KOALA: Estimating coalition probabilities in multi-party electoral systems

KOALA
KoalitionsAnalyse

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

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

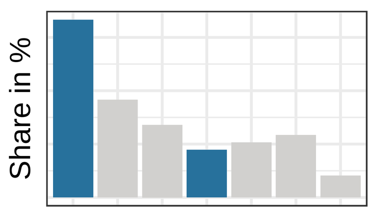
Alexander.Bauer@stat.uni-muenchen.de

Motivation

Election poll reporting

What's the status quo?

Typical election poll reporting:

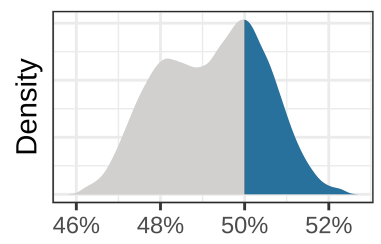


- is based on observed mean voter shares
- sets the focus on individual party achievements
- imparts sample uncertainty only insufficiently

Typical headline:
"The two parties jointly obtain 48% of all votes."

What do we propose?

Proposed type of reporting:



- focuses on specific events (e.g. potential majorities)
- naturally imparts sample uncertainty using probabilities
- prevents misunderstandings by using a holistic approach

Proposed headline:
"The two parties have a probability of 32% to jointly obtain a majority."

We aim to **shift the focus** from Incomprehensive reported party shares to Uncertainty-based probabilities of events (POEs)

Real-world 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 vote share to enter the 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 to miss a majority
- 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

Foundations of POE-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 for easily grasping the relevant information

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

* If the election was held today

Methods

Estimating POEs

1 **Multinomial-Dirichlet model** for the true party shares θ_p (Gelman et al., 2013):

$$(\theta_1, \dots, \theta_P)^T \sim Dirichlet(\alpha_1, \dots, \alpha_P), \quad \text{with} \quad \alpha_1 = \dots = \alpha_P = \frac{1}{2}$$

2 Given one survey, we obtain a **Dirichlet posterior** with $\alpha_p = x_p + \frac{1}{2}$ for each party $p = 1, \dots, P$ and its observed vote count x_p.

3 Using **Monte Carlo simulations** of election outcomes, we obtain specific POEs by calculating the events relative frequency of occurrence.

Pooling multiple polls

We aggregate the latest polls within a specific time window (e.g. 14 days) to reduce sample uncertainty. We adjust the uncertainty of the multinomially distributed summed number of votes per party by using an **effective sample size** (Hanley et al., 2003).

1 As polls from different polling agencies are correlated, **party-specific correlations** were estimated based on 20 surveys of polling agencies Emnid and Forsa, using

$$Cov(X_{Ap}, X_{Bp}) = \frac{1}{2} \cdot \left(Var(X_{Ap}) + Var(X_{Bp}) - Var(X_{Ap} - X_{Bp}) \right),$$

with

- X_{Ap}, X_{Bp} the observed vote counts for party p in surveys A and B,
- $Var(X_{Ap})$, $Var(X_{Bp})$ the theoretical variances of binomial distributions,
- $Var(X_{Ap} - X_{Bp})$ estimated from the party share differences.

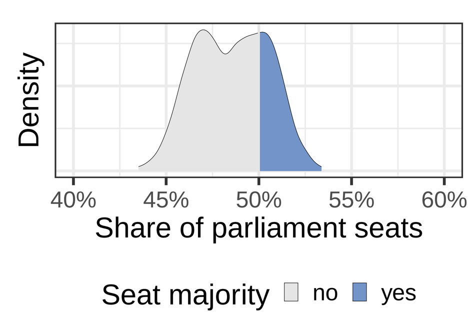
For simplicity, we set the correlation to a fixed value of 0.5.

2 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:

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

Visualization & Implementation

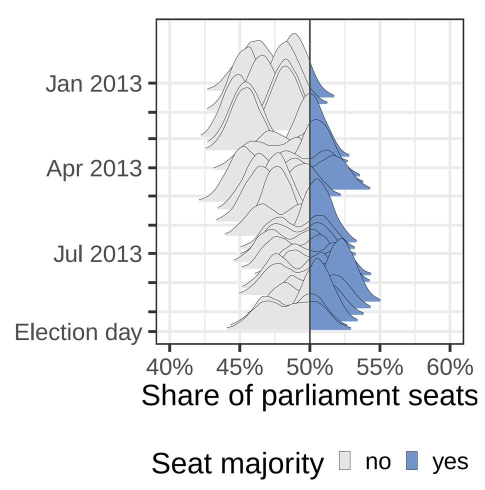
Selected visualizations



Density plots are used to **visualize POEs**, highlighting the area associated with simulations where the event of interest occurred.
Moreover, such plots highlight

- the uncertainty underlying the event of interest
- the range of possible outcomes

in a natural and intuitive way.



Ridgeline plots (Wilke, 2017) are used to depict the **development of POEs**, again visualizing the uncertainty underlying the event of interest in a natural way.

Implementation

KOALA
koala.stat.uni-muenchen.de

@KOALA_LMU

The **R package coalitions** (Bender and Bauer, 2018) includes all methods and allows for their application to any multi-party electoral system.

Our **dedicated website and Twitter channel** makes current POEs for selected elections accessible to the general public.

References

KOALA-Paper noch einfügen

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.

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