KOALA: Estimating coalition probabilities in multi-party electoral systems

Alexander Bauer[1], Andreas Bender[1], André Klima[1], Helmut Küchenhoff[1]

[1] *Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Germany*

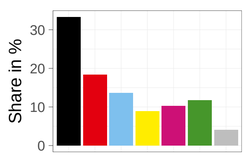
Alexander.Bauer@stat.uni-muenchen.de

Motivation

Election poll-based reporting

What's the status quo?

Typical election poll reporting:



- is based on observed mean voter shares
- sets the focus on individual party achievements
- imparts sample uncertainty only insufficiently

Example

Reporting on Union and FDP to jointly obtain a majority before the German federal election 2013

Last pre-election opinion poll: Source: Forsa, 20.09.2013

Union	SPD	Greens	FDP	The Left	AfD	Others
40%	26%	10%	5%	9%	4%	6%

After redistribution of party votes <5%
(i.e. the minimum hurdle to pass into German parliament)
Union-FDP jointly obtain exactly 50%.

Media headline:

"Union-FDP loses its majority"

Source: FAZ.net (2017). Umfrage zur Bundestagswahl: Schwarz-Gelb verliert die Mehrheit.http://archive.is/SuXVt. Accessed 26 April 2018.

Flaws of this type of reporting:

- Misleading conclusions are drawn
  A mean share of 50% only means that it's slightly more probable that a majority is missed
- Sample uncertainty is ignored
  E.g., with a mean voter share of 5%, FDP will only enter the parliament with ∼50%
- Redistribution of votes is ignored
  FAZ.net bases the conclusion on the observed voter share and not on the redistributed 50% share

What do we propose?

Good reporting:

- should impart findings in an easily graspable way
- should prevent potential misunderstandings
- should focus on the most relevant topics

We aim at **shifting the focus** from

Incomprehensive observed party shares to Uncertainty-based event probabilities

Foundations of KOALA-based reporting:

- Use event **probabilities** instead of voter shares
  Probabilities comprise sample uncertainty in a natural way and are less at risk to be misinterpreted
- Use **event** probabilities instead of voter shares
  Focusing on the main events allows the reader to easily grasp the big picture

KOALA headline:

"Union-FDP gains seat majority with 26%, FDP passes into parliament with 51%"*

* If the election was held today

① Event probability estimation

We use a **Multinomial-Dirichlet model** for the true party shares θ_j (Gelman et al., 2013):

$$(\theta_1, \ldots, \theta_k)^T \sim \mathit{Dirichlet}(\alpha_1, \ldots, \alpha_k), \quad \text{with} \quad \alpha_1 = \ldots = \alpha_k = \frac{1}{2}$$

Given one survey, the posterior also is a Dirichlet distribution with $\alpha_j = x_j + \frac{1}{2}$ for each party j and its observed vote counts x_j.

Using **Monte Carlo simulations** of election outcomes, we obtain obtain specific event probabilities by taking their relative frequency of occurence.

Pooling is used to summarize multiple polls to reduce sample uncertainty:

- We pool the most recent survey per polling agency within the past 14 days
- The summed number of votes per party are also multinomially distributed
- **But:** Polls from different polling agencies are correlated

Hence: We adjust the distribution by using an **effective sample size** (Hanley et al., 2003): Party-specific correlations were estimated based on 20 surveys of polling agencies Emnid and Forsa, using

$$\mathit{Cov}(X_{Aj}, X_{Bj}) = \frac{1}{2} \cdot \left(\mathit{Var}(X_{Aj}) + \mathit{Var}(X_{Bj}) - \mathit{Var}(X_{Aj} - X_{Bj})\right),$$

with

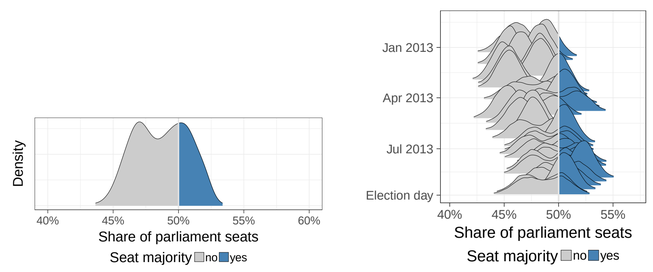
- $\mathit{Var}(X_{Aj})$, $\mathit{Var}(X_{Bj})$ the theoretical variances of binomial distributions,
- $\mathit{Var}(X_{Aj} - X_{Bj})$ estimated from the party share differences.

For simplicity, we set the correlation to a fixed value of 0.5. The effective sample size n_{eff} is then defined as the ratio between the estimated variance for the pooled sample and the theoretical variance for a sample of size one:

$$n_{\text{eff}} = \frac{\mathit{Var}(\text{pooled})}{\mathit{Var}(\text{sample of size one})}.$$

② Visualization

To visualize the development of such probabilities together with the underlying uncertainty for a specific coalition we use **ridgeline plots** (Wilke, 2017) for the simulated seat distributions:



Looking at the probabilities based on the last opinion poll before the German election 2013, the posterior distribution is bimodal, based on the distinction whether FDP and/or AfD pass the 5% hurdle. The resulting probability for a Union-FDP majority is 27.2%, based on 10,000 simulations.

③ Implementation and results communication

Results for selected elections are presented on koala.stat.uni-muenchen.de

The implementation is based on several points:

- Our approach is implemented in the R package coalitions
- The website is shiny-based
- The website update approach is automated
- Automatic tweets are sent in the case of new results
- For sharing our results we automatically export them to Google Sheets

References

Bender, A. and Bauer, A. (2018). coalitions: Coalition probabilities in multi-party democracies. *Journal of Open Source Software*, **3(23)**, 606, https://doi.org/10.21105/joss.00606.

Unser AStA-Paper no ois Technical Report?

Gelman, A. et al. (2013). *Bayesian Data Analysis, 3rd edition*. Boca Raton, FL: CRC press.

Hanley, J. A. et al. (2003). Statistical analysis of correlated data using gen- eralized estimating equations: an orientation. *American journal of epidemiology*, **157(4)**, 364–375.

Wilke C.O. (2017). *ggridges: Ridgeline Plots in 'ggplot2'*. R package version 0.4.1. URL https://CRAN.R-project.org/package=ggridges