A fully Bayesian approach for sample size determination

Nasser Safaie

Department of industrial and Manufacturing Engineering

Abstract

The sample size determination plays a significant role in design and analysis of many engineering problems plus data sampling projects. There is a number of sampling techniques including both classical as well as Bayesian techniques and only some of them consider the economical aspect. The objective of this paper is to offer an economic Bayesian approach for determination of sample size. Mathematical models are derived and used to establish implementation boundaries from economic and technical viewpoints. In addition, numerical examples are used to highlight the economic advantages of Bayesian approach. Using simulated data, the performance of the proposed approach is compared to the classical methods in terms of the number of sample size criterion.

1. Introduction

The Sample Size Determination (SSD) has a critical role in design and analysis of data sampling projects. The sample size impacts all steps of analysis explicitly or implicitly. The main purpose of this research is to explore an application of a fully Bayesian approach for SSD to verify a shift in the process average. This research also compares classical and Bayesian methods from number of sample size view point.

2. SSD Techniques

SSD methods are divided into two major groups [1]. These include both classical as well as Bayesian techniques. In classical method, two values are pre selected before sampling; the interval half width H and confidence level of the interval $(1-\alpha)\%100$. If a random sample of n observations are to be drawn from a normal distribution with mean $\square$ and variance $\square^2$, then the requirement that the unknown population mean $\square$ lies within $\pm$ H of the sample mean with probability $(1-\alpha)$, results in:

$$n \geq \left(\frac{z_{\alpha/2}}{H}\right)^2 \sigma^2 \qquad (1)$$

Where, the statistic $z_{\alpha/2}$ is obtained from the table of standard normal distribution and $\square^2$ is known [2]. This method has been criticized because the hidden uncertainty in $\square^2$ is ignored [3]. Calculating the sample size without considering the prior information and uncertainty, may result in a larger sample size than is necessary. The Bayesian approach is accompanied by uncertainty of prior knowledge and is a powerful tool for SSD. Berger discussed that the fully Bayesian treatment of SSD (FBSSD) involves utilizing loss function, and economical criteria [1]. The whole idea of FBSSD is to minimize cost and avoid oversampling. In the FBSSD method, the observations are taken one at a time and a decision is made after each observation, either to stop sampling and choose an action, or to take another observation [4]. Hence, one can collect the exact amount of observations for making a decision with desired accuracy. This attribute results in minimizing the sampling cost.

Choosing a loss function has a crucial role in FBSSD methods. Loss function is a mathematical function that explains the relationship between estimation error and the economic loss associated with that error. In this research, the Taguchi's quality loss function is used as:

$$L(x) = k(x-T)^2 \qquad (2)$$

Where, *K* is a positive constant, *x* is the value of the quality characteristic and *T* is the target value.

3. Mathematical Modeling Development

One of the discrete manufacturing of a single quality characteristic x is brought into consideration. The specification limits for x are expressed as $m \pm \square$ where $m \square$ is target and $\square \square$ is one half of the tolerance spread $\square \square$ It is also assumed that a state of statistical control has been established and that x follows a normal distribution with mean μ_0 which is off target and variance τ^2 with lot size of N units. Two actions could be imagined. The first (a_1) involves

taking no action and continue operation at an estimated per unit loss. The second (a_2) involves making corrective changes to re-center the process closer to the target while $\square^{\square}$ will not be changed. This would require an initial investment of \$W for setup and sampling cost of \$C per unit. However, operation at the new level is expected to reduce the per unit loss to \$$L_2$. Based on engineers' experiences, the probability of success for process improvement is p and the probability of failure is 1-p. This would result in the economical criterion which requires that the average risk under the second option be less than that under the first. i.e.

$$pN(L_2 - L_1) + (1-p)(W + nC) < N(L_1) \qquad (3)$$

Where, $L_1 = Nk\left[\tau^2 + (\mu_0 - m)^2\right]$, $L_2 = k(\tau^2 + \rho_n)$, and ρ_n is posterior variance. When the criterion (3) holds true, a_2 would represent the optimal Bayes rule. However, calculation of r (a_2) requires knowledge of the sample size n. Now, suppose that an unbiased gauge with variance $\square^2_g$ less than process variance will be used to collect the observations. This would result in a second criterion regarding the feasibility of improvement. That is, the value of N should be large enough to allow for n* to exceed zero. This can be expressed in the form of:

$$N > \frac{(1-p)}{p} \frac{C}{A} \frac{\Delta^2}{\sigma_g^2} \qquad (4)$$

This criterion should be satisfied before calculation of n*. Once the criteria (3) and (4) hold true, the sample size is computed by:

$$n^* = \sqrt{N \frac{A}{C} \frac{\sigma_g^2}{\Delta^2} \frac{p}{(1-p)}} \qquad (5)$$

4. Numerical Example

A control chart indicates that the current process for producing a component with dimension x is in a state of statistical control. The needed information about the process is given in Table 1.

Table 1 The given information of numerical example

N	$\square^{\square}$	$\square$	$\square^{\square}$	$\square_g^2$	m	W	C	p	1-p
200	0.0025	100	0.0004	0.000025	8.02	100	20	0.7	0.3

Utilizing the given information, the criteria (3) and (4) are satisfied. Thus n* is determined as 4.768 which could be approximated by 5. Applying the classical method, by substituting in the inequality (1), n is computed as 15.37 which is much larger compared to the sample size of n*=5.

5. Conclusion

In this research, a fully Bayesian approach for sample size determination was proposed. At the same time, a mathematical model accompanied with economical and technical criteria was developed. Utilizing the numerical example this method was compared to the classical. Classical methods do not take into account prior and loss information and consider the information of only the last sample. On the other hand, the Bayesian approach considers the information related to all earlier samples for making an inference. The Bayesian method is more efficient than traditional methods when the decision making is desired in terms of loss associated with every scenario and when prior knowledge is available. However, Bayesian analysis is not recommended when there is no good perception of the procedure, since the objectivity of the prior information is crucial in determining the sample size and decision making.

6. Acknowledgement

References
[1]Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis, 2nd ed.* New York: Springle-Verlag.
[2]Carlin, B. and Louis, T. (2009).*Bayesian Methods for Data Analysis*, Boca Roton, Fl: Chapman and Hall.
[3]Congdon, P. (2007), *Bayesian Statistical Modeling*, second edition, Chichester: John Wiley and Sons.
[4]Mendenhall, W. (2007). *Statistics for Engineering and the Sciences,5ed.* New Jercy: Pearson Prentice Hall.