A Markov chain Monte Carlo (MCMC) methodology with bootstrap percentile estimates for predicting presidential election results in Ghana

Ezekiel N. N. Nortey, Theophilus Ansah-Narh, Richard Asah-Asante and Richard Minkah

Abstract

Although, there exists numerous literature on the procedure for forecasting or predicting election results, in Ghana only opinion poll strategies have been used. To fill this gap, the paper develops Markov chain models for forecasting the 2016 presidential election results at the Regional, Zonal (i.e. Savannah, Coastal and Forest) and the National levels using past presidential election results of Ghana. The methodology develops a model for prediction of the 2016 presidential election results in Ghana using the Markov chains Monte Carlo (MCMC) methodology with bootstrap estimates. The results were that the ruling NDC may marginally win the 2016 Presidential Elections but would not obtain the more than 50% votes to be declared an outright winner. This means that there is going to be a run-off election between the two giant political parties: the ruling NDC and the major opposition party, NPP. The prediction for the 2016 Presidential run-off election between the NDC and the NPP was rather in favour of the major opposition party, the NPP with a little over the 50% votes obtained.

Keywords: Markov chains, Stochastic matrix, Elections, Forecasting, NDC, NPP, Ghana

Background

The prime concern for any political party is to map up strategies that would aid them to win an election particularly, the presidential election. This is of key interest to political analysts and the mass media as they would like to discuss and compare parties’ campaign strategies. There is the need therefore to study these political strategies and come up with a mathematical model to predict future elections. Most researchers (Wang et al. 2014; Boon 2012; Campbell and Lewis-Beck 2008) have published papers on election forecasting using opinion polls but not on Markov chain Monte Carlo (MCMC) approach. This research is motivated in introducing this statistical technique to predict the election results in Ghana.

Elections in Ghana can be classified as a random process and similar to the incremental methods, the knowledge of outcomes of previous elections can be used for predictions of future elections. In probability theory, Markov chains are an important type of processes used to study experiments in which the outcomes can be affected by the
outcomes of all previous experiments. What is more important about Markov chains is that the outcome of an experiment depends only on the previous experiment. The Ghana Presidential elections from the fourth republic often appear to “flip-flop” after two terms (i.e. a National Democratic Congress (NDC) candidate will win two terms and a National Patriotic Party (NPP) candidate will win the next two terms). MCs should therefore be a useful tool for predicting election results. However, the large literature on methods of predicting election results does not include Markov chain (MC) models in Ghana. One can find the studies on the US presidential elections and the British elections using Markov chains (see for example Wagner 2012; Certin and Bentli 2013).

This paper uses Markov chains generated from previous election data to predict the 2016 presidential elections in Ghana. Confidence intervals for these predictions are obtained from bootstrap percentiles.

Electoral history of Ghana
The country Ghana which was formerly called the Gold Coast came into existence after so many years of being under the British colony and German-Togo land territory. In 1957, Ghana gained independence under the leadership of Osagyefo Dr. Kwame Nkrumah and became the first West African country to have won freedom from its colonial masters. For over a decade, in 1966–69, 1972–79 and 1981–92 respectively (Asante and Gyimah-Boadi 2004) there had been numerous coup d’états which had affected the socio-economic processes of the new born country Ghana.

When Ft. Lt. Jerry John Rawlings took over power in 1981 (Rothschild 1985), he banned political parties until 1992 (Handley 2008) when he lifted the ban and restored the country Ghana to multiparty democracy and also introduced a new constitution. He later formed a new party called the National Democratic Congress (NDC) and was voted into power in 1992 and 1996 elections (Bimpong-Buta 2005).

After his 2nd term, a new opposition party by then known as the National Patriotic Party (NPP) was formed under the Dankwa-Busia tradition (Ayee 2009) and led by John Agyekum Kuffour also won for two terms, in 2000 and 2004 elections.

The NDC again is in its 2nd term (i.e. 2008-date) for the 2nd time and is currently led by John Dramani Mahama.

Since the introduction of the new constitution by Rawlings in 1992, voting patterns have been swindling and that’s why it is of key interest to researchers, political analysts and mass media as a whole, to find answers to why this phenomenon.

Ghana as displayed in Fig. 1 is spatially divided into three ecological zones, namely: the Savannah belt that consists of the Northern, Upper East and Upper West regions; the Forest or Middle belt consisting of Ashanti, Brong Ahafo and Eastern regions with the largest representation of the Akans and finally the Coastal belt which consists of the Western, Central, Greater Accra and Volta regions. It is believed that voting is actually characterized by ethnic sentiments and thus the study would want to find out if predicted results of the 2016 elections really follow that assertion.
Markov chains

Let $X = \{X_0, X_1, \ldots\}$ be a sequence of random variables taking values in some countable set $S = \{s_1, s_2, \ldots\}$ referred to as state space. The sequence $\{X_0, X_1, \ldots\}$ is called a Markov chain if

$$P(X_k = j | X_0 = x_0, \ldots, X_{k-1} = i) = P(X_k = j | X_{k-1} = i)$$

(1)

for all $k \geq 1$ and $x_0, \ldots, i, j$ in $S$. In addition, if

$$P(X_k = j | X_{k-1} = i) = p_{ij}$$

(2)

then the Markov chain is homogeneous. Here, $p_{ij}$ in Eq. (2) is referred to as the matrix of transition probabilities and it satisfies the following conditions:

$$0 \leq p_{ij} \leq 1$$

(3)

and

$$\sum_j p_{ij} = 1$$

(4)

Each transition is called a step. Any matrix satisfying Eqs. (2), (3) and (4) is referred to as a stochastic matrix. In addition if $\sum_i p_{ij} = 1$ then it is called a doubly stochastic matrix.

The first-order difference equation of a MC is expressed as

$$\phi_{r+1} = P\phi_r, \quad r = 1, 2, \ldots, m$$

(5)
where \( P \) is an \( m \times m \) square matrix.

**Theorem 1**  Let \( P \) be a matrix of transition probabilities of a Markov chain. The \( ij \)th element \( p^n_{ij} \) of the matrix \( P^n \) is the given probability that the Markov chain starting in state \( s_i \) will transition to state \( s_j \) after \( n \) steps.

If \( p_{ij} \) is regular, then there is a unique vector \( \phi_r \) such that, for any probability vector \( \phi_0 \) and for large values of \( r \),

\[
\lim_{r \to \infty} \phi_{r+1} = P^r \phi_0.
\]  

Here the vector \( \phi_r \) in Eq. (6) is called equilibrium or an ergodic vector of the MC. Therefore, we can compute probability vectors given that the transition matrix and the original probability vector are known (Lay 2011; Lial et al. 2012).

**Methodology**

In Ghana, the Presidential election results are determined by the Electoral Commission (EC) and the elections are carried out at various constituencies in each Region. In this paper, the Upper East, Upper West and Northern regions form the Savannah Zone; Brong-Ahafo, Ashanti and Eastern regions form the Forest Zone; and Western, Central, Greater Accra and Volta regions form the Coastal Zone. In the Ghana Presidential elections, each candidate receives a certain number of votes and the candidate with more than 50% of the total valid votes casted wins the presidential election in Ghana. Otherwise, a run-off election is organized for the two topmost candidates.

We used the 1992–2008 Presidential election results to generate a stochastic matrix and the 2012 Presidential results as the probability vector to predict the 2016 Presidential election results. Following the methodology of Wagner (2012), the transition probability matrices are created from the previous election results as depicted in Table 1.

We let \( \phi_i, i = 1, 2, ..., 8 \) represent the presidential election results for 1992, 1996,..., 2012. Thus, we have:

| No. | Year | NDC | NPP | Other | Rejected votes |
|-----|------|-----|-----|-------|----------------|
| 1   | 1992 | 58.4| 30.3| 11.30 | 0b             |
| 2   | 1996 | 57.4| 39.7| 1.37  | 1.53           |
| 3   | 2000 | 44.5| 48.17| 5.53 | 1.80           |
| 4   | 2000a| 43.10| 56.90| 0    | 0              |
| 5   | 2004 | 44.64| 52.45| 0.78 | 2.13           |
| 6   | 2008 | 47.92| 49.13| 0.55 | 2.4            |
| 7   | 2008a| 50.23| 49.77| 0    | 0              |
| 8   | 2012 | 50.70| 47.74| 1.21 | 0.35           |

Source: Ghana electoral commission certified results

*a* Indicates run-off votes

*b* Indicate a very negligible proportion close to zero
The stochastic matrix for the model is thus obtained by averaging the transformation of the previous election results. This is the so-called Average Transformation Method (ATM) of Wagner (2012). Let \( L_i \), \( i = 1, \ldots, 7 \) be the transformation matrix from \( i \)th to the \((i + 1)\)th election results such that \( L_i \phi_i = \phi_{i+1} \). For instance, \( L_1 \) is the transformation matrix of the Presidential Elections results from 1992 to 1996 is given by

\[
L_1 = \begin{bmatrix}
NDC & NPP & O & R \\
 l_{11} & l_{12} & l_{13} & l_{14} \\
 l_{21} & l_{22} & l_{23} & l_{24} \\
 l_{31} & l_{32} & l_{33} & l_{34} \\
 l_{41} & l_{42} & l_{43} & l_{44} \\
\end{bmatrix}
\]

where, \( O \) and \( R \) are Other parties and Rejected votes respectively. Here, \( L_1 \) is unknown but the probability vectors for the 1992 and 1996 elections are known and hence from Eq. (5), we have

\[
\begin{bmatrix}
l_{11} & l_{12} & l_{13} & l_{14} \\
l_{21} & l_{22} & l_{23} & l_{24} \\
l_{31} & l_{32} & l_{33} & l_{34} \\
l_{41} & l_{42} & l_{43} & l_{44} \\
\end{bmatrix} \begin{bmatrix}0.5840 \\ 0.3030 \\ 0.1130 \\ 0 \end{bmatrix} = \begin{bmatrix}0.5740 \\ 0.3970 \\ 0.0137 \\ 0.0153 \end{bmatrix}
\]

where, \( l_{11} \) is the percentage of people who voted for NDC in the 1992 PE that also voted for the same party in the 1996 PE. Similar explanations holds for \( l_{ij} \) for \( i, j = 1, 2, 3, 4 \).

For the use of MC analysis, the following assumptions were made:

1. Everyone who voted in the preceding election year voted in the following election year.
2. There is an equal probability for voting for another party in the following election year provided you did not vote for these parties in the preceding election year.
3. Other parties which did not take part in run-off elections were recorded zero.
4. There is no rejected votes in all run-off elections

Based on the first assumption,

\[
l_{11} = \frac{0.5740}{0.5840} = 0.9829,
\]

and

\[
l_{12} = l_{13} = l_{14} = 0.0057.
\]
Similarly, the percentage of other political parties \( l_{33} = 0.0137/0.01370.1130.0.1130 = 0.1212 \) and \( l_{31} = l_{32} = l_{34} = 0.2929 \). However, the percentage of NPP votes increased from 1992 to 1996, so we have \( l_{22} = 1 \) and \( l_{21} = l_{23} = l_{24} = 0 \).

In addition \( l_{44} = 1 \) and \( l_{41} = l_{42} = l_{43} = 0 \). Therefore, as specified in Eq. (8), we have:

\[
L_1 = \begin{pmatrix} 0.9829 & 0.0057 & 0.0057 & 0.0057 \\ 0 & 1 & 0 & 0 \\ 0.2929 & 0.2929 & 0.1212 & 0.2929 \\ 0 & 0 & 0 & 1 \end{pmatrix}
\]

The same procedure is followed to obtain the other transformation matrices \( L_2, \ldots, L_7 \).

The average of the transition matrices are obtained as \( P = \frac{1}{7} \sum L_i \).

Using the steady state property of Eq. (5), we obtain the following results as shown in Table 2.

Since no candidate is expected to obtain more than 50% in the 2016 Presidential votes by the model results: there will be no clear winner in the 2016 first round elections. Hence, a run-off vote between the two dominant parties i.e. the NDC and the NPP.

To model this, we follow assumptions 3 and 4 to modify Table 1 as follows:

Applying the procedure to the generated observations in Table 3 yields the predicted values as shown in Table 4. Figures 2 and 3 display respectively, the regional and ecological zone forecasts 2016 Presidential Election with Bootstrap estimates.

### Table 2 Predicted 2016 presidential elections with bootstrap standard errors

| Year | NDC | NPP | Other | Rejected |
|------|-----|-----|-------|----------|
| Percentage | 48.70 | 47.80 | 1.80 | 1.60 |
| SE | 4.80 | 6.4 | 0.26 | 0.80 |

Source: author’s computation
Similarly the same methodology was applied to the regional and ecological Presidential Election results to predict the run-off results in 2016. The results are as shown below:

Table 5 shows the model's predictions for the regional presidential election results for the 2016 presidential elections. The results show that the NDC is the popular choice of voters in the Western (50.64 %), Greater Accra (50.64 %), Volta (82.8 %), Northern (57.5 %), Upper East (65.09 %), and Upper West (63.35 %) whereas the NPP is popular in the Eastern (50.6 %) and Ashanti (70.58 %) regions.

The prediction of the Ecological zone presidential election results are presented in Table 6. The NDC has over 50 % of valid votes from the Savannah and Coastal belts whereas the NPP, their closest opponents remain the toast of the forest belt.
Table 5  Forecasted regional presidential election results for 2016

| Region         | NDC          | NPP          | Other | Rejected votes |
|----------------|--------------|--------------|-------|----------------|
|                | Main (%) Run-off (%) | Main (%) Run-off (%) | Main (%) | Main (%) Run-off (%) |
| National       | 48.70 48.30 | 47.80 51.70 | 1.80 1.60 | – |
| Western        | 51.62 53.27 | 43.81 46.73 | 2.21 2.36 | – |
| Central        | 48.95 50.95 | 45.44 49.05 | 2.55 3.08 | – |
| Greater Accra  | 50.64 52.44 | 45.73 47.56 | 1.73 1.90 | – |
| Volta          | 82.80 83.77 | 13.91 16.23 | 1.33 1.96 | – |
| Eastern        | 40.36 41.21 | 50.60 58.79 | 1.47 1.56 | – |
| Ashanti        | 26.66 27.63 | 70.58 72.37 | 1.29 1.48 | – |
| Brong Ahafo    | 48.35 49.38 | 48.58 50.62 | 1.56 1.51 | – |
| Northern       | 57.50 60.00 | 37.84 40.00 | 2.14 2.52 | – |
| Upper East     | 65.09 68.93 | 27.68 31.07 | 3.84 3.39 | – |
| Upper West     | 63.35 63.46 | 30.49 35.54 | 2.82 3.35 | – |

Source: author’s computation
– Rejected votes are not factored into the computations of the probabilities

Table 6  Forecasted ecological zone presidential election results for 2016

| Ecological zone | NDC          | NPP          | Other | Rejected votes |
|-----------------|--------------|--------------|-------|----------------|
|                 | Main election (%) Round-off (%) | Main election (%) Round-off (%) | Main election (%) | Main election (%) Round-off (%) |
| National        | 49.72 48.30 | 47.52 51.70 | 1.80 0.96 | – |
| Savannah        | 54.72 62.01 | 30.99 37.99 | 2.72 11.57 | – |
| Forest zone     | 34.51 38.42 | 61.47 61.58 | 1.96 2.06 | – |
| Coastal zone    | 52.36 52.66 | 39.83 47.34 | 3.89 3.95 | – |

Source: author’s computation

Forecasting the regional presidential votes for 2016

Western region

\[
\begin{bmatrix}
0.93350 & 0.02220 & 0.02220 \\
0.01330 & 0.96010 & 0.01330 \\
0.15500 & 0.15500 \times 0.53490 & 0.15500 \\
0.16670 & 0.16670 \times 0.16670 & 0.5000
\end{bmatrix}
\]

which equals [0.5162 0.4381 0.0221 0.0236]

Central region

\[
\begin{bmatrix}
0.9201 & 0.0266 & 0.0266 & 0.0266 \\
0.0136 & 0.9592 & 0.0136 & 0.0136 \\
0.1111 & 0.1111 & 0.6667 & 0.1111 \\
0.1667 & 0.1667 & 0.1667 & 0.5000
\end{bmatrix}
\]

which equals [0.4895 0.4544 0.0252 0.0308]
### Greater Accra region

\[
\begin{bmatrix}
0.5211 & 0.4711 & 0.0017 & 0.0061
\end{bmatrix}
\]

which equals \([0.5064\, 0.4573\, 0.0173\, 0.0190]\)

### Volta region

\[
\begin{bmatrix}
0.8446 & 0.1292 & 0.0041 & 0.0221
\end{bmatrix}
\]

which equals \([0.4036\, 0.5660\, 0.0147\, 0.0156]\)

### Eastern region

\[
\begin{bmatrix}
0.4261 & 0.5630 & 0.0020 & 0.0089
\end{bmatrix}
\]

which equals \([0.8280\, 0.1391\, 0.0133\, 0.0196]\)

### Ashanti region

\[
\begin{bmatrix}
0.2835 & 0.7080 & 0.0012 & 0.0073
\end{bmatrix}
\]

which equals \([0.2666\, 0.7058\, 0.0129\, 0.0148]\)

### Brong Ahafo region

\[
\begin{bmatrix}
0.5074 & 0.4900 & 0.0026 & 0.0000
\end{bmatrix}
\]

which equals \([0.4835\, 0.4858\, 0.0156\, 0.0151]\)
Northern region

\[
\begin{bmatrix}
0.9665 & 0.0112 & 0.0112 & 0.0112 \\
0.0202 & 0.9395 & 0.0202 & 0.0202 \\
0.1650 & 0.1650 & 0.5050 & 0.1650 \\
0.1667 & 0.1667 & 0.1667 & 0.5000
\end{bmatrix}
\]
which equals \([0.5750 \ 0.3784 \ 0.0214 \ 0.0252]\)

Upper east

\[
\begin{bmatrix}
0.9496 & 0.0168 & 0.0168 & 0.0168 \\
0.0404 & 0.8788 & 0.0404 & 0.0404 \\
0.1698 & 0.1698 & 0.4905 & 0.1698 \\
0.2173 & 0.2173 & 0.2173 & 0.3481
\end{bmatrix}
\]
which equals \([0.6509 \ 0.2768 \ 0.0384 \ 0.0339]\)

Upper West

\[
\begin{bmatrix}
0.9495 & 0.0168 & 0.0168 & 0.0168 \\
0.0085 & 0.9745 & 0.0085 & 0.0085 \\
0.1759 & 0.1759 & 0.4722 & 0.1759 \\
0.1619 & 0.1619 & 0.1619 & 0.5142
\end{bmatrix}
\]
which equals \([0.6335 \ 0.3049 \ 0.0282 \ 0.0335]\)

Forecasting ecological zone for presidential votes

Savannah zone

\[
\begin{bmatrix}
0.9590 & 0.0137 & 0.0137 & 0.0137 \\
0.0232 & 0.9305 & 0.0232 & 0.0232 \\
0.1774 & 0.1774 & 0.4677 & 0.1774 \\
0.0562 & 0.0562 & 0.0562 & 0.8315
\end{bmatrix}
\]
which equals \([0.5472 \ 0.3099 \ 0.0272 \ 0.1158]\)

Forest zone

\[
\begin{bmatrix}
0.9023 & 0.0326 & 0.0326 & 0.0326 \\
0.0096 & 0.9711 & 0.0096 & 0.0096 \\
0.1709 & 0.1709 & 0.4872 & 0.1709 \\
0.1351 & 0.1351 & 0.1351 & 0.5948
\end{bmatrix}
\]
which equals \([0.3451 \ 0.6147 \ 0.0196 \ 0.0206]\)
Coastal zone

\[
\begin{bmatrix}
0.8704 & 0.0432 & 0.0432 & 0.0432 \\
0.0281 & 0.9158 & 0.0281 & 0.0281 \\
0.1355 & 0.1355 & 0.5936 & 0.1355 \\
0.1241 & 0.1241 & 0.1241 & 0.6276
\end{bmatrix}
\]

which equals \([0.5236 0.3983 0.0389 0.0391]\)

Conclusion

The model used in this study predicted the party that will win the 2016 PE with NDC having 49.72 %, NPP (47.52 %) and Other parties and Rejected votes having 1.8 and 0.96 %. The overall average error in this prediction was estimated as \(\approx 2.4\ %\). This was determined by finding the absolute percentage differences between the predicted and the actual results for previous elections.

It is evidently clear that both NPP and NDC have approximately 47 % of loyal voters who would always vote for these parties on any day and any time. Therefore with more education on how to reduce rejected votes, certainly would show a significant effect in the 2016 PE. Thus, the party that would channel lots of resources into voter education could sway the results in its favour.

A further study on this research is to also use other sophisticated mathematical models like Bayesian Estimation to compare the results of this method.

Authors' contributions

ENNN conceptualized and designed the methodology of the study and also acquired the data and was part of the team who analyzed the data. TA-N worked on the literature review, both theoretical and empirical, and was also involved in writing the R-codes for the analysis. RA-A wrote some parts of the discussion of the manuscript. RM wrote the additional R-codes for the prediction and bootstrap percentile estimates. All authors agree to be accountable for all aspects of the work and jointly own the work. All authors read and approved the final manuscript.

Author details

1 Department of Statistics, University of Ghana, Box LG 115, Legon, Ghana. 2 Ghana Space Science and Technology Institute (GSSTI), Ghana Atomic Energy Commission (GAEC), Box AE 1, Atomic Kwabenya, Ghana. 3 Department of Political Science, University of Ghana, Legon, Ghana.

Acknowledgements

The authors are grateful to the Ghana Electoral Commission (EC) for allowing them to use their data sets of Presidential Election results in Ghana.

Compliance with ethical guidelines

Competing interests

We, the authors hereby certify that there is no conflict of interest with any organisation regarding the material and the research discussed in the manuscript. The research is also not financed by any entity and so remains the sole work of the authors.

Received: 27 April 2015   Accepted: 4 September 2015

Published online: 18 September 2015

References

Asante R and Gyimah-Boadi E (2004) Ethnic structure, inequality and governance of the public sector in Ghana, United Nations Research Institute for Social Development (UNRISD), pp 58–70

Ayee JR (2009) “The evolution and development of the new patriotic party in Ghana”, SAI, Occasional Paper, No 19

Bimpong-Buta SY (2005) “The role of the supreme court in the development of constitutional law in Ghana”, Doctor of Law—LLD Thesis, University of South Africa

Boon M (2012) Predicting elections, a wisdom of crowds approach. Int J Mark Res 54(4):465–483

Campbell JE, Lewis-Beck MS (2008) US presidential election forecasting: an introduction. Int J Forecast 24:189–192
Certin N, Bentli I (2013) Application of Markov Process to forecast elections outcomes by computer simulation, Epoka conference systems, First international conference on management and economics, pp 104–110

Handley A (2008) “The world bank made me do it: international factors and Ghana’s transition to democracy”, CDDRL Working Papers, No 82, Standford University, CA

Lay DC (2011) Linear algebra and its applications, 4th edn. Addison Wesley, USA, pp 253–259, 360–365

Lial ML, Greenwell RN, Ritchey NP (2012) Finite mathematics, 10th edn. Pearson Education, Inc, New York, pp 453–458

Rothschild D (1985) “The Rawlings revolution in Ghana: pragmatism with populist rhetoric”, CSIS Africa Notes, No 42, pp 1–6

Wagner CS (2012) “U.S. Presidential Election Forecasts: through the lens of linear algebra. Retrieved on 11/03/2014 from http://home2.fvcc.edu/~dhicketh/LinearAlgebra/studentprojects/spring2012/Cassia_Linear%20Algebra%20Project/Final%20Project.pdf

Wang W, Rothschild D, Goel S, Andrew G (2014) “Forecasting elections with non-representative polls” International journal of forecasting—Elsevier BV