}_1, \bar{x}_2, s_1^2, s_2^2 \right] \right) = 1 - \alpha,$$

i.e.

$$E\left(P\left[\frac{(n_1 + n_{01})(n_2 + n_{02})}{2\beta(n_1 + n_2 + n_{01} + n_{02})} \{\theta - E(\theta|x_1, x_2)\}^2 \leqslant \frac{l^2(n_1 + n_{01})(n_2 + n_{02})}{8\beta(n_1 + n_2 + n_{01} + n_{02})} \middle| \bar{x}_1, \bar{x}_2, s_1^2, s_2^2 \right] \right) = 1 - \alpha$$

so that

$$E\left[P\left\{ U_1 \leqslant \frac{l^2(n_1 + n_{01})(n_2 + n_{02})}{8\beta(n_1 + n_2 + n_{01} + n_{02})} \middle| U_2, U_3, U_4, U_5 \right\} \right] = 1 - \alpha,$$

which leads to inequality (14).

Appendix F: derivation of exact average length criterion solution for difference between two normal means with equal but unknown precision

Posterior to sampling, the length $l(x_1, x_2)$ of the highest posterior density region around the posterior mean of $\theta = \mu_2 - \mu_1$ of probability coverage $1 - \alpha$ is such that

$$P\left\{ |\theta - E(\theta|x_1, x_2)| \leqslant \frac{l(x_1, x_2)}{2} \middle| x_1, x_2 \right\} = 1 - \alpha.$$

Using arguments and notation similar to those in Appendix E, $l = l(x_1, x_2)$ is such that

$$P\left[\frac{(n_1 + n_{01})(n_2 + n_{02})}{2\beta(n_1 + n_2 + n_{01} + n_{02})} \{\theta - E(\theta|x_1, x_2)\}^2 \leqslant \frac{l^2(n_1 + n_{01})(n_2 + n_{02})}{8\beta(n_1 + n_2 + n_{01} + n_{02})} \middle| x_1, x_2 \right] = 1 - \alpha$$

i.e.

$$P\left\{U_1 \leqslant \frac{(n_1+n_{01})(n_2+n_{02})}{2\beta(n_1+n_2+n_{01}+n_{02})} \frac{l^2(U_2,U_3,U_4,U_5)}{4} \middle| U_2,U_3,U_4,U_5\right\} = 1-\alpha,$$

which leads to inequality (17).

Appendix G: preposterior distributions of $(\bar{x}_1, \bar{x}_2, n_1 s_1^2, n_2 s_2^2)$

To draw a random sample from the vector $(\bar{x}_1, \bar{x}_2, n_1 s_1^2, n_2 s_2^2)$, the following results can be used:

$$\bar{x}_1 \sim \mu_{01} + \sqrt{\left\{\frac{(n_1+n_{01})\beta}{n_1 n_{01} \nu}\right\}}\, t_{2\nu},$$

$$n_1 s_1^2 | \bar{x}_1 \sim \text{gamma} - \text{gamma}\left\{\frac{2\nu+1}{2}, 2\beta + \frac{n_1 n_{01}(\bar{x}_1-\mu_{01})^2}{n_1+n_{01}}, \frac{n_1-1}{2}\right\},$$

$$\bar{x}_2 | \bar{x}_1, n_1 s_1^2 \sim \mu_{02} + \sqrt{\left\{\frac{n_1 n_{01}(\bar{x}_1-\mu_{01})^2/(n_1+n_{01}) + n_1 s_1^2 + 2\beta}{(2\nu+n_1) n_2 n_{02}/(n_2+n_{02})}\right\}}\, t_{2\nu+n_1}$$

and

$$n_2 s_2^2 | \bar{x}_1, \bar{x}_2, n_1 s_1^2 \sim \text{gamma} - \text{gamma}\left\{\frac{n_1+1}{2} + \nu, n_1 s_1^2 + \frac{n_1 n_{01}(\bar{x}_1-\mu_{01})^2}{n_1+n_{01}} + \frac{n_2 n_{02}(\bar{x}_2-\mu_{02})^2}{n_2+n_{02}} + 2\beta, \frac{n_2-1}{2}\right\}.$$

References

Adcock, C. J. (1987) A Bayesian approach to calculating sample sizes for multinomial sampling. *Statistician*, **36**, 155–159.

———(1988) A Bayesian approach to calculating sample sizes. *Statistician*, **37**, 433–439.

———(1992) Bayesian approaches to the determination of sample sizes for binomial and multinomial sampling—some comments on the paper by Pham-Gia and Turkkan. *Statistician*, **41**, 399–404.

———(1993) An improved Bayesian procedure for calculating sample sizes in multinomial sampling. *Statistician*, **42**, 91–95.

———(1995) The Bayesian approach to determination of sample sizes—some comments on the paper by Joseph, Wolfson and du Berger. *Statistician*, **44**, 155–161.

Berger, J. (1985) *Statistical Decision Theory and Bayesian Analysis*, 2nd edn. New York: Springer.

Bernardo, J. M. and Smith, A. F. M. (1994) *Bayesian Theory*. Chichester: Wiley.

DasGupta, A. and Vidakovic, B. (1997) Sample size problems in ANOVA: Bayesian point of view. *J. Statist. Planng Inf.*, to be published.

Desu, M. M. and Raghavarao, D. (1990) *Sample Size Methodology*. Boston: Academic Press.

Goldstein, M. (1981) A Bayesian criterion for sample size. *Ann. Statist.*, **9**, 670–672.

Gould, A. L. (1993) Sample sizes for event rate equivalence trials using prior information. *Statist. Med.*, **12**, 1209–1223.

Hutton, J. L. and Owens, R. G. (1993) Bayesian sample size calculations and prior beliefs about child sexual abuse. *Statistician*, **42**, 399–404.

Joseph, L., du Berger, R. and Bélisle, P. (1996) Bayesian and mixed Bayesian/likelihood criteria for sample size determination. *Statist. Med.*, **15**, in the press.

Joseph, L., Wolfson, D. B. and du Berger, R. (1995a) Sample size calculations for binomial proportions via highest posterior density intervals. *Statistician*, **44**, 143–154.

———(1995b) Some comments on Bayesian sample size determination. *Statistician*, **44**, 167–171.

Lemeshow, S., Hosmer, Jr, D. W., Klar, J. and Lwanga, S. K. (1990) *Adequacy of Sample Size in Health Studies*. Chichester: Wiley.

Lipsey, M. W. (1990) *Design Sensitivity, Statistical Power for Experimental Research*. Newbury Park: Sage.

Pham-Gia, T. G. (1995) Sample size determination in Bayesian statistics—a commentary. *Statistician*, **44**, 163–166.

Pham-Gia, T. G. and Turkkan, N. (1992) Sample size determination in Bayesian analysis. *Statistician*, **41**, 389–397.

Spiegelhalter, D. J. and Freedman, L. S. (1986) A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion. *Statist. Med.*, **5**, 1–13.