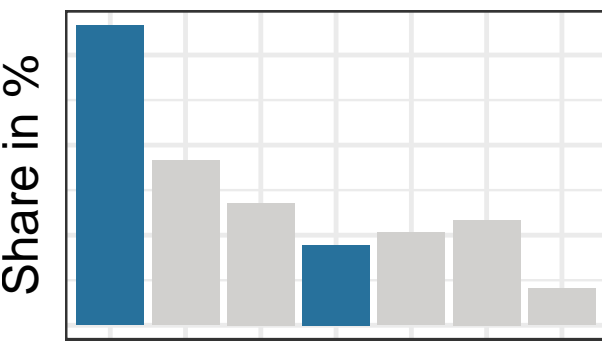


Motivation

Election poll reporting

What's the status quo?

Typical election poll reporting:

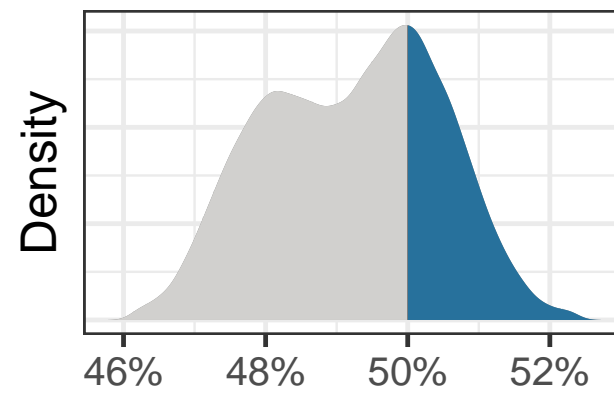


- is based on reported party shares
- sets the focus on individual party achievements
- imparts sample uncertainty only insufficiently

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%.

Typical media headline:  
"Union-FDP loses its majority"

Source: FAZ.net (2017). Umfrage zur Bundestagswahl: Schwarz-Gelb verliert die Mehrheit. <http://archive.is/l76o3>. Accessed 16 July 2018.

Major 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%

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

- Multinomial-Dirichlet model** for the true party shares  $\theta_p$  (Gelman et al., 2013):  
$$(\theta_1, \dots, \theta_P)^T \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_P), \text{ with } \alpha_1 = \dots = \alpha_P = \frac{1}{2}$$
- 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  $x_p$ .
- Using **Monte Carlo simulations** of election outcomes, we obtain specific POEs by calculating the events relative frequency of occurrence.

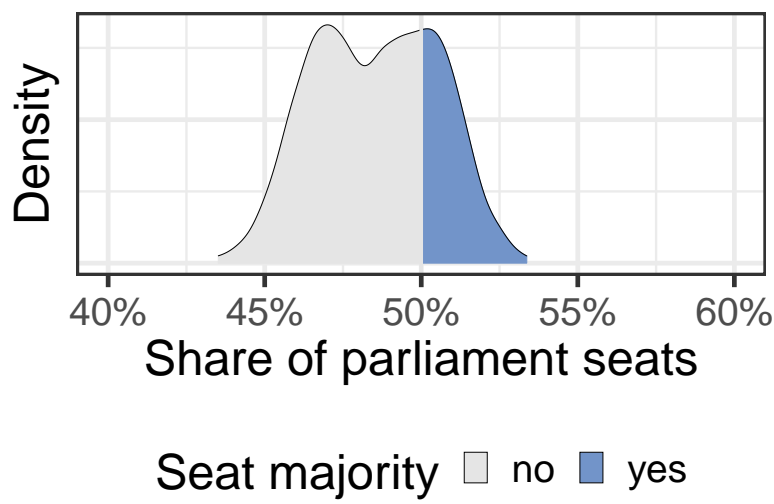
Pooling multiple polls

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).

- As polls from different polling agencies are correlated, **party-specific correlations** were estimated based on 20 surveys of polling agencies Emnid and Forsa, using  
$$\text{Cov}(X_{Ap}, X_{Bp}) = \frac{1}{2} \cdot (\text{Var}(X_{Ap}) + \text{Var}(X_{Bp}) - \text{Var}(X_{Ap} - X_{Bp})),$$
with
  - $X_{Ap}, X_{Bp}$  the observed vote counts for party  $p$  in surveys  $A$  and  $B$ ,
  - $\text{Var}(X_{Ap}), \text{Var}(X_{Bp})$  the theoretical variances of binomial distributions,
  - $\text{Var}(X_{Ap} - X_{Bp})$  estimated from the party share differences.For simplicity, we set the correlation to a fixed value of 0.5.
- The **effective sample size**  $n_{\text{eff}}$  is then defined as the ratio between the estimated variance for the pooled sample and the theoretical variance for a sample of size one:  
$$n_{\text{eff}} = \frac{\text{Var}(\text{pooled})}{\text{Var}(\text{sample of size one})}$$

Visualization & Implementation

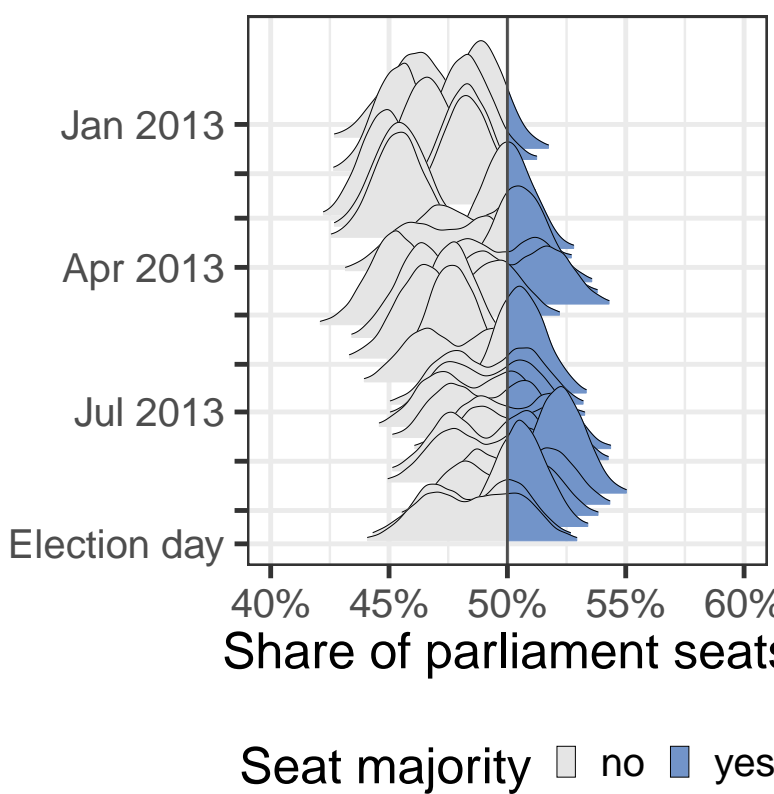
Selected visualizations



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** over time, again visualizing the uncertainty underlying the event of interest in a natural way.

Implementation

  
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

Bauer, A. et al. (2018). KOALA: A new paradigm for election coverage. *arXiv.org*, <http://arxiv.org/abs/1807.09665>

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 generalized 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=gggridges>

IWSM Bristol 2018