International Journal of Computational and Theoretical Statistics
ISSN (2210-1519)
Int. J. Comp. Theo. Stat. 2, No. 2 (Nov-2015)

A Bayesian Analysis for Repeated Measurements Adopting a More Informative New Prior for Updating Predicted Mortality Rate After the Cardiac Surgery Activity

Jamal A. Al-Saleh

Department of Statistics and Operation Research, Faculty of Science, Kuwait

Received 20 May 2015, Revised 26 September 2015, Accepted 15 October 2015, Published 1 November 2015

Abstract: In this paper, a repeated predicted measurements, using Bayesian approach is considered to filter non appropriate priors from a pool of prior information on (0,1) interval. We used data on mortality rates after the cardiac surgery activity in the UK and Ireland between the years 2003 to 2012. One of a member of this pool is a new proposed unit interval prior (uitp). In repeated Bayesian analysis, the most appropriate prior will be the one which is more informative for mortality rates update and adjustment. The performance of the proposed prior (utip) is found to be the most satisfactory working mechanism prior information, for the particular data set, before observing the data. By using Markov chain Monte Carlo simulation, posterior summaries required for an uncertain parameter are obtained for this particular data set. To show the performance and importance of proposed new prior, we considered Jeffreys' prior (non-informative), Kumaraswamy's prior (partly informative) and conventional beta prior (informative) and update the prior information with that of proposed new prior under the same conditions. This opens the way to richer, more reliable and consistent inference summaries and avoids the numerical problems that are encountered with non-Bayesian methods.

Keywords: Cardiac Surgery Activity; Data Analysis; Model Selection; Pool Of Priors; Mortality Rate; Repeated Bayesian Analysis; Unit Interval Type Prior

1. Introduction

In Bayesian analysis, the main goal is to obtain the most appropriate model selection and prior knowledge to produce posterior summaries. To achieve this goal we adopt a repeated measurements through Bayesian inference approach. It helps to filter non appropriate prior among a pool of prior information available from the experts, and applying different and more informative prior to choose the best information. This information will help us to get close to the true parameter. Usually the Bayesians prefer to consult a few experts/consultancies to get the prior information about a parameter. If a pool of prior information is available, it would help the Bayesian to select the best prior from the pool. However, in case of sensitive information, there are high chances that the information provided may be questionable. For example in cardiac surgery activity if the number of deaths are significant during and after the surgery, it is quite likely that the hospital may not provide with reliable information. In such cases, the researchers have no idea about the past history of the hospital and therefore, they have no choice but to go for a non-informative (flat) prior which expresses vague or general information about a variable. Non-informative priors, though do not provide any specific information, but still preferred because it provides the information such as the variable is positive or the equal probabilities are assigned to all possibilities.

Due to the flexibility of the beta distribution, it is the first choice to consider a prior in the interval (0,1). To increase the number of two parameters (say α and β), priors in (0,1) we define through construction a new unit interval type distribution (uitp) that has some of the same properties which the beta prior has. These properties are: it reduces to uniform and straight line (non-informative) which adds to versatility of the proposed distribution, and it is unimodel, hence more convenient to handle. The pdf of uitp, the shapes of the pdf for the different choices of its parameters are not available in the literature and hence have not been studied. The boundary behavior and a few special cases of uitp and beta distribution are common. One of the important feature of uitp is that it does not require special function either

E-mail: jamal.alsaleh@ku.edu.kw

in pdf or in distribution function, and has very simple normalizing constant. The absence of special function in the distribution function of uitp may find some favor in some quarters from the modeling perspective. In section 4, from model selection perspective, we showed the usefulness of uitp which the conventional beta distribution does not have. The proposed uitp can easily be generalized by introducing more parameters.

We have arranged the paper as follows: In section 2, the problem of repeated Bayesian (RBayesian) approach for the problem of mortality rates is introduced and the parameter p of success in a number of trials, with x number of success such that $x \sim b(n,p)$ a binomial distribution of index x and parameters p is considered. In sub section 3.1-3.3, RBayesian analysis by using some priors is defined along with conventional beta(α,β) as a prior and obtained the posterior estimate under the same conditions. In sub section 3.4, the construction of the pdf of new uitp(α,β) along with the cumulative distribution function, and the expressions of r-th moment, a few shapes of the density, along with figure for different choices of parameters, are shown which may give some insight to the scientists. We have also considered in sub section 3.4 the usefulness of the proposed distribution with the help of data of the mortality rates in all hospitals taken from the Society for Cardiothoracic Surgery in Great Britain & Ireland [see [1]] after the cardiac surgery activity and outcomes in the UK, are given by using well known MCMC, Gibbs sampling approach. The comparison of the posterior estimates obtained by using new prior, are made in sub section 3.5 and the section 4 has concluding remark.

2. The problem of repeated Bayesian approach for updating predicted mortality rates

The key idea to a repeated Bayesian (RBayesian) approach are the repeated predicted measurements, through consideration of the likelihood function which includes information about the mortality data, and the prior information knowledge which provides known information about the mortality before considering the data. The prior information knowledge and likelihood can be combined to form the posterior probability, which gives total information knowledge about the mortality after the data has been observed. Repeated statistical analysis with different prior information knowledge of the posterior, can be used to wash out all possible priors that are unsuitable and isolate the best prior of researcher's interest that help to draw substantive conclusions about the mortality rates.

We illustrate each of these steps of an RBayesian analysis using the data set in (Table 1) which represents the observed mortality rates from the period 2003 to 2012.

Table 1. The mortality rates from 2003-2012

Year	Number of Operations	Mortality rates (%)
2003	36,374	3.74
2004	37,453	3.62
2005	35,922	3.56
2006	36,654	3.58
2007	39,202	3.12
2008	39,239	3.21
2009	36,321	3.14
2010	34,737	3.25
2011	34,959	2.97
2012	34,174	2.98

The last column of the table is the mortality rate in percentage (for the year 2003 it is 0.0374 or 3.74% and for the year 2012 it is 0.0298 or 2.98%). The years given in the data represent a financial year. For example 2011 represents 1st April 2011 – 31st March 2012. It is worth to mention that though it is yearly data, yet it is not time series data. This data set will be analyzed using Openbugs[@] computer software. Our interest, therefore, would be in the probability p of success in a number of trials, which can result in success or failure. Suppose there is a fixed number of n trials, with x, number of successes such that $x \sim b(n,p)$ a binomial distribution of index x and parameters p. Now, let x_t and n_t be, the number of deaths, and the number of cardiac surgery performed in respective year t. This problem can be modeled as a binary response variable with true failure probabilities p_t. Thus x_t can follow $b(n_t,p_t)$ a binomial distribution where t=1, ..,10.

3. RBayesian analysis by using some selected priors

As mentioned in the introduction, in case of sensitive information, there are high chances that the information provided may not be truthful and hence one goes for a non-informative prior (flat prior) which expresses vague or general information about a variable. It is common practice that the Bayesians prefer to consult a few experts/consultancies or his own knowledge based on past data, to get the prior information about a parameter p. After selecting data model, a repeated Bayesian analysis by choosing a suitable prior, representing the current state of information is applied as explained in the section 2. We will start the RBayesian analysis using the non-informative prior, followed by partly-informative prior then informative prior using the most popular beta prior.

3.1 Jeffery's prior

We first assume that the true failure probabilities for each year [see Table 1] having a standard non informative prior distribution for the p_t's, The (table 2) represents the estimates of surgical mortality for each year with Jeffry's prior [beta(1/2,1/2)] along with mean, standard deviation, median and MC error. It can also be noted that the posterior medians for all years move drastically towards the right in comparison to the last column of (table 1). The maximum posterior median is 24.46% (true mortality =2.98%) and the posterior median for the minimum is 17.15% (true mortality =3.58%).

Table 2. Bayesian summary of surgical mortality for each year with non-informative prior beta(1/2 ,1/2)

Year	Mean	SD	MC error	2.5%	Median	97.5%
2003	0.3176	0.2869	0.002579	6.864E-4	0.2346	0.9364
2004	0.2685	0.2555	0.002679	6.682E-4	0.1849	0.8611
2005	0.2583	0.2491	0.002356	6.831E-4	0.1771	0.8523
2006	0.2562	0.2501	0.002716	6.219E-4	0.1715	0.8499
2007	0.2625	0.2529	0.002498	5.601E-4	0.1803	0.8614
2008	0.2706	0.2601	0.002173	5.907E-4	0.187	0.8803
2009	0.2712	0.2607	0.002494	5.703E-4	0.1863	0.8705
2010	0.2735	0.2627	0.002621	7.157E-4	0.1864	0.8857
2011	0.2718	0.2591	0.002247	6.284E-4	0.1866	0.8735
2012	0.3140	0.2733	0.002903	0.001467	0.2446	0.9074

3.2 Kumaraswamy's prior

Table 3 represents the estimates of surgical mortality for each year with partly-informative prior Kumaraswamy$(1,\beta)$ [[2] Jones] along with mean, standard deviation, median and MC error. The posterior medians for all years once again move drastically towards the right in comparison to the last column of (table 1). The maximum posterior median is 18.03% (true mortality = 3.74%) and the posterior median for the minimum is 15.84% (true mortality =3.12%).

Table 3. Bayesian summary of surgical mortality for each year with informative Kumaraswamy$(1,\beta)$ prior

Year	Mean	SD	MC error	2.5%	Median	97.5%
2003	0.2325	0.192	0.002228	0.007502	0.1803	0.704
2004	0.2118	0.1775	0.001962	0.007633	0.1641	0.662
2005	0.2069	0.1752	0.001983	0.006237	0.1596	0.6589
2006	0.2064	0.1743	0.002007	0.00648	0.1599	0.6479
2007	0.2063	0.1771	0.001939	0.006014	0.1584	0.6558
2008	0.2109	0.1767	0.001839	0.007442	0.1631	0.6548
2009	0.2135	0.1788	0.001976	0.006788	0.1666	0.6651
2010	0.212	0.1786	0.002141	0.006233	0.1653	0.6608
2011	0.2123	0.1796	0.002074	0.006588	0.1617	0.6642
2012	0.2256	0.1845	0.002159	0.007898	0.1795	0.6926

3.3 Hierarchical model

A more realistic model for the surgical data is to suggest the hierarchical model implemented as follows:

1. At the first stage we assume a prior belief that follows beta(α_t,β_t,) for the true failure probabilities p_t for each hospital *t*.
2. At the second stage, we will assume the following prior specification for the hyperparameters α_i and β_t; α_t~gamma(ϑ,Λ), and β_t~gamma(ϑ,Λ) independent.
3. At the third stage, we assume the following prior specification for the hyper-hyperparameters where ϑ~exponential(T), Λ~exponential(T), T=0.1 [Table 3].

A Markov Chain Monte Carlo (MCMC) Gibbs sampling approach implemented in using Openbugs[@] computer software can give an analysis of estimates of surgical mortality in each year *t*. A burn in of 1000 updates followed by a further 11000 updates give estimates of $p_{t,}$ for each year t=1,…,10.

Table 4 represents the estimates of surgical mortality for each year with informative prior beta(α,β) along with mean, standard deviation, median and MC error. On critical examination of the simulations (table 4), it is noticed that there is a dramatic shift to the right in the posterior means from the true mortality rate for all years (table 1). The maximum increase is for the year 2012, the mortality rate in table 4 is 7.42%, and the true value is 3.74%. Almost similar type of observations can be seen for the rest of the years from 2003 to 2011.

Table 4. Bayesian summary of surgical mortality for each year with informative prior beta(α,β)

Year	Mean	SD	MC error	2.5%	Median	97.5%
2003	0.1356	0.1788	0.002117	1.631E-7	0.06018	0.6506
2004	0.1233	0.1639	0.001986	1.354E-7	0.0525	0.5958
2005	0.1194	0.1607	0.00191	1.464E-7	0.04971	0.5763
2006	0.1179	0.1593	0.002084	1.535E-7	0.04739	0.5837
2007	0.119	0.1632	0.001833	6.071E-8	0.04903	0.6012
2008	0.1226	0.1647	0.002134	1.205E-7	0.04947	0.6056
2009	0.122	0.1652	0.001886	5.293E-8	0.04938	0.5955
2010	0.1248	0.1674	0.002183	4.81E-8	0.0522	0.5994
2011	0.1237	0.1679	0.001955	9.786E-8	0.05171	0.6084
2012	0.1456	0.1787	0.002071	4.72E-6	0.0742	0.6503

If we compare the posterior medians among the tables 2 to 4, we find that the trend is more or less same. However, the hierarchical models (table 4) give better results than that of using Jeffery's or Kumaraswamy's prior. But the posterior medians of all are almost more than two times than that of the true values even if informative prior beta(α,β) is used. Still all priors give inconsistent predicted measurements with true mortalities. The above results clearly show that Bayesian summaries obtained from the available priors in the literature for the interval (0,1) do not improve the estimates which to us is new information about the parameter *p*.

3.4 Construction of unit interval type prior (uitp)

In this sub section we will construct a new prior by using the results of tables 2 to 4, obtained in Section 3.1-3.3, as updated new information. The main motivation to develop new distributions was the distribution function of the beta prior in estimating the posteriors. One extended form of the beta prior is by considering the proportion $C\sum_0^1 \frac{\beta^j x^\beta}{(\alpha x^\beta+1)^{j+1}}$, where C is the normalizing constant. To find C, integrating to unity, that is $\int_0^1 C\sum_0^1 \frac{\beta^j x^\beta}{(\alpha x^\beta+1)^{j+1}}dx = 1$, which gives C=$\alpha$+1. Therefore, we can introduce the new pdf of the new unit interval type prior uitp(α,β) of a random variable X, with parameters $\alpha \geq 0$, $\beta \geq 0$, which is defined in eq (3.4.1) as follows:

(3.4.1) $$f(x) = \frac{(\alpha+1)X^\beta(\alpha x^\beta+\beta+1)}{(\alpha X^\beta+1)^2} \quad , \ x \in (0,1)$$

The cumulative distribution function of (3.4.1) is $F(\mathrm{x})=\frac{(\alpha+1)X^{\beta+1}}{\alpha x^{\beta}+1}$.

A few shapes of the density, for different values of parameters are shown in figure 1. From the figure 1, it can be noted that for α = 1, β = 0, the distribution reduces to uniform and straight line. The *r*th moment about origin of uitp(α,β) is given in eq (3.4.2):

$$(3.4.2)\quad \mu_r' = \frac{1}{\beta+r+1}\left((\alpha+1)(\beta+1)F''_{\frac{\beta+r+1}{\beta}}\right) + \frac{1}{2\beta+r+1}\left(\alpha(\alpha+1)F''_{\frac{2\beta+r+1}{\beta}}\right)$$

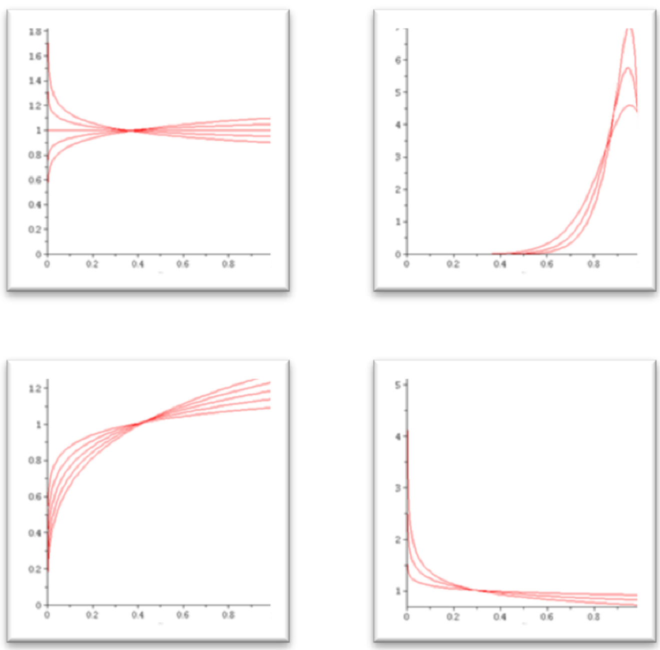

Figure 1. The pdf of eq (3.4.1) for different choices of *α*, *β*.

where $F_v'' = F([2,v];[1+v];[-\alpha])$, and F is the generalized hypergeometric series.[see [3] Andrews, Chapter 9, p 365], and is defined by:

$$(3.4.3)\quad F([n_1,n_2,\ldots,n_s];[d_1,d_2,\ldots,d_t]:z) = \sum_{k=0}^{\infty}\frac{\left(\prod_{i=0}^{s}\frac{\Gamma(n_i+k)}{\Gamma(n_i)}\right)z^k}{\left(\prod_{i=0}^{t}\frac{\Gamma(d_i+k)}{\Gamma(d_i)}\right)k!}.$$

Unlike the pdf eq(3.4.1) and F(x) of uitp(α,β), the r-th moments do have special functions in the eq (3.4.2). However, by using eq (3.4.2) and (3.4.3) we can compute the mean and the variance and/or the r-th moments which are not difficult to obtain. [4] Feng, Li, wrote an R function to calculate the generalized hypergeometric function for real numbers. With minor changes we can use his function to compute eqs(3.4.2) and (3.4.3).

Using the same hierarchical model implemented in sub section 3.3, Table 5 represents the estimates of surgical mortality for each year with informative priors uitp(α,β), along with mean, standard deviation, median and MC error. From Table 5, it can be noted that there is a marginal shift to the right in the estimates posterior medians from the true mortality rate for all years (table 1). The maximum increase is for the year 2003, the mortality rate is 6.19%, while the true value is 3.74% and the posterior median is 5.38% for the minimum true mean 2.97% for the year 2011. Almost similar type of observations can be seen for the rest of the years from 2003 to 2012. When we look into the MC error given in tables 2, 3 and 4, 5, it can be noticed that for uitp's as prior (tables 5) has the least values for all the years.

Table 5. Bayesian summary of surgical mortality for each year with informative prior uitp(α, β)

Year	Mean	SD	MC error	2.5%	Median	97.5%
2003	0.06198	0.04002	4.796E-4	0.0103	0.05387	0.156
2004	0.06034	0.03929	4.676E-4	0.00958	0.05266	0.1537
2005	0.05896	0.03888	4.361E-4	0.00937	0.05149	0.148
2006	0.06005	0.03778	4.498E-4	0.00911	0.05238	0.1526
2007	0.05459	0.03721	4.915E-4	0.008069	0.04708	0.1415
2008	0.05499	0.03575	3.906E-4	0.008503	0.04765	0.1411
2009	0.05564	0.03815	4.939E-4	0.007709	0.04792	0.1463
2010	0.05449	0.03722	4.55E-4	0.006958	0.047	0.1452
2011	0.05378	0.03611	4.343E-4	0.006993	0.04603	0.1425
2012	0.05581	0.03815	4.465E-4	0.007225	0.04768	0.1494

Table 6. Summary comparison of surgical mortality prediction of median in (%) for each year for different prior information

Year	Beta	Uitp	True Mort. Rates
2003	6.01	5.39	3.74
2004	5.25	5.27	3.62
2005	4.97	5.15	3.56
2006	4.74	5.24	3.58
2007	4.90	4.71	3.12
2008	4.95	4.77	3.21
2009	4.94	4.79	3.14
2010	5.22	4.70	3.25
2011	5.171	4.60	2.97
2012	7.42	4.77	2.98

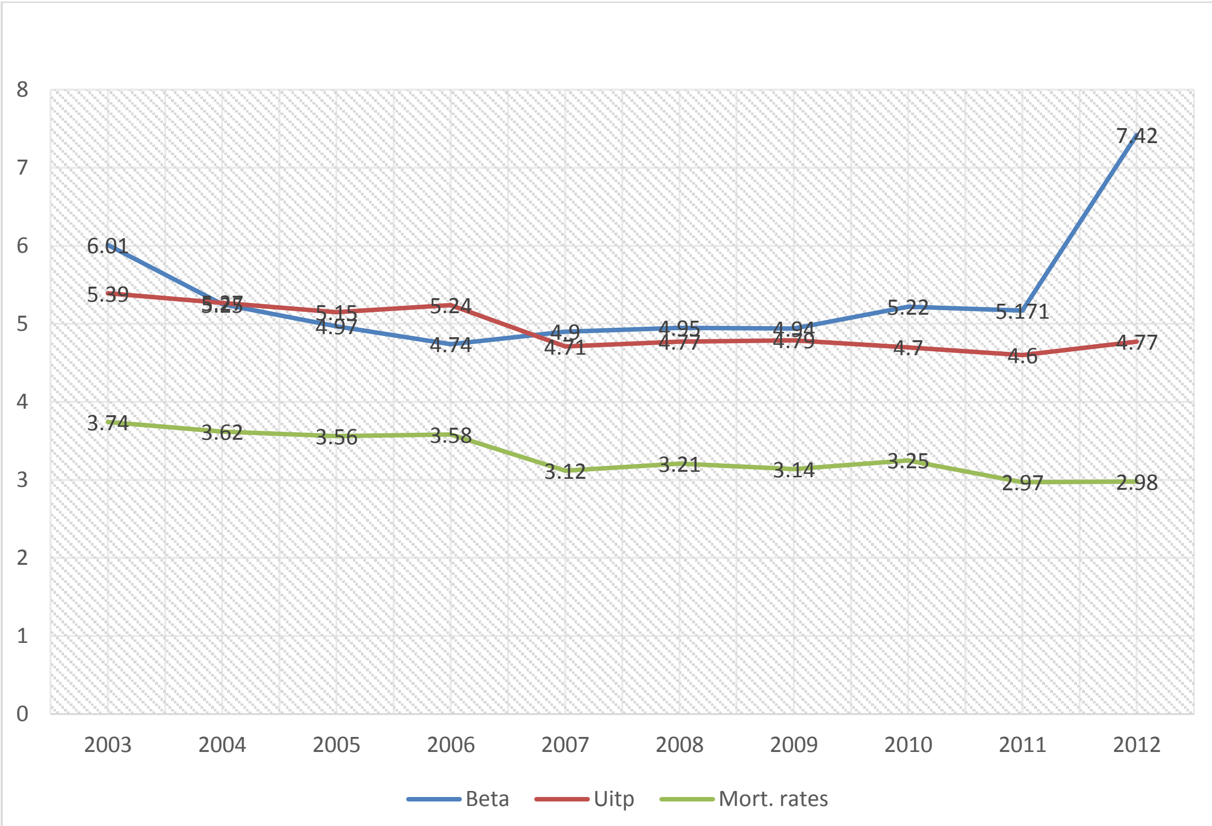

Figure 2. Chart of summary comparison of surgical mortality prediction in (%) for each year compared with mortality rates

3.5 Comparison of uitp prior's and Beta prior posterior summaries

After obtaining the posterior distribution, we analyze and summarize the inferential conclusions through the established framework of Bayesian approach by using point estimates either through mean or median or mode followed by specified percentiles. In table 6 and figure 2, we have summarized comparison of surgical mortality prediction of median in (%) for each year.

4. Concluding Remarks

The purpose of the present study is to consider mortality rates data applying RBayesian analysis to filter non appropriate priors using a pool of prior information on (0,1) interval. One of a member of this pool is the new construction (uitp) prior. The Table 6 and figure 2 give us an idea how useful the uitp is for surgical mortality prediction. We have shown that the proposed prior is a good working mechanism for the mortality rates data which is different and more informative for mortality update and adjustment. This opens the way to richer, more reliable and consistent inference summaries. Furthermore, we aim to create a pool of priors in the interval (0,1) and it might be worth consideration from time to time. The proposed prior have two important properties which are common with standard beta distribution. One of the advantages of uitp is that they have simple distribution functions as well as the probability density function. Another salient feature of the proposed distributions is that neither they depend on special functions in the density nor in the distribution function which makes computational calculation straightforward. The proposed distribution is likely to be useful for statistical modeling and model selection, and hence may find preference for scientific uses in more applications.

Acknowledgments

The support by the Research Administration, Kuwait University, Kuwait, is acknowledged.

References

[1] http://bluebook.scts.org

[2] Jones, M.C. "Kumaraswamy's distribution: A beta-type distribution with some tractability advantages" (2009). *Statistical Methodology* **6** (1): 70–81.

[3] L. C. Andrews, Special Functions of Mathematics for Engineers (1998). *Oxford University Press, Oxford, UK.*

[4] Feng Li., Code at (https://github.com/fengli/flutils/blob/master/math/ghypergeo.R). The function depends on the Pochhammer function which is available at (https://github.com/feng-li/flutils/blob/master/math/pochhammer.R).