# An efficient 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. Even though election results take around a week to be certified, in recent years there has been a widespread effort by government authorities to determine the winner of each election the day it takes place. For this purpose, 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 (2016) proposed simple Bayesian model 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 (2018) proposed some modifications to the model and used it to predict the winner of the 2018 election. Anzarut, González, and Ortiz (2018) proposed another (more complex) Bayesian model, and used it in the 2018 election as well.

The main idea behind all these Bayesian models is to calculate the posterior distribution of the proportion of votes that each candidate receives given the available voting data. Data from the polling stations arrive as field personnel finish counting the votes, which *(i)* depends on weather and geographical location and *(ii)* means that model learning should be done expeditiously and online. Point *(i)* means that models should be able to perform estimation in the presence of data that are missing not at random. The model in (Anzarut, González, and Ortiz, 2018) accounts for this fact but is computationally expensive—for the presidential election, it took well over 5 minutes to run. On the other hand, although the models proposed by (Mendoza and Nieto-Barajas, 2016; Di-Luvi, Orantes-Jordan, and Mendoza, 2018) are trivially fast to learn, they fail to account for missing patterns in the data. Furthermore, all models assume that the proportions of votes received by candidates are independent, which is not correct and can lead to bias in estimates.

Paragraph on proposed model. Dependence structure and missing data handling while being computationally efficient.

# 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.