Inference of Secret Voting's Distribution by Open Voting Data of the National Assembly

Changyoung Koh, Dongkwan Kim, Donggyun Kim

June 6, 2018

Disclaimer This work was done as part of Probabilistic Programming (KAIST CS492) project. All the names are anonymized and the results only work based on our strong assumptions.

1 Introduction

On May 21, a motion to arrest Hong Moon-chong and Yeom Dong-Yeol, members of the Jayuhangug Party, was rejected by the National Assembly. The National Assembly rejected the proposal for arresting Hong, who received embezzlement and bribery charges, with 129 consents, 141 dissents, 2 withdrawals and 3 absences. Yeom, who has been charged with abuse of power and obstructing his duties, has also voted against the bill with 98 consents, 172 dissents, 1 withdrawals and 4 absences. Since the Jayuhangug party has only 113 seats, it is presumed that a lot of 'sympathy signs' for fellow lawmakers have come out regardless of the political affiliation, especially from Deobuleominju, which is the ruling party. We want to find the 'traitors' from the Deobuleominju party, but the vote was conducted under anonymity, so we can't find out them explicitly.
However, National Assembly stores ballot results of every open votes in their own database and any people can access it. Thus we decided to find out 'traitors' of the secret votes based on the ballot data of 20^{th} (current) National Assembly.
First, we constructed a model about the consistency of vote results between two members, one for our key member (Hong or Yeom), the other for each remaining members who attended at the day. Then we inferred the probability to vote same as each key member in open votes for all members, and based on this, we finally inferred the probability to vote same as each key member in secret votes. Although we cannot verify our results, we found some interesting features in our project. We used Interacting Particle Markov Chain Monte Carlo inference algorithm for every Anglican query.

2 Inference

In this project, we have three assumptions. First, Mr. Hong and Yeom voted to dissent to their warrant requests. This is very reasonable. Second, a person who has a similar trend in an open vote with them will also show a similar trend in a secret vote with them. It is slightly strong but acceptable. Finally, all people make decisions independently except for Mr. Hong and Yeom. This is very strong but necessary because it is difficult to find and infer a suitable model with Anglican that does not include this assumption.
After encoding ballot result of open votes, we made a posterior distribution with the proportion of same vote results with each key member. Starting from $beta(a, b)$ whose a, b is drawn from $poisson(1) + 1$, update posterior with open voting data. We used $poisson(1) + 1$ one to make prior similar to uniform distribution which is $beta(1, 1)$. This posterior distribution, which we will denote it by P, means the probability of whether given member votes like as Mr. Hong or Yeom votes in open votes.

$$P \sim beta(a, b)$$

```
(defquery non-secret [prob]
  (let [a (+ 1 (sample (poisson 1)))
        b (+ 1 (sample (poisson 1)))]
    (observe (beta a b) prob)
    (sample (beta a b))))

(defn sample-prob [i n key-mem]
  (let [data (if (= key-mem 0) (vote-ith1 i) (vote-ith2 i))
        freq (reduce + 0 data)
        len (count data)
        prob (/ freq len)]
    (map :result
        (take n (take-nth
                    10 (drop 1000 (doquery :ipmcmc non-secret [prob])))))))
```

Since tendency to voting will different in open votes and secret votes, we set a random variable that is multiplied to previous probabilities to perturb them to match the secret votes. We assumed the modified truncated normal prior, where 'modified' means we gave all outer densities to two boundary points. At first, we sampled m, s from a uniform continuous distribution from zero to one.

$$T \sim \hat{N}(m, s)$$

Note that our assumption is Mr. Hong and Yeom vote dissent. So the summation of them will be the total number of dissents. Let S be the summation of samples of flip distribution with $P \cdot T$. Then we observed the number of dissents, 141 in the case of key member is Hong, from $N(S, 2)$. After doing this, $P \cdot T$ becomes the probability that a given member votes like as Mr. Hong or Yeom votes in a secret vote.

```
(defquery inference [probs key-mem]
  (let [a (sample (uniform-continuous 0 1))
        b (sample (uniform-continuous 0 1))
        pre-trans (sample (normal a b))
        trans (if (< pre-trans 0.5)
                    0.5
                    (if (> pre-trans 1.5) 1.5 pre-trans))
        real-sum (if (= key-mem 0) 172 141)]
    (loop [sum 0
           ps probs]
      (if (empty? ps)
        (do
          (observe (normal sum 2) real-sum)
          trans)
        (let [pre-p-con (* (first ps) trans)
              p-con (if (> pre-p-con 0.99) 0.99 pre-p-con)
              is-con (if (sample (flip p-con)) 1 0)]
          (recur (+ sum is-con)
                 (rest ps)))))))
```

3 Experiment

After inference of probability to give 'dissent' in the secret votes at May 21^{st}, we carried out experiments to find out 'traitors'. In the secret voting, total 275 people have participated. Here, the secret vote means that we dont even know names of the participants. However, we know those who attended assembly the day: 279 people. So we added 4 withdrawals to the result to mitigate missing 4 people. For the reliability, we collect 10 simulation results of votes for Mr. Hong and Yeom. Here, each simulation is done as explained previously: we simulate each members vote as a flip of $P \cdot T$ (true is dissent). Then we get a sum of the simulated dissents. If the sum is not reasonable, concretely, the difference between true dissent and simulated is larger than 2, then we discard the simulation.

```
(defquery final-experiment [probs trans]
  (loop [result '()
         ps probs]
    (if (empty? ps)
      result
      (let [pre-p-con (* (first ps) trans)
            p-con (if (> pre-p-con 0.99) 0.99 pre-p-con)
            is-con (if (sample (flip p-con)) 1 0)]
        (recur (cons is-con result)
               (rest ps))))))
```

So finally we get 10 reasonable simulations. From those simulated votes, we extracted members of the Deobuleominju party who made more than 7 dissents out of 10 total votes. Here, the parameter 7 is chosen through some tuning; we tuned it to make the number of extracted people close to those reported by news articles, which are based on the number of seats of each party.

4 Result

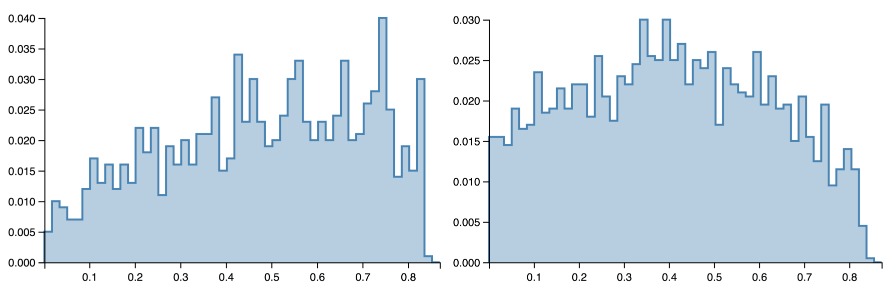

Figure 1: 2 Distributions from Mr. Hong's Case. Left is member K's distribution and Right is member C's.

For the simplicity, we will call the members of Deobuleominju party 'Blue people'.
First, on arrestment of Mr. Hong, we got 7 Blue people as result: namely `K, K, K, K, P, O, J`. In Figure 1, the histogram on the left shows Blue member K's distribution of $P \times T$, where K is one of the 7 people. On the contrary, the histogram on the right shows Blue member C's, who is not in the 7 people. We confirm that they are quite different from each other, just as our expectation.
As we mentioned earlier, there can be at most 135 total dissent votes from the right wing (108 from Red + 27 from Mint). However, the actual dissent was 141, which means votes from the left wings and independents are at least 6. The 6 yellows and 11 greens probably are consent to the arrestment as on the opposite wing. Hence, if we exclude the independent, then the result is consistent with the numbers.

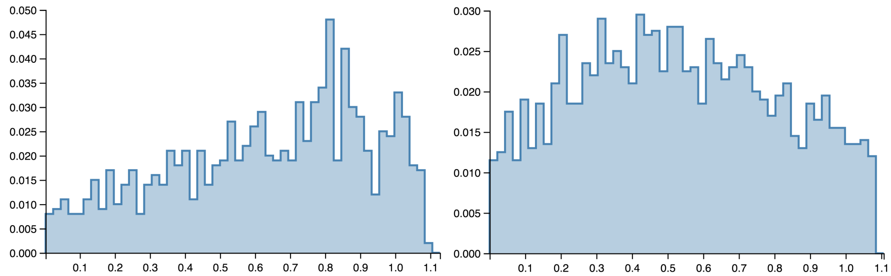

Figure 2: 2 Distributions from Mr. Yeom's Case. Left is member P's distribution and Right is member C's.

On Mr. Yeom's case, the total number of dissent votes was 172, which is ridiculously many. From this numbers, we got 25 Blue people as result: `K,K,K,K,K,K,K,D,P,P,P,S,S,S,E,O,W,Y,Y,L,J,J,C,J,C`. The distribution of a Blue member in 25 people and another Blue one not from 25 can be found in Figure 2.
Similarly to Mr. Hong's case, even though we assume every right-wing member voted to dissent, 37 are left. If we introduce strong assumption that members of Green and Independent voted to dissent, still 19 left and they are likely to be members of Blue party. From this observation, representative of the Blue party made a statement right after the secret vote: *"There were at least 20 dissent votes from our party. I feel responsibility."* Which makes our experiment more convincing. Moreover, 3 out of 7 people in Mr. Hong's case appears again in 25 people. It makes sense with our intuition, too.

5 Discussion

We could not find a meaningful relationship between party and voting tendency. From figure 3, each point is a member and each color represents a party. The x-axis is sample mean of P times T for Mr.Hong and y-axis is same for Mr. Yeom. There seems no relationship between color and position.
There are two possible interpretations of this issue. First, we made a strong independence assumption. Second, Mr. Hong and Yeom have so nice reputations from the wide range of assembly members, even even on the opposite side.

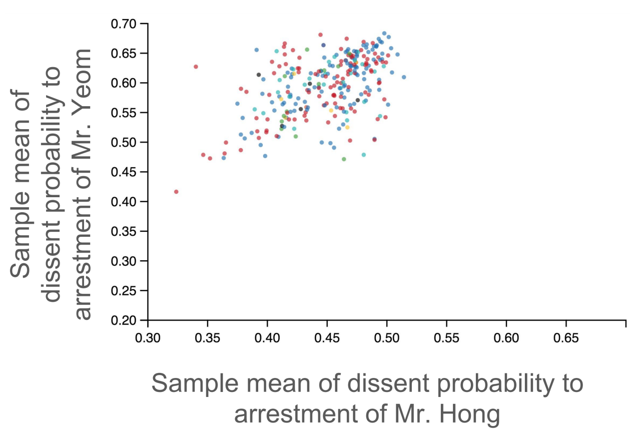

Figure 3: Scatter plot of members. Each point represents a member and its color represents party.

There are many interesting directions for future work. First, we can relax our assumptions to make more reliable and realistic inference. Second, incorporating more data such as contents of agendas will make us build a realistic model. Finally, we can check whether our model is consistent with open votes, evaluating the model with future open vote results.