Bayesian Analysis of Election Surveys and Forecasts: Learning from the State of Valencia Experience

Andrew Gelman, Department of Statistics,
University of California, Berkeley

ABSTRACT Election surveys have several purposes, including forecasting election outcomes and studying the distribution of votes as they vary over geographic, demographic, and political variables. Bayesian methods can be useful in the design and analysis of election surveys, with the details depending on specific features of the data collection and goals of inference. In many cases, the natural Bayesian analyses correspond to variants of standard methods in survey sampling. This article discusses the design and modeling choices for pre-election and exit polls discussed in Bernardo (1996), emphasizing (1) the relation of the analyses to the methods of data collection, (2) the extent to which his Bayesian approach differs from classical methods, (3) differences between election studies in Spain and the United States.

1 Introduction and overview

Bernardo (1996) is a beautiful paper with two especially nice features. First, it is a start-to-finish description of a statistical analysis—starting with the setting and the substantive questions of interest, moving through data collection, modeling, data analysis, auxiliary analyses, validation (model checking), and presentation and distribution of results—including important practical details, such as time constraints, that are often treated much too casually in this sort of report. Second, the reasoning is all done from first principles rather than relying on standard methods or accepted wisdom. This adds an air of intellectual excitement to the paper, as we are led through a rediscovery of some standard (and nonstandard) methods for the design and analysis of opinion polls and electoral data. The derivations are in the best spirit of Bayesian inference and follow in the tradition of Bernardo's earlier work.

In this discussion, I will go through Bernardo (1996), focusing on the following three issues: (1) the ways in which the methods of data collection influenced the data analysis, (2) connections to classical survey sampling methods, and (3) comparisons to election studies in the United States.

I would also like to express my appreciation to the organizers of the Bayesian Case Studies workshop for inviting this paper by Bernardo. Many times, I have heard the reaction from statisticians and mathematicians that studying elections is "fun," with the implication that it is not a serious topic. As a matter of fact, there are some important issues that are addressed by the study of elections—who's in power, why they're there, and what it took for them to get there. How stable and predictable are election results, and can office-seekers take advantage of this? This is a question that is important to the functioning of a democratic society, and it is answered using statistical methods—either crudely, by scanning and number-crunching election results and poll data, or using more advanced methods. I applaud Jose Bernardo for making contributions in this area as well as for showing some of the connections between his foundational work and applied statistics.

2 The electoral system

The Spanish electoral system is in some ways more complicated in other ways simpler to analyze and forecast compared to the United States. Obviously, the four parties make the modeling and analysis more difficult than with two parties. On the other hand, the proportional representation system means that if you can forecast statewide votes for the parties, you can immediately forecast their vote shares. In contrast, in the United States (and Britain), with winner-take-all elections in legislative districts, it is necessary to study the distribution of votes in districts. (For example, in Spain, if a party wins 45% of the vote, it will get very close to 45% of the seats, whereas in the United States, a party can theoretically win 45% of the vote in each district and win no seats. More realistically, a party that wins 45% of the vote in the United States can expect to win about 40% of the seats, on average; see, e.g., Gelman and King, 1994.)

One implication for survey design is that, in Spain, one can proceed directly to estimating the average vote proportions of the parties. The locations of voters in polling stations are useful as covariates to improve the efficiency of estimation, and for validation of the model, but are not necessary in defining estimands in the way that legislative districts are crucial in the United States.

3 Survey design

Pre-election polling is not the same as election forecasting. There is a large literature in political science about the fact that early opinion polls can be far off the mark; much better forecasts can be obtained using economic data and information about previous votes and party preferences in addition to popularity polls (see Rosenstone, 1984, and Gelman and King, 1993).

The survey design used by Bernardo is the standard method of *cluster sampling* (see, e.g., Kish, 1965). Cluster sampling is an economical design for personal interviewing, where the cost of getting to the interview location must be considered. In the United States, most national political surveys use telephone interviews, and so there is no need for cluster sampling; variants of simple random sampling are standard (see, e.g., Brady and Orren, 1992, and Voss, Gelman, and King, 1995). Exit polls, however, are done in person and would thus be amenable to the cluster sampling design ideas presented by Bernardo.

The optimal sampling method described by Bernardo is interesting. I wonder if it might be improved by two modifications. First, I don't understand why the average votes in the sampled electoral sections need to mirror the average vote in the state. As long as the differences between sample and population are predictable, they can easily be accounted for in the estimation procedure: this could be done using classical *regression estimation* (see, e.g., Cochran, 1977) or more formal Bayesian models, in either case conditioning on previous election results in the electoral sections. Second, it might be useful to formally stratify by geographic and socioeconomic factors so as to ensure a diverse sample that will be more likely to capture nonuniform swings in the electorate.

4 Survey data analysis

The assumption that "relevant functions" are "approximately sufficient" in Section 4 of Bernardo (1996) corresponds to the assumption of *ignorability*, which is basic to Bayesian analysis of sample surveys, experiments, and observational studies (for a recent review, see Gelman et al., 1995, ch. 7). The method of weighting groups based on their population proportions is standard in sample survey analysis and is called *poststratification* (see Kish, 1965, for a classical treatment, and Little, 1993, for a discussion from a Bayesian perspective). It would be interesting to compare Bernardo's work on reference prior distributions for this problem is interesting to the models discussed in Little (1993).

In analyzing the survey results, it would seem to make sense to combine the surveys data with past election results. Classically, this would be done using regression estimation. Bayesian methods based on linear models and noninformative prior distributions would give similar results; using hierarchical models would probably lead to dramatic improvements, as in Bernardo and Giron (1992) and Gelman and King (1993). On a related note, I think the inferential results would be more interpretable if also presented as changes from previous election results: for each party and region, the expected (and, after the election, observed) change compared to the previous election, after subtracting out the statewide swing. This sort of hierarchical *presentation* of results (following the ideas of Tukey, 1977) would be nice whether or not hierarchical *modeling* was used.

Finally, in the spirit of Bayesian inference, I suggest that forecasts be accompanied by measures of posterior uncertainty—posterior intervals or standard errors. This would eliminate hyper-precise estimates such as "34.9%" and "19.1%", whose decimal places must certainly outstrip their posterior precision.

5 Election night forecasting

My comments on the analysis and display of the election night forecasts echo those for the previous section: regression estimation methods such as Bernardo and Giron (1992) would seem natural here, and displays of swings would focus on the more interesting aspects of the predictions. There is also a great deal of scope here for model checking: a topic that is of interest to statisticians as well as, of course, to users of these numbers. Perhaps the display of the results could automatically flag those areas where the exit poll was far off from the forecast, after subtracting off statewide partisan swings.

A related question, particularly relevant to the Bayesian: do you trust your standard errors? It presumably would not be too hard to perform a validation study and see if 95% of the 95% intervals contained the true values—this check could be done both for votes and vote swings for each electoral section.

Finally, the graphical displays presented in Bernardo (1996) are nice and could be made even nicer by judicious displays of residuals. I would also suggest placing a single symbol in each region rather than shading in the regions, so as not to draw too much attention to the geographically large but less populous rural regions (see, e.g., Cleveland, 1985).

6 The day after

The analysis of the exit poll data was interesting but left me with the following concern: the exit poll gives no information on the nonvoters in the most recent election. The analysis must thus make some assumption about these nonvoters, but the assumptions are not described in Bernardo (1996). I would be interested in hearing in more detail how the nonvoters were modeled.

In addition, as before, it would be nice to see the estimates such as "61.8%" come with uncertainties attached: once a Bayesian analysis has been done, one might as well take full advantage of it.

7 Some general comments on election surveys and forecasts

I would like to conclude my discussion of this thought-provoking paper with some thoughts on the analysis of electoral data: why is it possible to forecast elections well, and what are the best ways of doing so?

Elections are predictable. The geographic pattern of votes for a set of parties does not change much from one election to another, after adjusting for statewide or nationwide swings. This has long been noted by political scientists in Britain and the United States and appears to be true in Spain as well. The relative stability of relative votes reduces election forecasting to a problem of predicting aggregate swings.

Classical methods of survey design and analysis are highly relevant to pre- and post-election polls. Cluster sampling is relevant for in-person surveys. Post-stratification is relevant for adjusting for nonresponse. Regression estimation is useful because previous election results in electoral districts are highly correlated with current survey and election results. It is important for the Bayesian to recognize the relevance of these classical ideas for (at least) three reasons: first, a great deal of useful and practical work has been done in classical survey design and analysis, and we should all make use of that expertise; second, it is useful to understand the standard methods so as to better communicate with survey practitioners; and third, because useful Bayesian approaches can be derived by generalizing classical methods.

Ideas of Bayesian modeling and inference are also highly relevant. Most notably, ignorability helps us understand how to model nonresponse and suggests natural classes of models (modeling votes conditional on geographic and socioeconomic variables that are believed to affect nonresponse), and the notion of the posterior distribution is crucial in summarizing uncertainty about inferences for predictions.

Finally, the natural hierarchical structure of electoral data—with regions, electoral sections, parties, and multiple elections—leads to hierarchical models and data display. These ideas are well understood by political practitioners as "vote swings" and also suggest natural ways of model validation.

Acknowledgments

The National Science Foundation provided partial financial support for this work through grant DMS-40305 and Young Investigator grant DMS-9457824.

References

Bernardo and Giron (1992). Robust sequential prediction from non-random samples: the election-night forecasting case (with discussion). In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 61–77. New York: Oxford University Press.

Bernardo, J. (1996). Probing public opinion: the state of Valencia experience. In this volume.

Brady, H. E., and Orren, G. R. (1992). Polling pitfalls: source of error in public opinion surveys. In *Media Polls in American Politics*, ed. T. E. Mann and G. R. Orren. Washington, D.C.: Brookings Institution.

Cleveland, W. S. (1985). *The Elements of Graphing Data.* Pacific Grove, Calif.: Wadsworth.

Cochran, W. G. (1977). *Sampling Techniques*, third edition. New York: Wiley.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis.* New York: Chapman and Hall.

Gelman, A., and King, G. (1993). Why are American Presidential election campaign polls so variable when votes are so predictable? *Brit. J. Pol. Sci.* **23**, 409–451.

Gelman, A., and King, G. (1994). A unified model for evaluating electoral systems and redistricting plans. *Amer. J. Pol. Sci.* **38**, 514.

Kish, L. (1965). *Survey Sampling.* New York: Wiley.

Little, R. J. A. (1993). Post-stratification: a modeler's perspective. *Journal of the American Statistical Association* **88**, 1001–1012.

Rosenstone, S. (1984). *Forecasting Presidential Elections.* New Haven: Yale University Press.

Tukey, J. W. (1977). *Exploratory Data Analysis.* New York: Addison-Wesley.

Voss, D. S., Gelman, A., and King, G. (1995). Pre-election survey methodology: details from nine polling organizations, 1988 and 1992. *Public Opinion Quarterly* **59**, 98–132.