Proceedings of Machine Learning Research 147:201–209, 2021 ISIPTA 2021

Towards Improving Electoral Forecasting by Including Undecided Voters and Interval-valued Prior Knowledge

Dominik Kreiss DOMINIK.KREISS@STAT.UNI-MUENCHEN.DE
Georg Schollmeyer GEORG.SCHOLLMEYER@STAT.UNI-MUENCHEN.DE
Thomas Augustin THOMAS.AUGUSTIN@STAT.UNI-MUENCHEN.DE
Institute of Statistics, Ludwig-Maximilians Universität München (LMU), Munich, Germany

Abstract

Increasing numbers of undecided voters constitute a severe challenge for conventional pre-election polls in multi-party systems. While these polls only provide the still pondering individuals with the options to either state a precise party or to drop out, we suggest to regard their valuable information in a set-valued way. The resulting consideration set, listing all the options the individual is still pondering between, can be interpreted under epistemic imprecision. Within this paper we extend the already existing approaches including this valuable information, by making first steps to utilize interval-valued prior information. Including background information is common in election forecasting while we focus on realistically obtainable and credible interval-valued prior information about transition probabilities from the undecided to the eventual choice. We introduce two approaches utilizing this interval-valued information, weighting the credibility against the precision of the results. For the first approach, we narrow the most cautious and wide so-called Dempster bounds by deploying the prior information on the transition probabilities as new worst and best case scenarios for each party. The second approach applies if these interval-valued results are still too wide for useful application. We hereby narrow them towards a good guess of the eventual choice, estimated by a further model-based source of information making use of the covariates. These single-valued estimates on the individual level are regarded as realizations of an underlying probability distribution, which we combine with the prior knowledge in a Bayesian way. The approach can thus be seen as an attempt to combine two, for the needed outcome by themselves inadequate, sources of information to obtain more concise results. We conduct a simulation study showing the applicability and virtues of the new approaches and compare them to conventional ones.

Keywords: epistemic imprecision, election forecasting, undecided respondents, survey methodology, imprecise probabilities

1. Introduction

As more and more voters are undecided in pre-election polls, methodology incorporating their valuable information is called for. To this end Plass et al. (2015); Oscarsson and Rosema (2019); Kreiss and Augustin (2020); Kreiss et al. (2020) suggested to regard an undecided voter as the set of options the individual is still pondering between, the from now on so-called *consideration sets*. Several arguments are put forward to substantiate this approach, like the reduction of nonresponse, the natural procedure or the more accurate representation of uncertainty. An election in a multi-party-system is hereby a separate choice of $\{1, \cdots, n\}$ individuals between a discrete set of, from the beginning on known, alternatives $\{1, \cdots, s\} = S$. But at the point in time of the pre-election poll, the undecideds' position can only be characterized by a combination of the original parties representing the options he or she is still pondering between, hence, an element from the state space of the power set of the original options $\mathscr{P}(S)$. This set-valued information about a still undecided can be seen as a container of the one true element he or she ends up voting for, thus as a coarse version of that true element. This is called the epistemic interpretation of the consideration set (e.g. Couso and Dubois (2014)), focussing on election forecasting.

Provided with this set-valued information, *transition probabilities* within the consideration sets of the undecided to the final choice can be assessed, in order to obtain overall forecasting together with the decided. As we are faced with inherent, non-stochastic uncertainty which element of their consideration set the individual ends up choosing, the resulting forecasts are naturally interval-valued if no further procedures are deployed. To this end, Kreiss and Augustin (2020); Kreiss et al. (2020) suggested preliminary approaches, reaching from the *Dempster bounds* as the most cautious to point-valued results based on strong additional assumptions.

Within this paper we introduce methodology to incorporate credible interval-valued prior information about transition probabilities as a natural further step to improve forecasting in this setting. Different sorts of background knowledge are commonly used in electoral research. (e.g. Linzer (2013)) Information in the form of probability intervals over the singletons can hereby either stem from experts and/or previous elections. We suggest two approaches utilizing prior information, weighting the credibility against the precision of the results. We hereby use the term prior information in an informal sense, comprising all kinds of background information beyond the data.

The first approach narrows the most cautious upper and lower Dempster bounds deploying the prior information as new best and worst case scenarios for each party. If the interval-valued prior information is credible, we obtain accurate but possibly wide intervals. If the resulting interval-valued forecasts are still too wide for useful application, the prior information alone is inadequate to obtain the necessary outcome. Hence, we have to include further knowledge about the process. This is a rather complicated task as due to the epistemic nature of the problem no entirely reliable information about the eventual choice is available. Therefore, we suggest to include information from another source, using covariates to estimate transition probabilities on an individual level. Hereby, we make use of the information within the covariates about the eventual choice with a working model reliant on presuppositions and trained on supplementary data. This can be seen as our best guess, which on the one hand is presumably biased but on the other hand carries the information within covariates. Towards this guess we can now narrow the interval from the prior knowledge. As the working model is trained on supplementary data, the single-valued estimates on the individual level are obtained by a deterministic function of the covariates. Hence, if the sample is identically independently distributed (i.i.d.), the resulting estimates can be seen as i.i.d. realizations of an underlying distribution characterizing the transition probabilities based on this working model. We combine the interval-valued prior information with the results of the precise working model for each party in each group of undecided separately in a Bayesian way. To achieve this, the interval-valued prior information is not seen infallible but also distributed with a certain variance. This variance determines the impact of the prior information within the overall forecasts.

Depending on distributional assumptions, the lower and upper bounds of the prior information can be respectively deployed as priors for Bayesian models, to obtain two posteriori distributions leading to estimates of the upper and the lower bound of the transition probabilities. As a consequence we obtain narrower bounds for the transition probabilities, combining two, for the needed outcome by themselves potentially inadequate, sources of knowledge. We conduct a simulation study for a simplified but realistic case, showing the applicability and virtues of the new approaches and comparing them to conventional ones.

This paper is structured as follows: First, we discuss the epistemic background of the set-valued information and the basis of overall forecasting with transition probabilities. Then, we introduce our two approaches based on interval-valued prior information and show the applicability and virtues with a simulation. In the concluding remarks we reflect on the approaches and possible further advancements in this particular field.

2. Methods

2.1. The Epistemic Interpretation of the Consideration Sets

The consideration sets $\ell = \mathcal{P}(S)$ characterizing the undecided individuals' position in the pre-election poll, contains all the elements the undecided is still pondering between. Thus, it can be seen as a coarse version of the one true element $l \in \ell$ contained in the set the individual ends up choosing, which is a particular interesting application of the theory about epistemic imprecision discussed by Couso et al. (2014). Hereby, the set-valued information is an imprecise version of something precise and only incomplete information about the phenomena of interest (the eventual choice) is provided by this consideration set. While we are looking for the random variable $Y(\omega)$ mapping from an underlying space of the population Ω to S, we are only provided with incomplete information in the sense that $\forall \omega \in \Omega$ only $Y(\omega) \in \ell = \mathcal{Y}(\omega)$ is observable, where $\mathcal{Y}$ is a multi-valued mapping $\Omega \to \mathcal{P}(S)$ representing the set of mappings $\{Y : \Omega \to S, Y(\omega) \in \mathcal{Y}(\omega)\ \forall \omega\}$. (Couso and Dubois, 2014, p. 1504) We therefore build an epistemic model of the random variable $Y(\omega)$, where for the undecided aside from covariates all that is known is $Y(\omega) \in \ell$.

Concerning forecasting one can either reflect the uncertainty of ℓ within the final results in an interval-valued manner, or incorporate further information or presuppositions to obtain more concise or even point-valued results. Thus, one has to ponder between imprecise results and the justifiability of assumptions leading to more precise statements.[1] Facing this tradeoff, one sometimes has to comply with an external specification of the maximal degree of imprecision for the results to be usefully applicable. In this case one has to find the most credible approach to comply with the provisions. We can asses (imprecise) *transition probabilities*, as an (imprecise) probability distribution over the elements of $l \in \ell$, to obtain forecasts as will be discussed in the following paragraph.

2.2. Forecasts Incorporating Consideration Sets

Within the sample of the population in the pre-election poll, the individuals are characterized by one element of the power set $\ell \in \mathcal{P}(S)$ and the values of the covariates in some space $\mathcal{X}$. Starting with the consideration sets and covariates of the $i \in \{1, \cdots, n\}$ participants, we want to estimate by

1. See also Manski's Law of Decreasing Credibility (Manski, 2003, p. 1)

an i.i.d. sample the expected frequency of each element of S within the population, using the generic variables Y and $\mathcal{Y}$. The individual's consideration set from the pre-election survey is written as an event $\{\mathcal{Y}_i = \ell\}$ with $\ell_i \in \mathcal{P}(S)$ and the possibly unknown choice on election day $\{Y = l\}$ with $l \in S$. Hence, we estimate the probability distribution $P(Y = l)\ \forall l \in S$ over the singletons, which can be seen as a *multinomial distribution* over the state space with $|S| - 1$ parameters. Hereby, the probability distribution can be factorized according to the chain rule into three parts like discussed in (Kreiss and Augustin, 2020, p. 244):

$$P(Y=l) = \sum_{(\ell,x)\,\in\,(\mathcal{P}(S)\times\mathcal{X})} P(Y=l,\mathcal{Y}=\ell,X=x) \quad (1)$$

$$= \sum_{(\ell,x)\,\in\,(\mathcal{P}(S)\times\mathcal{X})} \underbrace{P(Y=l|\mathcal{Y}=\ell,X=x)}_{\textit{Transition Probabilities}} \quad (2)$$

$$\cdot \underbrace{P(\mathcal{Y}=\ell|X=x)}_{\textit{Consideration Sets}} \cdot \underbrace{P(X=x)}_{\textit{Covariates}} \quad (3)$$

First, the *transition probabilities* determining the probability to vote for a specific party given the consideration set and covariates. Second, the probability of the consideration sets given the covariates and third, the one for the covariates. There are different approaches possible to estimate the second and third part of the factorization in (2) and (3). In this paper we focus on regression methodology. Even though one has to keep in mind sampling and modeling errors, there is sufficient information to estimate these factors right away with established procedures. Therefore, we treat these quantities as fixed and known in the sequel. The first part of the factorization on the other hand, reflects the epistemic problem previously discussed, as the value of $l \in Y$ among the options of ℓ is not observable for an undecided individual. For every decided individual, the transition probability is naturally one, as there is only one element to choose from, while for an undecided point- or interval-valued assessment is possible, allocating a specific range between 0 and 1 to every party in the consideration set. Hence, we concern ourselves with complex, non-stochastic, inherent uncertainty as there is no clear way to determine the resulting choice. The approaches suggested below distinguish themselves by the presuppositions and information utilized to estimate the transition probabilities from the undecided to the eventual choice, leading to overall different forecasts.

2.3. Approaches Incorporating Interval-Valued Prior Information

Due to the periodic nature of elections there is usually prior information about most properties available. This knowledge however is most of the times not precise, as despite overall continuity there are changes and no absolute certainty between the years. Nevertheless, there is usually at least some consistency as well as expertise, which is useful to improve estimation. In order for this prior information to be credible it has to be provided in an imprecise manner, reflecting the inherent uncertainty interval-valued (e.g. Augustin et al. (2014)). Hereby, the information is the least imprecise version for which we are convinced that it is accurate. Prior information, which can either stem from (a number of) experts or estimated from data of the previous election(s), is commonly used to forecast elections (e.g. Linzer (2013)). Hence, credible imprecise prior infomration is a natural way to improve (imprecise) election forecasting.

In our case, we employ prior information about the transition probabilities within the groups of undecided, containing information about the choice probabilities. Hereby, for practicability and modeling purposes, we regard the binary case of probabilities about choosing a specific party against choosing a different one. Furthermore, in this paper we only regard explicit probability intervals, stating a possible range over the singletons only. This somewhat restrictive kind of information can be particularly easy provided by experts or known from previous studies. Within each group defined by $\ell \in \mathcal{Y}$ we assume to be given imprecise knowledge about the probability that an individual chooses a given party, manifested in a probability interval. For example, we are certain that between 30% and 70% of the individuals undecided between the parties $\{A,B,C\}$ will end up voting for A. We can now denote the prior information about party A in the group $\{A,B,C\}$ as the interval between the extreme points, thus as $pr_{A,\{A,B,C\}} = [pr^{lower}_{A,\{A,B,C\}}\ ;\ pr^{upper}_{A,\{A,B,C\}}]$ or more generally as $pr_{l,\ell} = [pr^{lower}_{l,\ell}\ ;\ pr^{upper}_{l,\ell}]$. Therefore, we work with imprecise knowledge manifested as an interval for each party in every group respectively. In some modeling cases it is possible to only regard the upper and lower bound of the interval for all information, while in others it is not. This sort of imprecise and credible prior information will now be deployed to improve forecasting incorporating the undecided.

Approach One

The **first approach** builds on the *Dempster bounds* (e.g. Dempster (1967)) like mentioned above, only using the information within the data and not relying on any assumptions nor further information. Therefore, these bounds are completely credible but also reflect the entire uncertainty, hence the best and worst case for every party as interval-valued results. Hereby, the Dempster bounds assign the transition probabilities the entire interval between 0 and 1 to every undecided individual, as this is the only way to reflect the attached uncertainty.

With the first approach we deploy the interval-valued prior information, in order to obtain more concise results. This is at the cost of the assumption that this prior information is accurate, which depending on the source in few cases might be disputable. Hereby, we narrow the transition probabilities interval from the original 0 and 1 to the interval-valued prior information. As the prior information in this case directly provides the minimum and maximum of the proportions, we do no longer have to rely on the entire interval between 0 and 1. Under the assumption that the prior information is indeed accurate the transition probabilities can be narrowed, inserting the new values leading to the new transition probabilities:

$$P(Y = l|\mathcal{Y} = \ell) = [pr_{l,\ell}^{lower}\ ;\ pr_{l,\ell}^{upper}] \tag{4}$$

Therefore, the new best and worst case are provided with the prior information. In the case of no prior information available, the transition probabilities once again take the whole range between 0 and 1, while for a decided individual they remain 1.

Due to the independence from the covariates, these transition probabilities can directly be inserted in equation (1), leading to overall imprecise forecasts.

Overall, this approach relies on the accuracy of the interval-valued prior information, but if we assume the prior information to be completely reliable we can narrow the bounds without loosing any credibility at all. We hereby took a first step narrowing the Dempster bounds towards more concise results.

Approach Two

Prior information is frequently very imprecise, leading to possibly vague forecasts if we expect it to be entirely true like in approach one. Thus, one may be forced to further narrow the bounds in order for the results to be usefully applicable. Hence, the interval-valued prior informations alone are not enough to obtain a desired level of conciseness and it is necessary to take further measures. As due to the inherent non-stochastic uncertainty more concise results are not evident, we suggest to include another source of information exploiting the information from the covariates. Hereby, we utilize the information contained in the covariates, by training a working model in order to obtain single-valued estimates on the individual level for a first step. We can assume some information about the eventual choice to be within the covariates, even though this information might not be entirely reliable, as is the case in our example in section 3. The resulting estimates can thus be seen as some sort of best guess of single-valued transition probabilities, containing the information about the covariates. Even though this single-valued guess by itself is presumably biased, it carries the valuable additional information of the covariates and provides a direction with which we can achieve narrower results. We thus suggest an approach to combine two sources of knowledge, which by themselves are deficient to obtain an adequate outcome, to find a compromise which meets the external criteria of conciseness. To achieve this, the interval-valued prior information is not seen infallible but also distributed with a certain variance. With this variance we can determine the influence of the prior information on the overall results. We therefore still assume the prior information to be somewhat accurate, but unlike in the first approach it is not immediately used as the transition probabilities like in equation (4), but combined with the supplementary information.

We now suggest a Bayesian way to combine these sources of information in two steps. In the first step we describe why the predictions on the individual level based on a working model can be seen as i.i.d. data which within the second step can be combined with the prior information. To use the information structure of the covariates, we train a working model on supplementary data to predict the probability of a certain outcome on the individual level. This can either be achieved by regression or machine learning approaches like random forests, showing the working model training data for instance of previous elections or the decided. Proposals how to conduct this in the case of undecided voters reliant on different presuppositions are for example provided by Kreiss et al. (2020); Kreiss and Augustin (2020). We implement the working model below based on regression following equation (5). Hence, each undecided individual is assigned an explicit probability for each party in his or her consideration set by the working model, reflecting the information of the covariates. We therefore obtain the estimates in the form of a multinomial distribution between the elements of the consideration set for each individual. As the working model is trained on supplementary data, it is a deterministic function from the covariates to the predictions, preserving the i.i.d. structure of the sample. Thus, within each group of undecided we have identically and independently distributed observations, characterizing the transition probabilities based on the information of the covariates. We now suggest to regard these predictions as i.i.d. data themselves, following some unknown underlying distribution characterizing the transition probabilities for each group.

In the second step we propose a way to combine these – as data regarded – predictions with the interval-valued prior information. We hereby proceed for each party and group separately. As the working model provides us with estimates which sum up to one from an underlying multinomial distribution we can treat the parties separately, decomposing the multinomial distribution into party specific binomial ones.

Let's say we observe individuals for one specific group undecided between the three parties $\{A,B,C\}$ and are interested in the proportion of individuals who end up choosing A. We want to obtain an upper and lower probability how likely an individual in group $\{A,B,C\}$ ends up choosing A, combining both sources of information. From the working model we obtain a vector with probabilities how likely these individuals in group $\{A,B,C\}$ choose party A, reflecting the information of the covariates. These probabilities, as argued above, can be seen as i.i.d observations of an unknown underlying probability distribution. Furthermore, we have interval-valued prior information in the form $[pr^{lower}_{A,\{A,B,C\}} = 0.3\ ;\ pr^{upper}_{A,\{A,B,C\}} = 0.7]$. To obtain a posteriori from these two sources of information we can either make distributional assumption about the i.i.d. observations of the supplementary data, understanding the prior knowledge as information about the parameters, or we can work with the empirical probability distribution. (Gelman et al., 2013, ch. 20-23) Either way, we obtain a posteriori combining the information of our two sources, which fulfills two purposes: One, the resulting information now incorporates the information of the covariates and two, the resulting interval gets narrowed as it is combined with single-valued information. The new upper and lower bounds can be seen as a compromise between the sources of information, within the space between the single-valued prediction and the wide bounds.

Not relying on a distributional assumption about the i.i.d. supplementary data, we can work with the empirical probability distribution. Hereby, we can deploy a Dirichlet process Ferguson (1973) with beta distributed priors (as we are in the binary case) characterizing the problem in a nonparametric way. A posteriori distribution can be determined this way, resulting in overall forecasting.

Otherwise we could make a reasonable distribution assumption about the i.i.d. data. One intelligible assumption would be a beta distribution, characterizing the resulting probabilities in a natural way. The distribution is defined over the parameters α and β, where we rely on the parametrization: $f_X(x : \alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$. We now want to submit the interval-valued prior information of the experts as knowledge about the parameters. There is a number of distributional assumptions possible to realize this, while we focus on a (truncated) normal distribution, as priories for the parameters α and β. As the mean of the distribution is defined by $\frac{\alpha}{\alpha+\beta}$ we choose priori values for α and β to submit the desired mean. Hence, for a higher prior value we increase α and decrease β within the prior knowledge about the parameters. This can for example be achieved by demanding the mean values of the parameters to sum up to a given value, ensuring one decreases while the other increases. The parameters can hereby still be simulated independently. The advantage of a (truncated) normal distribution are both its stability concerning sampling as well as the intuitive and explicit variance parameter. Furthermore, we keep the variance parameter constant over the entire interval constituting the prior knowledge. From this the convenient attribute results, that we only have to consider the extreme points of the interval rather than also the entire points in between. As all values within the interval are truncated normal distributed with identical variance parameter, higher ones hold first order stochastic dominance over lower ones. As the expectation of the posteriori monotonically increases in α and decreases in β and higher priori values lead to higher α and lower β, the extreme points of the interval produce all of the interjacent values for the result. This simplifies the process, only demanding two separate models, one for the upper and one for the lower bound. With a high variance parameter we indicate low belief in our prior as with a variance parameter close to zero the priori almost determines the results. A beta distributed prior over the parameters is possible as well, but might not be as intuitive and computationally feasible. As those priors are non-conjugated, we can deploy a MCMC process, drawing samples from the posteriori for estimation. (Gelman et al., 2013, ch. 12). Such a process can for example be easily implemented with *Rstan* (Stan Development Team (2020)).

The resulting upper and lower bounds for the party wise calculated transition probabilities can be directly deployed to calculate overall forecasting with equation (1). We hence obtain overall forecasts using both sources of knowledge to make the most of the information about the undecided. The results are therefore between the extreme points of the initial bounds and the single-valued estimator exploiting the information in the covariates. We can regard this approach, narrowing the initial wide intervals, as a pragmatic attempt. On the other hand, we could also see the process as a regularization of the single-valued estimates towards more credible bounds. In both cases we suggest a tradeoff between, and a combination of, two sources of information. By adjusting the variance parameter, we can determine the influence of the prior information as well as the conciseness of the overall outcome. Hence, the variance parameter effects the accuracy-precision tradeoff, laying more emphasis on the one, or the other information. But as due to the epistemic nature of the problem, we do not know how accurate the estimates based on the covariates are, it is difficult to give general statements about the accuracy-precision tradeoff. Therefore, we described the process as taking steps in hopefully the right direction to make the results as concise as necessary, but do not really generalize how accurate the results are, as this differs from case to case.
In our case we further assume the interval-valued prior information to be overall accurate. But in different applications one would have to be aware of a possible bias from this source as well.

3. Simulation Study and Further Details

3.1. Specifying the Simulation

To illustrate the applicability and virtues of the two new approaches we conduct a simulation study, comparing them to the Dempster bounds as the most cautious and the point-valued approach neglecting the undecided overall. We consider a scenario in which three parties $\{A;B;C\}$ can be chosen at the election. Thus, within the pre-election poll, there are three groups of undecided voters $\{A,B\};\{B,C\};\{A,C\}$ resulting in overall six options including the decided voters. We choose a realistic sample size of 1000 individuals and ensure them to be an i.i.d. representation of the underlying truth.

For the first step we draw the individuals from an multinomial distribution describing the proportion of the groups within the population at the pre-election poll. We specified the parameters of this distribution $P(\mathcal{Y} = \ell)$ as follows:

$$\{p_{\{A\}} = 0.45\,;\; p_{\{B\}} = 0.2\,;\; p_{\{C\}} = 0.1,$$
$$p_{\{A,B\}} = 0.15\,;\; p_{\{A,C\}} = 0.05\,;\; p_{\{B,C\}} = 0.05\}$$

From the resulting data we specify the true transition probabilities with which we simulate the eventual choice of the undecided.

$$\{p(Y = A|\mathcal{Y} = \{A,B\}) = 0.5\,;$$
$$p(Y = A|\mathcal{Y} = \{A,C\}) = 0.9\,;$$
$$p(Y = B|\mathcal{Y} = \{B,C\}) = 0.5\}$$

This determines the true outcome in our underlying population to be:

$$\{p_A = 0.57\,;\; p_B = 0.30\,;\; p_C = 0.13\}$$

Furthermore, to mimic the exploitation of the information of the external data, we simulate in addition a covariate which is somehow correlated with the eventual choice. This continuous covariate therefore varies within the respective groups and eventual choices. We hereby use a normal distribution, which contains some, but biased information about the eventual choice. This resembles the realistic scenario in which covariates contain valuable information, but which is only by itself not adequate to produce a reliable prognosis. The variance is fixed within all groups resulting in the parameters in table (3.1). We thus obtain a simplified but

Choice	A	B	C	A	B	B	C	A	C
Set	A	B	C	A/B	A/B	B/C	B/C	A/C	A/C
μ	70	50	30	65	55	55	30	65	30
σ^2	20	20	20	20	20	20	20	20	20

Table 1: Parameters of the normal distribution of the continuous covariate amongst the different groups and parties

realistic sample with which we can estimate transition probabilities in order to obtain overall forecasts with equation (1).

Furthermore, we have imprecise prior knowledge in the form of probability intervals about the transition

A	Minimum	Maximum
A/B	0.3	0.7
A/C	0.6	0.9

B	Minimum	Maximum
B/A	0.3	0.7
B/C	0.4	0.6

Table 2: Prior knowledge about the probability to choose party A on the left, and to choose party B on the right sight, depending on the underlying groups.

probabilities from an expert illustrated in table (3.1). The prior knowledge concerning party C results from the probabilities of the complements. The interval-valued information is hereby wide enough to be realistically credible, as we expect expert to provide somewhat accurate information with confidence. Hence, the prior information in this case satisfies the realistic criteria to be accurate, but is very imprecise.

Provided with these samples and prior knowledge we can now apply the two approaches discussed above and compare them with the Dempster bounds and the conventional approach.

3.2. Applying the Approaches

We simulate and apply the approaches multiple times (50) and average over the results. Hereby, we first calculate the transition probabilities for all three approaches and determine the overall forecasting together with the decided according to equation (1). As we have a representative sample there are means to estimate the second and third part of the factorization in equation (1), we choose a logistic regression approach, estimating the conditional distribution. The prediction resulting form equation (5) is estimated via logistic regression as well. Within our approach we primarily focus on the non-stochastic, complex inherent uncertainty, not elaborating on the sampling and modeling errors induced and treat the estimated quantities as fixed. Nevertheless, an overview of variation between the different samples is provided within the appendix A.

As mentioned above, we regard the parties and groups separately. Thus, for every party in every group we are supplied with a vector of identically independently distributed probabilities from a working model utilizing the simulated covariate here. To achieve this, we follow (Kreiss and Augustin, 2020, p. 245) with a working model reliant on the presupposition that the undecided choose identical to the decided given their covariates and consideration sets. The transition probabilities on an individual level are hereby predicted resulting in

$$\hat{P}(Y = l|\mathcal{Y} = \ell, X = x) = \frac{\hat{P}(Y = l|X = x, I_d = 1)}{\sum_{a \in \ell} \hat{P}(Y = a|X = x, I_d = 1)} \quad (5)$$

with I_d as the indicator function for being decided.[2] This resulting best guess is now incorporated in a Bayesian way as discussed above. In our application we approximate these realizations in a natural way with a beta distribution. For feasible estimation of the parameters based on the i.i.d. data we need to specify the possible range of α and β setting it to $[0,10]$. Then, we incorporate the prior knowledge as information about the parameters α and β. To this end we choose a, strictly speaking truncated, normal distribution, only taking values in the possible range of α and β. With this (truncated) normal distribution we now specify the prior knowledge about α and β. The α and β parameters are hereby simulated independently. To increase one while decreasing the other parameter we demand their expectation to sum up to one. With this we can directly apply the overall expectation of the prior information. Furthermore, it is possible to only regard the upper and lower bounds of the interval, due to the constant variance parameter and the following properties of the (truncated) normal distribution concerning first order stochastic dominance. As an example, following the logic of above, the knowledge $[pr^{upper}_{A,\{A,B\}} = 0.7]$ is transferred into $\alpha \sim \text{Normal*}$[3]$(0.7,\sigma^2); \beta \sim \text{Normal*}(0.3,\sigma^2)$ with a (truncated) normal distribution controlling the strength of the prior knowledge with the variance parameter. The submitted knowledge therefore constitutes the targeted mean of $\frac{0.7}{0.7+0.3} = 0.7$. To give the priori reasonable weight we choose the variance parameter as $\sigma^2 = 0.05$. This precise value is admittedly chosen somewhat arbitrary as a subjective consideration concerning the accuracy-precision tradeoff. The approach for the upper bound with this specific prior knowledge can thus be written in a hierarchical way as:

$$\begin{aligned} \alpha, \beta &\in [0,10] \\ \alpha &\sim N^*(0.7, 0.05) \\ \beta &\sim N^*(0.3, 0.05) \\ \text{Likelihood} &: \text{Beta}(\alpha, \beta) \end{aligned}$$

The posteriori is calculated over a MCMC process implemented with *RStan* Stan Development Team (2020). From the expectation of the posteriori we obtain the overall estimate for the upper bound of the transition probability towards the specific party in the specific group. Hereby, we do not make a parametric assumption about the posteriori distribution but merely take the Monte Carlo expectation of the parameters. This process is repeated for each constellation of upper and lower bounds, groups and parties, which results in transition probabilities and overall forecasts.

2. There is a connection to the work of Heitjan and Rubin (1991), even though the assumption is somewhat different to Coarsening at Random.
3. Normal* or N* stands for the truncated normal distribution

The results of all approaches, additionally with the true parameters and the estimate reliant only on the decided are illustrated in figure (3.2). At the left upper

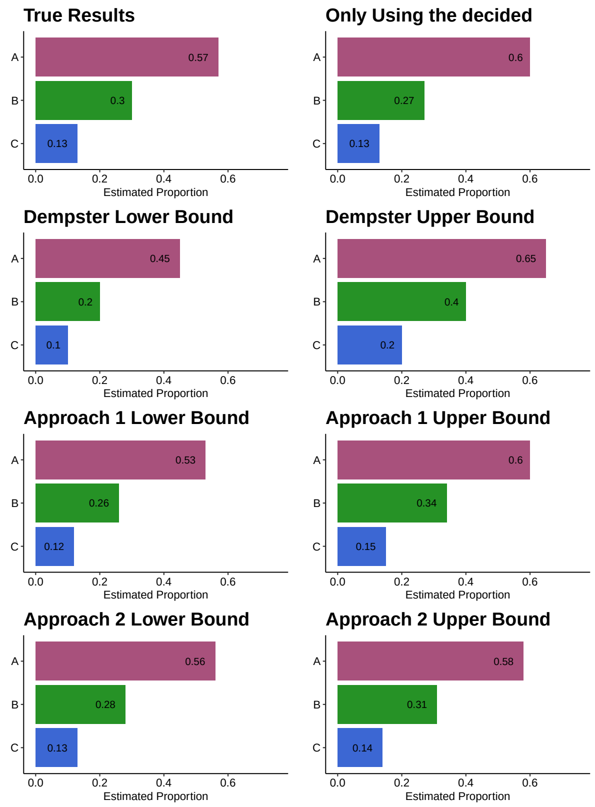

Figure 1: Mean of the result from 50 samples of $n = 1000$ for three approaches, alongside with the true values and the estimate neglecting the undecided overall in the first row. The variance parameter of the prior knowledge for approach 2 was set to 0.05.

corner we can see the true values the simulation is based on. On the upper right we can examine that the estimate overall neglecting the undecided like in conventional approaches leads to biased results.[4] While the Dempster upper and lower bounds in row two are very wide, they are substantially narrowed with the prior information in approach 1 in the third row. Finally, the second approach in the fourth row ensures the narrowest bounds, incorporating the two sources of information.

The approaches in rows two to four are all possible solutions of the tradeoff between precision of the results and credibility. The Dempster bounds are hereby the most credible but wide results, while approach 2 has the narrowed

4. Whenever in this case the undecided are *missing not at random*, as can usually be expected, such approaches end up biased.

bounds incorporating the information of the covariates. From a practical point of view, we suggest to use the most credible approach still satisfying the necessary criteria of conciseness. Decreasing the variance parameter of the priori within approach 2 would lead to even more concise results. With our realistically chosen parameters all three approaches overlap the true values, emphasizing the credibility of the approaches. The variation of the results reliant on different draws from the simulation are illustrated in the appendix A with a box plot. We can see that the dispersion is not too severe between the results of the different datasets.

Despite the desirable traits of approach 2, it is somewhat complicated to evaluate the accuracy, as it results from a combination of multiple sources of information. One has to choose whether the information provided by the covariates outweighs the potential bias introduced, which in this simplified but realistic scenario is definitively the case. The credibility furthermore depends on how the two sources of information are weighted. Within the simulation the prior information is accurate but quite imprecise. Examining different scenarios lead us to believe that small bias in the prior information does not effect the results severely. Approach two is definitively a strong tool to narrow the initial bounds, which with reasonable weighting of the sources of information should still overlap the true value, as shown in the exemplary simulation.

4. Concluding Remarks

Within this paper we introduced two approaches incorporating interval-valued prior knowledge in order to improve election forecasting including undecided voters. The first one provides narrower bounds in a straight forward manner, only reliant on accurate prior information. Narrowing these bounds further in a credible way is far more complicated, and we address this problem by including further information making use of the informations in the covariates in a Bayesian way. Hereby, we suggest and apply first methodology, regarding the single-valued predictions on an individual level by the covariates as i.i.d. data. The results are in between the initial bounds and the single-valued predictions, incorporating both sources of information. The first results are promising, achieving narrower bounds in a plausible way.

For further research following this train of thought, one could determine the variance parameters for the prior knowledge by demanding a specific precision of the resulting overall forecasts. This can be implemented recursively, increasing or decreasing the variance parameter to obtain more, or less concise overall results. With the extreme points we get the initial bounds or the point-valued estimate of the transition probabilities. This would be one way to explicitly account for the tradeoff between credibility and precision of the results (e.g. Manski (2003)) in the second approach. Furthermore, highlighting the implications of biased prior information in this context is interesting.

Additionally, there are plenty of directions possible to address undecided voters. For example, regarding the distributions overall, not decomposing it into binary cases and deploying Dirichlet processes for Bayesian modeling is interesting. Furthermore, we could combine the initial bounds with imprecise estimates based on the covariates. The same basic concepts apply, while the procedure is a little more complicated as more combinations arise. Also highlighting connections and differences to other approaches combining evidence, in particular the Dempster-Shafer theory of evidence Denoeux (2016) is interesting for further research.

Within this work we solely focussed on election systems in which the individual casts one vote like common in Europe. Instant-runoff-voting and different ranking voting systems are worthy of exploring further on. Some of the thoughts above can be adopted, but due to different ranking approaches the structure of the underlying state space and with it the methodology changes.

Overall, considering one source of information as i.i.d. data for Bayesian modeling has proven to be a useful measure to combine two, by themselves inadequate, sources of information. This basic concept could be transferred to multiple different applications and is especially useful concerning undecided voters in pre-election polls.

Acknowledgement. We are very grateful for the valuable and supportive remarks from the three anonymous reviewers. We further sincerely thank the LMU Mentoring Program, supporting young researchers. Furthermore, we would like to thank Malte Nalenz for the helpful discussions concerning RStan and Bayesian modeling.

References

T. Augustin, G. Walter, and F. Coolen. Statistical inference. In T. Augustin, F. Coolen, G. de Cooman, and M. Troffaes, editors, *Introduction to Imprecise Probabilities*, pages 135–189. Wiley, 2014.

Inés Couso and Didier Dubois. Statistical reasoning with set-valued information: Ontic vs. epistemic views. *International Journal of Approximate Reasoning*, 55(7): 1502–1518, 2014.

Inés Couso, Didier Dubois, and Luciano Sánchez. *Random sets and random fuzzy sets as lll-perceived random variables*. Springer, 2014.

Arthur Dempster. Upper and lower probabilities induced by a multivalued mapping. *The Annals of Mathematical Statistics*, 38(2):325–339, 1967.

Thierry Denoeux. 40 years of Dempster–Shafer theory. *International Journal of Approximate Reasoning*, 79: 1–6, 2016.

Thomas Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, 1:209–230, 1973.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. CRC press, 2013.

Daniel Heitjan and Donald Rubin. Ignorability and coarse data. *The Annals of Statistics*, pages 2244–2253, 1991.

Dominik Kreiss and Thomas Augustin. Undecided voters as set-valued information, towards forecasts under epistemic imprecision. In Jesse Davis and Karim Tabia, editors, *Scalable Uncertainty Management 2020*, pages 242–250. Springer, 2020.

Dominik Kreiss, Malte Nalenz, and Thomas Augustin. Undecided voters as set-valued information, machine learning approaches under complex uncertainty. In Eyke Huellermeier and Sebastian Destercke, editors, *ECML/PKDD 2020 Tutorial and Workshop on Uncertainty in Machine Learning*. 2020. URL https://drive.google.com/file/d/1abrLGZ154htGuYz8HzyLQzJ8vyc3kr2K/view.

Drew Linzer. Dynamic Bayesian forecasting of presidential elections in the states. *Journal of the American Statistical Association*, 108(501):124–134, 2013.

Charles Manski. *Partial identification of probability distributions*. Springer, 2003.

Henrik Oscarsson and Martin Rosema. Consideration set models of electoral choice: Theory, method, and application. *Electoral Studies*, 57:256–262, 2019.

Julia Plass, Paul Fink, Norbert Schoening, and Thomas Augustin. Statistical modelling in surveys without neglecting 'The undecided'. In Thomas Augustin, Serena Doria, Enrique Miranda, and Erik Quaeghebeur, editors, *ISIPTA 15*, pages 257–266. SIPTA, 2015.

Stan Development Team. RStan: the R interface to Stan, 2020. URL http://mc-stan.org/. R package version 2.21.2.

Appendix A. Boxplot of the simulations

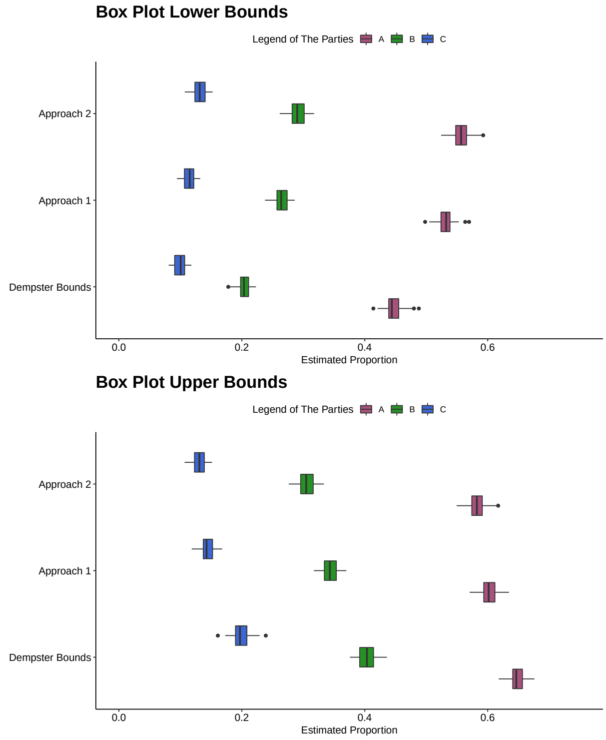

Figure 2: Box Plots illustrating the results from the different simulation iterations.