American Journal of Theoretical and Applied Statistics
2013; 2(1) : 7-11
Published online January 10, 2013 (http://www.sciencepublishinggroup.com/j/ajtas)
doi: 10.11648/j.ajtas.20130201.12

Comparative analysis of bayesian control chart estimation and conventional multivariate control chart

Adewara[1], J. Ademola[1], Ogundeji K. Rotimi[2]

[1]Distance Learning Institute, University of Lagos, Lagos, Nigeria
[2]Department of Mathematics, Faculty of Science, University of Lagos, Lagos, Nigeria

Email address:
adewaraja@yahoo.com (Adewara), jadewara@unilag.edu.ng (J. Ademola), rko3000@yahoo.com (O. K. Rorimi)

To cite this article:
Adewara, J. Ademola, Ogundeji K. Rorimi. Comparative Analysis of Bayesian Control Chart Estimation and Conventional Multivariate Control Chart, *American Journal of Theoretical and Applied Statistics*. Vol. 2, No. 1, 2013, pp. 7-11. doi: 10.11648/j.atjas.20130201.12

Abstract: Bayesian model or Beta-binomial conjugate using Bayesian sequential estimation method to estimate the proportion of different age groups is compared with the conventional multivariate control chart method. The parameters for the techniques were derived and applied. The result shows that the patients between the ages of 15-44 in 2009 and 44-64 and 64 and above in 2011 are out of control. This implies the Bayesian sequential estimation method is very efficient to notice any small shift that occurs among patients that make use of the hospital. Also the bracket mentioned above was very high among the people that used the hospital compared to others. The result of 2011shows that there was a high shift in the ages of the patients that attended the hospital for the ages between 44-64 and 64 and above respectively.

Keywords: Beta-Binomial, Sequential Estimation, Hyperparameters, Conjugates Beta-Binomial, Shrinkage Factor And Multivariate Random Variables

1. Introduction

Statistical process control (SPC) chart is an important tool in the control chart. It can be used to detect changes in production processes, assess process stability, and identify changes that indicate either improvement or deterioration in quality and also to measure increase in performances of a particular sector. The momentum is changing as result of the adoption of these SPC techniques in the healthcare system to aid in process understanding and to measure the delivery of the services rendered to the public. Hospitals, in particular, are a part of the health care service industry that routinely collect data but do not use it to the best advantage. Cases treated at times in hospital are both univariate and multivariate cases. The Univariate has only one variable or sickness at time and the multivariate analysis involves variables that have more than one quality characteristic or sicknesses [1]. These quality characteristics are clearly correlated and control chart for monitoring the individual quality characteristic may not be adequate for detecting changes in the overall quality of the product. The system experiences some challenges in the application of SPC to monitor performance systems which including identification of the best statistical model for the common cause variability, grouping of data, selection of type of control chart, the cost of false alarms and lack of signals, and difficulty in identifying the special causes when a change is signaled [2] [3] [4]. Nevertheless, carefully constructed control charts are powerful methods to monitor performance systems in a hospital. Control charts were introduced by Dr. Shewhart in1920's and involve two phases. In phase I, a set of historical data is analyzed to assess stability and identify special causes. If no special causes are present, the in-control process parameters are estimated and control limits are established. In phase II, the data are sequentially collected over time to assess whether the performance has changed from the estimated value [5] [6] [7]. The objective of this paper is to used a Bayesian sequential estimation control chart to determine a small shift that can easily show an out of control signals. Also the phase II approach which involves sequential collection of data over a period of time is adopted in this research using National Orthopaedic data and the result is compared with the conventional Hotellings' T2.

2. The Bayesian Sequential Methodology

If a set of observations x1 , x2 , x3 , . . . , xn generates a posterior distribution and, in a similar situation, additional

data are collected beyond these observations, then the posterior distribution found with earlier observations becomes the new prior distribution and the additional observations give a new posterior distribution and inference can be made from the second posterior distribution. This procedure can continue with newer and more observations. That is, the second posterior becomes the new prior, and the next set of observations give the next posterior from which the inference can be made [8]. This is the principle of Bayesian sequential methodology that we propose to estimate the proportion of counts data obtained from the hospital.

Based on the Bayesian approach described above, data were collected monthly and collated yearly for three years (2009, 2010 and 2012) from the hospital records. The population proportion of patients admitted for orthopaedic surgery is denoted by Po while the proportion of patients admitted for orthopaedic surgery in age group j is Pj (j = 1, 2, . . ., 5). Xij represents a random outcome of patient i examined in age group j.

$$Y_{ijk} = \begin{cases} 1 \text{ if ith Patient is admitted for orthopaedic surgery in age group j and in year k} \\ 0 \text{ Otherwise} \end{cases}$$

$Y_{jk} = \sum_{i=1}^{n_{jk}} Y_{ijk}$ = the total number of patients admitted for orthopaedic surgery in age group j in year k.

njk = the total number of patients admitted for treatments (both orthopaedic and non-orthopaedic surgeries) in the hospital in age group j in year k.

$P_{jk} = \frac{Y_{jk}}{n_{jk}}$ = the proportion of patients admitted for orthopaedic surgery in age group j and year k.

For each year in each age group, we computed sample proportions Pjk as follows:

In 2009 and age group j: $P_{j1} = \frac{Y_{j1}}{n_{j1}}$

In 2010 and age group j: $P_{j2} = \frac{Y_{j2}}{n_{j2}}$

In 2011 and age group j: $P_{j3} = \frac{Y_{j3}}{n_{j3}}$

Estimators of sample proportions: $\hat{P}_{jk} = \frac{Y_{jk}}{n_{jk}}$ and

$Var(\hat{P}) = \frac{\hat{P}_{jk}(1-\hat{P}_{jk})}{n_{jk}}$.

2.1. The Beta-Binomial Model

The EB model to be applied is a conjugate beta-binomial model where the binomial distribution represents the likelihood of the observed data likelihood and the beta distribution serves as the prior distribution of the binomial parameter. The posterior mean is

$$\tilde{P}_{jk} = \int P_{jk} f(P_{jk} \mid Y_{jk}, \eta) dP_{jk} \quad (1)$$

A key component of this integral is $f(P_{jk} \mid Y_{jk}, \eta)$, the posterior distribution of which is P_{jk}. Under the general Bayesian framework and using the beta conjugate prior plus the binomial likelihood, the posterior distribution of P_{jk} is:

$$f(P_{jk} \mid Y_{jk}, \eta) = \frac{\binom{n_{jk}}{Y_{jk}} P_{jk}^{Y_{jk}} (1-P_{jk})^{n_{jk}-Y_{jk}} \frac{1}{B(r,s)} P_{jk}^{r-1}(1-P_{jk})^{s-1}}{\int P_{jk}^{Y_{jk}} (1-P_{jk})^{n_{jk}-Y_{jk}} \frac{1}{B(r,s)} P_{jk}^{r-1}(1-P_{jk})^{s-1} dP_{jk}}, \quad (2)$$

$$\eta = (r, s)$$

There is need to estimate the hyperparameters r and s of the beta distribution in order to completely specify the prior. This can be achieved easily through re-parameterization of $f(P_{jk} \mid \eta)$, and using moment estimation [9]. Letting $P_o = \frac{r}{r+s}$; $M = r + s$ and using the prior distribution of P_{jk} ;

$E(P_{jk}) = P_o$ and $Var(P_{jk}) = \frac{rs}{(r+s+1)(r+s)^2} = \frac{P_o(1-P_o)}{M+1}$.

These are known as prior mean and variance respectively. Consequently,

$$f(P_i \mid Y_{jk}, \hat{\mu}, \hat{\mathrm{M}}) = \frac{1}{B(\alpha, \beta)} P_{jk}^{\alpha-1}(1-P_{jk})^{\beta-1} \quad (3)$$

Where

$$\hat{\alpha} = Y_{jk} + \hat{\mathrm{M}}\hat{P}_o ; \ \hat{\beta} = n_i - Y_{jk} + \hat{\mathrm{M}}(1-\hat{P}_o), \ \hat{P}_o = \frac{\sum Y_{jk}}{\sum n_{jk}}$$

And

$$\hat{\mathrm{M}} = \frac{\hat{P}_o(1-\hat{P}_o) - S_P^2}{S_P^2 - \frac{\hat{P}_o(1-\hat{P}_o)\sum \frac{1}{n_{jk}}}{N}}$$

where $S_P^2 = \frac{N \sum n_{jk} (\hat{P}_{jk} - \hat{P}_o)^2}{(N-1)\sum n_i}$.

With M and P_o estimated, then;

$$\tilde{P}_{EB} = E(P_{jk} / Y_{jk}, \hat{P}_o, \hat{\mathrm{M}}) = \frac{\alpha}{\alpha+\beta} = \frac{Y_{jk} + \hat{\mathrm{M}}\hat{P}_o}{n_{jk} + \hat{\mathrm{M}}} = \left(\frac{\hat{\mathrm{M}}}{n_{jk} + \hat{\mathrm{M}}}\right) P_o + \left(\frac{n_{jk}}{n_{jk} + \hat{\mathrm{M}}}\right) \frac{Y_{jk}}{n_{jk}} \quad (4)$$

$$Var(\tilde{P}_{EB}) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}. \quad (5)$$

Consequently, $\hat{\lambda} = \frac{\hat{\mathrm{M}}}{n_{jk} + \hat{\mathrm{M}}}$ and it can be readily seen

that where $\hat{M}$ (the scale factor) is large relative to n_{jk}, λ is large and $\hat{P}_o$ receives a larger weight than $\frac{Y_{jk}}{n_{jk}}$. But large $\hat{M}$ implies small prior variance. Thus, the estimate which is associated with smaller variance receives larger weight in determining the posterior mean $\tilde{P}_{EB}$. On the other hand, if $\hat{M}$ is small relative to n_{jk}, the sample mean receives more weight. We note that the posterior density for the overall age group proportion P_o is obtained by replacing Y_{jk} and n_{jk} in equation (3) with Y and N, respectively. Under conjugacy, the EB estimator of a proportion $\hat{P}_{ij}$ is a weighted mean of two estimators, the mean of the prior density P_o and the sample proportion estimator $\hat{P}_{ij}$. Thus,

$$\tilde{P}_{EB} = \lambda P_o + (1-\lambda)\hat{P}_{ij} \quad (6)$$

$\tilde{P}_{EB}$ is the empirical Bayes Estimators with λ as the shrinkage factor. λ is a function of the prior and sample estimator variance such that, if variance of sample estimator is large, the weight of $\hat{P}_o$ (i.e. λ) will be large and $\tilde{P}_{EB}$ will shrink towards $\hat{P}_o$. Two components of the above model λ and $\hat{P}_o$ are derived from the EB process, [10].

2.2. Multivariate Hotelling's T^2 Control Chart

Hotelling's T2 is a very versatile multivariate control chart statistic. It can be used not only to identify outliers in the historical data set but also to detect process shift using new incoming observation.

In the univariate test of means, the test statistic employed is Student t given by $t = \frac{\overline{X} - \mu}{s/\sqrt{n}}$

where $\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$ and $s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2$

This test statistic has a Student t distribution with n – 1 degrees of freedom. When the observed t exceeds a specified percentage point of the t distribution with n – 1 degrees of freedom, H_o is rejected.

The multivariate analogue of the square of t was proposed by Hotelling's in 1931, it was proposed for the 2-sample case as; [11]

$$t^2 = \frac{(\overline{X} - \mu)^2}{s/\sqrt{n}} = n((\overline{X} - \mu_0)' (s^2)^{-1}(\overline{X} - \mu_0)$$

Rejecting Ho when absolute value of t (|t|) is large is equivalent to rejecting Ho: if t^2, the squared distance from sample mean $\overline{x}$ to the test value μ_0, is large. When t^2 is generalized to p multivariate random variables, it becomes

$$T^2 = (\overline{X} - \mu_0)' \left(\frac{1}{n}S\right)^{-1} (\overline{X} - \mu_0)$$

$$T^2 = n(\overline{X} - \mu_0)' (S)^{-1} (\overline{X} - \mu_0)$$

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j ,$$

$$S_{(p\times p)} = \frac{1}{n-1}\sum_{j=1}^{n} (X_j - \overline{X})(X_j - \overline{X})'$$

3. Results

The results of the application of Beta-Binomial model and Bayesian sequential methods to the data of different age group patients for the three years (2009, 2010 and 2011) are presented in Table 1 and 2 below. The hyperparameters μ and M are estimated using sample information. These are subsequently used to determine the parameters of the posterior distributions α and β , thereby completely specifying them. In our analyses, we obtain the yearly results for Bayesian Sequential (see Table 1).The result is plotted as shown in Figure 1. Comparing the yearly basis estimated sample proportions and EB proportions as well as variances of estimated sample proportions and EB proportions.

Table 1. *Comparative Analysis of Estimated Sample Proportions and EB Proportions.*

Year:	**2009**		**2010**		**2011**	
Age group	**P_{j1}**	**P_{EB}**	**P_{j2}**	**P_{EB}**	**P_{j3}**	**P_{EB}**
< 1yr	0.4900332	0.4885043	0.547619	0.5430286	0.4608819	0.470019
1 - 14yrs	0.4288879	0.4294206	0.52383	0.5215909	0.5281195	0.527877
15 – 44yrs	0.3881102	0.3887539	0.4268812	0.4263775	0.4754009	0.474256
45 – 64yrs	0.5764463	0.574785	0.6481088	0.6463872	0.6910533	0.689386
> 64yrs	0.5823899	0.5772227	0.6392638	0.6352365	0.7375887	0.727948
Overall	0.4628868	0.4628868	0.5231361	0.52281	0.5593983	0.559072

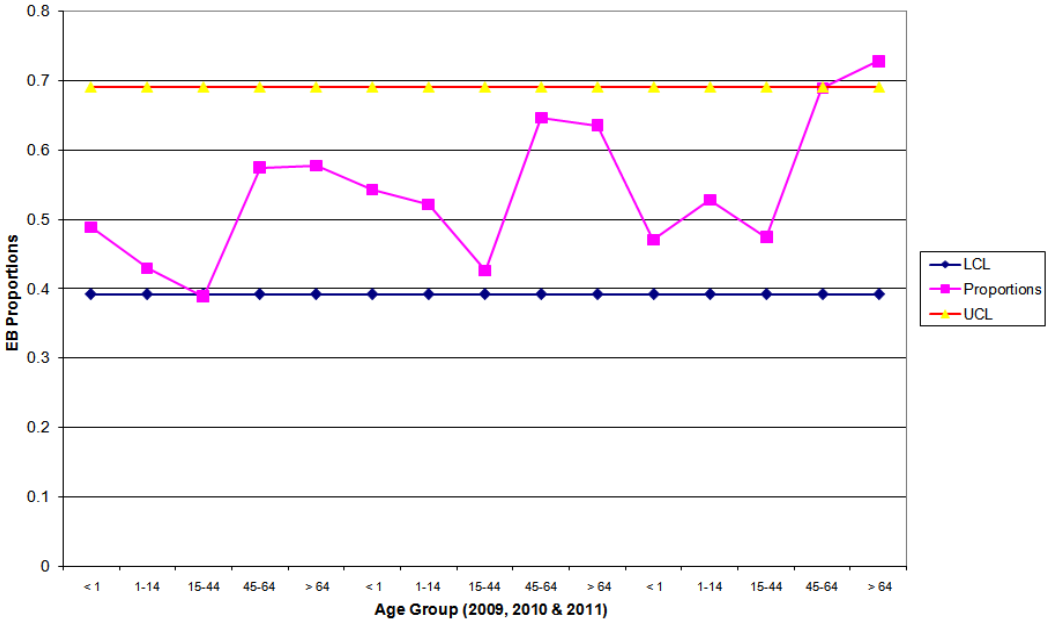

Figure 1.*The control chart for Bayesian Sequential of EB proportion.*

The result shows that the chart for patients between the ages of 15-44years (0.389) in 2009 is out of control. This implies that among the people that make use of the hospital the age bracket 15 – 44 records very high figure compared to others. In 2011 the result shows that there is a shift in the ages of the patients that attended the hospital from 15-44 years to 44-64 and 64 and above respectively (see Table 1 and Figure 1 above respectively). The result shows that this new approach is able to identify a small or slight shift that may occur among those that attended the hospital. Table 2 is the estimated values obtained for the covariance's and Table 3 is the variance and covariance values obtained from computation of Hotellings. Comparing the results of figures 1 and 2, figure 2 cannot identify any slight change that occur while the result of the sequential Bayesian analysis does. Also the values of the variances obtained from the sequential Bayesian analysis are better than that of Hotellings.

Table 2. *Comparative Analysis of Variances of Estimated Sample Proportions and EB Proportions.*

Year:	**2009**		**2010**		**2011**	
Age group	**Var(P_{j1})**	**Var(P_{EB})**	**Var(P_{j2})**	**Age group**	**Var(P_{j1})**	**Var(P_{EB})**
< 1yr	0.00042	0.00039	0.00037	0.00034	0.00035	0.00031
1 - 14yrs	0.00011	0.00011	0.00011	0.00010	0.00011	0.00011
15 - 44yrs	0.00006	0.00006	0.00006	0.00006	0.00007	0.00007
45 - 64yrs	0.00010	0.00010	0.00010	0.00009	0.00009	0.00009
> 64yrs	0.00031	0.00029	0.00028	0.00027	0.00023	0.00021
Overall	0.00002	0.00002	0.00002	0.00002	0.00003	0.00002

Table 3. *Computation of variance of Hotellings' T Square.*

	under 1yr	**1 - 14YRS**	**15 - 44YRS**	**45 - 64YRS**	**65YRS & ABOVE**
under 1yr	72.60635	77.31905	182.14286	152.22063	71.78571
1 - 14YRS	77.31905	692.25	652.9	291.04048	150.93571
15 - 44YRS	182.14286	652.9	2473	172.21429	-2.75714
45 - 64YRS	152.22063	291.04048	172.21429	843.97063	393.19286
65YRS & ABOVE	71.78571	150.93571	-2.75714	393.19286	264.53571

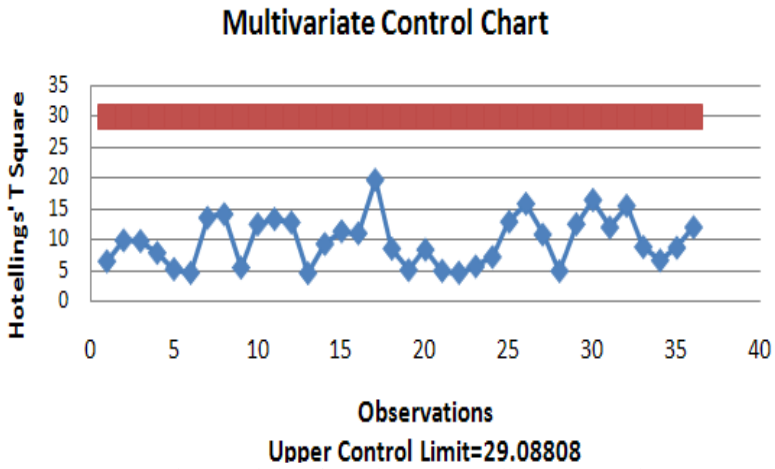

Figure 2. *The control chart for Multivariate HotellingsT24. Conclusion.*

This paper has been able to used a Bayesian sequential estimation control chart to determine a small shift that can easily show an out of control signals. Also the phase II approach which involves sequential collection of data over a period of time is adopted in this research using National Orthopaedic data and the result is compared with the conventional Hotellings' T^2. Bayesian sequential estimation of proportion is suitable to identify and slight change that occurs than the usual or conventional technique. Similarly, the overall variances of the proportions tend more to zero over the three years under review than that of the variance of Hotellings' T square. Thus, the results show that the EB estimators are better estimators on the basis of efficiency and consistency properties of good estimators.

References

[1] Resul Oduk (2012) Control Charts for Serially Dependent Multivariate Data Thesis submitted to the Department of Informatics and Mathematical Modeling at Technical University of Denmark in partial fulfillment of the requirements for the degree of Master of Science in Mathematical Modeling and Computation Technical University Of Denmark.

[2] Shewart W.A (1952) The Application of Statistics as an aid in maintaining quality of a manufactured product" Journal of American Statistical Association: 546-548.

[3] Shewhart WA, (1986), Statistical Method from the Viewpoint of Quality Control General Publishing Company, ISBN 0-486-65232-7.

[4] Shewhart WA, Economic control of quality of manufactured product (1931), Princeton, NJ:Reinhold Co.

[5] Woodall, W. H. (2000). "Controversies and Contradictions in Statistical Process Control" (with discussion). Journal of Quality Technology 32, pp. 341–378. (available at ww.asq.Org/pub/jqt).

[6] Woodall, W.H. Review of Improving Healthcare with Control Charts by Raymond G. Carey, Journal of Quality Technology, 36, 336-338 (2004). 23.

[7] Woodall, W.H. Use of control charts in health-care and public-health surveillance (with discussion), Journal of Quality Technology, 38, 89-104 (2006).

[8] Lee, P. (2004), Bayesian Statistics An Introduction, Hodder Arnold, New York.

[9] Brandel, J. (2004), Empirical Bayes Methods for missing data analysis, Department of Mathematics Uppsala University, Project Report.

[10] Carlin, B. P. and Louis, T. A. (2000b), Bayes and Empirical Bayes Methods for Data Analysis, Boca Raton, Florida: Chapman and Hall/CRC Press.

[11] Richard A.J. and Dean W.W.(1988) Second Edition Applied Multivariate Statistical Analysis Prentice, Hall International, Inc. 607.