

KOALA: A new paradigm for election coverage

An opinion poll based “now-cast” of probabilities of events in multi-party electoral systems

Alexander Bauer

Statistical Consulting Unit StaBLab, LMU Munich

DAGStat | March 20, 2019 | Munich

KOALA

KoalitionsAnalyse

Alexander Bauer

Statistical Consulting Unit StaBLab, LMU Munich

DAGStat | March 20, 2019 | Munich

KOALA

KoalitionsAnalyse

Collaborators

Dr. Andreas Bender

Nuffield Department of Clinical Medicine,
University of Oxford, United Kingdom

Dr. André Klima

StaBLab, LMU Munich

Prof. Dr. Helmut Küchenhoff

StaBLab, LMU Munich

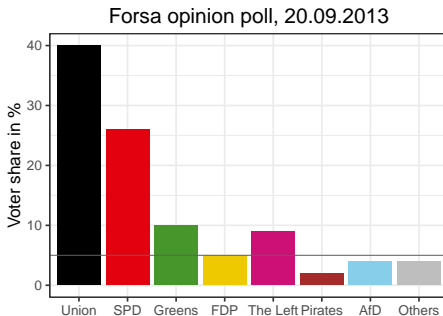
Outline

1. Motivation
2. Methods
3. Technical implementation
4. Results & Outlook

Outline

1. Motivation
2. Methods
3. Technical implementation
4. Results & Outlook

1 Motivation



Questions of interest

- Which parties will pass the 5% hurdle and enter the parliament?
- Which parties will form the governing coalition?
- Which party will have the third largest share of votes?

① Motivation

Reported voter shares

Union	SPD	Greens	FDP	The Left	Pirates	AfD	Others
40%	26%	10%	5%	9%	2%	4%	5%

Redistributed voter shares (based on 5% hurdle)

Union	SPD	Greens	FDP	The Left	Pirates	AfD	Others
44.44%	28.89%	11.11%	5.56%	10.00%	-	-	-

- Union-FDP have a joint seat share of exactly 50%
- Stating that Union-FDP would thus miss a joint majority would neglect sample uncertainty

⇒ We calculate event probabilities that fully reflect sample uncertainty

① Motivation

Reported voter shares

Union	SPD	Greens	FDP	The Left	Pirates	AfD	Others
40%	26%	10%	5%	9%	2%	4%	5%

Redistributed voter shares (based on 5% hurdle)

Union	SPD	Greens	FDP	The Left	Pirates	AfD	Others
44.44%	28.89%	11.11%	5.56%	10.00%	-	-	-

- Union-FDP have a joint seat share of exactly 50%
- Stating that Union-FDP would thus miss a joint majority would neglect sample uncertainty

⇒ We calculate event probabilities that fully reflect sample uncertainty

We aim to do now-casting

- We incorporate the uncertainty as reported by the polling agencies
- Potential house biases or an industry bias are not accounted for

We do not aim to do for-casting

- Our approach simply communicates sample uncertainty in a novel way
- Also, a relevant share of voters is still undecided shortly before election day (Küchenhoff et al., 2018)

① Motivation

We aim to do now-casting

- We incorporate the uncertainty as reported by the polling agencies
- Potential house biases or an industry bias are not accounted for

We do not aim to do for-casting

- Our approach simply communicates sample uncertainty in a novel way
- Also, a relevant share of voters is still undecided shortly before election day (Küchenhoff et al., 2018)

Outline

1. Motivation
2. Methods
3. Technical implementation
4. Results & Outlook

② Methods

Estimating probabilities of events (POEs)

Given one opinion poll with sample size n :

$$\mathbf{X} = (X_1, \dots, X_P)^T \sim \text{Multinomial}(n, \theta_1, \dots, \theta_P),$$

with voter counts X_j and the true percentage of voters θ_j per party j
(assuming a simple random sample, ignoring a possible bias)

Using an **uninformative Dirichlet prior** (Gelman et al., 2013)

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_P)^T \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_P),$$

$$\text{with} \quad \alpha_1 = \dots = \alpha_P = \frac{1}{2},$$

the resulting posterior distribution of $\boldsymbol{\theta}|\mathbf{x}$ is again Dirichlet:

$$\boldsymbol{\theta}|\mathbf{x} \sim \text{Dirichlet}(x_1 + 1/2, \dots, x_P + 1/2).$$

② Methods

Estimating probabilities of events (POEs)

Given one opinion poll with sample size n :

$$\mathbf{X} = (X_1, \dots, X_P)^T \sim \text{Multinomial}(n, \theta_1, \dots, \theta_P),$$

with voter counts X_j and the true percentage of voters θ_j per party j
(assuming a simple random sample, ignoring a possible bias)

Using an **uninformative Dirichlet prior** (Gelman et al., 2013)

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_P)^T \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_P),$$

$$\text{with} \quad \alpha_1 = \dots = \alpha_P = \frac{1}{2},$$

the resulting posterior distribution of $\boldsymbol{\theta}|\mathbf{x}$ is again Dirichlet:

$$\boldsymbol{\theta}|\mathbf{x} \sim \text{Dirichlet}(x_1 + 1/2, \dots, x_P + 1/2).$$

② Methods

Estimating probabilities of events (POEs)

Given the **posterior distribution of voter shares** we can use **Monte Carlo simulations** to estimate POEs:

1. Simulate 10 000 election outcomes from the posterior
2. If necessary: Redistribute voter shares to get obtained seats in parliament
3. POE = Percentage of simulations where event of interest occurred

Example

Given the Forsa poll, the coalition of Union-FDP obtained a majority of seats in 2633 of 10 000 simulations

⇒ POE \approx 26%

② Methods

Estimating probabilities of events (POEs)

Given the **posterior distribution of voter shares** we can use **Monte Carlo simulations** to estimate POEs:

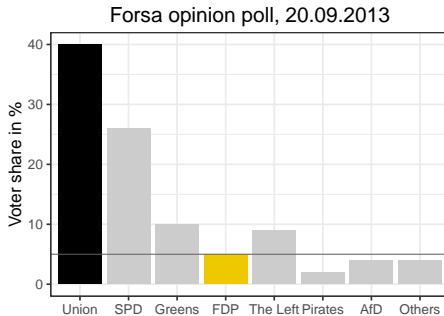
1. Simulate 10 000 election outcomes from the posterior
2. If necessary: Redistribute voter shares to get obtained seats in parliament
3. POE = Percentage of simulations where event of interest occurred

Example

Given the Forsa poll, the coalition of Union-FDP obtained a majority of seats in 2633 of 10 000 simulations

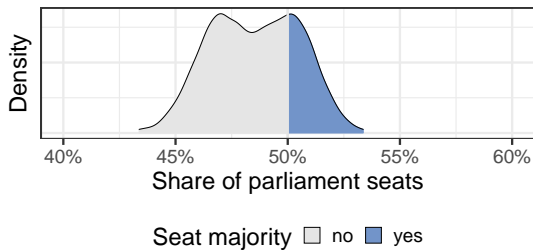
⇒ POE \approx 26%

Voter shares



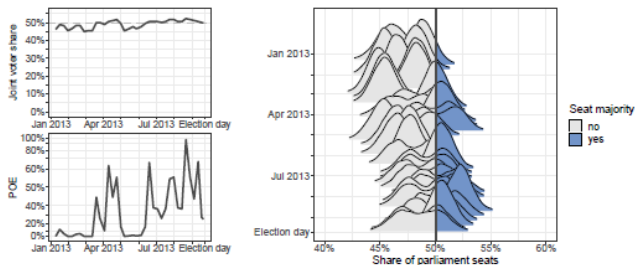
② Methods

Posterior distribution of joint CDU-FDP seat share



⇒ POE \approx 26%

② Methods



Plan:

1. Als Motivationsbeispiel die fruehste gepoolte Umfrage im 2013er Ridgeline Plot nehmen
2. Erstmal nur Ridgeline zeigen und Zeitverlauf animieren
3. Bei erster richtiger Bimodalitaet Animation anhalten und kurz daneben Union- und FDP-Stimmanteil-Zeitverlauf einblenden
4. Animation fertig laufen lassen (Union und FDP dabei wieder ausgeblendet)
5. Am Ende links den redistributed joint voter share und die POE-W'keiten einblenden

② Methods

Pooling

We aggregate multiple polls to reduce sample uncertainty.

In case of multiple random samples:

$$\left(\sum_i X_{i1}, \dots, \sum_i X_{iP} \right)^T \sim \text{Multinomial} \left(\sum_i n_i, \theta_1, \dots, \theta_P \right).$$

We account for correlations between polling agencies by using an **effective sample size** (Hanley et al., 2003).

⇒ **Example:** Pooling two polls with 1 500 and 2 000 respondents (where the strongest party obtained 40%), we get a conservative effective sample size of $n_{\text{eff}} = 2\,341$.

② Methods

Pooling

We aggregate multiple polls to reduce sample uncertainty.

In case of multiple random samples:

$$\left(\sum_i X_{i1}, \dots, \sum_i X_{iP} \right)^T \sim \text{Multinomial} \left(\sum_i n_i, \theta_1, \dots, \theta_P \right).$$

We account for correlations between polling agencies by using an **effective sample size** (Hanley et al., 2003).

⇒ **Example:** Pooling two polls with 1 500 and 2 000 respondents (where the strongest party obtained 40%), we get a conservative effective sample size of $n_{\text{eff}} = 2\,341$.

② Methods

Pooling in practice

- We only pool surveys published in the last 14 days
- We only include one survey per polling agency

Correction of rounding errors

Party shares are only published with a certain accuracy.

We add **uniformly distributed random noise** to avoid potential biases.

② Methods

Pooling in practice

- We only pool surveys published in the last 14 days
- We only include one survey per polling agency

Correction of rounding errors

Party shares are only published with a certain accuracy.

We add **uniformly distributed random noise** to avoid potential biases.

Outline

1. Motivation
2. Methods
3. Technical implementation
4. Results & Outlook

② Technical implementation


R package `coalitions`



... Code example: [More on GitHub](#)

② Technical implementation

Web-Interface

- `koala.stat.uni-muenchen.de`
- Blog
- @koala_lmu 
- based on Shiny
- automatic update scraping data from `wahlrecht.de`

Outline

1. Motivation
2. Methods
3. Technical implementation
4. Results & Outlook

⑤ Results

content

References

Topic

Doe J, Mustermann M (2019) This is the paper title. Journal, 19(2–3), 1–19

Doe J, Mustermann M (2019) This is the paper title. Journal, 19(2–3), 1–19

Another topic

Doe J, Mustermann M (2019) This is the paper title. Journal, 19(2–3), 1–19

Hanley et al., 2003

References

One more topic

Doe J, Mustermann M (2019) This is the paper title. Journal, 19(2–3), 1–19