Journal of the Korean Data & Information Science Society
2010, **21**(6), 1343–1351

한국데이터정보과학회지

Two Bayesian methods for sample size determination in clinical trials[†]

Sang Gyu Kwak [1] · Dal Ho Kim[2]· Im Hee Shin [3] · Ho Gak Kim[4]
Sang Gyung Kim[5]

[12]Department of Statistics, Kyungpook National University
[345]School of Medicine, Catholic University of Daegu

Received 23 October 2010, revised 23 October 2010, accepted 15 November 2010

Abstract

Sample size determination is very important part in clinical trials because it influences the time and the cost of the experimental studies. In this article, we consider the Bayesian methods for sample size determination based on hypothesis testing. Specifically we compare the usual Bayesian method using Bayes factor with the decision theoretic method using Bayesian reference criterion in mean difference problem for the normal case with known variances. We illustrate two procedures numerically as well as graphically.

Keywords: Bayes factor, Bayesian reference criterion, clinical trial, hypothesis testing, normal case, sample size.

1. Introduction

Sample size determination would be one of the most general tasks in clinical trials. There are two approaches in sample size determination problem. One is the frequentist one and the other is the Bayesian one. We can determine the sample size based on confidence interval, hypothesis test, and many others. The frequentist approach is seeking the smallest sample size that is sufficient to achieve a desired power at a specified significance level. The Bayesian approach is finding the smallest sample size that is necessary to obtain a desired rate of

[†] This study was supported by a grant of the Korea Healthcare technology R&D Project, Ministry for Health & Welfare, Republic of Korea (Project No.: A084177).

[1] Doctor of philosophy student, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea.
[2] Corresponding author: Professor, Department of Statistics, Kyungpook National University, Daegu 702-701, Korea. E-mail: dalkim@knu.ac.kr
[3] Professor, Department of Medical Statistics, School of Medicine, Catholic University of Daegu, Daegu 705-718, Korea.
[4] Professor, Division of Gastroenterology, Department of Internal Medicinel, School of Medicine, Catholic University of Daegu, Daegu 705-718, Korea.
[5] Professor, Department of Laboratory Medicine, School of Medicine, Catholic University of Daegu, Daegu 705-718, Korea.

correct classification of the hypothesis as true or false. The purpose of this paper is to explore Bayesian methods in which one uses Bayes factor and the other uses Bayesian reference criterion in hypothesis testing. We apply those methodologies to the normal case with known variance.

George and Desu (1974), Wu*et al* . (1980) and Sozu *et al* . (2006) reviewed and investigated the sample size determination problem in clinical trials. Adcock (1997), Weiss (1997), Desu and Raghavarao (1990), Florens and Mouchart (1993) studied the sample size determination problem in either Bayesian or frequentist hypothesis testing setup. Inoue *et al* . (2005) investigate the relationship between Bayesian and frequentist sample size determination.

The paper is organised as follows. In Section 2 we consider Bayesian sample size determination. In Bayesian hypothesis testing we compares the posterior probability of H_0 (the null hypothesis) and the posterior probability of H_1 (the alternative hypothesis) or equivalently, the Bayes factor. The sample size determination based on Bayes factor requires the desired rate for correctly identifying a hypothesis as true or false. It ensures a minimum rate r^* of correct classification. Also the decision theoretic approach to hypothesis testing using Bayesian reference criterion is considered. We calculates the logarithmic discrepancy, the corresponding reference posterior expectation, and decides some constant d^* for reject the any hypothesis. But this task needs to decide prior probability.

Two approaches to sample size determination based on Bayes factor and Bayesian reference criterion differ in philosophy and goals. However, both provide algorithms for determining a sample size from specified criteria and inputs. In Section 3, we utilize both approaches to the normal case with known variance. Some discussion is given in Section 4.

2. Methodologies

2.1. Sample size determination with Bayes factor

Using traditional notation, let H_0 and H_1 denotethe null and alternative hypothesis respectively. Type I and Type II error rates are denoted by α and β respectively and z_α denote the α-quantile of the standard normal distribution, that is, $\Phi(z_\alpha) = P(Z \leq z_\alpha) = \alpha$. We formalize the sample size determination goal by specifying a Bayesian goal function $G_B(n, \mathbf{v})$. And $\mathbf{v} = (\pi, \alpha, \delta, \sigma)$ is the Bayesian input which needs for sample size determination. This function must have information, mean squared prediction error, size of probability interval and classification error.

We want to test the hypothesis $H_0 : \theta = \theta_0$ verse $H_1 : \theta = \theta_1$, where $\theta_0 < \theta_1$. Under the Bayesian approach, uncertainty about θ is presented by its prior distribution. Assume a priori that $P(\theta_0) = 1 - P(\theta_1) = \pi$. A Bayesian decision between H_0 and H_1 is based on their posterior probabilities.

Suppose that the null hypothesis H_0 is not rejected if the posterior probability of the null hypothesis H_0 is at least $1/(1 + K)$. This cutoff-point for the posterior probability is consistent with a $0 - 1 - K$ loss function shown in Table 2.1 and minimizes the posterior expected loss (Berger, 1985).

Table 2.1 0-1-K loss function

	H_0 is True	H_1 is True	Sum
Not Reject H_0	0	1	1
Reject H_0	K	0	K
Sum	K	1	K+1

: true
: false

We can obtain cutoff-point by following expression.

$$\frac{KP(H_0|\boldsymbol{x})}{P(H_1|\boldsymbol{x})} \geq 1. \tag{2.1}$$

Since theorem of conditional probability, we can express (2.1) in inequation difference form

$$\frac{P(\boldsymbol{x}|H_1)P(\theta_1)}{P(\boldsymbol{x}|H_0)P(\theta_0)} \leq K \tag{2.2}$$

Then we choose the Bayesian goal function to be the rate of correctly classifying a hypothesis as true or false. The Bayesian goal function can be formally calculated as follows.

$$\begin{aligned} G_B(n, \boldsymbol{v}) = & KP(H_0)P(correct\ decision|H_0) \\ & + P(H_1)P(correct\ decision|H_1). \end{aligned} \tag{2.3}$$

A Bayesian finds the sample size to ensure a minimum rate r^* of correct classification.

2.2. Sample size determination with Bayesian reference criterion

To decide whether or not some data $\boldsymbol{x}$ are compatible with the null hypothesis (H_0) $\theta = \theta_0$, assuming that the data have been generated from the model

$$p_{\boldsymbol{x}}(\cdot|\theta, \sigma),\ \theta \in \Theta, \sigma \in \Omega \tag{2.4}$$

steps are as follows (Bernardo and Rueda, 2002):

(i) Compute the logarithmic discrepancy,

$$\delta(\theta_0, \theta, \sigma) = inf_{\sigma \in \Omega} \int p_{\boldsymbol{x}}(\boldsymbol{y}|\theta, \sigma) \log \frac{p_{\boldsymbol{x}}(\boldsymbol{y}|\theta, \sigma)}{p_{\boldsymbol{x}}(\boldsymbol{y}|\theta_0, \sigma)} d\boldsymbol{y} \tag{2.5}$$

between the assumed model and its closest approximation under the null hypothesis (H_0) $\theta = \theta_0$.

(ii) Derive the corresponding reference posterior expectation

$$d_r(\boldsymbol{x}, \theta_0) = \int \int \delta(\theta_0, \theta, \sigma) \pi_\delta(\theta, \sigma|\boldsymbol{x}) d\theta d\sigma \tag{2.6}$$

(iii) For some d^*, reject the hypothesis $\theta = \theta_0$, if and only if, $d_r(\boldsymbol{x}, \theta_0) > d^*$, where values such as $d^* = 2.5$ (mild evidence against θ_0) or $d^* = 5$ (significant evidence against θ_0) may conveniently be chosen for scientific communication.

The choice of d^* is formally determined by the utility gain which may be expected by using the null model when it is true. The larger that gain, the larger d^*. The analysis above suggests that $d_r(\boldsymbol{x}, \theta_0)$ close to 1 may be expected if H_0 is true, and that d_r-values over 5 should typically be regarded as significant evidence against the suitability of using H_0 as a proxy to H_1.

If $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ is a sufficiently large random sample from a regular model $p(x|\theta, \sigma)$, the posterior distribution of (θ, σ) will concentrate on their maximum likelihood estimates $(\widehat{\theta}, \widehat{\sigma})$, and thus the expected posterior discrepancy $d_r(\boldsymbol{x}, \theta_0)$ will be close to $\delta(\theta_0, \widehat{\theta}, \widehat{\sigma})$, the logarithmic discrepancy between the model identified by $(\widehat{\theta},\widehat{\sigma})$ and its closest approximation under the null hypothesis.

Moreover, if $\boldsymbol{x} = \{x_1, \cdots, x_n\}$ is random sample from a model $p_x(x|\theta)$, where θ is one-dimensional and there no nuisance parameters, then $\delta(\theta_0, \theta)$ will typically be a piecewise invertible function of θ and hence the relevant reference prior will simple be Jeffrey's prior, that is $\pi_\delta(\theta) \propto i(\theta)^{1/2}$, where $i(\theta)$ is Fisher's information function. Thus, in terms of the natural parameterization, defined as $\phi = \phi(\theta) = \int^\theta i(\theta)^{1/2} d\theta$, the reference prior $\pi_\delta(\theta)$ will be uniform. For large sample sizes, the corresponding reference posterior distribution of ϕ will then be approximately normal $\pi_\delta(\phi|\boldsymbol{x}) \approx N(\phi|\widehat{\phi}, 1/\sqrt{n})$, and will only depend on the data through its mle $\widehat{\phi}$. Moreover, the sampling distribution of $\widehat{\phi}$, $p(\widehat{\phi}, \phi)$ will also be approximately normal, $p(\widehat{\phi}|\phi) \approx N(\phi|\widehat{\phi}, 1/\sqrt{n})$. Since the discrepancy function is invariant under one-to-one reparametrization, and hence $\delta(\phi_0, \phi) = \delta(\theta_0, \theta)$ one obtains, after some algebra,

$$d_r(\boldsymbol{x}, \theta_0) \approx \frac{1}{2}[1 + z^2(\widehat{\theta}, \theta_0)], \tag{2.7}$$

where $z(\widehat{\theta}, \theta_0) = \sqrt{n}[\phi(\widehat{\theta}) - \phi(\theta)]$.

This type of approximation may be extended to multivariate situation, with or without nuisance parameters.

3. The normal case

We want to test the hypothesis $H_0 : \theta = \theta_0$ verse $H_1 : \theta = \theta_1$, where $\theta_0 < \theta_1$. The setting is normal likelihood with known variance. Let's apply two methods for sample size determination to the normal case.

3.1. Sample size using Bayes factor

Suppose that $x_1, \cdots, x_n \sim N(\theta, \sigma^2)$, σ is known and it desired to test the hypothesis $H_0 : \theta = \theta_0$ verse $H_1 : \theta = \theta_1$, where $\theta_0 < \theta_1$.

And suppose that the hypothesis H_0 is not rejected of the posterior probabilities of the null hypothesis is at least $1/(1 + K)$. And the loss function follow Table1. First, we can

obtain the cutoff point by

$$\frac{P(\boldsymbol{x}|H_1)P(\theta_1)}{P(\boldsymbol{x}|H_0)P(\theta_0)} \leq K \tag{3.1}$$

The prior probability is $P(\theta_0) = \pi,\ P(\theta_1) = 1-\pi$ respectively, and likelihood is $P(\boldsymbol{x}|\theta) \sim N(\theta, \sigma^2)$. Then (3.1) can be rewritten as

$$\bar{x} \leq \frac{\sigma^2 \log\left(K\dfrac{\pi}{1-\pi}\right)}{n(\theta_1 - \theta_0)} + \frac{(\theta_1 + \theta_0)}{2} \tag{3.2}$$

This is cut-off point for the posterior probability is consistent with a $0-1-K$ loss function shown in Table 1 and minimizes the posterior expected loss (Berger 1985). Moreover, this cutoff point imples that the null hypothesis is not rejected if (3.4). Next, we can obtain the Bayesian goal function can be formally calculated as

$$\begin{aligned} G_B(n, \boldsymbol{v}) = {} & KP(H_0)P(\textit{correct decision}|H_0) \\ & + P(H_1)P(\textit{correct decision}|H_1). \end{aligned} \tag{3.3}$$

Using (3.4) $G_B(n, \boldsymbol{v})$ can be written as

$$\begin{aligned} G_B(n, \boldsymbol{v}) = {} & K\pi P\left(Z \leq \frac{\sigma \log\left(K\dfrac{\pi}{1-\pi}\right)}{\sqrt{n}\delta} + \frac{\delta\sqrt{n}}{2\sigma}\right) \\ & + (1-\pi)P\left(Z \geq \frac{\sigma \log\left(K\dfrac{\pi}{1-\pi}\right)}{\sqrt{n}\delta} - \frac{\delta\sqrt{n}}{2\sigma}\right), \end{aligned} \tag{3.4}$$

where Z is the standard normal, $\delta = \theta_1 - \theta_0$.

Then, we can obtain the sample size by using the input $\boldsymbol{v_1} = (\pi, \alpha, \delta, \sigma)$. If $\boldsymbol{v_1} = (0.5, 0.5, 0.1, 1)$ then the sample size is 857. And if $\boldsymbol{v_1} = (0.5, 0.5, 0.05, 1)$ then the sample size is 3,426. Table 3.1 and Figure 3.1 are showing the sample size with the changes of δ, r^* and fixed $\pi = .5, \sigma = 1$, $\alpha = 0.5$.

Table 3.1 The sample sizes using Bayes factor

σ	δ / r^*	**0.5**	**0.75**	**1.00**	**1.25**	**1.50**	**1.75**	**2.00**
	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.6	1.0	0.5	0.3	0.2	0.1	0.1	0.1
1	**0.7**	4.4	2.0	1.1	0.7	0.5	0.4	0.3
	0.8	11.3	5.0	2.8	1.8	1.3	0.9	0.7
	0.9	26.3	11.7	6.6	4.2	2.9	2.1	1.6
	0.5	0	0	0	0	0	0	0
	0.6	9.2	4.1	2.3	1.5	1	0.8	0.6
3	**0.7**	39.6	17.6	9.9	6.3	4.4	3.2	2.5
	0.8	102	45.3	25.5	16.3	11.3	8.3	6.4
	0.9	236.5	105.1	59.1	37.8	26.3	19.3	14.8
	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.6	25.7	11.4	6.4	4.1	2.9	2.1	1.6
5	**0.7**	110.0	48.9	27.5	17.6	12.2	9.0	6.9
	0.8	283.3	125.9	70.8	45.3	31.5	23.1	17.7
	0.9	657.0	292.0	164.2	105.1	73.0	53.6	41.1

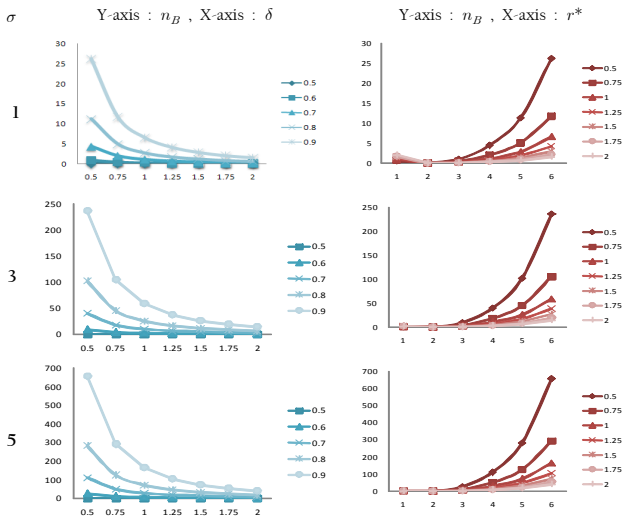

Figure 3.1 Graphs of sample sizes using Bayes factor

3.2. Sample size using Bayesian reference criterion

In this section we will be discuss the sample size determination using Bayesian reference criterion. The setting is same as Section 3.1. We will calculate the sample size following the steps given in Section 2.2. First, we can compute the logarithmic discrepancy,

$$\delta(\theta_0, \theta, \sigma) = inf \int N(\boldsymbol{x}|\theta, \sigma) \log \frac{N(\boldsymbol{x}|\theta, \sigma)}{N(\boldsymbol{x}|\theta_0, \sigma)} d\boldsymbol{x} \tag{3.5}$$

Since σ is known, the infimum is not available anymore.

We know that $\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) dx = E(X) = \theta$, and $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right) dx = 1$. So, (3.8) can be re-written as

$$\delta(\theta_0, \theta, \sigma) = \frac{n}{2\sigma^2}(\theta - \theta_0)^2 = \frac{n}{2}\left(\frac{\theta - \theta_0}{\sigma}\right)^2 \tag{3.6}$$

Next, we can derive the corresponding reference posterior expectation,

$$d_r(x, \theta_0) = \frac{n}{2}\int \left(\frac{\theta - \theta_0}{\sigma}\right)^2 N\left(\theta|\bar{x}, \frac{\sigma^2}{n}\right) d\mu = \frac{1}{2}\left(1 + \frac{(\bar{x} - \theta_0)^2}{\sigma^2/n}\right). \tag{3.7}$$

Therefore, $d_r(x, \theta_0) \equiv (1 + z^2)/2$, where $z = (\bar{x} - \theta_0)/(\sigma/\sqrt{n})$.

Then we can obtain the sample size by using the input $\boldsymbol{v}_2 = (\delta^*, \bar{x}, \theta_0, \sigma)$. If $\boldsymbol{v}_2 = (2.5, 1, 0, 1)$ then the sample size is 4. Table 3.2 and Figure 3.2 are showing the sample size with the changes of d^*, $\bar{x}$, and σ. Here we put $\theta_0 = 0$.

Table 3.2 The sample sizes under Bayesian reference criterion

σ	$\frac{\lvert\bar{x}\rvert}{d^*}$	**0.5**	**0.75**	**1.00**	**1.25**	**1.50**	**1.75**	**2.00**
	2	12.0	5.3	3.0	1.9	1.3	1.0	0.8
	3	20.0	8.9	5.0	3.2	2.2	1.6	1.3
1	**4**	28.0	12.4	7.0	4.5	3.1	2.3	1.8
	5	36.0	16.0	9.0	5.8	4.0	2.9	2.3
	6	44.0	19.6	11.0	7.0	4.9	3.6	2.8
	2	108.0	48.0	27.0	17.3	12.0	8.8	6.8
	3	180.0	80.0	45.0	28.8	20.0	14.7	11.3
3	**4**	252.0	112.0	63.0	40.3	28.0	20.6	15.8
	5	324.0	144.0	81.0	51.8	36.0	26.4	20.3
	6	396.0	176.0	99.0	63.4	44.0	32.3	24.8
	2	300.0	133.3	75.0	48.0	33.3	24.5	18.8
	3	500.0	222.2	125.0	80.0	55.6	40.8	31.3
5	**4**	700.0	311.1	175.0	112.0	77.8	57.1	43.8
	5	900.0	400.0	225.0	144.0	100.0	73.5	56.3
	6	1100.0	488.9	275.0	176.0	122.2	89.8	68.8

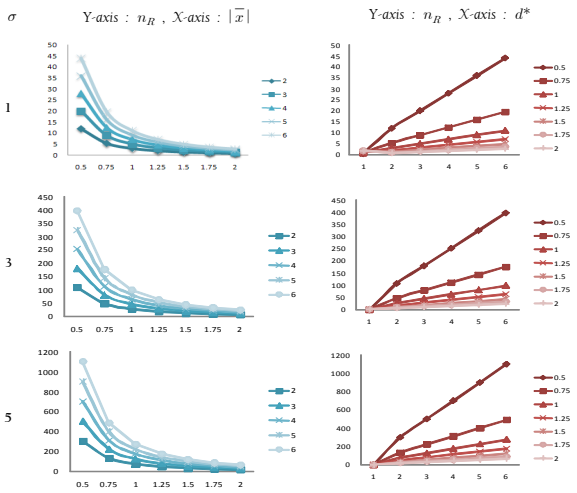

Figure 3.2 Graphs of sample sizes under Bayesian reference criterion

4. Discussion

In this paper, we described two methods for the determination of sample size in clinical trials. We also apply two approaches to the normal case with known variance. Under the method using Bayes factor, we can look around some interesting information. First, the sample size is in direct proportion to r^*. Second, the sample size is in inverse proportion to δ. Under the method using Bayesian reference criterion, we can also observe some interesting information. First, the sample size is in direct proportion to d^*. Second, the sample size is in inverse proportion to absolute $\bar{x}$. Table 4 shows that the comparison of the relations in two methods for sample size determinations. The rate of correct classification (r^*) and the cut-off value for reject hypothesis (d^*) are direct proportion to sample size. This mean that if we want to determine small sample size then we must decide small r^* and d^*. And the difference between the null and the alternative hypothesis (δ) and the absolute $\bar{x}$ ($|\bar{x}|$) are inverse proportion to sample size. This mean that if we want to determine small sample size then we must decide big δ and $|\bar{x}|$.

Table 4.1 The relation of, and r^*, d^*, σ, δ, $|\bar{x}|$.

Relation	Direct proportion	Inverse proportion
n_B	r^*, σ	δ
n_R	d^*, σ	$\|\bar{x}\|$

These methods needs so complex calculation. But Bayesian method includes more information than frequentist one. Furthermore, we can pursue these methods for sample size determination in normal case with unknown variance.

References

Adcock, C. J. (1997). Sample size determination: A review. *The Statistician*, **46**, 261-283.

Berger, J. O. (1985). *Statistical decision theory and Bayesian analysis*, 2nd Ed., Springer-Verlag, New York.

Bernardo, J. M. and Rueda, R. (2002). Bayesian hypothesis testing: A reference approach. *International Statistical Review*, **70**, 351-372.

Desu, M. and Raghavarao, D. (1990). *Sample size methodology*, Academic Press, New York.

Florens, J. P. and Mouchart, M. (1993). Bayesian testing and testing Bayesians. *Handbook of Statistics (G. S. Maddala and C. R. Rao, eds.)*, **11**, 303-334.

George, S. L. and Desu, W. W. (1974). Planning the size and duration of a clinical trial studying the time to some critical event. *Journal of Chronic Diseases*, **27**, 15-24.

Inoue, L. Y. T., Berry, D. A. and Parmigiani, G. (2005). Relationship between Bayesian and frequentist sample size determination. *American Statistician*, **59**, 79-87.

Sozu, T., Sugimoto, T. and Hamasaki, T. (2010). Sample size determination in clinical trials with multiple co-primary binary endpoints. *Statistics in Medicine*, **29**, 2169-2196.

Weiss, R. (1997). Bayesian sample size calculations for hypothesis testing. *The Statistican*, **46**, 185-191.

Wu, M., Fisher, M. and DeMets, D. (1980). Sample sizes for long term medical trials with time-dependent dropout and events rates. *Controlled Clinical Trials*, **1**, 111-123.