
Sequential Bayesian Analysis of Bernoulli Opinion Polls; a Simulation-Based Approach

Jeremiah Kiingati*, Samuel Mwalili, Anthony Waititu

Department of Statistics and Actuarial Sciences, Jomo Kenyatta University of Agriculture and Technology, Nairobi, Kenya

Email address:

jkiingati@gmail.com (J. Kiingati), samuel.mwalili@gmail.com (S. Mwalili), agwaititu@gmail.com (A. Waititu)

*Corresponding author

To cite this article:

Jeremiah Kiingati, Samuel Mwalili, Anthony Waititu. Sequential Bayesian Analysis of Bernoulli Opinion Polls; a Simulation-Based Approach. *International Journal of Data Science and Analysis*. Vol. 6, No. 4, 2020, pp. 113-119. doi: 10.11648/j.ijdsa.20200604.13

Received: August 17, 2020; **Accepted:** September 5, 2020; **Published:** September 19, 2020

Abstract: In this paper we apply sequential Bayesian approach to compare the outcome of the presidential polls in Kenya. We use the previous polls to form the prior for the current polls. Even though several authors have used non-Bayesian models for countrywide polling data to forecast the outcome of the presidential race we propose a Bayesian approach in this case. As such the question of how to treat the previous and current pre-election polls data is inevitable. Some researchers consider only the most recent poll others Combine all previous polls up the present time and treat it as a single sample, weighting only by sample size, while others Combine all previous polls but adjust the sample size according to a weight function depending on the day the poll is taken. In this paper we apply a sequential Bayesian model (as an advancement of the latter which is time sensitive) where the previous measure is used as the prior of the current measure. Our concern is to model the proportion of votes between two candidates, incumbent and challenger. A Bayesian model of our binomial variable of interest will be applied sequentially to the Kenya opinion poll data sets in order to arrive at a posterior probability statement. The simulation results show that the eventual winner must lead consistently and constantly in at least 60% of the opinions polls. In addition, a candidate demonstrating high variability is more likely to lose the polls.

Keywords: Sequential Bayesian Analysis, Bernoulli Opinion Polls, Election Forecasting, Simulation-Approach

1. Introduction

1.1. Background

The problem of understanding and predicting election outcomes has long been part of political science research. However, lack of pre-election poll data especially in developing countries is one of the main challenges associated with forecasting election outcomes [1-8]. Nevertheless, in developed countries opinion polls are now easily accessible through online polling. The online polling is overhauling traditional phone polls, according to [9] analysis of the 2016 US presidential election campaign.

The seminal work of inflation [10, 11], spurred numerous studies which examined the evolution of voting intentions, as measured by opinion polls, and in particular the relationship between political popularity, ethnicity, youth factor and economic variables such as inflation, gross domestic product, personal producer index and unemployment. See for example

[12-14]. An empirical issue of particular relevance to the present study is the degree of persistence in political popularity. Building on the rational expectations' version of the permanent income hypothesis due to [11] argued that the effect of news about the economy on voting intentions would be permanent. The practical implication of their model is that the time series of opinion data should behave like a random walk, with the autoregressive-moving average (ARMA) representation of the time series containing an autoregressive root of unity. Such processes are nonstationary, and exhibit no mean-reversion tendencies.

Further analysis by [15] rejected the unit root hypothesis in favour of stationary ARMA models, although with autoregressive coefficients close to unity. Such models would imply that the effect of news on voting intentions, although it could be quite persistent in practice, is in principle transitory. As a consequence of aggregating heterogeneous poll responses under certain assumptions about the evolution of individual opinion, Byers [16] concluded that the time series

of poll data should exhibit long memory characteristics. In an analysis of the monthly Gallup data on party support in the UK, Byers [16] confirmed that the series are long memory, and virtually pure ‘fractional noise’ processes.

However, even though this time series approach is appealing it requires data observed over a long period of time which is a limitation to us. The alternative approach is the frequentist regression modelling which does argument/update opinion polls as series of observations. However, holds our model to be updated once the data set is updated sequentially from time to time. In other words our expression must include the past information which serves as a prior information. It follows therefore that a Sequential Bayesian Analysis is the best candidate for this type of model.

Basically, a simple model of political popularity, as recorded by opinion polls of voting intentions, is proposed; in particular, the Sequential Bayesian Analysis.

1.2. Motivation for Bayesian Approach

Bayesian estimation and inference has a number of advantages in statistical modelling and data analysis. These includes:- (a) Provision of confidence intervals on parameters and probability values on hypotheses that are more in line with commonsense interpretations; (b) provision of a way of formalizing the process of learning from data to update beliefs in accord with recent notions of knowledge synthesis; (c) assessing the probabilities on both nested and non-nested models (unlike classical approaches) and; (4) using modern sampling methods, is readily adapted to complex random effects models that are more difficult to fit using classical methods [17].

Unlike in the past when statistical analysis based on the Bayes theorem was often daunting due to the numerical integrations needed. Recently developed computer-intensive sampling methods of estimation have revolutionised the application of Bayesian methods, and such methods now offer a comprehensive approach to complex model estimation, for instance in hierarchical models with nested random effects [18-21]. They provide a way of improving estimation in sparse datasets by borrowing strength [22] and allow finite sample inferences without appeal to large sample arguments as in maximum likelihood and other classical methods. Sampling-based methods of Bayesian estimation provide a full density profile of a parameter so that any clear non-normality is apparent, and allow a range of hypotheses about the parameters to be simply assessed using the collection of parameter samples from the posterior.

Bayesian methods may also improve on classical estimators in terms of the precision of estimates. This happens because specifying the prior brings extra information or data based on accumulated knowledge, and the posterior estimate in being based on the combined sources of information (prior and likelihood) therefore has greater precision. Indeed a prior can often be expressed in terms of an equivalent ‘sample size’.

The relative influence of the prior and data on updated beliefs depends on how much weight is given to the prior

(how ‘informative’ the prior is) and the strength of the data. For example, a large data sample would tend to have a predominant influence on updated beliefs unless the prior was informative. If the sample was small and combined with a prior that was informative, then the prior distribution would have a relatively greater influence on the updated belief:

How to choose the prior density or information is an important issue in Bayesian inference, together with the sensitivity or robustness of the inferences to the choice of prior, and the possibility of conflict between prior and data [23-25].

In some situations it may be possible to base the prior density for θ on cumulative evidence using a formal or informal meta-analysis of existing studies. A range of other methods exist to determine or elicit subjective priors [23-27]. A simple technique known as the histogram method divides the range of θ into a set of intervals (or ‘bins’) and elicits prior probabilities that θ is located in each interval; from this set of probabilities, $p(\theta / C)$ may be represented as a discrete prior or converted to a smooth density. Another technique uses prior estimates of moments along with symmetry assumptions to derive a normal $N(\mu, \sigma^2)$ prior density including estimates μ and σ^2 of the mean and variance. Other forms of prior can be re-parameterised in the form of a mean and variance (or precision); for example beta priors $\text{Be}(\alpha, \beta)$ for probabilities can be expressed as $\text{Be}(m\tau, (1 - m)\tau)$ where m is an estimate of the mean probability and τ is the estimated precision (degree of confidence in) that prior mean.

2. Materials and Method

2.1. Binomial Data

Consider a binary outcome variable T defined as;

$$Z_i = \begin{cases} 1 & \text{if the } i^{\text{th}} \text{ respondent voted for the incumbent} \\ 0 & \text{if the } i^{\text{th}} \text{ respondent voted for the Challenger} \end{cases}$$

Therefore in an opinion poll of size n where x respondents voted for the incumbent and $n-x$ for the challenge, the random variable $X = \sum_{i=1}^n Z_i$ has a binomial distribution with parameter (i.e. the probability that respondent i will vote for the incumbent) θ . The probability density function of X given θ is as below

$$p(X / \theta) = {}^n C_x \theta^x (1 - \theta)^{n-x} \quad x = 0, 1, 2, \dots, n$$

2.2. The Sequential Binomial Model

We can use a distribution to represent our prior knowledge and uncertainty regarding unknown parameter θ . An appropriate and a conjugate prior distribution for our unknown parameter θ is a beta distribution denoted by $\text{Be}(\alpha, \beta)$. The probability density function of a beta distribution is:

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

where $\Gamma(\alpha)$ is the gamma function applied to α and $0 < \theta < 1$. The parameters α and β can be thought of as prior “successes” and “failures,” respectively.

This prior density can also be expressed using the proportionality sign as;

$$\pi(\theta) \propto \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

The prior mean $E(\theta) = \mu = \frac{\alpha}{\alpha + \beta}$ and variance $\text{var}(\theta) = \sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$

The posterior density of θ is a product of a prior and the

$$p(\theta / X_{(2)}) = P(X = x_{(2)} / \theta) p(\theta / X_{(1)}) = \text{Be}(\alpha + x_1 + x_2, \beta + n_1 + n_2 - x_1 - x_2)$$

Recursively, for the k^{th} sample we have the posterior for $X_{(k)}$ as

$$p(\theta / X_{(k)}) = P(X = x_{(k)} / \theta) p(\theta / X_{(k-1)}) = \text{Be}(\alpha + \sum_{i=1}^k x_i, \beta + \sum_{i=1}^k n_i - \sum_{i=1}^k x_i)$$

This gives a better estimate than the one obtained by just aggregating all the previous pre-election polls in a single prior.

The choice of α and β for our prior distribution depends on at least two factors: (1) the amount of information about the parameter θ available prior to this poll; (2) the amount of stock we want to put into this prior information. Contrary to the view that this is a limitation of Bayesian statistics, the incorporation of prior information can actually be an advantage and provides us considerable flexibility. If we have little or no prior information, or we want to put very little stock in the information we have, we can choose values for α and β that reduce the distribution to a uniform distribution. For instance, choosing $\alpha = 1$ and $\beta = 1$, we get $p(\theta / \alpha, \beta) \propto \theta^0 (1-\theta)^0 = 1$ which is proportional to a uniform distribution on the allowable interval for θ . That is, the prior distribution is flat, not producing greater a priori

likelihood. Implying a beta distribution with parameters $\alpha + x$ and $\beta + n - x$, specifically:

$$p(\theta / X) = P(X / \theta) \pi(\theta) = \text{Be}(\alpha + x, \beta + n - x)$$

$$p(\theta / X) \propto \theta^{\alpha+x-1} (1-\theta)^{\beta+n-x-1}$$

We shall denote this posterior by $p(\theta / X_{(1)})$

Now, if we observe another sample $X_{(2)}$ then the posterior becomes

weight for any value of θ over another. Thus, the prior distribution will have little effect on the posterior distribution. For this reason, this type of prior is called “noninformative.”

On the other hand, if we have considerable prior information that we wish to weigh heavily relative to the current data, large values of α and β are used. A little massage of the formula for the variance reveals that, as α and β increase, the variance decreases, which makes sense, because adding additional prior information ought to reduce our uncertainty about the parameter. Thus, adding more prior successes and failures (increasing both parameters) reduces prior uncertainty about the parameter of interest (θ).

Lastly, if we have considerable prior information that we do not wish to weigh heavily in the posterior distribution, moderate values of α and β are chosen that yield a mean that is consistent with the previous research but that also produce a variance around that mean that is broad.

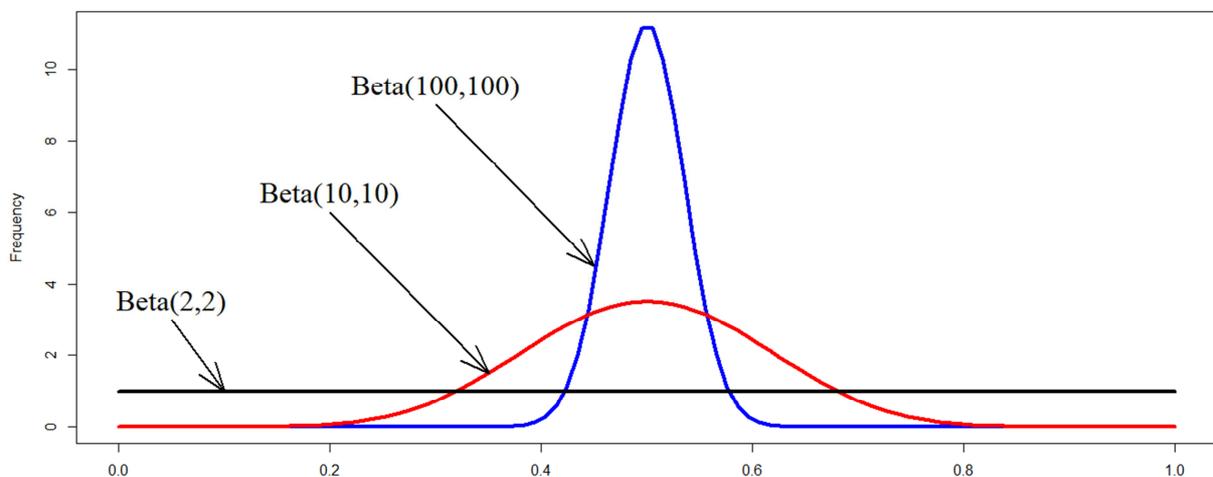


Figure 1. Various beta distributions with mean =0.5 for various choice of parameters.

In order to clarify these ideas, we illustrate using beta distributions plots with different values of α and β . All the three beta distributions, displayed in Figure 1, have a mean of 0.5; but different variances as a result of having α and β parameters of different magnitude. The most-peaked beta distribution has parameters $\alpha = \beta = 100$. The least-peaked distribution is almost flat—uniform—with parameters $\alpha = \beta = 2$. As with the binomial distribution, the beta distribution becomes skewed if α and β are unequal, but the basic idea is the same: the larger the parameters, the more prior information and the narrower the density

Throughout the fall of every general election year in Kenya, many pollsters conduct a number of polls attempting to predict whether candidate A or candidate B would win the presidential election. One of the hotly contested general election was the

2007 elections the battleground predominantly between the incumbent (here demoted as K) and the challenger (here denoted as R). The polls leading up to the election showed the two candidates claiming proportions of the votes that were statistically indistinguishable in the nation.

Figure 2 shows the prior, likelihood, and posterior densities. The likelihood function has been normalized as a proper density for θ , rather than X . Clearly the posterior density is a compromise between the prior distribution and the likelihood (current data). The posterior is between the prior distribution and the likelihood, but closer to the prior. The reason the posterior is closer to the prior is that the prior contained more information than the likelihood: There were 1,950 previously sampled persons and only 1,067 in the current sample.

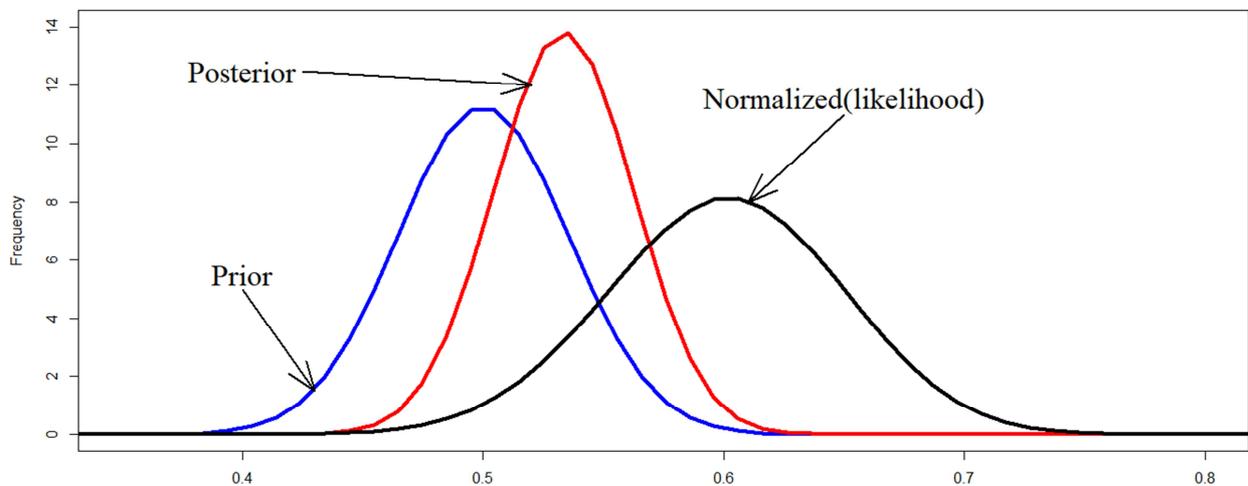


Figure 2. Prior, likelihood, and posterior for 2007 polling data for Kenya.

With the posterior density determined, we now can summarize our updated knowledge about θ the proportion of voters who will vote for incumbent, and answer our question of interest: What is the probability that the incumbent would win? A number of summaries are possible, given that we have a posterior distribution with a known form (a beta density). First, the mean of incumbent K is $1498/(1498 + 1519) = 0.497$, and the median is also 0.497. The variance of this beta distribution is .00008283 (standard deviation=.0091). If we assume that this beta distribution is approximately normal, then the approximate a 95% confidence interval of K is [0.479-0.515].

3. Simulation Results

In order to understand the concept of sequential Bayesian analysis, we will consider a case of two candidates (incumbent denoted by K and challenger denoted by R) with four different scenarios forming our simulation set ups.

Scenario one

To begin with, let us consider the case where the two candidates have roughly equal popularity proportions but with some observable fluctuations. The first 100 iterations

yielded results shown in Figure 3. Even after 20 iterations, the chain tends to the true value.

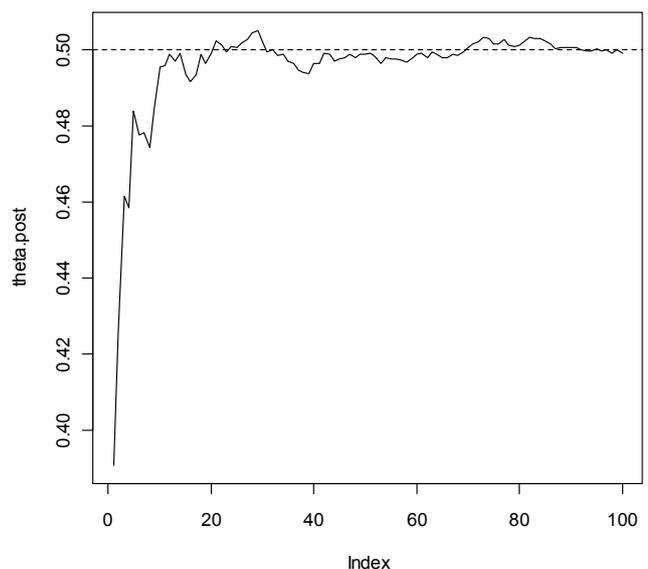


Figure 3. History plot for the first 100 iterations to demonstrate initiation of the chain.

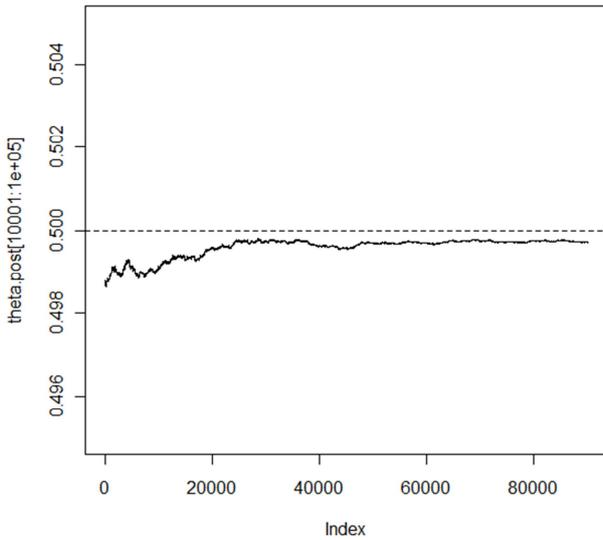


Figure 4. Trace plot after burn-in of 10,000 iterations predicting the success rate of the incumbent.

From the graph it's clear that the election will be tightly contested for instance the 2007 presidential elections in Kenya between K (incumbent) and R (challenger).

Scenario two

We now consider the case where the popularity of one candidate (say the challenger) is increasing implying that the popularity of the other is decreasing over time. As shown in Figure certainly the incumbent will win the presidential race as the posterior probability is well above 0.5.

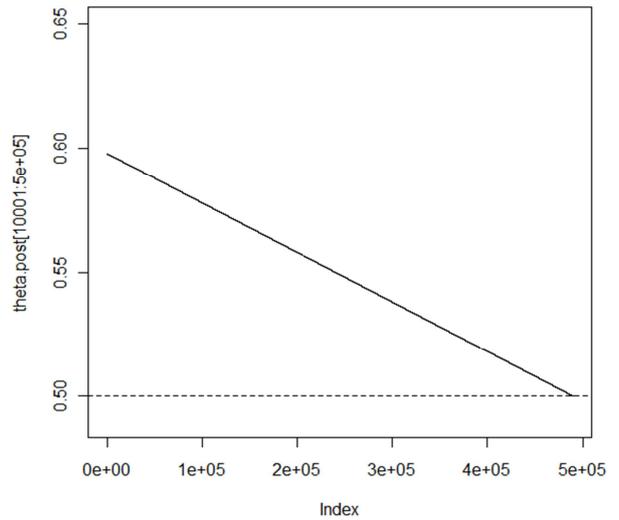


Figure 5. Trace plot after 10,000 burn-in where the probability of the incumbent is assumed over 0.5.

Scenario three

Thirdly, we consider the case where the popularity of one candidate (say the challenger) being constantly slightly higher but with misclassification in favour of the other (say the Incumbent). The misclassification for this scenario is as follows

- 1) No misclassification
- 2) Low misclassification: $p_{01}=.05$ $p_{10}=.10$
- 3) Misclassification: $p_{01}=.05$ $p_{10}=.15$
- 4) Misclassification: $p_{01}=.10$ $p_{10}=.10$

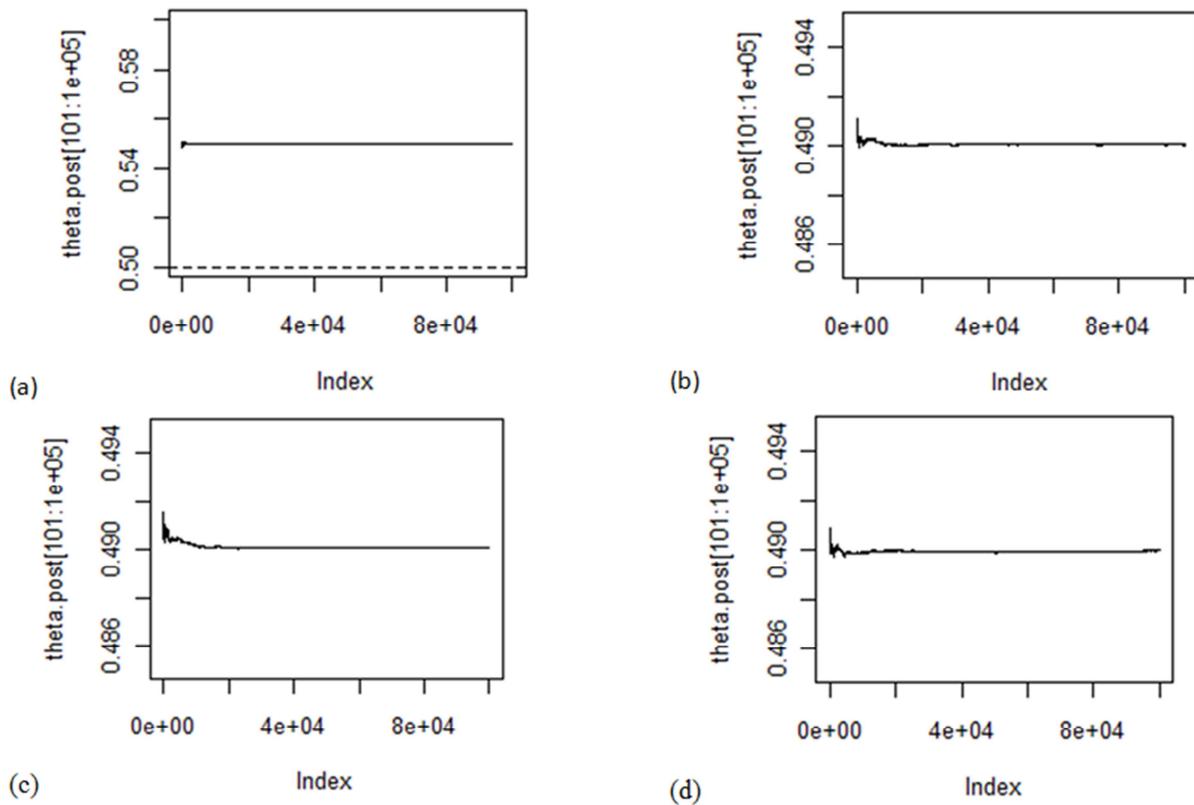


Figure 6. History plot showing the effect of misclassification (panels a-d).

The simulations results show that without the misclassification, panel (a), the challenger will win the election with a good margin. However, with misclassification, panels (b) to (d), the challenger will narrowly lose the election as his popularity eventually stabilizes around 0.491.

4. Conclusions and Recommendations

In this paper, we have developed the basis of the Bayesian approach to statistical inference. Bayesian approach handles various scenario in the fall projection including the aspect of misclassification. Further, even where data are scanty, it incorporates prior distribution to express the model uncertainty.

In this work, we have provided a flexible way of comparing the two leading candidates since in most election there is always two candidates who lead the pack. Our approach, though applied to Kenyan opinion polls, can be applied anyway in the world. This work can be extended to the case of multiple candidates.

Acknowledgements

We acknowledge the Kenyan media houses and opinion pollsters who were the making sources of these Kenyan opinion polls data.

References

- [1] L. Bursztyrn, D. Cantoni, P. Funk, and N. Yuchtman, "Polls, the Press, and Political Participation: The Effects of Anticipated Election Closeness on Voter Turnout," National Bureau of Economic Research, 2017, doi: 10.3386/w23490.
- [2] W. F. Christensen and L. W. Florence, "Predicting presidential and other multistage election outcomes using state-level pre-election polls," *American Statistician*, 2008, doi: 10.1198/000313008X267820.
- [3] H. Keun Lee and Y. Woon Kim, "Public opinion by a poll process: Model study and Bayesian view," *Journal of Statistical Mechanics: Theory and Experiment*, 2018, doi: 10.1088/1742-5468/aabbc5.
- [4] L. F. Stoetzer, M. Neunhoeffler, T. Gschwend, S. Munzert, and S. Sternberg, "Forecasting Elections in Multiparty Systems: A Bayesian Approach Combining Polls and Fundamentals," *Political Analysis*, 2019, doi: 10.1017/pan.2018.49.
- [5] D. P. CHRISTENSON and C. D. SMIDT, "Polls and Elections: Still Part of the Conversation: Iowa and New Hampshire's Say within the Invisible Primary," *Presidential Studies Quarterly*, 2012, doi: 10.1111/j.1741-5705.2012.03994.x.
- [6] P. Selb and S. Munzert, "Forecasting the 2013 german bundestag election using many polls and historical election results," *German Politics*, 2016, doi: 10.1080/09644008.2015.1121454.
- [7] P. Selb and S. Munzert, "Forecasting the 2013 Bundestag Election Using Data from Various Polls," *SSRN Electronic Journal*, 2013, doi: 10.2139/ssrn.2313845.
- [8] R. McDonald and X. Mao, "Forecasting the 2015 General Election with Internet Big Data: An Application of the TRUST Framework," Working Papers, 2015.
- [9] C. Spike and P. Vernon, *US election analysis 2016 : Media, voters and the campaign*. 2016.
- [10] P. C. Ordeshook and T. R. Palfrey, "Agendas, Strategic Voting, and Signaling with Incomplete Information," *American Journal of Political Science*, 1988, doi: 10.2307/2111131.
- [11] M. Haspel and H. Gibbs Knotts, "Location, location, location: Precinct placement and the costs of voting," *Journal of Politics*, 2005, doi: 10.1111/j.1468-2508.2005.00329.x.
- [12] M. Henn and N. Foard, "Social differentiation in young people's political participation: The impact of social and educational factors on youth political engagement in Britain," *Journal of Youth Studies*, 2014, doi: 10.1080/13676261.2013.830704.
- [13] M. Henn, M. Weinstein, and S. Forrest, "Uninterested youth? Young people's attitudes towards party politics in Britain," *Political Studies*, 2005, doi: 10.1111/j.1467-9248.2005.00544.x.
- [14] A. M. Williams, C. Jephcote, H. Janta, and G. Li, "The migration intentions of young adults in Europe: A comparative, multilevel analysis," 2018, doi: 10.1002/psp.2123.
- [15] J. D. Byers, "Is political popularity a random walk?" *Applied Economics*, 1991, doi: 10.1080/00036849100000045.
- [16] D. Byers, J. Davidson, and D. Peel, "Modelling political popularity: An analysis of long-range dependence in opinion poll series," *Journal of the Royal Statistical Society. Series A: Statistics in Society*, 1997, doi: 10.1111/j.1467-985X.1997.00075.x.
- [17] J. B. Carlin, "A case study on the choice, interpretation and checking of multilevel models for longitudinal binary outcomes," *Biostatistics*, 2001, doi: 10.1093/biostatistics/2.4.397.
- [18] A. Gelman, *Bayesian data analysis* Gelman. 2013.
- [19] X. Puig and J. Ginebra, "A Bayesian cluster analysis of election results," *Journal of Applied Statistics*, 2014, doi: 10.1080/02664763.2013.830088.
- [20] K. Lock and A. Gelman, "Bayesian combination of state polls and election forecasts," *Political Analysis*, 2010, doi: 10.1093/pan/mpq002.
- [21] J. Mwanyekange, S. M. Mwalili, and O. Ngesa, "Bayesian Joint Models for Longitudinal and Multi-state Survival Data," *International Journal of Statistics and Probability*, 2019, doi: 10.5539/ijsp.v8n2p34.
- [22] P. Selb and S. Munzert, "Estimating constituency preferences from sparse survey data using auxiliary geographic information," *Political Analysis*, 2011, doi: 10.1093/pan/mpr034.
- [23] J. O. Berger, "Robust Bayesian analysis: sensitivity to the prior," *Journal of Statistical Planning and Inference*, 1990, doi: 10.1016/0378-3758(90)90079-A.

- [24] M. Ghosh and J. Berger, "Stastical Decision Theory and Bayesian Analysis.," Journal of the American Statistical Association, 1988, doi: 10.2307/2288950.
- [25] J. O. Berger, "Bayesian Analysis: A Look at Today and Thoughts of Tomorrow," Journal of the American Statistical Association, 2000, doi: 10.1080/01621459.2000.10474328.
- [26] J. O. Berger, "Bayesian analysis: A look at today and thoughts of tomorrow," in Statistics in the 21st Century, 2001.
- [27] D. Barber, Bayesian Reasoning and Machine Learning. 2011.