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

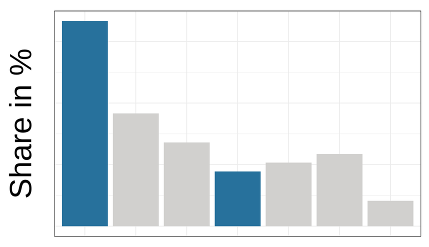
KoalitionsAnalyse

Motivation

Election poll-based reporting

What's the status quo?

Typical election poll reporting:

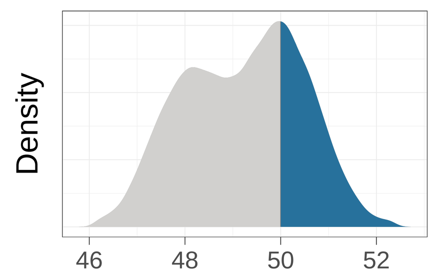


- is based on observed mean voter shares
- sets the focus on individual party achievements
- imparts sample uncertainty only insufficiently

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 this holistic approach

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

Real-world Example

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

We aim to **shift the focus** from Incompetensive observed party shares to Uncertainty-based event probabilities

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 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 relevant information

KOALA headline:

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

* If the election was held today

Event probability estimation

Estimating event probabilities

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

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

2 Given one survey, we obtain a **Dirichlet posterior** with $\alpha_j = x_j + \frac{1}{2}$ for each party $j = 1, \dots, k$ and its observed vote counts x_j.

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

Pooling multiple surveys

We pool the most recent surveys within the past 14 days (one per polling agency) 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

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

with

- X_{Sj} the observed votes for party j in survey S,
- $Var(X_{Aj})$, $Var(X_{Bj})$ the theoretical variances of binomial distributions,
- $Var(X_{Aj} - X_{Bj})$ estimated from the party share differences.

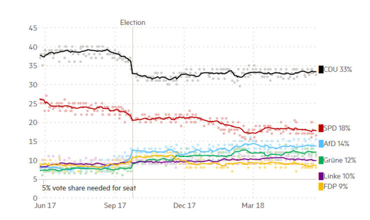
For simplicity, we set the correlation to a fixed value of 0.5.

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:

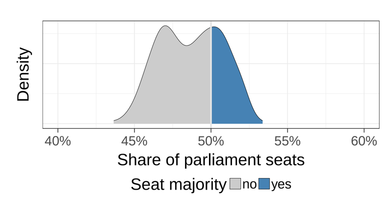
$$n_{\text{eff}} = \frac{Var(\text{pooled})}{Var(\text{sample of size one})}$$

Visualization & Implementation

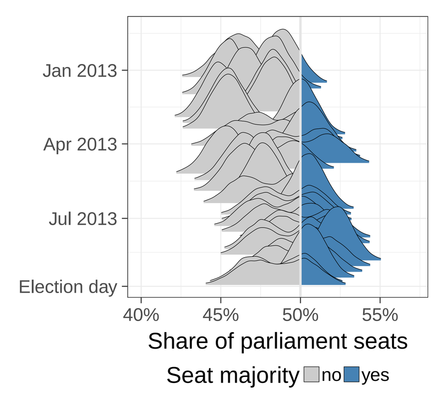
Selected visualizations



Adjusted line plots are used to visualize the pooled voter shares are visualized, showing both the mean share and the corresponding uncertainty



Density plots are used to depict one simulated seat distribution



Ridgeline plots (Wilke, 2017) are used to depict the simulated seat distribution development over time

Implementation

KOALA
koala.stat.uni-muenchen.de

@KOALA_LMU

Major building blocks

- The accompanying R package `coalitions`
- An automated fetch-and-update process for the website
- An automated bot tweeting new results

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.