# Forecasting the Final Results in Election Day: A Bayesian Analysis

José M. Bernardo

**Abstract** In parliamentary elections, the important final result is not the set percentages of votes obtained by each party, but the number of seats finally allocated to them by application of the appropriate electoral law. This usually involves the use of an algorithm which includes some restrictions on the minimum percentage of votes to qualify (often 3 percent), and a procedure to allocate seats from the vote distribution (often d'Hondt rule). The appropriate forecast for the final electoral result is clearly a probability distribution on the possible allocations of seats in the Parliament among the competing parties. From a Bayesian viewpoint, this may be obtained as a probability transformation of the posterior probability distribution which serves to forecast the election results in terms of vote distribution. No appropriate solution to this problem seems available from a conventional, frequentist viewpoint.

**Key words:** Bayesian reference analysis, Election forecasting, Logarithmic divergence, Noninformative prior, Posterior distribution, Probability forecasting, Seat allocation.

## 1 The Problem

In parliamentary elections, the important final result is *not* the set percentages of votes obtained by each party or coalition, but the number of seats finally allocated to them by application of the appropriate electoral law. Thus, the result of an appropriate forecast on the outcome of the election should focus of the expected seat distribution. On the other hand, standard foundational arguments (see, *e.g*, [3] and references therein) may be invoked to establish that the *only* form of describing uncertainty which is compatible with a sensible set of coherence axioms is by means of a probability distribution. If follows that an appropriate forecast for the result of a parliamentary election should take the form of a probability distribution over the set

José M. Bernardo
Universitat de València, Spain; e-mail: jose.m.bernardo@uv.es

of possible allocations among the parties of the parliament seats. This is an impossible task for a conventional frequentist viewpoint, but it is relatively straightforward from a Bayesian perspective.

In most political systems, a number of seats are allocated to each region as a function of their electoral results, and the final results are simply obtained by the addition of all the regional results. It follows that any sensible statistical analysis has to be performed a regional level. From a Bayesian viewpoint this requires the determination of a probability distribution over the set of possible allocations among the parties of the seats associated to each region; the final forecast is then obtained as the convolution of all the regional forecasts.

This paper focuses on the derivation of a forecast based on actual results observed on a small set of appropriately chosen polling stations. Most of the material, however, applies to the related problem of forecasting the electoral results based on the outcome of an opinion poll.

As regard the **notation**, let $s$ be the number of seats in the parliament, and let $m$ be the number of political parties which take part in the election. Consider that the country is divided into $r$ regions, and let $s_j$ be the number of seats elected by region $j$, so that $s = \sum_{j=1}^{r} s_j$. Let $D = \{D_1, \ldots, D_r\}$ be the available data set, so that $D_r$ is the set of available data in region $j$. Any possible electoral result in region $j$ has the form $\theta_j = \{s_{1j}, \ldots, s_{mj}\}$, where $s_{ij}$ is the number of seats allocated to party $i$ in region $j$; obviously, the $s_{ij}$'s must be nonnegative integers such that $s_j = \sum_{i=1}^{m} s_{ij}$; more formally, $\theta_j$ is a solution of the diophantic equation $s_j = \sum_{i=1}^{m} s_{ij}$, $s_{ij} \in \mathcal{N}$. Thus, the forecast for region $j$ should take the form$\{p(\theta_j|D_j), \theta_j \in \Theta_j\}$, a probability distribution over the set $\Theta_j$ of possible seat allocation within region $j$. The final electoral result must be of the form $\theta = \sum_{j=1}^{r} \theta_j$ and, hence, the final forecast should be of the form $\{p(\theta|D), \theta \in \Theta\}$, a probability distribution over the set $\Theta$ of possible seat allocations in the parliament, to be obtained as a convolution of the $r$ partial forecasts, $p(\theta_j|D_j)$, $j = 1, \ldots, r$.

## 2 The Design

Using detailed past election results, it is possible to determine within each region a set of polling stations whose political behavior is as close as possible to the political behavior of the region as a whole. The idea is to find those polling stations which systematically yielded a vote distribution among the parties similar to that which resulting the region, for an appropriately defined notion of probability divergence.

A number of arguments (see, *e.g.*, [4], references therein) may be invoked to propose the logarithmic divergence (often referred to as Kullback-Leibler directed divergence) $\kappa\{q|p\} = \sum_{i=1}^{m} p_j \log(p_j/q_j)$ as an appropriate measure of the divergence of the probability distribution $q = \{q_1, \ldots, q_m\}$ from the probability distribution of reference $p = \{p_1, \ldots, p_m\}$. For each region $j$, one may identify the set of those $t_j$ polling stations whose results minimized their average logarithmic divergences from the regional results in the past elections considered. It is an interesting empirical fact, that these specially representative polling stations are remarkably constant over time and type of election, which means that they may be safely considered as a good representation of the the whole region. Thus, the data $D_j$ would typically consist of either the intention of vote of the citizens selected in a random sample from within the physical areas associated to these polling stations (in the case of opinion polls), or they may consist of actual results (all the results or just a sample of them) in those polling stations (in the case of forecasting with early actual results).

In the case of opinion polls, this procedure yields very good, relatively cheap sample designs, for the citizens in the sample are clustered within each of the (often very small) geographical areas associated to the $t_j$ selected polling stations. In the case of early actual results, this means that $t_j$ agents sitting at the (typically public) vote counting may easily phone the results of the selected polling stations (or even the results of the, say first 200 votes counted) to the analysis office for immediate derivation of an extremely precise forecast for region $j$.

## 3 A Bayesian Hierarchical Model

The data $D_j$ in region $j$ consists on the voting behavior of a sample of citizens in each of the $t_j$ polling stations selected in region $j$. Thus, $D_j = \{n_{lij}\}$, with $l = 1,\ldots,t_j$ and $i = 1,\ldots,m$, typically consists of a matrix with $t_j$ rows, which contains the voting pattern in each of the selected polling stations. Thus, $n_{lij}$ is the number of votes for party $i$ in polling station $l$ of region $j$.

Each of those rows may be considered as a random sample from a multinomial distribution with parameter vector $\omega_{lj} = \{\omega_{1lj},\ldots\omega_{mlj}\}$, where $\omega_{ilj}$ is the (unknown) probability that a citizen in polling station $l$ from region $j$ will vote for party $i$, so that $p(D_j|\omega_j) = \prod_{l=1}^{t_j} \mathrm{Mu}(n_{lj}|\omega_{lj})$. Moreover, since all polling stations are chosen to be representative of the area in the sense described above, the $\omega_{lj}$'s may themselves be safely considered as a random sample of size $t_j$ from representative polling stations in region $j$. A Dirichlet hyperprior is a useful choice at this level, so that $p(\omega|\omega_0) = \prod_{l=1} t_j \mathrm{Di}(\omega_{lj}|\omega_0)$. Finally, for an objective Bayesian analysis, a *reference* (noninformative) prior $\pi(\omega_0)$ (which only depends of the model chosen, with no added subjective elements) has to be chosen for the hyperparameter vector $\omega_0$. For details of the choice of the reference prior, see [1, 2] and references therein.

Standard use of Bayes theorem will then yield $\pi(\omega_j|D_j)$, the reference posterior density of $\omega_j$, the unknown vote distribution vector for region $j$. This encapsulates all available information on the voting in region $j$, and this would be the final result of the analysis for region $j$ if the distribution of the vote among the parties was all what was required. However, as stated before, what it is really required is the probability distribution for the seat allocation. This is now however, easily implemented.

Each electoral law provides a particular algorithm that associates a seat distribution to each possible electoral result. Thus, if $s_j$ seats are to be allocated in region $j$, there is a well defined function $F$ of the form $F(\omega_j|s_j) = \theta_j$, with $\theta_j = \{s_{1j},\ldots,s_{mj}\}$, which allocates $s_{ij}$ seats to party $i$, with $\sum_{i=1}^m s_{ij} = s_j$ if the vote distribution in the region has been $\omega_j = \{\omega_{1j},\ldots,\omega_{mj}\}$, with $\sum_{i=1}^m \omega_{ij} = 1$. Notice that $F$ is a mathematical function, in the sense that to each vote distribution $\omega_j$ associates one, and only one, seat distribution $\theta_j$, but this is typically *not* an analytical expression but a relatively complex algorithm. In many European countries, including Spain, the algorithm stated in the electoral law is the rule d'Hondt. Since the seat distribution $\theta_j$ is a mathematical function of the vote distribution $\omega_j$, standard probability theory may be used to obtain the reference posterior distribution of the seat allocation $\pi(\theta_j|D_j)$, which is simply that implied by the reference posterior density of the vote distribution, $\pi(\omega_j|D_j)$. The reference posterior distributions obtained for the each of the $r$ regions may finally be combined, by convolution, to obtain the reference posterior distribution $\pi(\theta|D_1,\ldots,D_r)$ for the seat allocation in the parliament, which is the desired result. The marginal distributions of $\pi(\theta|D_1,\ldots,D_r)$ directly provide a precise forecast for the number of seats that, given the data, each of the parties may be expected to obtain.

Since generally $F$ is not a simple analytical function, but a rather complex algorithm, all these computations must be carried out numerically, typically using Monte Carlo integration techniques.

## 4 Example

The procedure described above has been systematically used in Spain for many years, for all types of elections: European, general, state and local. The table below reproduces the outcome obtained in the province of Valencia in a particular general election, where 15 seats were to be distributed among the parties which reached the legal 3% threshold. This was done using the first 100 counted votes of the best 20 polling stations in province of Valencia. The results were obtained at 9 pm, just one hour after the polling stations closed, many hours before any other accurate estimate for the results was published. Bold numbers indicate the final official result.

**Table 1:** *Area seat distribution probability forecast*

| Probability | Party 1 | Party 2 | Party 2 | Party 4 | Party 5 |
|---|---|---|---|---|---|
| 0.75 | **10** | **5** | **0** | **0** | **0** |
| 0.12 | 9 | 5 | 0 | 0 | 1 |
| 0.11 | 9 | 6 | 0 | 0 | 0 |
| 0.02 | 11 | 4 | 0 | 0 | 0 |

Thus, the final configuration was given a reference posterior probability of 0.75, and the analysis shoved that there was a probability of 0.12 that Party 1 lost one seat to Party 5, and a probability of 0.11 that Party 1 lost one seat to Party 2, indicating the existence of a non-secured seat, which would probably depend on a handful of votes.

## References


1. J. M. Bernardo. Reference posterior distributions for bayesian inference. *Journal of the Royal Statistical Society, Series B*, 41:113–147, 1979.
2. J. M. Bernardo. Reference analysis. *Handbook of Statistics 25*, 2005.
3. J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. Wiley, 1994.
4. S. Kullback. *Information theory and Statistics*. Dover, 1994.