Mathematics Internal Assessment

Using Bayesian Analysis to Predict Election Results

February 7, 2023

Contents

1 Introduction

For as long as I can remember, I have been fascinated by politics, from the power dynamics that have shaped recent history to the magnificent system in which we live, a democracy. Although democracies are not without their flaws, particularly when we consider the current voting system used in Canada, they are arguably the best political system ever created by mankind.

A highly interesting event that results from a democratic election is the night right after where the nation awaits for the final results, slowly receiving updates for the current ballots counts for different constituencies. While this is happening, news agencies are trying to use their current data to predict the final results. This process of highly confidently predicting the final results of constantly updating data while trying to make that prediction as soon as possible has long been a source of interrogation for me. Impressively, news agencies are ridiculously fast at forming their predictions, like when Radio-Canada successfully predicted that the Coalition Avenir Québec would form a majority government less than 11 minutes after results started to come in for the Québec 2022 election [8]. Furthermore, although they occasionally make wrong predictions [4], this is exceedingly rare.

In short, I started to wonder about how news agencies could be so fast and so accurate. This paper will be my attempt at exploring a model to make electoral predictions, so that I can better understand the seemingly magical tools that are used. It is to be noted that my goal here is not to reverse engineer how existing systems work, as I do not have access to the same data that news agencies have.

As a Canadian citizen, the “first-past-the-post” election system is the one I am most familiar with and will therefore use for this paper. In general, Canadian elections are divided into many constituencies, smaller districts of similar size in terms of population, where each constituency elects the most popular candidate within it. This is determined by looking at which candidate had the most votes amongst all running inside of it.

Out of all the possible ways to approach such a problem, the one I found the most interesting was to model the situation as a conditional probability problem, as it is a very theoretical approach and I was curious to know if it could accurately represent the real world. Other approaches, such as regression or hypothesis testing, would be quite interesting extensions to this paper.

However, as building a full model for electoral predictions is a very complex task that requires many steps, I will focus on only one of them for this paper: quantifying the probability that each candidate in a constituency has to receive the next vote.

The process we will follow in order to find those probabilities is generally referred to as *Bayesian analysis* [11]. First, we will quantify our prior beliefs (what we think about the probabilities we are searching for) before we observed the evidence (the current votes count) using information from surveys. Then, we will consider the evidence we have and its likelihood of happening. Finally, we will combine both of these pieces of information together to get our posterior beliefs (what we think the probabilities are after observing the evidence). Doing these steps will allow us to find the probabilities we are searching for.

2 Building the Model

As with any mathematical problem, a considerable portion of building the model is simply to lay down our assumptions and to split the task into multiple, more specific, problems.

To approach this using the tools of conditional probability, we first need to understand why predicting election results even involves random events. The fundamental assumption we need to do here, from which all of the mathematics will follow, is that we can consider each individual casting its vote as an independent random event where the different possibilities are the different candidates in the constituency, with each candidate having a different probability of receiving a vote.

Let's unpack this. Essentially, we can imagine that the probability that a voter will vote for a given candidate (what we are searching for) is the final proportion of votes that that candidate will have received in the final results.[1] Furthermore, each vote would be independent of the other ones, because election results aren't shown until every polling booth is closed.[2]

Let's start by defining a few variables. Let n be the number of candidates in the constituency. Let $v = \{v_1, v_2, v_3, \ldots, v_n\}$ be the set of the current vote counts for the different candidates, ordered from largest to smallest, where v_1 is the number of votes for candidate 1, v_2 is the number of votes for candidate 2, etc. In general, when discussing a certain candidate, I will refer to it as the kth-candidate. For example, I consider the candidate k to currently have v_k votes. Also, let $v_t = \sum_{k=1}^{n} v_k$ be the total number of votes.

Finally, let $D = \{D_1, D_2, D_3, \ldots, D_n\}$ be the set of results we are searching for. This is essentially the set of our posterior beliefs about the probability each candidate has to win the next vote. One caveat to address is that the elements of the set D will not truly be probabilities, but rather probability distributions representing how probable each of the infinitely many possible probabilities each candidate has to receive the next vote are. This idea of probabilities of probabilities will be detailed below in Section 2.1.

Due to the usefulness of specific, visual examples when trying to investigate probability questions, let's use the following variables as a simple and concrete example:

$$n = 5$$

$$v = \{60, 50, 36, 34, 20\}$$

$$v_t = 60 + 50 + 36 + 34 + 20 = 200$$

[1]This assumes that the election runs indefinitely so that the limited quantity of votes in a constituency can be ignored. This turns out to be a quite good approximation as most electoral predictions will be made when relatively few votes have been counted.

[2]For federal elections, due to the large timezone differences, the results of some of the Eastern provinces are compiled before polls close in some of the Western provinces. However, there is, overall, very little overlap.

This means that we will be looking at a 5-candidate election (n), where the leading candidate currently has 60 votes (v_1) and where the last candidates currently has only 20 votes (v_5).

Although this set of data will be used for numerical and graphical example, this paper will not focus on the computation of specific numerical examples, as the endgoal is to have a generalized computer model. Furthermore, due to their nature, many of the computations discussed here have no analytical solutions, which is why computer based approximations will be favoured.

2.1 Probability of Probabilities

A recurrent theme in this paper will be the idea of *probability of probabilities*. Although this may seem like an utterly nonsensical statement at first, it is actually at the root of many advanced concepts in conditional probability. In order to explore this idea, let's use an example situation.

Considering a biased coin whose mathematical weight (bias) is unknown, after observing 90 heads and 10 tails out of 100 trials, what should we expect the bias to be?

One might argue that the answer is trivial: to find the weight, we divide the number of observed heads (or tails) by the number of throws. This goes with the idea of the *Law of large numbers* [15], that the more trials we observe, the more the observed frequency will approach the theoretical (the real) probability.

However, I would argue that this reasoning is flawed. Yes, $\frac{90}{100} = 0.9$ is the most likely probability, but it is possible that the *true* probability is 0.1, 0.99 or any other value between 0 and 1 (exclusively). An event being unlikely does not mean it is impossible.

The better approach is therefore to use probability distributions: instead of trying to define the weight of the coin with a single number, we can define a probability distribution that represents how likely each of the infinitely many possible values of the bias are. That probability distribution would most likely be a beta distribution, which we will explore below, in Section 2.2.

Coming back to our objective, finding the probability each candidate has of receiving the next vote, I hope it is now clear how answering this question with probability distributions (the set D) showing how some probabilities of receiving a vote are more likely than others for a given candidate is actually providing much more useful information about the probability each candidate has to receive the next vote than simply summarizing the entire distribution in a single number, the mean of the distribution for example.

2.2 Understanding the Beta Distribution

As we will heavily rely on it, it is important that we understand the beta distribution. Two reasons make it ideal for representing probabilities of probabilities: its domain is $[0,1]$ and the area under a beta distribution's Probability Density Function (PDF) over its range is 1. This means that any value on the x-axis represents a possible probability and that the y-value of the distribution at that point represents the probability density that that probability is the true one.

Furthermore, the beta distribution can take a variety of shapes, as its PDF is, most commonly, defined in terms of two shape parameters, α and β, both being positive non-null real numbers. Its definition is based on the beta function, here denoted $\mathcal{B}$ [12]. Let's define a distribution X such that $X \sim \text{Be}(\alpha, \beta)$, where Be is the beta distribution.

$$P(X = x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathcal{B}(\alpha, \beta)}, x \in [0, 1]$$

This distribution would have a mean of [12]:

$$E(X) = \mu_X = \frac{\alpha}{\alpha + \beta}$$

Dividing by the beta function has the effect of scaling the numerator in order to make the area under the beta distribution's PDF equal to 1. It is therefore equal to the integral of the numerator.

$$\mathcal{B}(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, \mathrm{d}x$$

However, it is more commonly defined as follows, where Γ is the gamma function [13]:

$$\mathcal{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

It is to be noted that both these definitions are equivalent.

Finally, the gamma function can be viewed as an expansion of the factorials to the Reals (except for integers smaller than 0) while respecting the following identity [14], n being a positive integer[3]:

$$\Gamma(n) = (n-1)!$$

Here are a few beta distributions plotted, demonstrating some of the various shapes it can take:

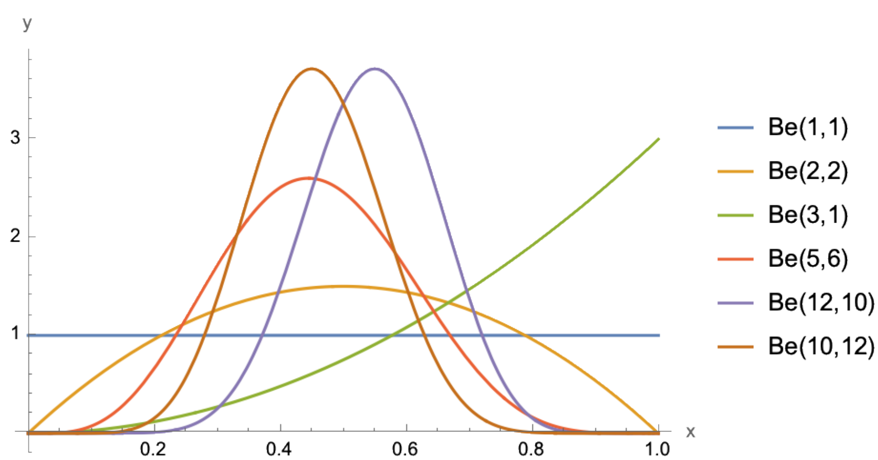

Figure 1: A few beta distributions

In Figure 1, we can see multiple interesting things, notably that a Be$(1,1)$ distribution is equivalent to a Uniform$(0,1)$ distribution [16][4] and that the beta distribution can be both

[3]A more detailed explanation of the gamma function has been deemed outside of the scope of this investigation.

[4]A uniform distribution is a distribution where all values in a given interval (in this case, $[0,1]$) are equally likely.

symmetric and highly asymmetric about the average. Now that we understand the beta distribution, we can go back to building the model.

2.3 Building Prior Beliefs

Our prior beliefs, as the name implies, is what we believe the probability distribution to be before seeing the evidence (the partial election results, in our context). We express them in the form of a probability distribution. In our context, there are two ways we can approach this: prior ignorance and substantial prior knowledge [3]. This process of quantifying our prior beliefs is often referred to as *prior elicitation* [2].

Prior ignorance is really quite easy: we assume we know nothing before the election. Therefore, we need a distribution illustrating that we consider all probabilities to be equally likely. This is the perfect use for the uniform distribution, so we would say that our prior beliefs about the probability distribution representing the probability a certain candidate has of receiving the next vote (D_k) follows a Uniform$(0, 1)$ distribution (also known as a Be$(1, 1)$ distribution).

Substantial prior knowledge is quite a bit less trivial. First, let's define exactly what this means. In Bayesian analysis, we say we have substantial prior knowledge "[when] expert opinion, for example, gives us good reason to believe that some values in a permissible range for [the probability] are more likely to occur than others." [2] In our case, expert opinions could be the polls and surveys from firms like LÉGER, who usually publish there results a few weeks before any major election. An example of such a report could be LÉGER's *ÉLECTIONS PROVINCIALES : MONTRÉAL ET LAVAL* [5], which contains two key pieces of information:

- The voting intentions (what percentage of people plan to vote for each of the parties).
- The firmness of the intentions (for each party, what percentage of people don't expect to change their minds).

For example, suppose we knew from a report that 35% of the citizens intended to vote for a given party, and that 45% of those people are quite firm about their decision, how could we transform this into a probability distribution? For the reasons outlined in Section 2.2, it seems reasonable to try building a beta distribution. Let's therefore define our prior beliefs distribution as $U \sim \text{Be}(\alpha, \beta)$.

First, we know that our expected value (the mean of the distribution) should be 35 % (0.35). Then, we could define "quite firm" as being within ±5% of the mean. The probability of landing in that range must therefore be equal to 45 % (0.45). This is equivalent to stating that the area under the PDF of our distribution in the range $[0.30, 0.40]$ should be equal to 0.45. Let's write a system of equation using both of these facts:

$$\begin{aligned} 0.35 &= E(U) \\ &= \mu_U \\ &= \frac{\alpha}{\alpha+\beta} \end{aligned} \qquad \begin{aligned} 0.45 &= \int_{0.30}^{0.40} P(U = x)\,\mathrm{d}x \\ &= \int_{0.30}^{0.40} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{\mathcal{B}(\alpha, \beta)}\,\mathrm{d}x \end{aligned}$$

We should also keep in mind that both α and β need to be positive in order to satisfy the requirements of the beta distribution. As there is no trivial analytical solution to this system of equations, the most efficient solution is to resort to numerical approximation to solve for α and β. It is to be noted that this system of equations may not always yield a solution when considering extreme requirements, like having an exceedingly small margin around the mean for the definition of "quite firm." This, however, is not really an issue as these cases would lead to such certain prior beliefs that any evidence would hardly be relevant.

Using Wolfram Mathematica [17] or similar software, we can find that this system is solved by $\alpha \approx 11.486$ and $\beta \approx 21.330$. This gives us $U \sim \text{Be}(11.486, 21.330)$ for the probability distribution representing our prior beliefs. This distribution is illustrated in Figure 2.

It is important to keep in mind that this process is quite subjective. In fact, we chose to define "quite firm" as being within ±5% of the mean, but we could have chosen ±7%, ±3%

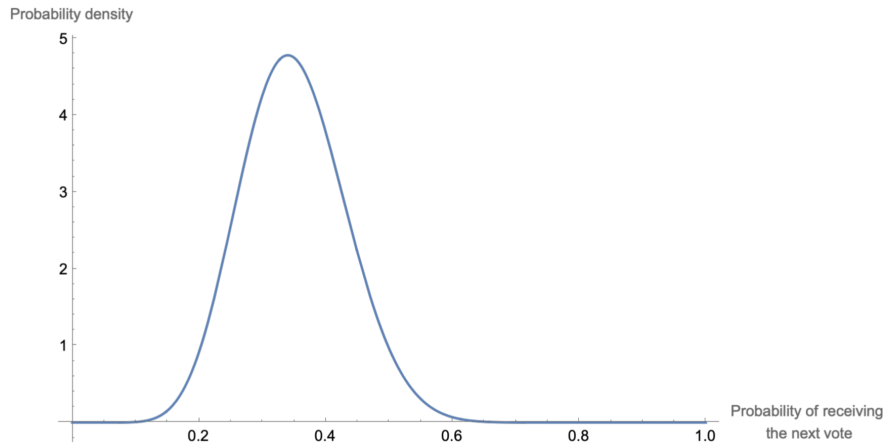

Figure 2: Plot of the probability distribution built from prior knowledge

or any other value. This is the main weakness of this process: our biases can easily sneak into our statistics if we are not careful.

If this margin around the mean is too small, our prior beliefs will be so concentrated that the evidence, the actual vote counts during the elections, will not really matter. On the opposite side, if it is too wide, our prior beliefs will become insignificant and approach prior ignorance. The margin $\pm 5\%$ was chosen as a compromise between these two extremes.

As our prior beliefs can be represented as a beta distribution no matter if we have prior ignorance or prior substantial knowledge, it makes sense to define our prior beliefs for the candidate k as $D_k \sim \mathrm{Be}(a_k, b_k)$ before seeing any of the evidence. For the rest of this investigation, all of our prior knowledge about the candidate k will be referred to with the variables a_k and b_k shaping this distribution. We can now write our prior beliefs as follows:

$$P(D_k = p) = \frac{p^{a_k - 1}(1-p)^{b_k - 1}}{\mathcal{B}(a_k, b_k)}$$

As we can see, we can summarize our entire prior beliefs about a candidate's probability of receiving the next vote (which is equivalent to its share of votes in the end) using only two parameters: a_k and b_k. By repeating the process of prior elicitation for each of the candidates, we could obtain the values of these two parameters for each of them.

Finally, it is to be noted that we have not yet found the actual distribution D_k just yet. This only represents our prior beliefs about it. To figure out the actual distribution D_k, we will also need to consider the evidence (the number of votes).

2.4 Building the Likelihood Function

Seeing this from the perspective of each of the candidates, we can consider the number of votes received over the total number of votes as a binomial experiment, where a *success* is defined as a vote for that candidate and a *failure* as a vote given to any other. As a reminder, the Probability Mass Function [7] (PMF), the discrete analogue of the PDF, for a binomial distribution Y, $Y \sim \mathrm{B}(m, p)$,[5] would be the following, where p is the probability of the event happening and m is the total number of trials:

$$P(Y = x) = \binom{m}{x} p^x (1-p)^{m-x}, x \in \{0, 1, 2, \ldots, m\}$$

In our case, we know both the number of successful trials, v_k, (the current number of votes for the candidate) and the total number of trials, v_t, (the current total number of votes). This means that, for the candidate k, with number of votes v_k, the unknown left is the probability, here p, of receiving a vote, which is distributed from the unknown distribution D_k, D_k being the distribution representing the probability that the candidate has a certain probability of receiving the next vote. We can therefore rewrite the above equation by building a binomial distribution $V_k \sim \mathrm{B}(v_t, p)$.

$$P(V_k = v_k \mid D_k = p) = \binom{v_t}{v_k} p^{v_k} (1-p)^{v_t - v_k}$$

However, as the distribution V_k is not really important as we only care about the distribution D_k, we could also represent the above as follows. This notation is quite common in Bayesian

[5] Here, m is used instead of the typical n in order to avoid confusion with the number of candidates in the constituency.

analysis.

$$P(v_k \mid D_k = p) = \binom{v_t}{v_k} p^{v_k} (1-p)^{v_t - v_k}$$

As what really interests us is the unknown distribution D_k, we can rewrite this as its likelihood function [11], $L_{D_k}(p)$, which will answer the question: Based solely on the evidence, how likely is it that a certain value of the probability p is the true probability that lead to the observed events?

$$\begin{aligned} L_{D_k}(p) &= P(v_k \mid D_k = p) \\ &= \binom{v_t}{v_k} p^{v_k} (1-p)^{v_t - v_k} \end{aligned}$$

Here is the plot of this function for the leading candidate ($k = 1$) in our example, considering it currently has $v_k = v_1 = 60$ votes and that the total number of votes is $v_t = 200$:

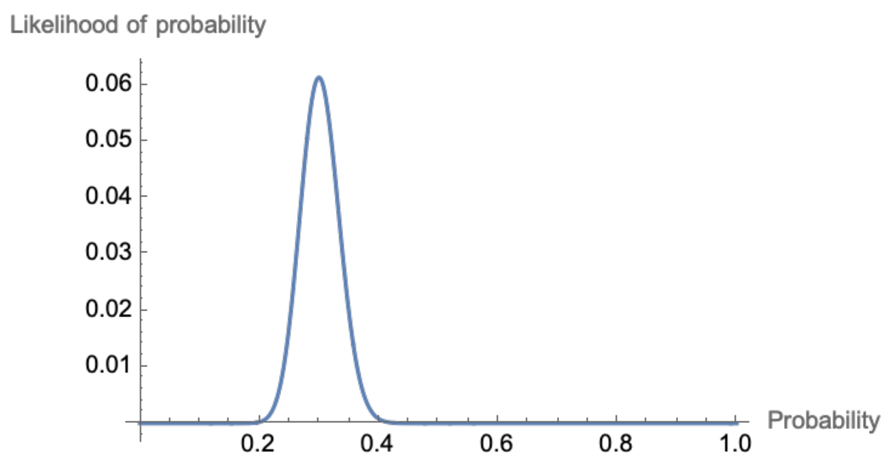

Figure 3: Plot of the likelihood function for the leading candidate

Referring back to Section 2.1, this is an example of a probability distribution representing an unknown probability. We should still expect the mode of our distribution, its maximum, to be the simple frequency calculation $\frac{v_1}{v_t} = \frac{60}{200} = 0.3$, which we can verify to be true in Figure 3.

2.5 Combining Prior Beliefs and Likelihood

Now that we know how to form our prior beliefs and our likelihood function representing the observed evidence, it is time to combine them into the probability distribution for the share of votes of a candidate, our posterior beliefs.

This is where Bayes' theorem comes in. In fact, this theorem gives a systematic method to mix prior beliefs and observed evidence (summarized by the likelihood function) into posterior beliefs[6]. As a reminder, here is the formula for said theorem [10], where A and B are independent random events:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

However, I dislike this depiction of Bayes' theorem as it abstracts and hides its true beauty. Exploring each of the terms leads us to the following:

$\boldsymbol{P(A \mid B)}$ This represents our *posterior beliefs* about A, considering that B happened.

$\boldsymbol{P(B \mid A)}$ This represents the *likelihood* that A happens given the observed evidence for B.

$\boldsymbol{P(A)}$ This represents our *prior beliefs* about A.

$\boldsymbol{P(B)}$ This represents the total probability of B. Essentially, this has the effect of scaling the probability of $A|B$ such that it lands between 0 and 1. In the case of probability distributions, this ensures that the area under the distribution's curve equals 1 [1].

It is also interesting to note that $P(B \mid A)$ and $P(A)$ can not only be probabilities, but also probability distributions, making $P(A \mid B)$ into one too.

As $P(B)$ is simply a scaling constant, we can rewrite this formula as

$$P(A \mid B) \propto P(B \mid A)P(A)$$

However, I believe that the following is a much more elegant way to describe Bayes'

[6] A justification for Bayes' theorem has been deemed outside of the scope of this investigation.

theorem [1]:

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

The beauty of this lies in how clearly it highlights how evidence (likelihood) doesn't replace our prior beliefs, but rather updates them to form our posterior beliefs [9].

But how could we apply this to our situation? Let's rewrite this in terms of our variables and explore each of the terms, keeping in mind $p \in [0, 1]$:

$$P(D_k = p \mid v_k) \propto P(v_k \mid D_k = p)P(D_k = p)$$

$\boldsymbol{P(D_k = p \mid v_k)}$ This is the probability distribution D_k (as a function of p) we are searching for.

$\boldsymbol{P(v_k \mid D_k = p)}$ This is the likelihood function we derived earlier, $L_{D_k}(p)$.

$\boldsymbol{P(D_k = p)}$ This is the prior beliefs distribution we derived earlier.

As we can see, all of our work is really coming together. Let's substitute the terms with our findings from the previous subsections. As we are working in a proportionality context, we can ignore all the terms that are not a function of p.

$$\begin{aligned}
P(D_k = p \mid v_k) &\propto P(v_k \mid D_k = p)P(D_k = p) \\
&\propto \binom{v_t}{v_k} p^{v_k}(1-p)^{v_t - v_k} \cdot \frac{p^{a_k - 1}(1-p)^{b_k - 1}}{\mathcal{B}(a, b)} \\
&\propto p^{v_k}(1-p)^{v_t - v_k} \cdot p^{a_k - 1}(1-p)^{b_k - 1} \\
&\propto p^{v_k + a_k - 1}(1-p)^{v_t - v_k + b_k - 1}
\end{aligned}$$

There are three things to notice and recall here: (I) As this distribution represents possible values of a probability p, its domain is $[0, 1]$. (II) As with any other continuous probability distribution, its area over its range (here, $[0, 1]$) must be equal to 1. (III) The beta distribution matches both the form of the equation we obtained and the above two criteria.

Finding the beta distribution corresponding to our above equation is simply a question of identifying the values of the unknown parameters. In a beta distribution $\text{Be}(\alpha, \beta)$ whose PDF is expressed as a function of x, x is raised to the power of $\alpha - 1$ and $1 - x$ is raised to the power of $\beta - 1$. Applying this to our example, where the distribution is expressed as a function of p, we get the following coefficients and, therefore, the following distribution:

$$\alpha - 1 = v_k + a_k - 1 \qquad\qquad \beta - 1 = v_t - v_k + b_k - 1$$

$$\alpha = v_k + a_k \qquad\qquad \beta = v_t - v_k + b_k$$

Therefore

$$D_k \mid v_k \sim \text{Be}(v_k + a_k, v_t - v_k + b_k)$$

As a reminder, D_k is the distribution representing the probability that the candidate k will receive the next vote.

It is interesting to note that both our prior and posterior beliefs are beta distributions when the likelihood function comes from a binomial distribution. In Bayesian analysis terminology, we would describe this by saying that the beta distribution is a conjugate prior to the binomial distribution [6].

As detailed polls before elections are often not existent, especially in more rural areas, the case where we have to assume prior ignorance is quite common. As a reminder, prior ignorance about the probability of candidate k of receive the next vote is equivalent to setting the parameters summarizing our prior beliefs about its probability, a_k and b_k, equal to one.

$$D_k \mid v_k \sim \text{Be}(v_k + 1, v_t - v_k + 1)$$

This equation is therefore the representation of the specific case where there is no substantial prior knowledge.

For the sake of visual understanding, let's visualize our findings for each of the candidates in our example in Figure 4 (this graph was plotted assuming prior ignorance for all the candidates).

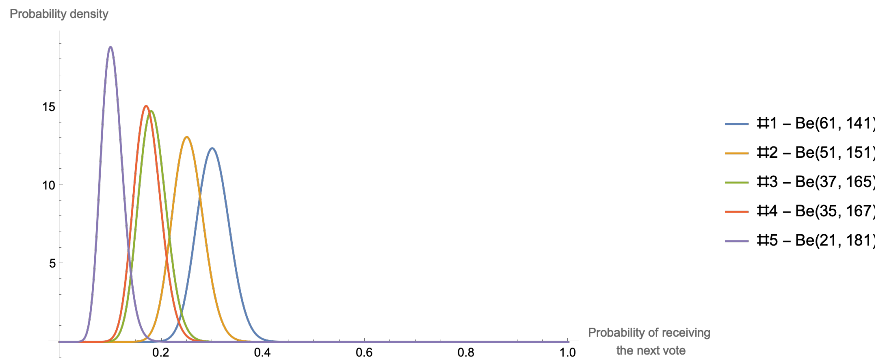

Figure 4: The set of distributions D

These curves show us the probability density that a certain candidate has a certain probability of receiving the next vote. As we can see, the leading candidate (illustrated in blue) has the rightmost probability distribution indicating that he is the most likely to receive the next vote, just as we would have expected. More generally, we can see how the number of votes relates to the position of the probability distributions: the more votes a candidate has, the more to the right its distribution is. We can also see how the spread of the distributions is influenced by the relative vote counts. For example, the last candidate (#5) is very unlikely to receive the next votes as he only has 20 votes out of 200, which gives it a probability distribution quite concentrated to the left.

3 Conclusion

In conclusion, when watching an election night, we can compute the probability distribution representing the probability that a certain candidate in a given constituency has to receive the next vote to be counted.

To do so using conditional probability and Bayesian analysis, we first need to quantify our prior beliefs about the candidate's chance of receiving a certain share of the votes, which can be done with the help of surveys. Those prior beliefs can be summarized using the two parameters (a_k and b_k, for a candidate k) shaping the beta distribution representing them. In the case where we do not have access to survey data, we need to assume prior ignorance, which we can do by setting both of these parameters equal to one.

Then, we can combine these prior beliefs and the observed evidence (the current vote counts for the different candidates in the constituency) to form our posterior beliefs (what we believe about the probability that the candidate has to receive the next vote, considering the evidence). From this, we were able to derive the following formula, where D_k is the distribution we are searching for, v_k is the vote count for the candidate we are interested in and v_t is the total vote count:

$$D_k \mid v_k \sim \mathrm{Be}(v_k + a_k, v_t - v_k + b_k)$$

Our beliefs about the probability a candidate has of receing the next vote are represented using a probability distribution instead of a simple probability, because, just like we can never truly be certain about the bias of an unfair coin, it is impossible to determine it with full certainty. A probability distribution instead shows us how likely all the infinitely many possibilities for the value of that probability are.

Using the mathematics described in this paper, I was able to build and deploy a web form that generates the posterior distributions on the fly based on the vote counts and survey data input by the user. It is available here:

https://www.wolframcloud.com/obj/
cae4ebda-250f-46fe-bde8-05327580b6ad

The above distribution gives us insight about the probability that the candidate has of receiving the next vote. This information would be invaluable when building a full model to predict election results based on Bayesian analysis, as it is the very first step required to do so.

The next steps would be to devise an algorithm to compare probability distributions, to extend the distribution we found above to take into account the number of votes left to be counted in the election and to calibrate the model using real-world data. It would be an utmost interesting extension to this paper to carry on the work started here to complete a full model based on these steps.

4 Bibliography

[1] L. Fawcett. "MAS2317/3317: Introduction to Bayesian Statistics — Chapter 2: Bayes' Theorem for Distributions — Lecture Notes." (2015), [Online]. Available: https://www.mas.ncl.ac.uk/~nlf8/teaching/mas2317/notes/chapter2.pdf (visited on 12/18/2022).

[2] L. Fawcett. "MAS2317/3317: Introduction to Bayesian Statistics — Chapter 3: Prior Elicitation — Lecture Notes." (2015), [Online]. Available: https://www.mas.ncl.ac.uk/~nlf8/teaching/mas2317/notes/chapter3.pdf (visited on 12/18/2022).

[3] L. Fawcett. "MAS2317/3317: Introduction to Bayesian Statistics — Chapter 3: Prior Elicitation — Slides." (2015), [Online]. Available: https://www.mas.ncl.ac.uk/~nlf8/teaching/mas2317/notes/slides3.pdf (visited on 12/18/2022).

[4] L. Gagné, "Marie-Josée Savard déclarée gagnante par erreur : mais que s'est-il passé? — Élections municipales au Québec 2021," *Radio-Canada.ca*, [Online]. Available: https://ici.radio-canada.ca/nouvelle/1838339/erreur-prediction-victoire-course-mairie-quebec-medias-radio-canada-tva-marie-josee-savard-bruno-marchand (visited on 11/27/2022).

[5] Léger. "Élections provinciales : Montréal et laval - 10 septembre 2022." (Sep. 10, 2022), [Online]. Available: https://leger360.com/fr/intentions-de-votes/elections-provinciales-montreal-et-laval-10-septembre-2022/ (visited on 12/19/2022).

[6] J. Orloff and J. Bloom. "Beta Distributions." (n.d.), [Online]. Available: https://math.mit.edu/~dav/05.dir/class14-prep-a.pdf (visited on 12/22/2022).

[7] PennState: Statistics Online Courses. "7.2 - Probability Mass Functions — STAT 414." (n.d.), [Online]. Available: https://online.stat.psu.edu/stat414/lesson/7/7.2 (visited on 12/18/2022).

[8] Radio-Canada Info. "Élections Québec 2022 : La soirée électorale." (Oct. 3, 2022), [Online]. Available: https://www.youtube.com/watch?v=EkV28Wew-1Q (visited on 12/17/2022).

[9] Veritasium. "How To Update Your Beliefs Systematically - Bayes' Theorem." (Apr. 5, 2017), [Online]. Available: https://www.youtube.com/watch?v=R13BD8qKeTg (visited on 12/23/2022).

[10] E. W. Weisstein. "Bayes' theorem." (n.d.), [Online]. Available: https://mathworld.wolfram.com/BayesTheorem.html (visited on 12/19/2022).

[11] E. W. Weisstein. "Bayesian analysis." (n.d.), [Online]. Available: https://mathworld.wolfram.com/BayesianAnalysis.html (visited on 12/18/2022).

[12] E. W. Weisstein. "Beta distribution." (n.d.), [Online]. Available: https://mathworld.wolfram.com/BetaDistribution.html (visited on 12/18/2022).

[13] E. W. Weisstein. "Beta function." (n.d.), [Online]. Available: https://mathworld.wolfram.com/BetaFunction.html (visited on 12/18/2022).

[14] E. W. Weisstein. "Gamma function." (n.d.), [Online]. Available: https://mathworld.wolfram.com/GammaFunction.html (visited on 12/18/2022).

[15] E. W. Weisstein. "Law of Large Numbers." (n.d.), [Online]. Available: https://mathworld.wolfram.com/LawofLargeNumbers.html (visited on 12/23/2022).

[16] E. W. Weisstein. "Uniform distribution." (n.d.), [Online]. Available: https://mathworld.wolfram.com/UniformDistribution.html (visited on 12/18/2022).

[17] "Wolfram Mathematica: Modern Technical Computing." (n.d.), [Online]. Available: https://www.wolfram.com/mathematica/ (visited on 12/19/2022).