Journal of the Nigerian Association of Mathematical Physics
Volume27, (*July, 2014*), pp 235 – 240

Applying Bayesian Model Averaging in Reliability Concept

Iyiegbu T.I[1], Asiribo O.E[2] and Dikko H.G[1]

[1]Department of Mathematic, Ahmadu Bello University, Zaria
[2]Department of Statistics, Federal University of Agriculture, Abeokuta.

Abstract

Reliability is the probability that a system will perform its intended function for a specified period of time under a given set of conditions. Bayesian techniques have always proved to be a good method in analyzing data from experimental design. This paper treats reliability as a single fixed number whose unknown value is to be estimated. Reliability has uncertainty associated with it, this thus require us to treat it as a random variable and therefore discuss it using probability distribution, f(R). Also in this paper, Bayesian model selection in R package is used to average all the 2^k possible models and thereby identify the correct model that best define the reliability improvement. The superior analytical power of this approach compared to other methods like OLS was shown.

Keywords: Fuzzy time series, Autoregressive integrated moving average, Theil's regression, Mean squared forecast error, Root mean square forecast error and Coefficient of determination

1.0 Introduction

Reliability is the probability that a product performs its function or does not breakdown upon demand. More precisely, reliability is the probability that a system will perform its intended function for a specified period of time under a given set of conditions.

Prior to the 1960s, there were many instances when experimental performance failed out rightly when exposed to environmental stress. The problems were largely due to the fact that reliability was not yet formally recognized as a performance characteristic. Hamada [1] defined reliability as the probability of equipment or process functioning without failure when operated as prescribed for a given interval of time, under stated conditions. Setting reliability requirement is a cornerstone of any reliability strategy. Statistically, design of experiment can be used to identify factors that affect reliability and to recommend factor level that leads to improved reliability. Based on this line of thought, Hamada [1] in his paper related his work to the findings of Taguchi [2]who recommended that noise factor be considered in any experiment that seeks to improve reliability, whenever practicable to do so.

From a reliability point of view, classical calculations can be thought of as a subset of Bayesian calculations. This paper will apply Bayesian techniques to a reliability data with the objective of improving its performance. Reliability will be treated as a single fixed number, in this paper, whose unknown is to be estimated. And because reliability has uncertainty associated with it, this will require us to treat it as a random variable and therefore discuss it using probability distribution, f(R).

It makes a great deal of practical sense to use all the information available, old and/or new, when making decisions under uncertainty which is exactly the situation one has with many systems in the field. Prior knowledge is not used except to suggest the choice of a particular population model to "fit" to the data, and this choice is later checked against the data for reasonableness. Bayesian statistics is used to incorporate this knowledge quantitatively into reliability calculations. Lifetime or repair models using frequentist method have one or more unknown parameters. The Bayesian approach treats these population model parameters as random, not fixed quantities. Before looking at the current data, use is made of old information, or even subjective judgments, to construct a prior distribution model for these parameters. Then use is made of the current data (via Bayes' formula) to revise the starting assessment, deriving what is called the posterior probability distribution model for the population model parameters.

Reliability improvement requires effective application of experimental design methodology to build quality into products and processes so that repair cost can be minimized and customer satisfaction can be enhanced. Improving product reliability is an

Corresponding author: Iyiegbu T.I E-mail:tochukwuifechukwu@yahoo.com, Tel.: +2348039354219

attempt to guarantee the highest quality of products over a certain time period. Existing methods such as minute accumulating analysis (NMA) of Taguchi [3] and the quick & dirty method mentioned by Hamada and Wu [4] are inadequate for analyzing such censored data from highly fractionated experiments, because the best combination of levels in the process variables is often suggested wrongly. Hamada and Wu [4] proposed an iterative "Expectation Modeling Maximization" (EMM) procedure to compare many models simultaneously. Their EMM procedure requires the existence of the maximum likelihood (ML) estimates of parameters in candidate models. This method does not work in many models which include second or higher order interactions because Maximum Likelihood Estimate often does not exist.

2.0 Methodology

The Bayesian approach to statistical problems is fundamentally probabilistic. A joint probability distribution is used to describe the relationships between all the unknowns and the data. Inference is then based on the conditional probability distribution of the unknown given the observed data (posterior distribution). The joint distribution is obtained by introducing the prior distribution on all the unknowns, Conditioning on the data then induces a posterior distribution of model uncertainty that can be used for model selection. However, two major challenges confront its practical implementations – the specification of the prior distributions and the calculation of the posterior.

The foundation of Bayesian statistics is Bayes Theorem. Suppose we observe a random variable y and wish to make references about another random variable θ, where θ is drawn from some distribution P (θ). From the definition of conditional probability;

$$\Pr(\theta, y) = \Pr(\theta / y) P(y) \tag{2.1}$$

Again from the definition conditional probability, we can express the joint probability by conditioning on θ to give

$$\Pr(y, \theta) = \Pr(y / \theta) \Pr(\theta) \tag{2.2}$$

Putting these together, gives Bayes Theorem.

$$\Pr(\theta / y) = \frac{\Pr(y / \theta) \Pr(\theta)}{\Pr(y)} \tag{2.3}$$

$$= \frac{\Pr(y / \theta)}{\sum_{i=1}^{n} \Pr(\theta) \Pr(y / \theta)} \tag{2.4}$$

Pr (θ) is the prior distribution of θ, while Pr(θ/y) is the posterior distribution of θ given the observe data y.

Bayesian model averaging (BMA) is an empirical tool to deal with model uncertainty in various analyses of applied sciences. Typically, though not always, BMA focuses on which regressors to include in the analysis. The attraction of BMA is that one can quickly determine models, or more specially, sets of explanatory variables, which possess high likelihoods. By averaging across a large set of models one can determine those variables which are relevant to the data generating process for a given set of priors used in the analysis. Each model (a set of variables) receives a weight and the final estimates are constructed as a weighted average of the parameter estimates from each of the models. BMA includes all of the variables within the analysis, but shrinks the impact of certain variables towards zero through the model weights. These weights are the key features for estimation via BMA and will also depend upon a number of other key features of the averaging exercise including the choice of priors. The implementation of BMA, which was first proposed by Leamer [6], for linear regression models are as follows: Consider a linear regression model with a constant term, β_0, and k potential explanatory variables $X_1, X_{2,\ldots,} X_k$.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon. \tag{2.5}$$

given the number of regressors, we will have 2^k different combinations of right hand side variable indexed by M_j for j = 1,2,3,. . . ,2^k . Once the model space has been constructed, the posterior distribution for any coefficient of interest, say β_h , given the data D is

$$\Pr(\beta_h | D) = \sum \Pr(\beta_h / M_j) \Pr(M_j / D) \tag{2.6}$$

BMA uses each models posterior probability, $\Pr(M_j|D)$, as weights. The posterior model probability of M_j is given by

$$\Pr(M_1 / D) = \Pr(D / M_j)\frac{\Pr(M_j)}{\Pr(D)} \tag{2.7}$$

$$= \frac{\Pr(D / M_j)\Pr(M_j)}{\sum_{j=1}^{2^k}\Pr(D / M_j)\Pr(M_j)}$$

where

$$\Pr(D / M_j) = \int \Pr(D / \beta^j, M_j)\Pr(\beta^j, M_j)\,\partial\beta^j \tag{2.8}$$

And β^j is the vector of parameters from model M_j, $\Pr(\beta^j / M_i)$ is a prior probability distribution assigned to the parameters of model M_j , and Pr (M_j) is the prior probability that M_j is the true model. The estimated posterior means and standard deviations of $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ are then constructed as:

$$E\left(\hat{\beta}/ D\right) = \sum_{j=1}^{2^k}\hat{\beta}\Pr(M_j / D), \tag{2.9}$$

$$V\left(\hat{\beta}/ D\right) = \sum_{j=1}^{2^k}\left(Var\left(\hat{\beta}/ D, M_j\right) + \hat{\beta}^2\right)\Pr(M_j / D) - E\left(\hat{\beta}/ D\right) \tag{2.10}$$

3.0 Bayesian Linear Models and Zellner's g prior

The literature standard is to use a 'Bayesian regression' linear model with a specific prior structure called 'Zellner's g prior'. The need to obtain posterior distributions requires specifying the priors on the model parameters. Here, 'improper' priors were placed on constant and error variance, which means they are evenly distributed over their domain:p (α_γ) ∝1, i.e. complete prior uncertainty where the prior is located. Similarly, set p (σ) ∝ σ^{-1}. The crucial prior is the one on regression coefficients β_γ: Before looking into the data (y, X), we formulate our prior beliefs on coefficients into a normal distribution with a specified mean and variance. It is common to assume a conservative prior mean of zero for the coefficients to reflect that, not much is known about them. Their variance structure is defined according to Zellner's gas $\sigma^2\left(\frac{1}{g}X_\gamma{}'X_\gamma\right)$:

$$\beta_\gamma / g \sim N\left(0, \sigma^2\left(\frac{1}{g}X_\gamma{}'X_\gamma\right)^{-1}\right)$$

This means that the coefficients are zero and that their variance-covariance structure is broadly in line with that of the data X. The hyper parameter g embodies how certain we are that coefficients are indeed zero: A small g means few prior coefficient variances and therefore implies we are quite certain (or conservative) that the coefficients are indeed zero. In contrast, a large g means that we are very uncertain that coefficients are zero. The posterior distribution of the coefficients reflects prior uncertainty: Given g, it follows a t-distribution with expected value

$$E(\beta\gamma / y, X, M\gamma) = \frac{g}{g+1}\hat{\beta}\gamma \tag{3.1}$$

where "$\hat{\beta}_\gamma$"is the standard OLS estimator for the model. The expected value of coefficients is thus a convex combination of OLS estimator and prior mean (zero). The more conservative (smaller) g, the more important is the prior, and the more the expected value of coefficients is shrunk toward the prior mean zero. As g→∞, the coefficient estimator approaches the OLS estimator. Similarly, the posterior variance of β_γ is affected by the choice of g:

$$Cov(\beta\gamma / y, X, M\gamma) = \frac{\left(y-\bar{y}\right)'\left(y-\bar{y}\right)}{N-3}\frac{g}{g+1}\left(1-\frac{g}{g+1}R_\gamma{}^2\right)\left(X_\gamma{}'X_\gamma\right)^{-1} \tag{3.2}$$

The posterior covariance is similar to that of the OLS estimator; times a factor that includes g and R^2, the OLS R-squared for model ϒ. The crucial choice here concerns the form of the hyper parameter g. A popular 'default' approach is the 'unit information prior' (UIP), which sets g = N(N is the number of runs (observations) in an experimental design) commonly for all models and thus attributes about the same information to the prior as is contained in one observation.Virtually all BMA applications rely on the presented framework with Zellner's g prior, and the bulk of them rely on specifying a fixed g. The value of gcorresponds to the degree of prior uncertainty: A low g renders the prior coefficient distributiontight around a zero mean, while a large g implies large prior coefficient variance and thusdecreases the importance of the coefficient prior.In Bayesian Model Selection, any fixed g-prior may be specified directly by submitting its value to the BMSfunction argument gusing R package.

4.0 Data presentation and analysis

The data used in these study is a simulated data by Rusco[7]. He made use of experimental design to select the best factors that affect the response variable (trap speed) but this paper re-analyzed the data using BMA method. It started by first creating the fractional factorial design with 5 factors and 16 runs with no replication. An OLS method (ANOVA) was used to first check the factors that affect the response variable. BMA method was also applied on the data with Zellner g- prior (UIP), which yielded the coefficient table that was used to analyze the data. Model check was run to know the model that contains the highest Posterior Model Probability (for Exact and Moving Chain Monte Carlo) or the model that best fit the data using R package version 3.0.2 [8].

Rusco[7] used Burnout-Championship Drag Racing as the simulation Software and Design – Expert software (www.statease.com) to generate the actual experimental design and analysis. The following main factors were identified in Table 4.1:

Table 4.1: Experimental Factors

factor	Name	Type	Low Actual	High Actual
A	Tires	Categorical	Stock Street	Race Only
B	Wing	Categorical	Low Force	High Force
C	Transmission	Categorical	2 Spd Auto	4 Spd Auto
D	Engine	Categorical	402 hp	599 hp
E	Centre of Gravity	Numeric	50.00	60.00

The main reason of the experiment is to find out among the important variables, which one has a major impact on the speed of the race car using "Trap Speed" as the response variable for this experiment.This paper applied a Bayesian Model Averaging procedure to the designed data with the aimof identifying the best model and the factors that have major impact on the response variable.

The Low Actual and High actual in table 4.1 were redefine as -1 and +1as shown in Table 4.2:

Table 4.2: Factor Settings

	A	B	C	D	E
1	-1	-1	-1	-1	-1
2	1	1	1	1	1

The design generators are E =ABCD

Table 4.1 contains 5 variables, which implies 2^5factor combinations. This paper made use of UIP g – prior where g = N = 16. However, our expected prior model parameter size is k/2 = 2.5. We use BMS library in R package to analyze our design.

The coefficient result generated using R package is given in Table 4.3:

Table 4.3: Coefficient of Bayesian Model Averaging

	PIP	Post Mean	Post SD	Cond.Pos.Sign	Idx
A	0.9997918	15.1733115	2.437263	1	1
D	0.9163742	6.3607153	2.967404	1	4
C	0.3385396	-0.9160485	1.867779	0	3
B	0.3245052	0.8398957	1.802675	1	2
E	0.1951941	0.0000000	1.072932	0	5

Table 4.3 shows the factors with their corresponding statistics: The Post Mean values are coefficients averaged over all models, including models that do not contain the factor (implying azero coefficient). The factors A and D havecomparatively large coefficientsindicating their relative importance. The importance of the factors in explaining the data is also given in the first column, PIP which represents posterior inclusion probabilities, that is, the sum of the Posterior Model Probabilities for all models that include a given factor. It shows that with 99.9%, virtually all of posterior model mass rests on models that include A. in contrast to 91% which include D and 33% or less which include others that do not matter much.The coefficient sign can also be inferred from the fourth column Cond.Pos.Sign, as 'sign certainty'. Here, it also showsthat in all encountered models containing these variables, the expected values of coefficients for factor A, D and B were positive. Finally, the last column idx denotes the index of the factors' appearance in the original data set, as the results are obviously sorted by PIP.

4.1 The Models

Table 4.4 below shows that the output includes the posterior model probability for each model. The best model, with 32% posterior model probability, is the one that only includes A and D. however the second best model includes C, A and D, with a PMP of 16% and third includes B, A and D, which has a PMP of 15%.

Table 4.4:**Models**

	12	16	1a
A	1.0000000	1.0000000	1.000000
B	0.0000000	0.0000000	1.000000
C	0.0000000	1.0000000	0.000000
D	1.0000000	1.0000000	1.000000
E	0.0000000	0.0000000	0.000000
PMP (Exact)	0.3270239	0.1673879	0.156617
PMP (MCMC)	0.3270239	0.1673879	0.156617

5.0 Conclusion

Bayesian method helps to build a nice frame work in which the data can be well analyzed. R package enable quick and easy data analysis.It is worthwhile noting that though our final results prove to be the same with using OLS method, BMA method showed more detailed output and emphasis were made on vital aspect of the experiment compare to using OLS method. The beauty of BMA method is that it takes cognizance of all factors and their models, averaging over all the models and identifying models that contain factors with highest PMP as the best model. Our paper also justifies the superiority of BMA method as it is more concise and involves both the factors and their model in the analysis. Analyzing data to improve reliability with BMA method is quite new and future research will also take cognizance of interaction effect.

References

[1.] Hamada, M. (1994), "Analysis of Experiment for Reliability Improvement and RobustReliability," University of Waterloo Institute for Improvement In quality and ProductivityResearch Report RR-94-07

[2.] Taguchi,G. (1986), "Introduction to Quality Engineering," Tokyo, Japan: AsianProductivity Organization.

[3.] Taguchi, G. (1987), System of Experimental Design, Vol 1 & 2, White Plains, New York:Unipub/Kraus International.

[4.] Hamada, M. and Wu, C.E.J. (1991), "Analysis of Censored Data from HighlyFractionated Experiments," Technometrics, 33, 25 – 38.

[5.] Raftery, A. E. (1995), `Bayesian model selection in social research', Sociological Methodology 25, 111{163.

[6.] Leamer, E. (1978), Specification searches: Ad hoc inference with non experimental data. Wiley New York.

[7.] Mark .S. Rusco (2002)Google: "Using Simulation Software for Design of Experiment Training".Ferris State University, Applied Technology Center

[8.] R Development Core Team (2010), R: A Language and Environment for Statistical Computing, R Foundation forStatistical Computing, Vienna, Austria. ISBN 3-900051-07-0.URL: http://www.R-project.org