A MONTE CARLO BASED COMPARISON OF BAYESIAN AND FREQUENTIST ESTIMATORS OF A BINOMIAL PROPORTION

BY

[*]OLUSANYA E. OLUBUSOYE
DEPARTMENT OF STATISTICS
ADDIS ABABA UNIVERSITY
ADDIS ABABA, ETHIOPIA
oe.olubusoye@mail.ui.edu.ng

AND

OSOWOLE OYEDEJI I.
DEPARTMENT OF STATISTICS
UNIVERSITY OF IBADAN
IBADAN, NIGERIA
dosowole@yahoo.com

Abstract

The long-time debate between Bayesians and Frequentists is unusual among philosophical arguments. This paper used a Monte Carlo approach to evaluate the performance of Bayesian and Frequentist estimators of a binomial proportion. The result showed that the Bayesian estimator outperformed the Frequentist estimator even when judged by the Frequentist criteria of mean squared error.

[*] The correspondence author

1.0 **Introduction**

At the present time, there exist two very different approaches to statistics. These are the traditional “Classical’ or “Frequentist” approach and the “Bayesian” approach. The main difference centered on the notion of probability they employ. The objectivity of the Frequentist statistics is obtained by disregarding any prior knowledge about the process being measured. Bayesian statistics uses both sources of information; the prior information we have about the process and the information about the process contained in the data.

In the past, Bayesian methods were of limited practical use, since analytic solutions for the Bayesian posterior distributions were only possible in few cases, and the numerical calculation of the posterior often was not feasible because of lack of computer power. Recent developments in computational innovations and the development of Markov chain methods for sampling from the posterior have made Bayesian methods possible, even in very complicated models. The Bayesian approach claims several advantages over the classical approach. A complete and more persuasive listing of the advantages of the Bayesian approach can be found in Zellner (1974).

In recent times, there has been a vicious controversy among the “believers” of these two approaches. Some references suitable for exploring the Frequentist/Bayesian controversy are Efron (1986) and associated commentary, Poirier (1988) and related discussion. Weber (1973) examines the history of the Bayesian controversy. Koop (2005) is an excellent exposition of progress in applied Bayesian econometrics, with particular emphasis on computational considerations.

As a contribution to this raging controversy, this paper attempts to compare the performance of the Frequentist and the Bayesian estimators of a binomial proportion based on a Monte Carlo Experiment. Judging from the mean square error criterion, the result shows that the Bayesian estimator (using Beta (1,1) prior) is better than the Frequentist estimator.

The rest of the paper is divided into four sections. In section 2.0, we derived the Bayesian estimator of binomial proportion. In section 3.0, we described the Monte Carlo experiment. The result of the experiment is presented and discussed in section 4.0 and finally the concluding remark is given in section 5.0.

2.0 **Bayesian Estimation of Binomial Proportion**

In this section, we followed Bolstad (2004) approach to derive the Bayesian estimator of a binomial proportion. Often we deal with a large population where π, a proportion of the population has some attribute. We take a random sample from the population and let Y be the observed number in the sample having the attribute. We are counting the total number of “successes” in n independent trials where each trial has two possible outcomes, “success” and “failure”. Success on trial i means the item drawn on trial i has the attribute. The probability of success on any single trial is π, the proportion in the population having the attribute. This proportion remains constant over all trials because the population is large.

The conditional distribution of the observation Y, the total number of successes in n trials given the parameter π, is *binomial(n, π)*. The conditional probability function for y given π is given by

$$f(y/\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y} \text{ for } y = 1, \ldots, n.$$

Here we are holding π fixed, and looking at the probability distribution of y over its possible values.

If we look at this same relationship between π and y, but hold y fixed at the number of successes we observed, and let π vary over its possible values, we have the likelihood function given by

$$f(y/\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y} \text{ for } 0 \le \pi \le 1$$

We see that we are looking at the same relationship as the distribution of the observation y given the parameter π, but the subject of the formula has changed to the parameter, for the observation held at the value that actually occurred.

To use Bayes' theorem, we need a prior distribution g(π) that gives our belief about the possible values of the parameter π before taking the data. It is important to realize that the prior must not be constructed from the data. Bayes' theorem is summarized by *posterior is proportional to the prior times the likelihood:*

$$g\left(\pi / y\right) \propto g\left(\pi\right) \times f\left(y / \pi\right)$$

This gives us the shape of the posterior density, but not the exact posterior density itself. To get the actual posterior, we need to divide this by some constant *k* to make sure it is a probability distribution, meaning that the area under the posterior integrates to 1. We find *k* by integrating $g\left(\pi\right) \times f\left(y / \pi\right)$ over the whole range. So, in general,

$$g\left(\pi / y\right) = \frac{g\left(\pi\right) \times f\left(y / \pi\right)}{\int_0^1 g\left(\pi\right) \times f\left(y / \pi\right) d\pi}$$

which requires an integration. Depending on the prior *g(* π *)* chosen, there may not necessarily be a closed form for the integral, so it may be necessary to the integration numerically.

Suppose a *beta(a,b)* prior density is used for π:

$$g(\pi; a, b) = \frac{\Gamma\left(a+b\right)}{\Gamma\left(a\right)\Gamma\left(b\right)} \pi^{a-1} \left(1-\pi\right)^{b-1} \text{ for } 0 \le \pi \le 1$$

The posterior is proportional to prior times likelihood. We can ignore the constants in the prior and likelihood that don't depend on the parameter, since we know multiplying either the prior or the likelihood by a constant won't affect the results of Bayes' theorem. This gives

$$g\left(\pi / y\right) \propto \pi^{\alpha+y-1} \left(1-\pi\right)^{b+n-y-1} \text{ for } 0 \le \pi \le 1$$

which is the shape of the posterior as a function of π. We recognize that this is the beta distribution with parameters $a' = a + y$ and $b' = b + n - y$. That is, we add the number of successes to *a* and number of failures to *b*:

$$g(\pi / y) = \frac{\Gamma(n+a+b)}{\Gamma(y+a)\Gamma(n-y+b)} \pi^{y+a-1}(1-\pi)^{n-y+b-1} \text{ for } 0 \le \pi \le 1.$$

We note that the uniform prior is a special case of the *beta(a,b)* prior where *a=1* and *b=1*.

The posterior mean is a very frequently used measure of location. It is the expected value, or mean of the posterior distribution.

$$m' = \int_0^1 \pi g(\pi / y) d\pi$$

When the posterior $g(\pi / y)$ is $beta(a', b')$ the posterior mean equals

$$m' = \frac{a'}{a'+b'}$$

Now, suppose we use the posterior mean as the Bayesian estimate for π, where we use the *beta(1,1)* prior (uniform prior). The estimator is the posterior mean, so

$$\hat{\pi}_B = m' = \frac{a'}{a'+b'}$$

Where $a' = 1 + y$ and $b' = 1 + n - y$. We can write this as a function of *y*, the number of successes in the *n* trials:

$$\hat{\pi}_B = \frac{y+1}{n+2}$$

3.0 Monte Carlo Experiment

We performed a Monte Carlo study approximating the sampling distribution of two estimators of π. The Frequentist estimator for π is the **sample proportion,** $\hat{\pi}_f = \frac{y}{n}$, where *y* is the number of successes in the *n* trials. The Bayesian estimator used is

$\hat{\pi}_B = \dfrac{y+1}{n+2}$ which equals the posterior mean when we used a uniform prior for π. We compared the sampling distributions (in terms of bias, variance, and mean squared error) of the two estimators over a range of π values from 0 to 1. For each of the parameter values, we approximated the sampling distribution of the estimator by an empirical distribution based on 2000 samples drawn when that is the parameter value. The true characteristics of the sampling distribution (mean, variance, mean square error) are approximated by the sample equivalent from the empirical distribution.

For π = 0.1, 0.2, 0.3, …, 0.9

i. We drew 2000 random samples from *binomial (n=10,* π).

ii. We calculated the Frequentist estimator $\hat{\pi}_f = \dfrac{y}{n}$ for each of the 2000 samples.

iii. We calculated the Bayesian estimator $\hat{\pi}_B = \dfrac{y+1}{n+2}$ for each of the 2000 samples

iv. We calculated the means of these estimators over the 2000 samples, and subtract π to give the *biases* of the two estimators. $bias(\hat{\pi}) = E(\hat{\pi}) - \pi$.

v. We calculated the variances of these estimators over the 2000 samples.

vi. We calculated the mean square error (MSE) of these estimators over the 2000 samples.

$$MSE(\hat{\pi}) = (bias(\hat{\pi}))^2 + Var(\hat{\pi})$$

4.0 Discussion of the Results

Tables 1-4 show the mean, bias, variance and mean squared error respectively for the two estimators over the 2000 samples. From table 1, it is observed that the mean of $\hat{\pi}_f$ is smaller than $\hat{\pi}_B$ for π =0.1, 0.2, 0.3, 0.4 but the reverse is the case for higher values of π. Similar pattern is observed for the bias of the estimators as shown in table 2. However, in terms of the spread of the estimates about the mean, table 3 reveals that $\hat{\pi}_B$ is closer. Generally, the variance of $\hat{\pi}_B$ is smaller than the variance of $\hat{\pi}_f$ for all π. A clear picture emerges in table 4 and the graph in figure 1. The Bayesian estimator has smaller mean squared errors than frequentist estimator. In other words, on average, it is closer to the true value. Figure 1 shows the mean squared error for the Bayesian estimator and the frequentist estimator as a function of π. We see that over most (but not all) of the range, the Bayesian estimator is better than the frequentist estimator.

Table 1: Mean of Frequentist and Bayesian Estimates Over 2000 Samples

Estimator	*Binomial Proportion*								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
$\hat{\pi}_f$	**0.104**	**0.207**	**0.299**	**0.406**	**0.501**	**0.602**	**0.699**	**0.796**	**0.899**
$\hat{\pi}_B$	**0.170**	**0.256**	**0.332**	**0.422**	**0.501**	**0.585**	**0.666**	**0.747**	**0.833**

Table 2: Bias of Frequentist and Bayesian Estimates Over 2000 Samples

Estimator	*Binomial Proportion*								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
$\hat{\pi}_f$	**0.004**	**0.007**	**-0.001**	**0.006**	**0.002**	**0.002**	**-0.001**	**-0.004**	**-0.0003**
$\hat{\pi}_B$	**0.070**	**0.056**	**0.032**	**0.022**	**0.001**	**-0.015**	**-0.034**	**-0.053**	**-0.067**

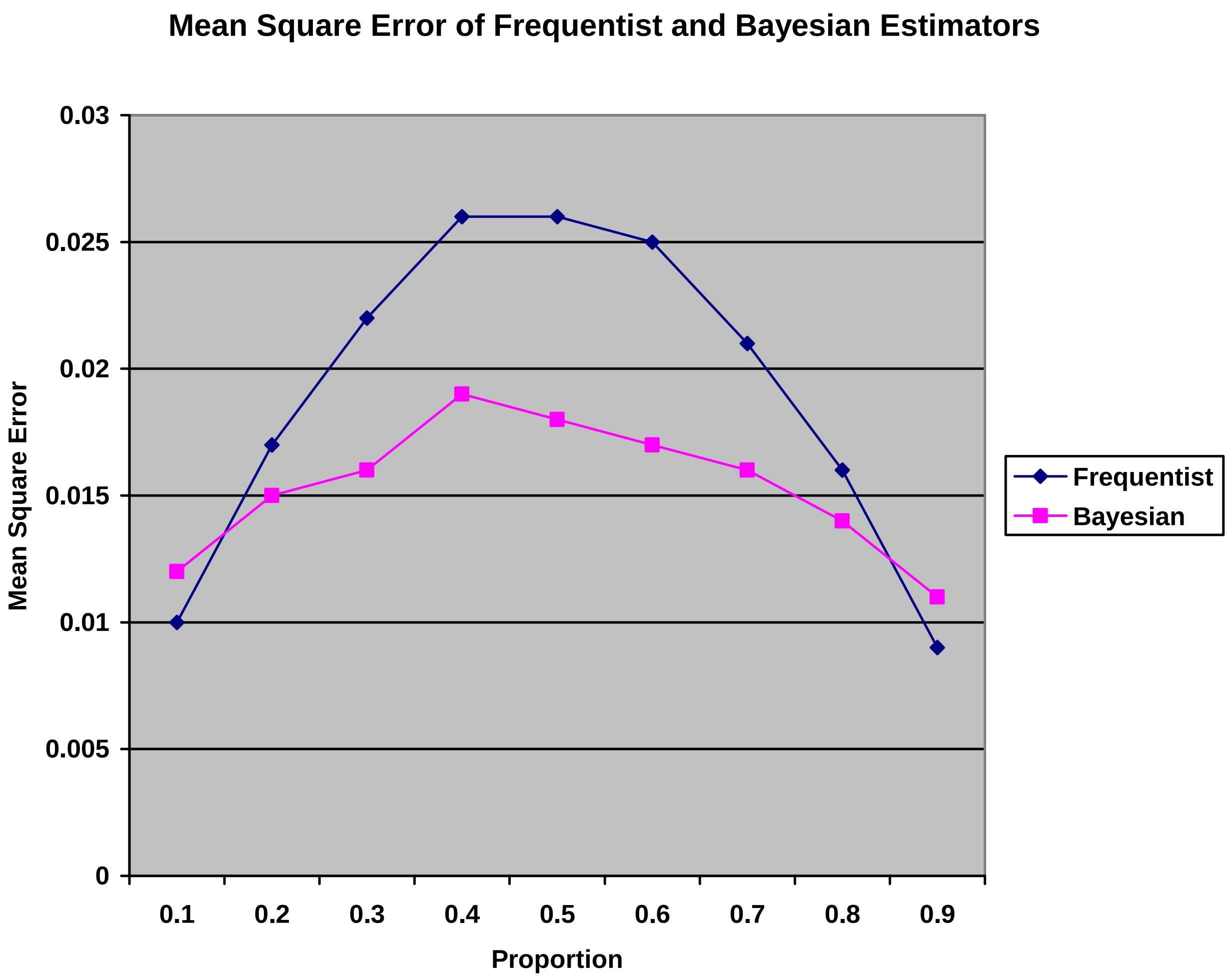
Table 3: Variance of Frequentist and Bayesian Estimates Over 2000 Samples

Estimator	*Binomial Proportion*								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
$\hat{\pi}_f$	**0.010**	**0.017**	**0.022**	**0.026**	**0.026**	**0.025**	**0.021**	**0.016**	**0.009**
$\hat{\pi}_B$	**0.007**	**0.012**	**0.015**	**0.018**	**0.018**	**0.017**	**0.015**	**0.011**	**0.006**

Table 4: Mean Square Error (MSE) of Frequentist and Bayesian Estimates Over 2000 Samples

Estimator	*Binomial Proportion*								
	0.1	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**
$\hat{\pi}_f$	**0.010**	**0.017**	**0.022**	**0.026**	**0.026**	**0.025**	**0.021**	**0.016**	**0.009**
$\hat{\pi}_B$	**0.012**	**0.015**	**0.016**	**0.019**	**0.018**	**0.017**	**0.016**	**0.014**	**0.011**

Figure 1



5.0 **Conclusion**

In this paper, a Monte Carlo experiment was used to compare the Bayesian and the Frequentist estimators of a binomial proportion. Judging from the mean squared error criterion, the Bayesian estimator performed better.

References

Bolstad, W. M. (2004). *Introduction to Bayesian Statistics*, John Wiley & Sons, Inc..

Efron, B. (1986). *Why isn't everyone a Bayesian,* The American Statistician 40 (1), 1-11.

Koop, G. (1999). *Bayesian Analysis, Computation and Communication Software.* Journal of Applied Econometrics 14, 677-689.

Poirier, D.J. (1988). *Frequentist and Subjectivist Perspectives on the Problems of Model Building in Economics.* Journal of Economic Perspectives 2, 121-144 (commentary, 145-170)

Weber, J.D. (1973). *Historical Aspects of the Bayesian Controversy.* Tucson: Division of Economic and Business Research, University of Arizona.

Zellner, A. (1974). *The Bayesian Approach and Alternatives in Econometrics.* In S.E. Fienberg an A. Zellner (eds), *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage.* Amsterdam: North Holland, 39-54. Reprinted in A.Zellner (1984), *Basic Issues in Econometrics.* Chicago: University of Chicago press, 187-200.