International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:4, No:5, 2010

Estimation of Bayesian Sample Size for Binomial Proportions Using Areas P-tolerance with Lowest Posterior Loss

H. Bevrani and N. Najafi

***Abstract*—**This paper uses p-tolerance with the lowest posterior loss, quadratic loss function, average length criteria, average coverage criteria, and worst outcome criterion for computing of sample size to estimate proportion in Binomial probability function with Beta prior distribution. The proposed methodology is examined, and its effectiveness is shown.

***Keywords*—**Bayesian inference, Beta-binomial Distribution, LPL criteria, quadratic loss function.

I. Introduction

ESTIMATION of sample size is one of the questions that we often encounter in research and applied designs. Various methods have been proposed to estimate Bayesian sample size. Freedman and Spiegelhalter [11] and Spiegelhalte et al. [12], used Bayesian approach for anticipating hypothesis testing power. Adcock [1], Pham Gia and Turkkan [10] used it for interval estimation based on normal approximations of posterior densities or intervals based on means and posterior variances.

One of the goals of estimating sample size is to make inference or decision about uncertain parameter θ. In classic approach which has been summarized by Desu and Raghavarao in [4], the main problem is to find point estimate $\hat{\theta}$ for unknown θ. Since we have no information about θ behavior in this method, it is suppose to be constant, thus the estimated sample size will encounter more error. Whereas Bayesian approach makes it possible to use prior distribution of θ instead of its point estimate.

The present paper deals with Bayesian sample size estimation based on lowest posterior loss (LPL) intervals. The LPL intervals are those intervals that posterior risk of points inside this area is less than posterior risk of points outside this area. Bernardo [3] calculated LPL intervals for binomial proportion, using real loss function. In the present paper, applying the theory of decision and using loss function through three methods i.e. average length, average coverage, and worst outcome criteria, the optimal sample size for parameter θ in binomial probability function with parameters n and θ through Beta prior distribution with parameters α and β will be obtained.

Joseph et al. [6] used the three methods for determining Bayesian sample size of binomial parameter based on HPD intervals. The first paper based on decision theory for determining sample size, according to utility function was presented by Grundy et al. [5], and it was developed by Lindley in [8].

H. Bevrani is with the University of Tabriz, Tabriz, CO 566661111 IRAN (corresponding author phone: 411-339-2871; fax: 411- 334-2102; e-mail: bevrani@ tabrizu.ac.ir).

N. Najafi is with the AZAD University of Makoo, , CO 80523 IRAN (e-mail: nargesnajafi@gmail.com).

II. Bayesian Sample Size Methods for Binomial Proportion

Suppose that random variable X has binomial distribution with parameters n and θ, i.e. $f(x \setminus \theta) = Bi(n,\theta)$, in which n refers to sample size. Moreover, assume that θ has Beta prior distribution with parameters α and β, $\pi(\theta) = Be(\theta \setminus \alpha, \beta)$.

Using Baye's theorem, θ posterior distribution, Beta distribution,

$$\pi(\theta \setminus x,n,\alpha,\beta) = Be(\theta \setminus x+\alpha, n-x+\beta) \quad (1)$$

and predictor density function of X, will be Beta-binomial distribution which is defined as follows:

$$P_X(x \setminus n,\alpha,\beta) = \binom{n}{x} \frac{B(\alpha+x, n-x+\beta)}{B(\alpha,\beta)}, \quad (2)$$

for x= 0,1,2,…,n. Here, $B(\alpha,\beta)$ indicates the beta function with parameters α and β.

A. LPL Tolerance Regions for Binomial Parameter

It seems natural to define p-tolerance lowest posterior loss (LPL) region estimators for any loss function of $L(\theta,\delta(x))$, [3]. This region with p probability, contain $\delta(x)$ values whose

expected loss $L(\delta(x)\backslash x)$, is smaller than that of any $\delta(x)$ values outside the region.

Let assume the quadratic loss function

$$L(\theta,\delta(x)) = (\delta(x)-\theta)^2 \tag{3}$$

be an error, then the posterior loss will be defined as:

$$R(\theta,\delta(x)) = \int_{\Theta} L(\theta,\delta(x))\pi(\theta\backslash x)d\theta. \tag{4}$$

Hence, a lowest posterior loss p-tolerance region for binomial distribution is a subset of the parameter space $R_p^l = R_p^l(\underline{x},\Theta)$ if we have:

$$\int_{R_p^l} Be(\theta\backslash x+\alpha, n-x+\beta)d\theta = p, \tag{5}$$

$$R(\theta_i,\delta(x)) \le R(\theta_j,\delta(x)), \quad \forall\theta_i \in R_p^l,\ \forall\theta_j \notin R_p^l \tag{6}$$

then, will be as follow:

$$R(\theta,\delta(x)) = \int_0^1 Be(\theta\backslash x+\alpha, n-x+\beta)(\delta(x)-\theta)^2 d\theta \tag{7}$$

Now, let $LPL_L(x,n,\alpha,\beta,l) = (h,g)$, $h<g$, be the corresponding LPL interval for of given length l and let $LPL_c(x,n,\alpha,\beta,p)$ be an LPL interval for θ of given posterior coverage p (Lan and et al., 2008). Define:

$$l_p'(x,n) = \int_{LPL_c(x,n,\alpha,\beta,p)} d\theta, \tag{8}$$

$$p_l'(x,n) = \int_{LPL_l(x,n,\alpha,\beta,l)} \pi(\theta\backslash x,n,p,\beta)d\theta. \tag{9}$$

In which, (8) indicates the actual length LPL interval with posterior coverage p, and (9) indicates actual posterior coverage LPL interval with known length l for known values x and n.

B. ALC for Binomial Parameter

For a given fixed LPL interval coverage p, find the minimum sample size n such that the expected length is utmost l, i.e. average length criterion (ALC), seeks the smallest n such that :

$$\sum_{x=0}^{n}\left[\int_{LPL_c(x,n,\alpha,\beta,p)} d\theta\right]\binom{n}{x}\frac{B(\alpha+x,n-x+\beta)}{B(\alpha,\beta)} \le l \tag{10}$$

In which l is the prespecified average length. Left side inequality (10), is the mean length LPL interval for various values of x [6].

C. ACC for Binomial Parameter

In contrast to the ALC, an average coverage criterion (ACC) seeks the minimum sample size n in such a way that we can have:

$$\sum_{x=0}^{n}\left[\int_{LPL_l(x,n,\alpha,\beta,l)} Be(\theta\backslash x+\alpha,n-x+\beta)d\theta\right]\binom{n}{x}\frac{B(\alpha+x,n-x+\beta)}{B(\alpha,\beta)} \ge p. \tag{11}$$

In other words, this method, by fixing the length of the LPL, will provide the probability of posterior coverage for different values of x and minimum n in such a way that the average of this posterior coverage at least becomes p, in which p is a definite value [6].

D. WOC for Binomial Parameter

Two criterions, ACC and ALC, only calculate the average of lengths or coverage. They have no guarantee for any particular. Another conservative approach which assures that expected and desirable convergence probability and length should be created on every single observation is worst outcome criteria (WOC). The worst outcome method finds the least n so that, we have:

$$\inf_{0\le x\le n}\left[\int_{LPL_l(x,n,\alpha,\beta,l)} \pi(\theta\backslash x,n,p,\beta)d\theta\right] \ge p. \tag{12}$$

where, p and l are constant values [6].

III. Simulation

For prior distribution of Beta with $\alpha=1, \beta=1, n=10$, and $x=2$, posterior risk has been shown in Fig. 1, and LPL region has been shown in Fig. 2. As these diagrams show, total algorithm for acquiring LPL areas is that we find a constant value from posterior risk, in such a way that posterior risk of points inside this area are minimum, and the level under this diagram for this area on posterior density function, is 0.95.

The values of estimated sample size for known parameters $\alpha=1$ and $\beta=1$ using three criterion, namely average coverage, average length and worst outcome have been presented on the basis of various values of p and length l in tables 1, 2, and 3, respectively. First line of these tables refers to various values of length and first column of left side refers to various values of coverage rate. For example, with coverage rate of 0.9 and length of 0.3, acquired sample sizes through ACC, ALC and WOC are 19, 15, and 30, respectively.

Fig. 3 shows sample size variations with mentioned three methods for l=0.3 and various values of p, Fig. 4 shows sample size variations with these three methods for p=0.9 and

various values of l. As it can be seen, sample size acquired through WOC method is larger than the other two methods. In Fig. 3, for constant length (l), by increasing coverage (p), sample size increases, and in Fig. 4 for constant p by increasing l, sample size decreases.

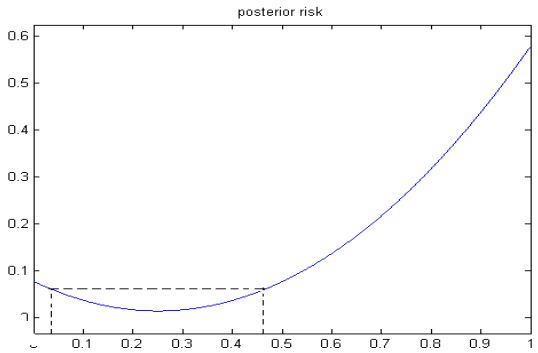

Fig. 1 Posterior risk for a binomial parameter with $\alpha = 1$ and $\beta = 1$

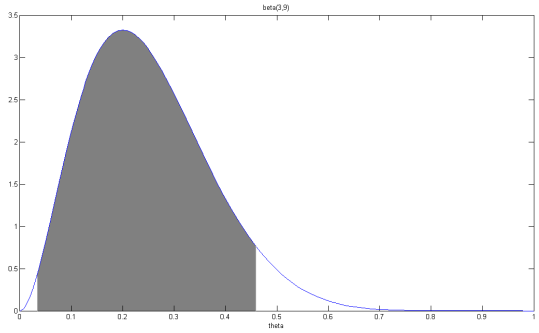

Fig 2. LPL 0.95-credible region for a binomial parameter with $\alpha = 1$ and $\beta = 1$

TABLE I
ACC SAMPLE SIZE FOR VARIOUS LENGTHS AND COVERAGE WITH $\alpha = 1$ AND $\beta = 1$

p \ l	0.2	0.25	0.3	0.4	0.5
0.5	6	4	3	1	1
0.8	26	16	10	5	2
0.85	34	21	14	7	4
0.9	46	29	19	10	5
0.95	70	44	29	15	9
0.99	132	83	55	30	18

TABLE II
ALC SAMPLE SIZE FOR VARIOUS LENGTHS AND COVERAGE WITH $\alpha = 1$ AND $\beta = 1$

p \ l	0.2	0.25	0.3	0.4	0.5
0.5	7	4	3	1	1
0.8	23	14	9	5	3
0.85	28	17	12	6	3
0.9	37	23	15	8	4
0.95	52	32	22	11	6
0.99	82	60	35	18	11

TABLE I
WOC SAMPLE SIZE FOR VARIOUS LENGTHS AND COVERAGE WITH $\alpha = 1$ AND $\beta = 1$

p \ l	0.2	0.25	0.3	0.4	0.5
0.5	11	7	3	2	1
0.8	44	27	18	9	5
0.85	55	34	23	11	7
0.9	73	45	30	15	9
0.95	103	64	43	17	13
0.99	132	119	74	39	23

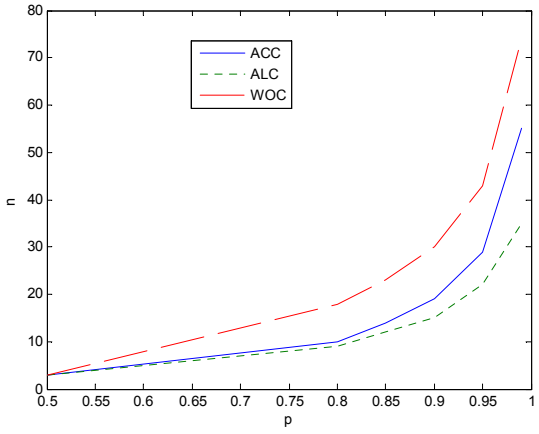

Fig 3. ACC, ALC and WOC sample size for various coverage and $l = 0.3$

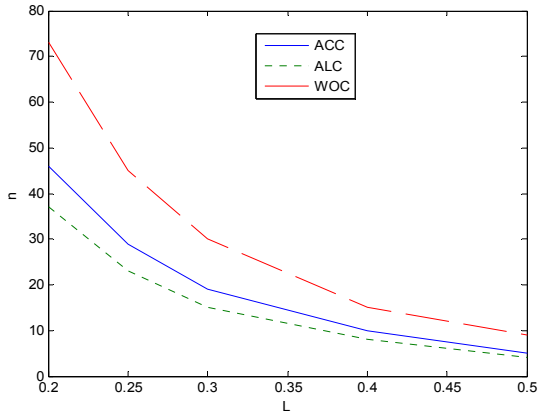

Fig 4. ACC, ALC and WOC sample size for various length and $p = 0.9$

IV. Conclusion

This paper deals with several computational methods to determine Bayesian sample size. As mentioned earlier, Joseph et al. used these methods for HPD intervals of binomial probability function. In this paper, we have described LPL areas, then we have used LPL areas instead of HPD intervals. Although HPD intervals are shorter than LPL intervals, the advantage that LPL intervals show the application of loss function in computing these intervals. In tables 1, 2, and 3, sample size of various values of l and p, and prior constant parameters have been acquired using ALC, ACC and WOC, respectively. As it is observed, sample size acquired using WOC method is larger than other methods. For constant coverage, the sample size increases when the length decreases, and with constant length, the sample size increases when coverage increases.

References

[1] C. J. Adcock, "A Bayesian Approach to Calculating Sample Sizes", The Statistician: Journal of the Institute of Statisticians, vol. 37, pp. 433-439,1988.
[2] C. J. Adcock, "Sample Size Determination: A Review", The Statistician: Journal of the Institute of Statisticians, vol. 46, pp. 261-283, 1997.
[3] J. M. Bernardo, "Intrinsic credible regions: an objective Bayesian approach to interval estimation", Test, vol. 14, pp. 317- 384, 2005.
[4] M. M. Desu, And D. Raghavarao, "Sample Size Methodology", Boston: Academic Press, 1990.
[5] P. M. Grundy, M. J. R. Healy and D. H Rees, "Economic choice of the amount of experimentation", J. R. Statist. Soc. A, vol. 18, pp.32-48, 1956.
[6] L. Joseph, D. B. Wolfson, and R. D. Berger, "Sample Size Calculations for Binomial Proportions Via Highest Posterior Density Intervals", The Statistician: Journal of the Institute of Statisticians, vol. 44, pp. 143-154, 1995.
[7] L. Joseph, P. Belisle and P. Bélisle, "Bayesian Sample Size Determination for Normal Means and Differences between Normal Means", The Statistician: Journal of the Institute of Statisticians, vol. 46, pp. 209-226, 1997a.
[8] D. V. Lindley, "The choice of Sample size", Statistician, vol. 46, pp. 129-138.
[9] C. E. M´lan, L. Joseph and D. B. Wolfson, "Bayesian Sample Size Determination for Binomial Proportions", Journal of the Bayesian Analysis, vol. 2, pp.269-296, 2008.
[10] T. Pham-Gia and N. Turkkan, "Sample Size Determination in Bayesian Analysis" (Disc: P399-404), The Statistician: Journal of the Institute of Statisticians, vol. 41, pp.389-397, 1992.
[11] D. J. Spiegelhalter and L. S Freedman, "A predictive approach to selecting the size of a clinical trial, based on subjective clinical opinion", Statist. Med., vol. 5, pp.1-13, 1986.
[12] D. J. Spiegelhalter, L. S. Freedman and M. K. B. Parmar, "Bayesian approaches to randomized trials (with discussion)", J. R. Statist. Soc. A, vol. 157, pp. 357-416, 1994.

H. Bevrani received the B. Sc, M. Sc and Ph.D degrees from Shahid Beheshti University (Iran), Teacher Training University (Iran) and Moscow State University (Russia), all in statistics in 1991, 1994 and 2005, respectively. Since 1997, he has been with Department of Statistics at University of Tabriz, Iran. His current research interests include the Limit theorem and random summation, Reliability and availability, Nonparametric statistics, and Statistical simulation methods. He is a member of Iranian Mathematical Society and Iranian Statistical Society.