Multi-level Ideal Point Modeling on Political Polls

Yilin He*, Shuying Wang

{he-yilin, shuying.wang}@utexas.edu
Department of Statistics & Data Science
*: Department of Information, Risk and Operations Management

May 14, 2019

Abstract

In this work, we fit a Bayesian ideal point model on the political polls from the 1988 U.S. presidential election. The model takes the respondents' state of residence, education level, age, gender, ethnicity and weight into account and intends to predict the likelihood of voting George H.W. Bush (41st) as an individual with the above features. We first apply Bayesian hierarchies on the first 5 categorical variables to encode each of them from one-hot vector to a corresponding latent scalar "score", after that, employ a probit model to regress the likelihood of voting Bush. Designed experiments display that in this way, we are able to beat vanilla probit regression by about 0.25 in fitting accuracy, and have a better understanding of how much each feature is related to the probability of supporting Bush in the election.

Keywords ideal point modeling; Bayesian hierarchical models; latent factor models; Markov Chain Monte Carlo

1 Introduction

Probit models [1] are widely adopted in regression problems when the dependent variable takes only two values. By mapping the entire real axis to $(0, 1)$, it can be interpreted as the likelihood of the regression target to take one value against the other. Considering the statistical meaning of probit models, they can either be fit in a frequentist fashion, *i.e.*, MLE, or a Baysian way, *i.e.*, Gibbs Sampling. Thanks to James H. Albert and Sidhhartha Chib who introduced data augmentation trick for binary data [2], we are now able to estimate parameters of a probit model in a Gibbs sampling framework, which is free from onerous Metropolis Hastings and therefore achieves lower computational complexity.

In most of contexts where a probit model is applied, the response variable is regressed on a $\Phi(\cdot)$ transformation of a linear combination of all the features of the data. If the data includes categorical features, the traditional approach is dummy coding, *i.e.*, creating a block of dichotomous variables out of one categorical variable then entering the block as predictors. Probit models constructed such way are referred to as "vanilla probit models" in the remaining contents. It could be straightforward to interpret the coefficients in these models, however, they are not trouble free, one of the gravest issues around vanilla probit models is the low prediction accuracy.

Such phenomenon is believed to attribute to two factors. On the model level, dummy coding confines the capacity of the model. For categorical covariates, the variation of the output is limited to the combination of category levels, making it difficult for coefficients to pick up some value that fits all the data. On the data level, dummy coding fails to let one "be different" with those have a shared feature, which is, however, natural in the real world. In order to remedy for the above aspects, we need to present these categorical variables in a more flexible manner.

Suggested by A. Aiken and James G. Scott [4], we use a Bayesian ideal point model to replace the vanilla probit model, where all the categorical variables are encoded to latent scores. For instance, the variable "female" is originally a 2-level categorical variable and will be a 2-element one-hot vector in dummy coding, whereas only a numerical scalar score in our model. We also model after the authors' way to generate the scores by setting a prior for each level, then sample the scores through Gibbs sampling. As there are only two levels involved, the scores can be appropriately scaled by the corresponding coefficient, so the only thing to mention in the prior

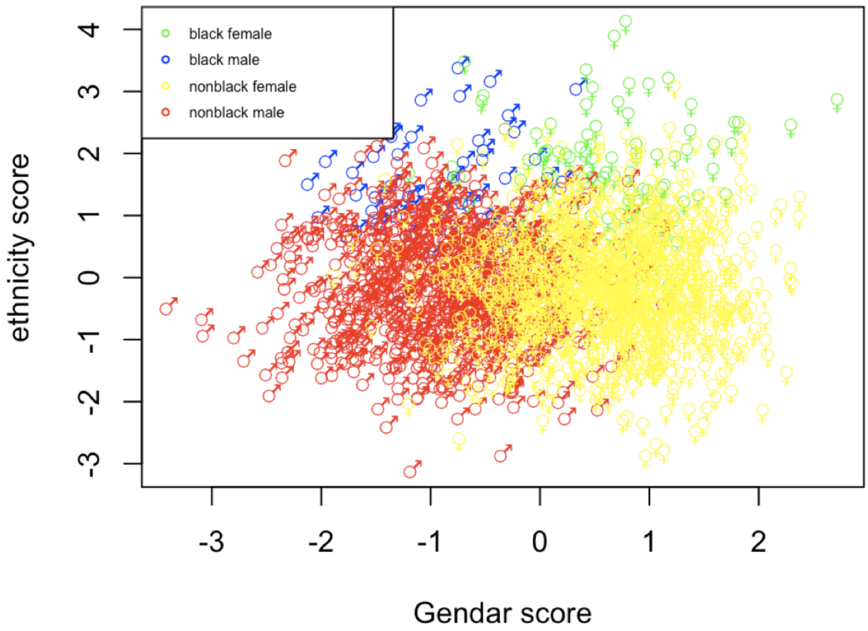

Figure 1: The latent scores' spatial distribution in the "gender-ethnicity" distributions. From the figure we can find inter-level overlap phenomenon does exist, but scores from the same feature level are generally clustered together, which to some extent reflects the concurrence of individual and community characteristics in human nature.

setup is to make the priors discernible, *i.e.*, making the mean value of each normal prior at least two standard deviations away from the other.

When a categorical variable has more than two levels, assigning priors for each level manually is not viable, given that the relative distance between the priors is beyond our knowledge. Specifically, if the prior for each state's score is a Gaussian distribution with some mean and standard deviation 1, it could be questionable sensible to determine how big the mean deviation for California and Texas should be, compared with that between Texas and Rhode Island. We address this problem with a Bayesian hierarchy that, instead of directly appoint values of the priors' parameters, we assign them non-informative priors. As a result, sampling the priors' parameters as well as model coefficients (beta's) and latent scores (predictors, a.k.a. x's) can be fit into one Gibbs sampling framework. This part of work will be elaborated in section 2. We will present in section 3 that in this way, we are able to improve the prediction accuracy by around 25%, compared with the vanilla probit model.

There is an extra credit for encoding categorical variables into latent numerical scalar scores. Since the features construct a multi-dimensional numerical space, the parameter vector is in the same vector space with the feature vector, we can apply the ideal point theory [3] to model interpretation. Related analyses will be carried out in section 3.

2 The Algorithm

The data we use is the results of several political polls from the 1988 U.S. presidential election. The binary outcome of interest is whether someone plans to vote for George Bush ($y = 0$ for "not voting" and 1 "voting"). There are several potentially relevant demographic features here, including the respondent's state of residence, education level, age, gender, ethnicity and weight. Except for weight, all the other five features are categorical variables, so we encode them to a corresponding latent scalar "score", and then apply a probit model to eistimate the likelihood of voting Bush.

The model is as follows:

$$\begin{aligned} P(y=1) &= \Phi(\beta_0 + \beta_1 x_{state} + \beta_2 x_{edu} + \beta_3 x_{age} \\ &\quad + \beta_4 x_{female} + \beta_5 x_{black} + \beta_6 x_{weight}) \\ &= \Phi(\mathbf{X}^T \boldsymbol{\beta}), \end{aligned} \tag{1}$$

where $\boldsymbol{\beta} = (\beta_0,\ \beta_1,\ \beta_2,\ \beta_3,\ \beta_4,\ \beta_5,\ \beta_6)$, and

$$\mathbf{X} = (1,\ \mathbf{x}_{state},\ \mathbf{x}_{edu},\ \mathbf{x}_{age},\ \mathbf{x}_{female},\ \mathbf{x}_{black},\ \mathbf{x}_{weight}).$$

Here β_0, the intercept, can be interpreted as the overall popularity of the candidate; for β_1

through β_5, a positive coefficient indicates individual with higher score favors Bush more than those with lower scores, and the reverse when the coefficient takes a negative value.

In the following session, we use abbreviation:

$$\mathbf{X} = (1,\ \mathbf{x}_s,\ \mathbf{x}_e,\ \mathbf{x}_a,\ \mathbf{x}_f,\ \mathbf{x}_b,\ \mathbf{x}_w).$$

Considering the x scores for each respondent as well as all the β's are part of the Gibbs sampling, therefore appropriate priors need to be assigned to these parameters. To begin with, the distribution for x scores are arranged as follows:

$$\begin{aligned}
&x_s \sim \mathrm{N}(\mu_s,\ 1),\\
&\quad s = \text{“AL”, “AR”}, \cdots, \text{“WY”};\\
&x_e \sim \mathrm{N}(\mu_e,\ 1),\\
&\quad e = \text{“Bacc”, “HS”, “NoHS”, “SomeColl”};\\
&x_a \sim \mathrm{N}(\mu_{a_j},\ 1),\\
&\quad a = \text{“18to29”, “30to44”, “45to64”, “65plus”};\\
&x_f \sim \mathrm{N}(-1,\ 1) \text{ for males,}\\
&\quad \sim \mathrm{N}(1,\ 1) \text{ for females;}\\
&x_b \sim \mathrm{N}(-1,\ 1) \text{ for non-blacks,}\\
&\quad \sim \mathrm{N}(1,\ 1) \text{ for blacks.}
\end{aligned} \tag{2}$$

As for x_w, since the weight data is numerical, we simply normalize the original data, *i.e.*, each weight value is subtracted by the sample mean and divided by sample standard deviation. As mentioned in section 1, rather than directly assign values to μ_s, μ_e and μ_a, we give them uniform non-informative priors so that they can be sampled along with other parameters in every Gibbs sampling iteration.

Next, let define the prior for coefficients $\boldsymbol{\beta}$:

$$\beta_j \sim \mathrm{N}(\phi,\ \sigma^2),\ j = 0, 1, \cdots, 6.$$

We can vectorize the above expression as:

$$\boldsymbol{\beta} \sim \mathrm{MVN}(\boldsymbol{\phi},\ \sigma^2\mathbf{I}), \tag{3}$$

where $\boldsymbol{\phi} = (\phi,\ \phi,\ \phi,\ \phi,\ \phi,\ \phi,\ \phi)$. In the model, we choose uniform non-informative priors for ϕ , and Gamma($\frac{1}{2}, \frac{1}{2}$) prior for the precision $\lambda = \frac{1}{\sigma^2}$.

To facilitate the inference of full conditional distributions of the previously mentioned parameters in our probit model, we adopt the Albert and Chib trick [2]. Latent variables z_i's are involved and act as a “bridge” between model parameters and regression targets. The relationship between $z_i, \mathbf{X}_i$ and y_i can be summarized as:

$$\begin{aligned}
&z_i \sim \mathrm{N}(\mathbf{X}_i^T\boldsymbol{\beta}, 1),\\
&y_i = 1 \text{ if } z_i > 0,\\
&y_i = 0 \text{ if } z_i \leq 0,\\
&i = 1, 2, \cdots, N.
\end{aligned} \tag{4}$$

Given all the defined priors of the parameters and introduction of z_i's, we can derive the posterior full conditionals used for Gibbs sampling.

Start from the latent variable z_i's:

$$\begin{aligned}
&z_i|- \sim \mathrm{N}(\mathbf{X_i}^T\boldsymbol{\beta}, 1),\\
&\quad \text{truncated at the left by 0, if } y_i = 1,\\
&\quad \text{truncated at the right by 0, if } y_i = 0.
\end{aligned} \tag{5}$$

Then the model coefficients $\boldsymbol{\beta}$:

$$\begin{aligned}
f(\boldsymbol{\beta}|-) &\propto \exp\Big[-\frac{1}{2}\sum_{i=1}^{N}(z_i - \mathbf{X}_i^T\boldsymbol{\beta})^2\\
&\quad - \frac{\lambda}{2}(\boldsymbol{\beta} - \boldsymbol{\phi})^T(\boldsymbol{\beta} - \boldsymbol{\phi})\Big],\\
&\Longrightarrow\\
\boldsymbol{\beta}|- &\sim \mathrm{MVN}(\mathbf{A}^{-1}\mathbf{B},\ \mathbf{A}^{-1}),
\end{aligned} \tag{6}$$

where

$$\mathbf{A} = \sum_{i=1}^{N}\mathbf{X_i}\mathbf{X_i}^T + \lambda\mathbf{I},$$

and

$$\mathbf{B} = \sum_{i=1}^{N} z_i\mathbf{X_i} + \lambda\boldsymbol{\phi}.$$

Next, latent scores $\mathbf{X}$. Let

$$\begin{aligned}
\mathbf{x}_i^* &= (x_{s,i},\ x_{e,i},\ x_{a,i},\ x_{f,i},\ x_{b,i}),\\
\boldsymbol{\beta}^* &= (\beta_1,\ \beta_2,\ \beta_3,\ \beta_4,\ \beta_5),\\
\boldsymbol{\mu}_i &= (\mu_{s_i},\ \mu_{e_i},\ \mu_{a_i},\ \pm 1,\ \pm 1),
\end{aligned}$$

where denoted by s_i is the state of the i^{th} observation, $i = 0,\ 1,\ 2,\ ...,\ N$. We have

$$\begin{aligned}
&f(\mathbf{x}_i^*|-)\\
&\propto \exp\Big[-\frac{1}{2}(z_i - \beta_0 - \beta_6 x_{w,i} - \mathbf{x}_i^{*T}\boldsymbol{\beta}^*)^2\\
&\quad - \frac{1}{2}(\mathbf{x}_i^* - \boldsymbol{\mu}_i)^T(\mathbf{x}_i^* - \boldsymbol{\mu}_i)\Big]\\
&\Longrightarrow\\
&\mathbf{x}_i^*|- \sim \mathrm{MVN}(\mathbf{P}^{-1}\mathbf{H_i},\ \mathbf{P}^{-1}),
\end{aligned} \tag{7}$$

where

$$\mathbf{P} = \boldsymbol{\beta}^*\boldsymbol{\beta}^{*T} + \mathbf{I},$$

and

$$\mathbf{H_i} = (z_i - \beta_0 - \beta_6 x_{w,i})\boldsymbol{\beta}^* + \boldsymbol{\mu}_i.$$

After that, the prior means for state, education and age scores:

$$\begin{aligned}
f(\mu_s|-) &\propto \exp\Big[-\frac{1}{2}\sum_{s_i=s}(x_{s,i} - \mu_s)^2\Big]\\
&\Longrightarrow\\
\mu_s|- &\sim \mathrm{N}(\frac{1}{m_s}\sum_{s_i=s} x_{s,i},\ \frac{1}{m_s}),
\end{aligned} \tag{8}$$

where m_s is the number of observations from state s.

$$f(\mu_e|-) \propto \exp\Big[-\frac{1}{2}\sum_{e_i=e}(x_{e,i}-\mu_e)^2\Big]$$
$$\Longrightarrow \tag{9}$$
$$\mu_e|- \sim \mathrm{N}(\frac{1}{m_e}\sum_{e_i=e}x_{e,i},\ \frac{1}{m_e}),$$

where m_e is the number of observations of education level e.

$$f(\mu_a|-) \propto \exp\Big[-\frac{1}{2}\sum_{a_i=a}(x_{a,i}-\mu_a)^2\Big]$$
$$\Longrightarrow \tag{10}$$
$$\mu_a|- \sim \mathrm{N}(\frac{1}{m_a}\sum_{a_i=a}x_{a,i},\ \frac{1}{m_a}),$$

where m_a is the number of observations of age a.

Finally, the mean and precision for the model coefficients $\boldsymbol{\beta}$:

$$f(\phi|-) \propto \exp\Big[-\frac{\lambda}{2}\sum_{j=0}^{6}(\beta_j-\phi)^2\Big]$$
$$\Longrightarrow \tag{11}$$
$$\phi|- \sim \mathrm{N}\Big(\frac{1}{7}\sum_{j=0}^{6}\beta_j,\ \frac{1}{7\lambda}\Big).$$

$$f(\lambda|-) \sim \mathrm{Ga}\Big(4,\ \frac{1}{2}+\frac{1}{2}\sum_{j=0}^{6}(\beta_j-\phi)^2\Big). \tag{12}$$

In a word, all the parameters of the model, including the latent scores $\mathbf{X}$ and coefficients $\boldsymbol{\beta}$, are obtained through algorithm 1. R scripts of algorithm 1 and further analyses are accessible through url: https://github.com/hylBen/Courses/tree/master/Learning/SDS-383D-STATS_MODELING_II/final-project.

```
Result: Sampled model coefficients $\boldsymbol{\beta}$ and
        latent scores $\mathbf{X}$
initialization;
$t \leftarrow 0$;
while the chain not converge do
    $t \leftarrow t+1$ sample $z_i^{(t)}$ from $f(z_i|-)$ (Eq.
      5), $i = 1, 2, \cdots, N$;
    sample $\boldsymbol{\beta}^{(t)}$ from $f(\boldsymbol{\beta}|-)$ (Eq. 6);
    sample $\mathbf{x}_i^{(t)}$ from $f(\mathbf{x}_i|-)$ (Eq. 7),
      $i = 1, 2, \cdots, N$;
    sample $\mu_s^{(t)}$ from $f(\mu_s|-)$ (Eq. 8),
      $s$ = “AL”, “AR”, $\cdots$, “WY”;
    sample $\mu_e^{(t)}$ from $f(\mu_e|-)$ (Eq. 9),
      $e$ = “Bacc”, “HS”, “NoHS”, “SomeColl”;
    sample $\mu_a^{(t)}$ from $f(\mu_a|-)$ (Eq. 10), $a$ =
      “18to29”, “30to44”, “45to64”, “65plus”;
    sample $\phi^{(t)}$ from $f(\phi|-)$ (Eq. 11);
    sample $\lambda^{(t)}$ from $f(\lambda|-)$ (Eq. 12);
end
```

Algorithm 1: The Gibbs sampler for all parameters in model 1.

3 Experiments & Results

We tested aforementioned algorithm on dataset “Pulls” [5]. We first eliminate data samples with incomplete features, then run the Gibbs sampler 1 for $5,000$ iterations and collect the last $1,000$ burnt-in samples for model evaluation and other analyses. Taking all the latent scores inclusive, our model is populated with parameters, which greatly enhances the model's data fitting. The average prediction accuracy for our model in iteration $4,000-5,000$ reached as high as **92.57%**, which surpasses the vanilla probit model's 68% accuracy after $15,000$ iterations by about 24.6%. With such a huge amount of parameters updated simultaneously, our model enjoys a surprisingly fast convergence: **it only takes less than 50 iterations for the prediction accuracy to exceed 85%**.

Once obtaining the latent scores, we subtract them by the sample mean $\hat{\mu}$ then divide them by sample standard deviation $\hat{\sigma}$, both $\hat{\mu}$ and $\hat{\sigma}$ are in the current iteration term, not for the entire extracted $1,000$ iterations. Meanwhile, except for the slope, the β's are multiplied by the sample standard deviation of the corresponding scores. The slopes are adjusted in a way that maintains the outputs of our probit model unchanged. In doing this, we reformed the latent scores for each covariate to the same scale, which isolates the calibration function of the β coefficients from the information they hold about the model, thus helps model interpretation.

With the model interpreted as in section 2, as well as the parameters obtained during the experiment, a positive $\hat{\beta}_0$ implies voters are supporting Bush, by and large. Additionally, given that the coefficient of state is on average negative, we can conclude that an individual with a lower state score has a higher likelihood of voting Bush. Displayed in figure 2 is the plot of the ground truth support rate in the 1988 election against $-\mu_s$ samples (in the $5,000^{th}$ iteration) averaged over states. In spite of some noisy data whose sample sizes are radically small and drift faraway from the main, a general positive correlation is explicit, indicating that the state scores we gained are meaningful.

Applying the same logic to the rest of scores, we may profile a voter who is highly inclined to supporting Bush as, a non-black male, at an age between 18 and 29, having attended some college, and under the residency of a state within (WY, VT, NV, NH, SC). Furthermore, if the score vector of an individual has uniformed length, we can depict "the most determined voter" with a set of quantified scores, which is exactly the dot in the "score space" where the direction of $\boldsymbol{\beta}$ points to. Related analyses are within the scope of ideal point theory.

We also evaluate the marginal effect of each covariate to see which features are more influential to the voting probability. Marginal effect of x_j measures the expected movement of y given an unit increase in x_j, holding all else as constants. It can be attained by taking derivative of $\mathbb{E}(y)$ with respect to x_j, *i.e.*,

$$\begin{aligned} \frac{\partial \mathbb{E}(y_i)}{\partial x_j} &= \frac{\partial P(y_i = 1)}{\partial x_j} \\ &= \frac{\partial \Phi(\mathbf{X_i}^T \boldsymbol{\beta})}{\partial x_j} \\ &= \beta_j \phi(\mathbf{X_i}^T \boldsymbol{\beta}). \end{aligned} \tag{13}$$

Here $\phi(\cdot)$ denotes the density of the standard normal distribution. We choose the samples from the last iteration, and compute the average marginal effects for all the observations, which is showed in Table 1. We can see that the education level, age, and ethnicity have relatively stronger marginal effects on voting, while for gender and weight, the marginal effects are weaker. Noting that covariates with strong marginal effects are always linking to tipping factors, masterminds of object specific political propaganda may be interested in related studies.

To sum up, the results of our model meet up with our expectations, in terms of the prediction accuracy, and general tendency of scores. Some interesting topics about marginal effects will be discussed in section 4.

4 Discussions

There's no evident clue of the interpretation of the scores, one option is that the magnitude of a score somewhat measures the sensitivity of a person about topics associated with some covariate, *i.e.*, education, and the sign of a score suggests his/her attitude towards the Bush government's promises (on these issues). Following discussions will lie on the premise that the above assumption is suitable.

There are achievements as well as remaining defects about the Bayesian multi-level ideal point model. To begin with, we have proposed a mechanism that could work out an adequate latent representation of political election, *i.e.*, the coefficients $\boldsymbol{\beta}$, and latent representations for people's "political concerns", namely, the scores $\mathbf{x}$. The process is statistically interpretable, and results, in most cases, are consistent with ground truth. With the information, our model can be used in predictive tasks, *i.e.*, estimating a respondent's tendency of supporting Bush, or generative tasks, *i.e.*, completing missing data.

Apart from these ordinary tasks, the information obtained by our model may be useful for some political purposes. Based on the marginal effects, it is totally feasible to inference what people care most about topics related to what theme, which at least provides the politician running the election with some guideline about what to do to optimally satisfy different people, not to mention the information's value in political propaganda. For example, education level, age and ethnicity may play important roles in the decision of voters, so candidates can focus on related issues in their governance outlines, wisely make decisions to gain broader support from people, especially those used to play neutral.

Nevertheless, validity of marginal effects acquired in this way is quite sensitive to data collinearity, which, unfortunately, is exactly what the dataset "Polls" suffers from. At the beginning, we detected an anomaly in the summary of marginal effects 1: according to the "Polls" dataset, the support rate of different age groups are (58.45%, 52.89%, 58.33%, 54.61%), and the support rate of different genders are (54.73%, 57.28%). Though the differences of support rate between age groups are a bit larger than that between genders, they cannot explain why the absolute value of marginal effect of age is more than ten times larger than that of gender. In another word, the difference between marginal effects is disproportional to the divergence of voting probability between different demographic groups, which is unexpected regarding that the movements of $\mathbf{x}$ scores are manually aligned beforehand.

We performed a permutation test upon the original dataset to check if the feature "female" is correlated to some other feature. We use the geometry average of male-female ratio as the test statistic, *i.e.*, for the test of independence be-

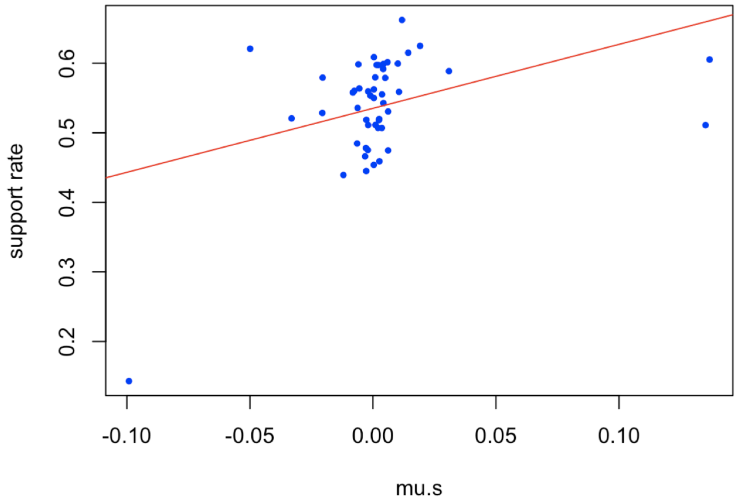

Figure 2: The ground truth support rate of each state v.s. sampled μ_s's.

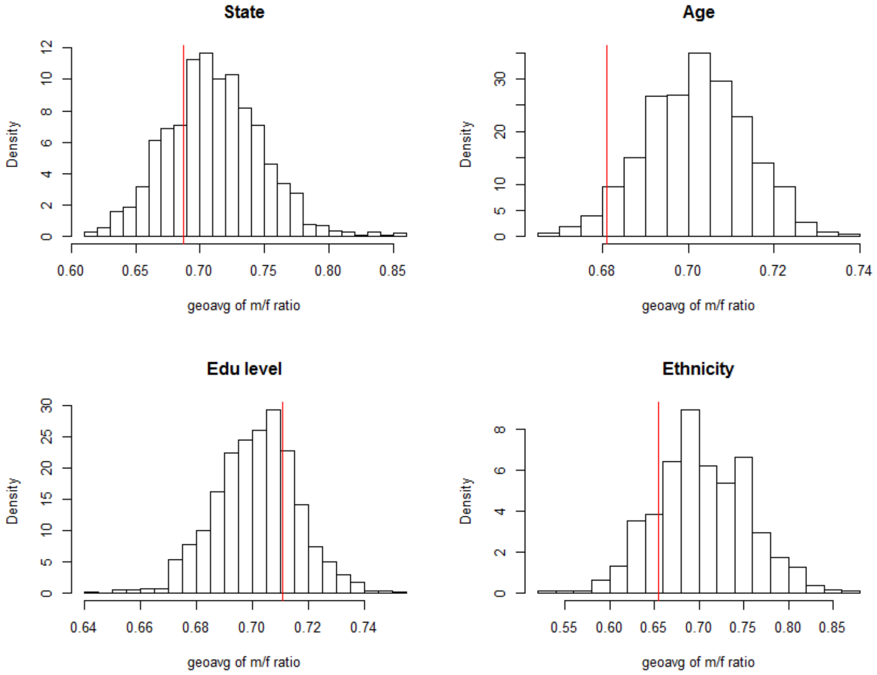

Figure 3: The result of permutation test where the red vertical lines are the values of test statistics of the original "Polls" dataset. Apparently, the observation of the test statistic for "age-gender" test is drifting from normal.

Covariate	Marginal effects	Avg. Coef. $\bar{\beta}$	Covariate stddev.
(intercept)	/	0.576	/
state	-0.087	-0.993	1.025
edu	0.267	2.270	0.999
age	-0.146	-2.333	0.991
gender	-0.012	-0.232	1.401
ethnicity	-0.195	-2.483	1.120
weight	0.001	0.019	

Table 1: The marginal effect of each covariate, sample standard deviation of corresponding scores and average value of sampled corresponding coefficients. Here the marginal effect and score sample stddev. are computed with samples obtained in the last iteration.

tween gender and ethnicity, it will be

$$\rho_b := \sqrt{\frac{\#\text{black male}}{\#\text{black female}} \times \frac{\#\text{non-black male}}{\#\text{non-black female}}}. \tag{14}$$

The null hypothesis is that the feature "female" is totally independent with any other feature, so after calculating the value of test statistics ρ's in the dataset, we randomly shuffle the values of "female" in the dataset $1,000$ times then record histograms of the test statistics computed out of the shuffled dataset, from where we can judge whether the observation of the ρ's are normal under the null. The result of permutation test is shown in figure 3, which suggests a correlation between age and gender in the dataset. Such correlation is further verified with the empirical p-value of "age-gender" statistic – 0.074, a much smaller value compared with the other three. The collinearity between age and gender accounts for the anomaly in sampled β's, but also compromises the soundness of obtained marginal effects. Some future works can be associated with improving the model by reducing collinearity.

5 Conclusions

Based on [4], we proposed a Bayesian hierarchical model that enables Bayesian ideal point modeling work on multi-level categorical features, which achieves enormous improvement in fitting accuracy. The model also brings about a mechanism that extracts interpretable latent representations with respect to the election event and voter orientations, which could offer critical information for political studies. Except for dedicated efforts for this project, there are potential works including but not limited to checking the model's predictive & generative ability on validation sets, and experimenting collinearity-supressing data preprocessing methods.

An extra claim to make is that the idea to substitute Bayesian hierarchy for hand-designed prior parameters is solely inspired by the course materials of SDS 383D. Due to the limited amount of time, we are not able to fulfill an exhaustive literature research about related works. If there are any coincidental collisions with exist ideas, we will be appreciated for your kindly remind.

References

[1] "The Method of Probits." Bliss, C. I. (1934). In *Science*. 79 (2037): 38–39.

[2] "Bayesian Analysis of Binary and Polychotomous Response Data." James H. Albert and Siddhartha Chib. *Journal of the American Statistical Association*, Vol. 88, No. 422 (Jun., 1993), pp. 669-679

[3] " Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking." Jackman S. *Polit Anal.* 2001;9(3):227–41.

[4] "Family Planning Policy in the United States: The Converging Politics of Abortion and Contraception." A. Aiken and James G. Scott. *Contraception* 93(5): 412–20, 2016.

[5] Political Polls results of 1988 presidential election. Url: `https://github.com/jgscott/SDS383D_Spring2019/blob/master/data/polls.csv`