

# 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. 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] proposed some modifications to the 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, new data are supplied and models are retrained every 5 minutes. 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 5 minutes to run. On the other hand, although the models proposed by [3, 2] are easy to train, they rely on a tractable posterior distribution that limits their flexibility. 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.

In this work, we propose a Bayesian model for quick counts in Mexico that *(i)* is flexible and accurate for small sample sizes and *(ii)* is fast to train. The number of votes for all candidates will be modeled with a multivariate Normal distribution, allowing for arbitrary correlations between candidates. The prior distribution for the proportion of votes in favor of each candidate is based on the additive logratio transform between the probabilistic simplex  $\Delta^{J-1}$ , where proportions of votes reside naturally, and  $\mathbb{R}^{J-1}$ . Furthermore, we leverage the STAN probabilistic programming language [4] to expeditiously generate samples from the posterior distribution.

## 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 & Infinitos* 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.
- [4] Stan Development Team. *RStan: the R interface to Stan*. R package version 2.21.2. 2020. URL: <http://mc-stan.org/>.