A Bayesian model for estimating election results in Mexico

Gian Carlo Di-Luvi

In Mexico, presidential and gubernatorial elections take place every six years. Mexico has a multi-party electoral system where the candidate that receives the largest amount of votes wins the election. Because election results take a week to be certified, a sample of polling stations is usually selected before the election and statistical estimates based on those polling stations are computed as the results become available on election day—a process called a *quick count.*

Multiple Bayesian models have been used in Mexico's quick counts. Mendoza and Nieto-Barajas [3] proposed a simple Bayesian model with a tractable posterior to estimate the proportion of votes received by each candidate; the model was successfully used to predict the winners of the 2006 and 2012 elections. Di-Luvi, Orantes-Jordan, and Mendoza [2] modified that same model and used it to predict the results of the 2018 election. Anzarut, González, and Ortiz [1] proposed a Bayesian multilevel regression model and used it in the 2018 election as well.

Data from the polling stations arrive as field personnel finish counting the votes. On election day, every 15 minutes new data are supplied and models are retrained. Thus, models should be rich enough to do accurate inference with small sample sizes while allowing for fast online inference. The model in [1] accounts for data missing not at random but is computationally expensive—for the presidential election, it took well over 15 minutes to run even when parallelising over 48 CPU cores. On the other hand, although the model proposed by [3] is easy to train, they rely on a tractable posterior distribution that limits its flexibility. Indeed, Di-Luvi, Orantes-Jordan, and Mendoza [2] modified the prior distribution but did not account for the corresponding change in the posterior.

In this work, we *(i)* propose a new model that accounts for the correlation structure between candidates—something not done before—and compare the results with other models, specifically in the setting with small sample sizes; and *(ii)* design and implement an MCMC procedure to sample from the correct posterior distribution in Di-Luvi, Orantes-Jordan, and Mendoza [2]'s model, and compare it with the "wrong" posterior originally used.

References

[1]   Michelle Anzarut, Luis Felipe González, and María Teresa Ortiz. "A Heavy-Tailed Multilevel Mixture Model for the Quick Count in the Mexican Elections of 2018". In: *National Statistics Forum (FNE) and Latin-American Congress of Statistical Societies (CLATSE)*. Springer. 2018, pp. 1–13.

[2]   Gian Carlo Di-Luvi, Gerardo Orantes-Jordan, and Manuel Mendoza. "Statistics in the 2018 Mexican General Election Quick Counts". In: *Laberintos & Infinitoss* 48 (2018), pp. 29–37.

[3]   Manuel Mendoza and Luis E Nieto-Barajas. "Quick Counts in the Mexican Presidential Elections: A Bayesian Approach". In: *Electoral Studies* 43 (2016), pp. 124–132.