

Classification of Trump Supporters with Dirichlet Mixture Models

Alex Leutenegger

6/1/2016

Background and Motivation

The purpose of this paper is to learn more about Donald Trump supporters. First, this is important because Trump may win the general election and become President of the United States. Second, even if he does not win, his supporters (and his message) will live on.

My hypothesis is that Trump represents a new movement in the recent history of American politics. Trump and his supporters have very different beliefs from those of their “fellow” Republicans. Trump has taken the Republican ticket, but I believe that many of his supporters should not be classified as Republican.

I chose to research Trump for another class last quarter. For the class, I collected survey data from the unofficial Trump subreddit. I did not conduct any quantitative analysis at that time. I will now do so in this paper.

Survey Methodology

SurveyMonkey poll of 10 questions. Posted 2/2/2016. 8 questions with responses “Strongly Agree”, “Agree”, “Disagree”, and “Strongly Disagree.” One question for age and another for gender.

Questions (Codes):

1. Cultural Marxism exists and it is a threat to the United States (Marx)
2. Sharia Law is a threat to the United States (Sharia)
3. Social security should be improved (SS)
4. Medicare coverage should be expanded (Medi)
5. China is threat to national security (ChinaWar)
6. China is “winning” on trade (ChinaMoney)
7. Traditional culture is under attack (Trad)
8. I support Donald Trump more than any other candidate (Support)

Gender question coded “Sex.” Age question coded “Age,” with responses “17 or younger”, “18-20”, “21-29”, “30-39”, “40-49”, “50-59”, and “60 or older.”

Additional response “Cultural Marxism does not exist” was available for Marx question.

Data Results

Introduction

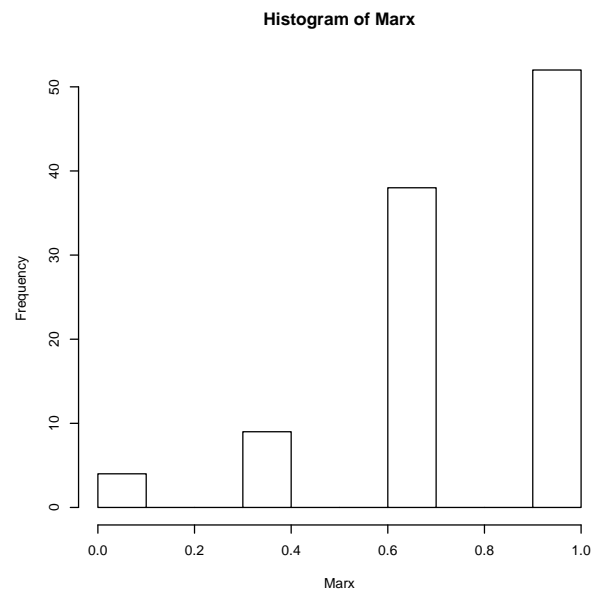
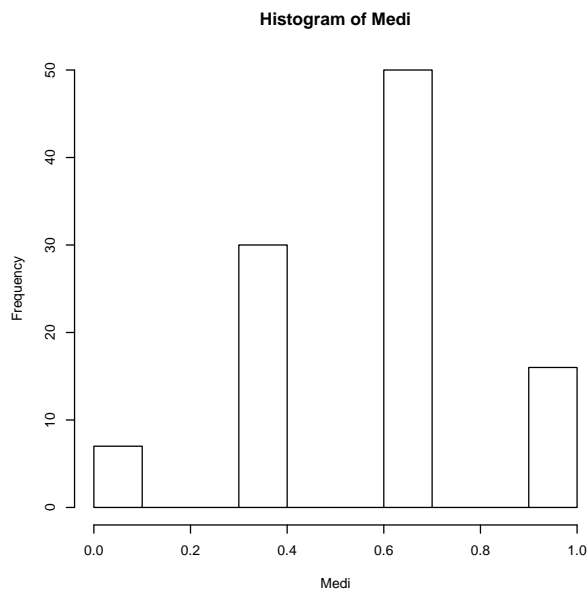
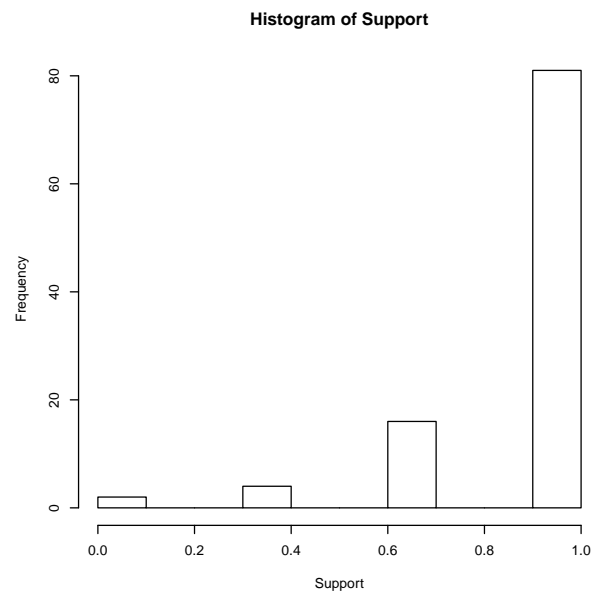
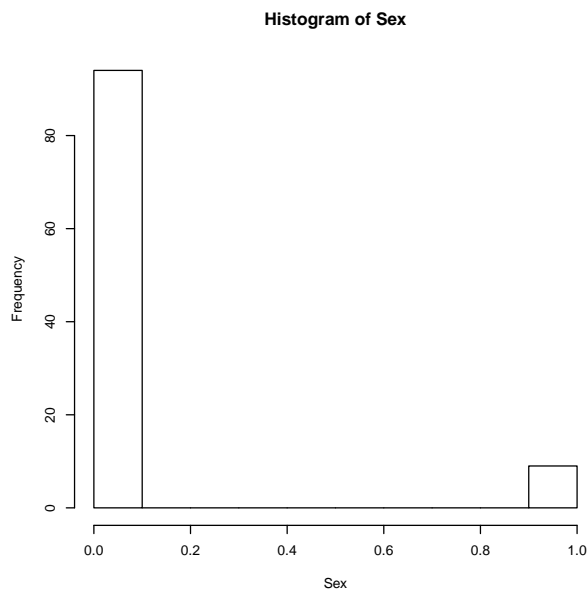
There were 103 survey respondents who responded from 2/2 to 2/3/2016. Response to the poll in Reddit comments was generally positive. There was only one response for “Cultural Marxism does not exist” out of all 103 responses. The question was thrown out and the participant’s response was converted to “Strongly Disagree.”

Cleaning

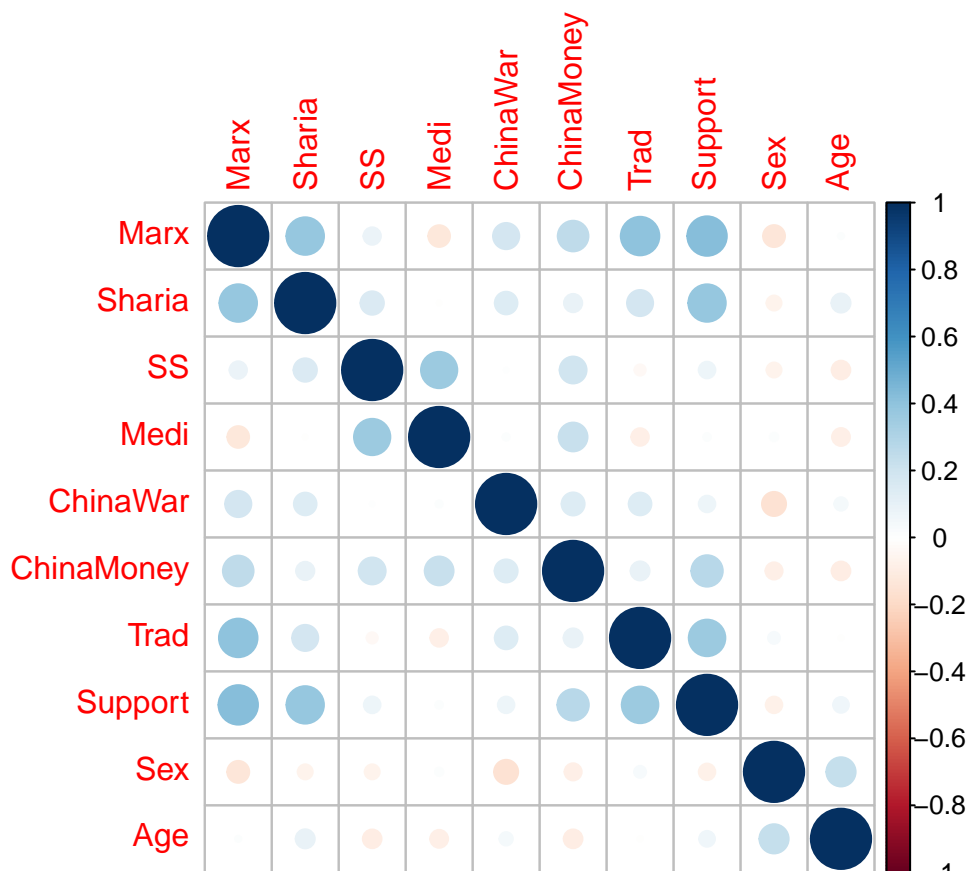
Data was converted to matrix form with the following procedure:

- 1) Views converted to assumed equivalent values $\in [0, 1]$: “Strongly Agree” = 1, “Agree” = $2/3$, “Disagree” = $1/3$, “Strongly Disagree” = 0.
- 2) Gender converted: Male = 0, Female = 1.
- 3) Age converted to midpoints of ranges. “17 or younger” was assumed to be 15, and “60 or older” was assumed to be 65.

Frequency of Notable Variables



Correlation of Variables



Classification

Fitting Number of Clusters

To best fit a number of clusters on response data I used an infinite Gaussian mixture model. Essentially, you can view the distribution of data like a Dirichlet distribution composed of smaller independent multivariate Gaussian distributions. See references for a better explanation.

I used an Expectation-maximization technique to fit my data to this distribution using the scikit-learn Python library. I repeated 10000 trials and found that the algorithm classified my data as being composed of either two or three groups. The third group (if present) contained at most one member. For that reason, I decided my data was best represented by a mixture model of two independent Gaussian distributions.

Principal Component Analysis

Having performed multivariate clustering of the data, I then sought to compare these clusters to results of univariate clustering of one factor.

PCA of the data revealed multiple components that seemed to have identifiable political meaning. The first component had the largest sdev. Here are the weights of the first component (you can see Trump support is correlated with an affirmative response to cultural questions):

##	Marx	Sharia	SS	Medi	ChinaWar	ChinaMoney
----	------	--------	----	------	----------	------------

```
## 0.50574363 0.41020368 0.15659940 0.02716980 0.23459946 0.31395783
##      Trad      Support      Sex      Age
## 0.38340240 0.48075089 -0.13760550 0.00585148
```

Bayesian Mixture Modeling

Now, having fit the number of clusters using EM, I used the BayesMix package to cluster the PC1 component data assuming two uncertain independence priors.

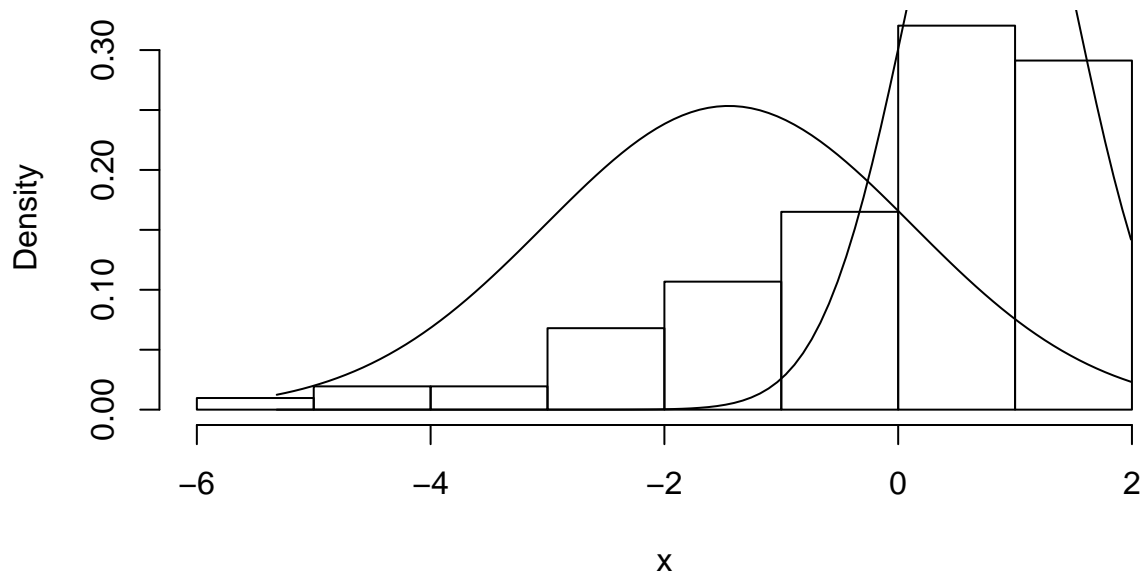
```
library(bayesmix)

get_mm=function(vec, p_string) {
  model = BMMmodel(
    vec, k = 2,
    priors = list(kind = "independence", parameter=p_string),
    no.empty.classes=TRUE
  )
  control = JAGScontrol(variables = c("mu", "tau", "eta", "S"), burn.in = 1000, n.iter = 5000, seed = 1)

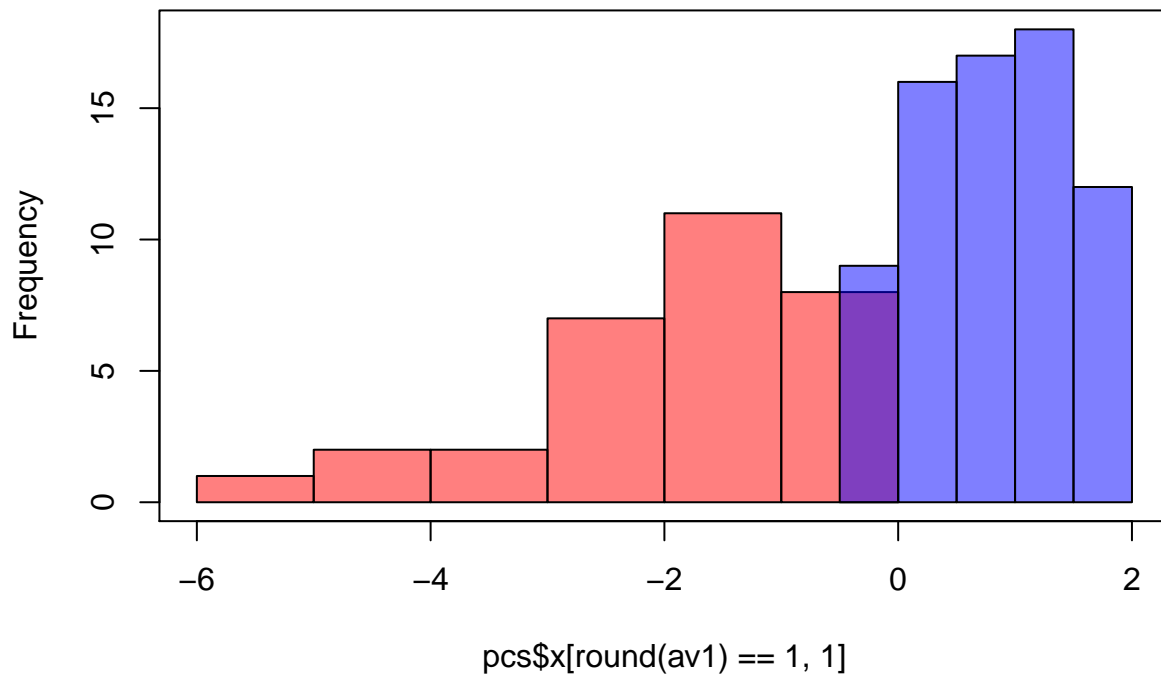
  z = JAGSrun(vec, model = model, control = control, initialValues=list(S0=4))
  return(z)
}

mm1 = get_mm(pcs$x[,1], "priorsUncertain")
```

Frequency of PC1 and Cluster Fits



Frequency of PC1 values by cluster



Monte Carlo Verification

We can verify our clusters through their differential entropy (determinant of covariance times a dimensional scaling factor.) We will compare this to a Monte Carlo simulation of groups from the same population. Smaller entropy means more correlated data.

```
## [1] "Monte Carlo Split Draws"
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.3473  2.3170  2.6760  2.6470  3.0050  4.0500
```

```
## [1] "99% Conf:"
```

```
##      0.5%    99.5%
## 1.314885 3.728122
```

```
## [1] "Group 1 DE: 4.10330660772681"
```

```
## [1] "Group 2 DE: 0.442575424990931"
```

Conclusion

Our observed differential entropy values for both clusters are outside the 99% confidence interval for 5000 cases of differential entropy of random group splittings. This implies these groups are both stastically different from the base population.

Group 1 is purposed to be a group of general “Republicans” with more diverse beliefs than Trump. Group 2 is purposed to be a group of hardcore Trump supporters with highly correlated beliefs. The second group has a differential entropy value of -0.09 which means they are extrodinarily homogeneous when compared to an average group which has a mean entropy of 2.6660.

References

1. <http://blog.echen.me/2012/03/20/infinite-mixture-models-with-nonparametric-bayes-and-the-dirichlet-process/>
2. <https://github.com/a-leut/bayes-present>

All code and data is availible for download from the second link.