A Nonparametric Bayesian Model for Gradual Structural Changes: The Intergenerational Chinese Restaurant Processes

Nuannuan Xiang* Yuki Shiraito†‡

This draft: February 22, 2021
First draft: July 7, 2020

Abstract

Many social changes occur gradually over time, and social scientists are often interested in such changes of unobserved heterogeneity. However, existing methods for estimating structural changes have failed to model continuous processes through which a data generating process evolves. This paper proposes a novel nonparametric Bayesian model to flexibly estimate changing heterogeneous data generating processes. By introducing a time dynamic to the Dirichlet process mixture model, the proposed intergenerational Chinese restaurant process (IgCRP) model categorizes units into groups and allows the group memberships to evolve as a Markov process. In the IgCRP, the group assigned to a unit in a time period follows the standard Chinese restaurant process conditional on the group assignments in the previous time period. A distinctive feature of the proposed approach is that it models a process in which multiple groups emerge and diminish as a continuing process rather than a one-time structural change. The method is illustrated by reanalyzing the data set of a study on the evolution of party positions on civil rights in the United States from the 1930s to the 1960s.

*Ph.D. candidate, Department of Political Science, University of Michigan. Email: nnxiang@umich.edu.

†Assistant Professor, Department of Political Science, University of Michigan. URL: shiraito.github.io.

‡Correspondence: Center for Political Studies, 4259 Institute for Social Research, 426 Thompson Street, Ann Arbor, MI 48104-2321. Phone: 734-615-5165, Email: shiraito@umich.edu.

1 Introduction

Latent group detection is a common problem in quantitative data analysis in political science. For example, a finite mixture regression model is used for testing competing theories, where a latent group is the group of units to which one of the competing theories applies (Imai and Tingley, 2012). Because scholars cannot observe which units share a common causal mechanism, they use a model in which data points are clustered into latent groups and these groups are estimated along with the model parameters shared within each group. Statistical models with latent groups or clusters are used to analyze a variety of data types and substantive topics (Park, 2012; Kim et al., 2020; Gill and Casella, 2009; Traunmller et al., 2015; Spirling and Quinn, 2010; Kyung et al., 2011).

In addition to latent heterogeneity across units, researchers often consider temporal heterogeneity and its dynamics. The relationship between income inequality and economic development in capitalist countries may shift over time. As countries experience different paths of economic development, groups of the countries that share common trade partners should change. To explicitly model such temporal dynamics, existing models (Park, 2011, 2012; Pang et al., 2012; Blackwell, 2018; Fox et al., 2011) assume that latent group assignments follow a hidden Markov process whose state space consists of potentially heterogeneous data generating processes (DGPs).

However, these existing models have either one of the following two caveats. First, existing models using parametric hidden Markov models (Park, 2012, 2011; Pang et al., 2012) require a pre-specified number of latent groups. However, researchers do not necessarily have sufficient prior knowledge about latent heterogeneity in their data. Second, most existing models including nonparametric variants (Fox et al., 2011; Blackwell, 2018) do not properly model structural changes that occur gradually over time. When a societal change happens, not all units switch from one state to another simultaneously as modeled in Markov processes. Rather, some units switch to a new state but the others stay in the current state in one period, and then in later periods they will follow the predecessors by moving to the new state. In existing dynamic models, temporal changes are revolutions—that is, the DGP shifts for all the units in the data set.

To address this issue, we develop a novel nonparametric Bayesian model to flexibly estimate changing heterogeneous structure of DGPs. By introducing a time dynamic to the Dirichlet process mixture model, the proposed intergenerational Chinese restaurant process (IgCRP) model categorizes cross-sectional units into groups and allows the group memberships to evolve as a Markov process. In the IgCRP, the group assigned to a unit in a time period follows the standard Chinese restaurant process Blackwell and MacQueen (1973) conditional on the group assignments in the previous time period.

A distinctive feature of the proposed approach is that it models a process through which multiple groups emerge and diminish as a continuing process rather than a one-time structural change. Moreover, because the IgCRP uses the Chinese restaurant process, it takes into account not only temporal dynamics but also across-unit heterogeneity nonparametrically. Due to the

Chinese restaurant process component, the proposed IgCRP is a nonparametric model where the number of mixture components does not need to be set prior to the analysis.

The reminder of the paper proceeds as follows. The next section introduces our motivating empirical example, Democratic and Republican legislators' behavior on civil rights issues. Section 3 first describes the proposed model using the stick-breaking definition of the Dirichlet process. Then, the section illustrates the Chinese restaurant metaphor for the proposed model, which provides the intuition of how latent groups evolve over time in the proposed model. Section 5 presents empirical analysis of the motivating example, followed by concluding remarks.

2 Motivating Example: Shifting Party Coalitions for Civil Rights in the United States

Our motivating example is the shifting party coalitions for civil rights in the US from the 1930s to the 1970s analyzed by Eric Schickler's book of *Racial Realignment: The Transformation of American Liberalism, 1932-1965* (Schickler, 2016). In his book, Schickler argues that the Democratic and Republican parties' positions on civil rights started changing in the 1930s, much earlier than the conventional wisdom.

The two parties in the US switched their positions on racial issues in the 20th century. Today, the Democratic Party is associated with racial liberalism, supporting government efforts to redress racial inequality, while the Republican party is associated with racial conservatism, resisting governmental interventions in racial issues. Before the 1930s, however, the two parties' positions on racial issues were the reverse. The Republican Party used to be more identified with African Americans than the Democratic Party.

Scholars often view the reversal in the two parties' positions as a sudden structural break that occurred in the 1960s (e.g., Garmines and Stimson, 1989). According to this view, party elites in Washington, D.C. led the change. The "critical juncture" arrived during the presidential election of 1964, when Democratic candidate Lyndon B. Johnson and Republican candidate Barry Goldwater took sharply different positions on civil rights issues. Local party activists then followed national leaders to change their racial positions.

In contrast to the structural-break view of the reversal, Schickler (2016) makes a critical observation that the change of parties' racial positions was a gradual process that started in the mid-1930s. Rather than national party elites, locally oriented rank-and-file party members drove the gradual change.

Specifically, Schickler (2016) emphasizes that northern Democrats at the local level took the lead in making the change.[1] Local northern Democrats were gradually transformed by the New Deal coalition with the Congress of Industrial Organizations (CIO), African Americans, and other urban liberals. Initially, the New Deal coalition, which was made possible by the shared interests

[1]Following Schickler (2016), we define "southern" states or the "South" as the 11 Confederate states plus Kentucky and Oklahoma. "Northern" states or the "North" refers to all other states.

of economic liberalism, had little to do with race. As Democrats' nonpartisan allies increased their civil rights advocacy, so were northern Democrats. While national party elites had a strong interest in maintaining the solidarity of the party, pressures from local party activists forced elites in Washington, D.C. to break the traditional North-South coalition. At the same time, the Republican party, which used to be more supportive of civil rights, was gradually divided over racial issues, as the demand for pursuing civil rights in some of its constituencies waned.

In the book, Schickler provides rich data to assess his argument, including historical survey data to trace the configuration of economic and racial liberalism at the macro level, data of state party platforms to understand how state Democratic and Republican parties positioned themselves on civil rights, and data concerning congressional action on civil rights to measure House members' positions on racial issues. In this paper, we reanalyze the congressional data to understand how House members made coalitions across states and party lines, and how the coalitions changed gradually over the years.

The data of congressional action includes roll-call votes, signing discharge petitions to advance civil rights bills, floor speech, as well as bill sponsorship. Through analyzing the four types of data, Schickler finds that ever since the 1930s, more and more northern Democrats became advocates of civil rights. While at the beginning of the 1930s, northern Republicans were a bit more supportive of civil rights than northern Democrats, over the years, northern Democrats became more liberal while northern Republicans became more conservative. Since the 1940s, the *average* differences between the two parties' positions became larger and larger.[2]

Insightful as Schickler's findings are, the average difference between the two parties' positions reveals limited information on how old racial coalitions dissolved and new coalitions emerged. As "shifting partisan coalitions" played a central role in the transformation of racial politics, it is important to know how congressional members in different states and parties made coalitions with each other, and how their coalitions changed. For instance, in what states did congressional Democrats first join the new racial coalition? Democrats of what states were late joiners? Within a state, when did congressional Democrats and Republicans act together? When did they start to take opposite positions?

Also, the average racial position of a party says little about the distribution of racial positions within the party. Schickler himself admits that "these findings still leave open the question of *which* northern Democrats and Republicans were particularly likely to back civil rights initiatives"(p187). As racial realignment is a gradual procedure that initialized by locally oriented rank-and-file members, it is necessary to understand how members in individual states changed their positions over the years.

The proposed methodology provides an analytical tool to solve the questions listed above. We regard House members of a party in one state as the unit of analysis. By categorizing the party-

[2]The major findings of analyzing congressional data is in Chapter 8. Schickler also shows the results of analyzing the same four types of data in a earlier paper coauthored with Kathryn Pearson and Brain D. Feinstein (Schickler et al., 2010).

state units into different groups and analyzing how the group membership changes dynamically over time, we can trace the gradual shift of racial coalitions as well as the gradual changes of party positions of individual states.

3 Model and Inference

3.1 The Model

The proposed IgCRP model is a nonparametric and dynamic Bayesian clustering model. It is developed by extending the Dirichlet process (DP) mixture model (Ferguson 1973. See also Teh 2010). Both DP and IgCRP are models for the latent group membership of observations, and both are nonparametric in the sense that the number of latent clusters is not specified *a priori*. However, while the DP mixture model clusters cross-sectional units, IgCRP is a model in which units move across groups over time. IgCRP is a dynamic model since the group membership of each unit changes over time, and the temporal shift of the membership is modeled as a Markov process. In each time period, a unit remains in the same group as the previous period or moves to a different group. If the unit moves, the group to which it is switching is determined by the Dirichlet process conditional on the group memberships in the previous period. Therefore, IgCRP is considered as a dynamic extension of DP.

To formally describe the model, let $i \in \{1, \dots, N\}$ denote a unit and $t \in \{1, \dots, T\}$ denote a time period in a panel data set. Also, let Y_{it} be observed (possibly multivariate) measurement for unit i in time t. It is assumed that unit i in time t belongs to a latent group $g[it]$, where $g[it] \in \{1, 2, \dots\}$ denotes the latent group index of unit i in time t, and that Y_{it} is generated from a probabilistic model f with parameter $\Theta_{g[it]}$:

$$Y_{it} \sim f\left(\Theta_{g[it]}\right). \tag{1}$$

That is, Y_{it} and $Y_{i't'}$ share a common DGP if $g[it] = g[i't']$, but they may follow different DGPs otherwise. An appropriate prior distributions are placed on the model parameters Θ. A simple example of f is the Gaussian regression model with group-specific parameters:

$$Y_{it} \overset{\text{indep.}}{\sim} \mathcal{N}\left(X_{it}^{\top}\boldsymbol{\beta}_{g[it]}, \sigma_{g[it]}^2\right)$$

In this example, $\Theta_g = (\boldsymbol{\beta}_g, \sigma_g)$ and the parameters vary across latent groups. Thus, this regression model accounts for unobserved heterogeneity of the data generating process.

IgCPR models the generative process of group assignment, $g[it]$, in the generic data model defined in equation (1). While IgCPR is a dynamic extension of DP as briefly explained above, the group membership in the initial period ($t = 1$) is generated by DP. Using the stick-breaking construction (Sethuraman, 1994), the generative process of $g[i1], i = 1, \dots, N$, is hierarchically defined as:

$$g[i1] \overset{\text{i.i.d.}}{\sim} \mathsf{Discrete}(\{q_k\}_{k=1}^{\infty}) \tag{2}$$

$$q_k = \pi_k \prod_{l=1}^{k-1}(1 - \pi_l) \tag{3}$$

$$\pi_k \overset{\text{i.i.d.}}{\sim} \mathsf{Beta}(1, \gamma) \tag{4}$$

where γ is the concentration parameter of DP.

The key innovation of IgCRP is the dynamic generative process of latent groups for periods $t = 2, \ldots, T$. Specifically, $g[it]$ follows a Markov process conditional on $(g[1, t-1], \ldots, g[N, t-1])$. On one hand, similar to the sticky HDP-HMM model (Fox et al., 2011), unit i in time t stays in the same group as it was in time $t-1$ with probability p. Formally,

$$\Pr\left(g[it] = g[i, t-1] \mid g[i, t-1]\right) = p \tag{5}$$

where

$$p \sim \mathsf{Beta}(\alpha_p, \beta_p). \tag{6}$$

On the other hand, with probability $1-p$, unit i's group in time t follows the Dirichlet process, but conditional on the cluster assignments in $t-1$. That is,

$$g[it] \overset{\text{i.i.d.}}{\sim} \mathsf{Discrete}(\{q_k^t\}_{k=1}^{\infty}) \tag{7}$$

$$q_k^t = \pi_k^t \prod_{l=1}^{k-1}(1 - \pi_l^t) \tag{8}$$

$$\pi_k^t \sim \mathsf{Beta}(1 + n_k^{t-1}, \gamma + N - \sum_{l=1}^{k} n_l^{t-1}) \tag{9}$$

where n_k^{t-1} is the number of units in latent cluster k in period $t-1$.

3.2 Intergenerational Chinese Restaurant Metaphor

The generative process for the latent groups is intuitively illustrated by the Chinese restaurant representation (Blackwell and MacQueen, 1973) of the Dirichlet process. The Dirichlet process is known to be equivalent to two constructive representation: The stick-breaking process and the Chinese restaurant process. The model described above is built on the former, because it is mathematically simpler. However, the Chinese restaurant representation provides far more intuitive illustration of the identical generative process. The name of the proposed model, the intergenerational Chinese restaurant process, comes from this representation.

Figure 1 illustrates the Chinese restraurant representation of the generative process in $t = 1$.

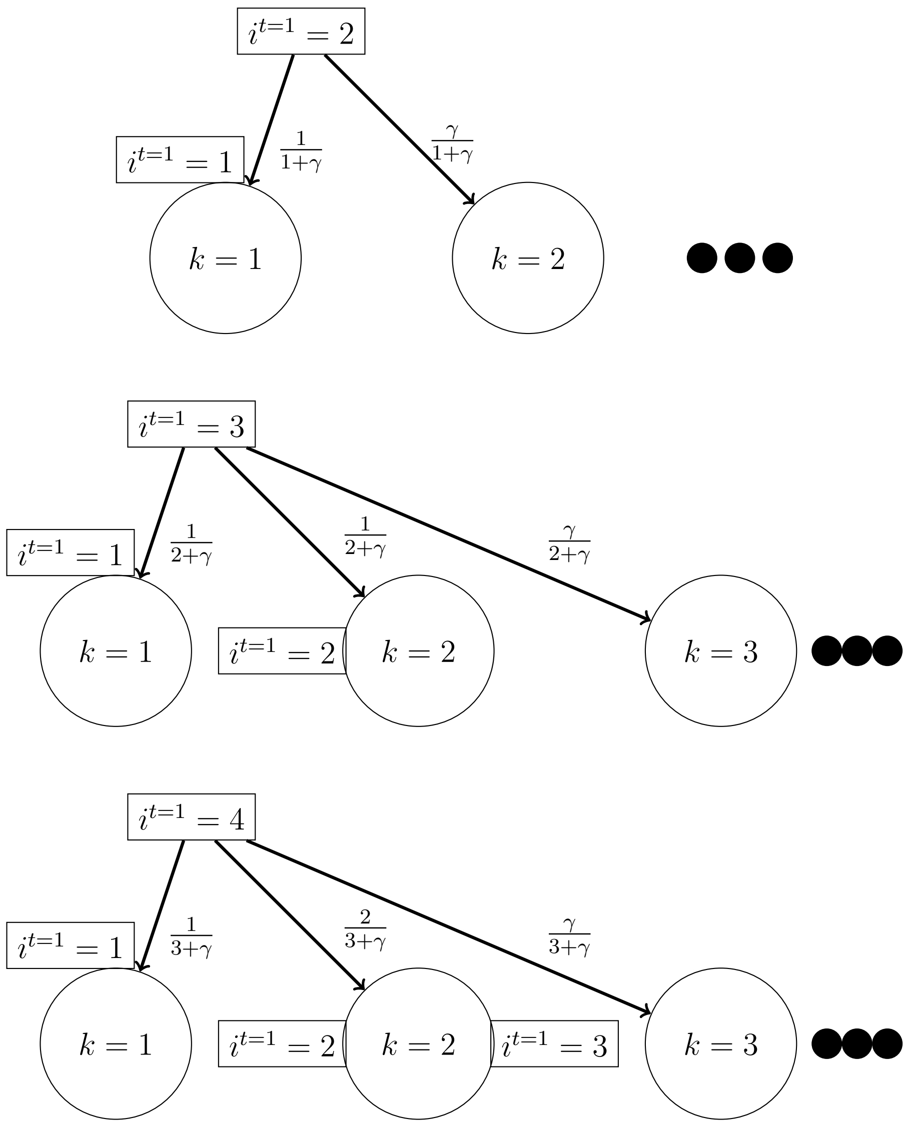


Figure 1: Chinese Restaurant Process. This figure illustrates the generative process of the latent cluster membership in $t = 1$. The fractions near arrows are the probabilities of cluster assignment. The top, the middle and the bottom panel show the probability that unit 2, 3, and 4 are assigned to each cluster, respectively. The probability that a unit assigned to a cluster is proportional to the number of previously assigned units, and the probability that a unit creates a new cluster is proportional to parameter γ.

In this representation, units are assigned to latent groups sequentially. At the beginning, unit 1 is (arbitrarily) assigned to group 1. For each subsequent unit, the probability that the unit assigned to a group is proportional to the number of the units already assigned to that group, and the probability that a unit creates a new cluster is proportional to prior parameter γ. For example, unit 2 goes to group 1 with probability $1/(1+\gamma)$ or creates a new cluster (cluster 2) with probability $\gamma/(1+\gamma)$ (the top panel of Figure 1). If, unit 2 is assigned to group 2 in the realized state, the next unit (unit 3) is assigned to group 1 or 2 with probability $1/(2+\gamma)$, or forms a new group (group 3) with probabilty $\gamma/(2+\gamma)$ (the middle panel of Figure 1). Furthermore, given that unit 1 is assigned to group 1 and units 2 and 3 are assigned to group 2, unit 4 is assigned to group 1 with probability $1/(3+\gamma)$, group 2 with probability $2/(3+\gamma)$, and group 3 with probability $\gamma/(3+\gamma)$, respectively (the bottom panel of Figure 1). This process is equivalent to the model defined by equations (2) through (4) for latent groups in $t = 1$.

The dynamic process in the proposed model defined by equations (5) through (9) has analogous interpretation. Figure 2 illustrates the generative process of latent groups in $t = 2$. This process is the Chinese restaurant process given the group assignments in $t = 1$ with some stickiness added.

The top panel of Figure 2 illustrates the group assignment of the first unit in period 2. Here, only for the purpose of illustration, it is assumed that two units are in group 1 and the other two units are in group 2 in period 1. In period 2, unit 1 has some stickiness to group 1, because it is in group 1 in the previous period. This stickiness is represented by probability parameter p. However, to model possible change in group assignment for unit 1, unit 1 enters the Chinese restaurant process given group assignments in period 1 with probability $1 - p$. This Chinese restaurant process works as follows: the probability of unit 1 being assigned to group k is proportional to the number of the units assigned to group k in period 1. For example, the probability that unit 1 will be assigned to group 2 is $(1-p) \times 2/(4+\gamma)$, because unit 1 enters the Chinese restaurant process with probability $1 - p$ and if it does, it will be assigned to group 2 with probability proportional to 2, which is the number of the units assigned to group 2 in period 1. On the other hand, unit 1 may still stay in group 1 even if it enters the Chinese restaurant process. The probability of that happening is $(1-p) \times 2/(4+\gamma)$, because it follows the Chinese restaurant process with probability $1 - p$ and the conditional probability of group 1 is proportional to 2, which is the number of the units in group 1 in period 1. Since unit 1 is sticky to group 1 with probability p, the total probability of unit 1 being assigned to group 1 is $p + (1 - p)2/(4 + \gamma)$.

The group assignment of the other units again follows the Chinese restaurant process conditional on the period 1 group assignment, with stickiness parameter p. The bottom panel of Figure 2 illustrates the group assignment of unit 2 in period 2, assuming that unit 1's assignment in period 2 is group 2. Again, unit 2 is sticky to the cluster to which the same unit was assigned in period 1. Therefore, the probability that unit 2 goes to cluster 2 is p plus the probability determined by the Chinese restaurant process. The difference from the case of unit 1 is that the Chinese restaurant process for unit 2 involves unit 1 in period 2. That is, the probability that

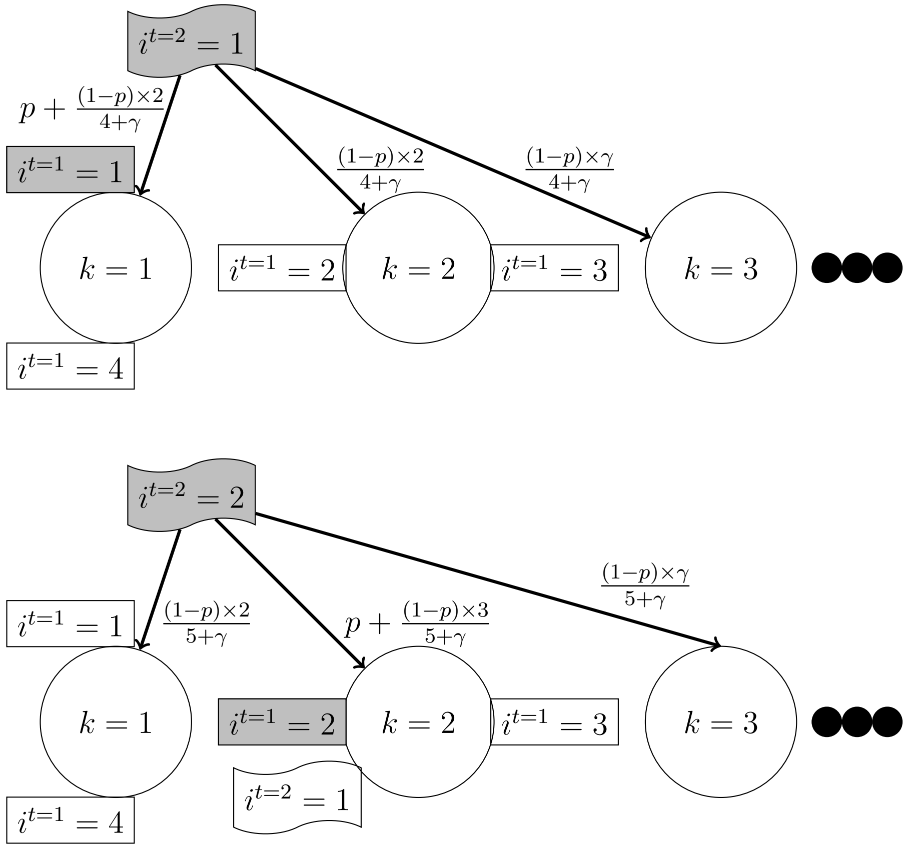


Figure 2: Intergenerational Chinese Restaurant Process. This figure illustrates the generative process of the latent cluster membership in $t = 2$. The fractions near arrows are the probabilities of cluster assignment. The top and bottom panels present the probability that unit 1 and 2 in period 2 are assigned to each cluster, respectively. The probability that a unit assigned to a cluster depends on cluster assignments in $t = 1$ and units already assigned to clusters in $t = 2$. The probability is proportional to $1-p$ times the number of units in $t = 1$ *and* 2, and the probability that a unit creates a new cluster is proportional to $1 - p$ times parameter γ. In addition, there is stickiness that stickiness parameter p is added to the probability that a unit is assigned to the cluster it occupied in $t = 1$.

unit 2 is assigned to cluster 2 should be $p + (1-p) \times 3/(5+\gamma)$, because two units in period 1 and unit 1 in period 2 are assigned to cluster 2.

The probability of each of the other groups is simply determined by the number of units in that group. Since group 1 has two units from period 1, the conditional probability that unit 2 belongs to group 1 is $2/(5+\gamma)$. On the other hand, the conditional probability of a new group formation (group 3 in the figure) is $\gamma/(5+\gamma)$. The marginal probabilities are obtained by multiplying the conditional probability by $1-p$, the probability that the unit enters the Chinese restaurant process.

3.3 Markov Chain Monte Carlo Algorithm for Estimation

In this part, we introduce the Markov Chain Monte Carlo (MCMC) algorithm to estimate the model constructed through the dynamic stick-breaking process. In general, we use the blocked Gibbs sampling algorithm. It includes two major steps. One step is to sample the group-specific parameter Θ_g that defines the data generating function $f(\Theta_k)$ for $k = 1, 2, ...,$ conditioned on the current group assignment $g[it] = k$. Sampling Θ_k from its posterior distribution is simple when the prior distribution for Θ_k is conjugate. The second step is to sample group assignments $g[it]$ for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$ conditioned on Θ_k for $k = 1, 2, ...$ Sampling $g[it]$ is relatively complicated, as the posterior distribution is conditioned on both cross-sectional and cross-time relationships with other units. We combine the forward-backward algorithm designed for change-point models (Chib, 1998) and the truncation approximation approach developed for stick-breaking priors (Ishwaran and James, 2001) to sample $g[it]$. In what follows, we describe the details for this MCMC algorithm.

Following the truncation approximation approach of Ishwaran and James (2001), at the beginning of the algorithm, we set an arbitrarily large number K at which we truncate the number of groups. In the posterior distributions of $g[it]$ for all $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, the probabilities for most groups will be zero. In this way, the model estimates the number of groups automatically.

With K set, we initialize the starting values of $g[it]$ for all $i = 1, ..., N$ and $t = 1, ..., T$. Then, each iteration of the Gibbs sampler proceeds as follows.

First, we update Θ_k for $k = 1, 2, ..., K$. The posterior distribution of Θ_k is only conditioned on X_{it} and Y_{it} for units in group k. The sampling algorithm for Θ_k is specific to the form of function $f(\Theta_{g[it]})$. If possible, set the prior for Θ_k a conjugate prior to make the sampling simple.

Second, we sample p, the probability that a unit stays in the same group as before and does not go through a new round of group assigning process. To make the sampling of p simpler, we first introduce a series of intervening dummy variables d_{it} for $i = 1, 2, ..., N$ and $t = 1, 2, 3, ...T$ to indicate whether a specific unit direly stays in the same group as before. The posterior distribution of d_{it} is only conditioned on $g[it]$, $g[i, t-1]$, and p. When $g[it] \neq g[i, t-1]$, d_{it} must be 0; when $g[it] = g[i, t-1]$, unit i may directly stays in the same group as before, or may be assigned to the same group through the new round of group assigning process. Specifically,

$$p(d_{it} = 1) = \begin{cases} 0 & \text{if } g[it] \neq g[i, t-1]; \\ \frac{p}{p+(1-p)q_k^t} & \text{if } g[it] = g[i, t-1] = k. \end{cases}$$

As we set the prior distribution for p a Beta distribution, conditioned on d_{it} for all $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, the posterior distribution of p is also a Beta distribution. Specifically, we sample p as follows:

$$p \sim \mathsf{Beta}(\alpha_p + N_1, \beta_p + N_2)$$

$$N_1 = \sum_{i=1}^{N} \sum_{t=2}^{T} d_{it}$$

$$N_2 = \sum_{i=1}^{N} \sum_{t=2}^{T} (1 - d_{it})$$

Third, we update the stick-breaking weight π_k^t and q_k^t for $k = 1, 2, ...K$ and $t = 1, 2, ..., T$. As we set K to be an arbitrarily large number to approximate the stick-breaking prior with infinite number of groups, the posterior distribution of π_k^t simply becomes a Beta distribution:

$$\pi_k^t \sim \mathsf{Beta}(1 + n_k^{t-1} + n_k^t, \gamma + \sum_{l=k+1}^{K} n_l^{t-1} + \sum_{l=k+1}^{K} n_l^t)$$

$$n_k^{t-1} = \sum_{i=1}^{N} \mathbb{I}(g[i, t-1] = k)$$

$$n_k^t = \sum_{i=1}^{N} (1 - d_{it})\mathbb{I}(g[it] = k)$$

Once π_k^t is updated, calculate q_k^t:

$$q_k^t = \pi_k^t \prod_{l=1}^{k-1} (1 - \pi_l^t)$$

The last step is to sample group assignments $g[it]$ for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$. Let $\mathbf{g}[t] \equiv (g[1t], g[2t], ..., g[Nt])'$, $\mathbf{q}^t \equiv (q_1^t, q_2^t, ..., q_K^t)'$, $\mathbf{Y}_t \equiv (Y_{1t}, Y_{2t}, ..., Y_{Nt})'$. Following Chib (1998), we sample $\mathbf{g}[T]$,$\mathbf{g}[T-2]$, ...,$\mathbf{g}[1]$ in turn:

$$Pr(g[it] = k | \mathbf{g}[T], ..., \mathbf{g}[t+1], p, \mathbf{q}^T, ..., \mathbf{q}^1, \mathbf{Y}_T, ..., \mathbf{Y}_1, \Theta_1, ..., \Theta_K) \tag{10}$$

$$\propto \underbrace{Pr(g[i, t+1] | p, g[it] = k, \mathbf{q}^{t+1})}_{\text{part 1}} \underbrace{Pr(g[it] = k | p, \mathbf{q}^t, ..., \mathbf{q}^1, \mathbf{Y}_t, ..., \mathbf{Y}_1, \Theta_1, ..., \Theta_K)}_{\text{part 2}} \tag{11}$$

The equation above shows a decomposition of the conditional posterior distribution of $g[it]$ given the observed data, the model parameters, and group assignments for period $t+1$ through T. Part 1 is the probability of $g[i, t+1]$ given that $g[i, t]$ is in group k. Then part 1 is :

$$Pr(g[i, t+1] = l | p, g[it] = k, q_l^{t+1}) = (1-p)q_l^{t+1} + p\mathbb{I}(l = k)$$

Part 2 is the conditional probability of $g[it] = k$ given the model parameters and the data from period 1 up to period t. Part 2 can be further decomposed as:

$$Pr(g[it] = k | p, \mathbf{q}^t, ..., \mathbf{q}^1, \mathbf{Y}_t, ..., \mathbf{Y}_1, \Theta_1, ..., \Theta_K))$$

$$\propto \underbrace{f(Y_{it} | g[it] = k, \mathbf{Y}_{t-1}, ..., \mathbf{Y}_1, \Theta_1, ..., \Theta_K)}_{\text{part } a} \underbrace{Pr(g[it] = k | p, \mathbf{q}^t, ..., \mathbf{q}^1, \mathbf{Y}_{t-1}, ..., \mathbf{Y}_1, \Theta_1, ..., \Theta_K)}_{\text{part } b}$$

Part a is the distribution of outcome variable Y_{it} conditioned on contemporary group assignment $g[it]$, outcome variables of former periods, and group-specific parameters Θ_k for $k = 1, 2, ..., K$. Given the group assignment $g[it] = k$ and parameter for the group Θ_k, the distribution of Θ_k is conditionally independent of the data in the other periods:

$$f(Y_{it}|g[it]=k, \mathbf{Y}_{t-1}, ..., \mathbf{Y}_1, \Theta_1, ..., \Theta_K) = f(Y_{it}|\Theta_k)$$

Compared to the conditional distribution of $g[it]$ in Part 2, in Part b $g[it]$ is no longer conditioned on $\mathbf{Y}_t$. Part b can be further decomposed as:

$$Pr(g[it]=k|p, \mathbf{q}^t, ..., \mathbf{q}^1, \mathbf{Y}_{t-1}, ..., \mathbf{Y}_1, \Theta_1, ..., \Theta_K)$$

$$= \sum_l \underbrace{Pr(g[it]=k|p, q_k^t, g[i,t-1]=l)}_{\text{part } c} \underbrace{Pr(g[i,t-1]=l|p, \mathbf{q}^{t-1}, ..., \mathbf{q}^1, \mathbf{Y}_{t-1}, ..., \mathbf{Y}_1, \Theta_1, ..., \Theta_K)}_{\text{part } d}$$

Now one can immediately observe the recursive structure between the equation above and part 1 and part 2 in equation (11). Therefore, the standard forward recursion algorithm computes the conditional posterior given by equation (10), and the backward sampling algorithm generates MCMC draws from the conditional posterior of the group assignment.

4 Simulation Study

In this section, we describe two simulations to highlight two kinds of group membership changes over time. In one simulation, we generate group memberships based on a structural break model; in the other simulation, we generate group memberships based on a gradual change model. We will show that the proposed method works well in both situations.

4.1 Data Generating Process

Recall the question of changing voting group we described in the motivating example. To simplify, we set up the question as the following: there are $T = 30$ parliamentary sessions, $N = 50$ representatives, and $M = 4$ issues to vote in each session. There are different voting groups in the parliament. For representative i in a voting group g, the probability of voting "yea" for issue j in session t follows a Bernoulli distribution, $\mathsf{Bernoulli}(\theta_{gjt})$. We use different ways to generate group memberships in simulation 1 and simulation 2.

Simulation 1: a structural break model. Two transition points at $t = 11$ and 21 separate the 30 parliamentary sessions into three periods. In the first period, there are 3 groups, with 20 representatives in group 1, 20 representatives in group 2, and 10 representatives in group 3; entering into the second period, 5 representatives in group 1 change to group 2 and 10 representatives in group 2 shift to group 1; in the last period, 5 representatives in group 1 change to group 2, 5 representatives in group 3 move to group 1, and the other 5 representatives in group 3 move to group 2.

To summarize, there are 3 groups in the first and second periods, and only 2 groups in the last period. For each group, we generate the parameter θ_{gjt} of the Bernoulli distribution that models the voting outcomes from a uniform distribution. For group 1, the four uniform distributions for the four voting issues are **Uniform**$(0.8, 1)$, **Uniform**$(0.7, 1)$, **Uniform**$(0, 0.2)$ and **Uniform**$(0, 0.3)$; for group 2, they are **Uniform**$(0, 0.2)$, **Uniform**$(0, 0.3)$, **Uniform**$(0.8, 1)$ and **Uniform**$(0.7, 1)$; for group 3, they are **Uniform**$(0.7, 1)$, **Uniform**$(0, 0.2)$, **Uniform**$(0, 0.3)$ and **Uniform**$(0.8, 1)$.

Simulation 2: a gradual change model. In the first 5 parliamentary sessions, there are 3 groups with 20, 20, and 10 representatives in each group. From $t = 6$ to $t = 25$, representatives in group 1 may change to group 2 with a probability of 0.5, and this change may happen at any time during this period; similarly, with a probability of 0.5, representatives in group 2 may shift to group 1 at a random time; for representatives in group 3, they will shift to group 1 with a probability of 0.5 and otherwise they will shift to group 2. The process to generate θ_{gjt} for each group g and each issue j in session t is the same as the process in simulation 1.

4.2 The Model

In section 3, we only describe a general version of the proposed method. Here, we introduce the detailed model we use to analyze the simulated data.

Let $g[it]$ represent the group of representative i in session t. Then,

$$g \sim \mathsf{IgCRP}(\gamma, \alpha_{\mathsf{p}}, \beta_{\mathsf{p}})$$

For V_{ijt}, the vote of issue j that representative i in sessions t casts, we assume it follows a Bernoulli distribution $\mathsf{Bernoulli}(\theta_{\mathsf{jk}})$ for $g[it] = k$. Unlike the data generating process, we assume that θ_{jk} does not change with t. As we will show, the model still works well under this assumption.

$$V_{ijt} \sim \mathsf{Bernoulli}(\theta_{\mathsf{j,g[it]}})$$

For $g[it] = 1, 2, ..., k, ...$, we assume:

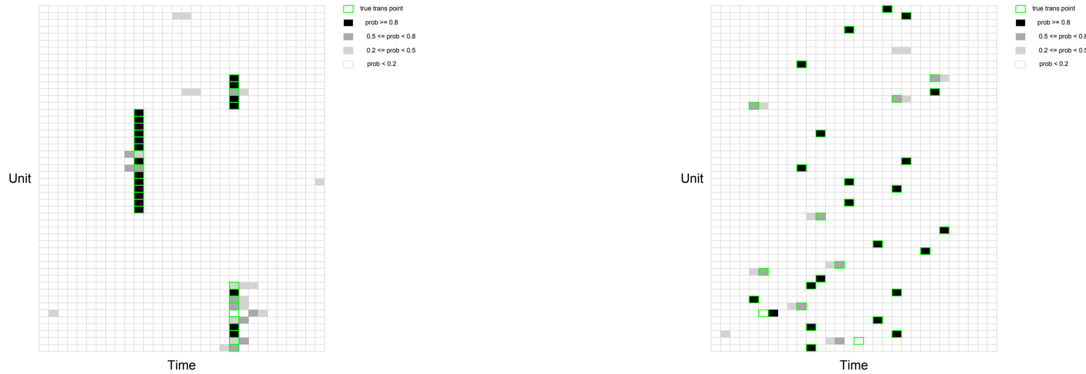
$$\theta_{jk} \sim \mathsf{Beta}(\alpha_{\theta,\beta_\theta}).$$

The MCMC algorithm for this model is a simplified version of the MCMC algorithm we use for the empirical application. Thus, we skip the detailed algorithm here.

4.3 Results

We first investigate whether the proposed method can detect the true transition points. For each unit, we calculate the probability that the unit in the current time and in the former time are in different groups. A probability approaching 1 indicates a transition point. As shown in Figure 3, the proposed method detects almost all transition points, successfully recovering both the structural break model and the gradual change model.

Besides checking whether the proposed method can detect the true transition points, we also investigate whether our method recovers true group memberships across units and over time together. For this purpose, we calculate the probability that two observations (they are either



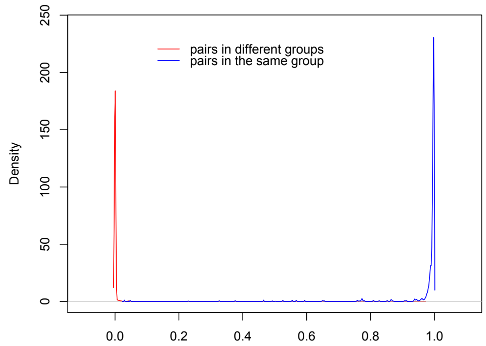
(1) Recover Structural Break Model (2) Recover Gradual Change Model

Figure 3: Probabilities of Changing to a New Group. In the left figure, data is generated through a structural break model; in the right figure, data is generated through a gradual change model. A square represents the probability that the unit in the current time changes to a different group. The true transition points are circulated with green lines. This figure shows that no matter the data is generated through a structural break model or a gradual change model, the proposed method works well in identifying the transition point.

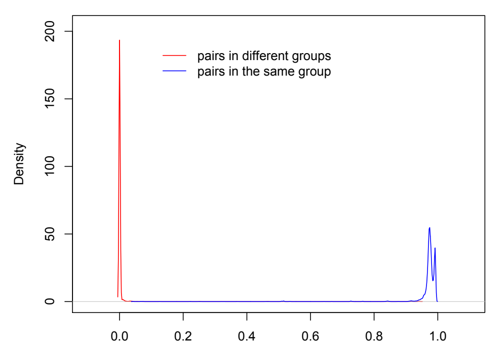
from different units,or in different time points, or both) are in the same group for all possible pairs. As we know the true group memberships in simulation studies, we separate pairs in different groups from pairs in the same group. For pairs in different groups, we expect the density of probabilities that two observations are in the same group to center around 0; for pairs in the same, we expect the density of probabilities to center around 1. As shown in Figure 4, the proposed method discovers the true group memberships for both the structural break model and the gradual transition model.

5 Empirical Application

In this section, we apply the proposed methodology to study shifting party coalitions on civil rights. First, we introduce the four measures of civil rights positions Schickler (2016) uses. Then, we explain how to apply the IgCRP model to the empirical example. We introduce the specific statistical model that we use to analyze the data of civil rights positions. After this, we show the results of statistical analysis. Besides confirming Schickler's original argument that racial realignment was a gradual procedure that started ever since the New Deal period, the new results provide more information on how old coalitions dissolved and new coalitions emerged. The new results also provide inference on how positions of congressional members in individual states changed over the years.



(1) Recover Structural Break Model



(2) Recover Gradual Change Model

Figure 4: Densities of Probabilities that Two Observations are in the Same Groups. As we know the true memberships, we can separate pairs in the same groups from pairs in different groups and calculate the densities separately.

5.1 Measures of Civil Rights Positions and Data

The original data set of Schickler (2016) consists of four types of data measuring House members' positions on civil rights. Roll-call votes are most commonly used to measure members' preferences. However, Schickler points out that roll calls could be a misleading measure. Only a very small portion of civil rights bills reached the floor. Most civil rights bills were blocked by the House Rules Committee, which was dominated by southern senior Democrats. Instead, Schickler suggests using other legislative behaviors to complement roll-call votes, including signing discharge petitions to advance civil rights bills, floor speeches in supporting civil rights, as well as sponsoring a civil right bill.

Roll-call Votes. Following the original work of Schickler (2016), we use a dummy variable to indicate whether a House member supported a civil rights bill or not. "1" means that a member voted "yes" for a bill, and "0" indicates that the member either voted "no" or was absent. In Table 1, we list the number of roll calls about civil rights from the 73rd Congresses (1933-35) to the 92nd Congress (1971-73). It shows that during the period of this study, not every Congress had roll calls of civil rights issues. The table also presents the means and standard deviations of the roll-call variable by parties and years. The standard deviations are as large as the means, meaning the within-party heterogeneity is high.

Discharge Petition Signatures. If a legislative committee or the Rules Committee has blocked a bill from reaching the floor for a long time, House members can sign a petition to ask for advancing the bill.[3] We use a dummy variable to measure whether a member signed a certain petition. Table 1 shows the number of petitions in each Congress. It also shows the means and standard deviations of the petition variable by parties and by years. On average, a Congress

[3]For a bill in a legislative committee, the waiting period is 20 days; for a bill in the Rules Committee, the waiting period is 7 days Schickler et al. (2010, p. 676).

member was less likely to sign a petition than to vote "yes" for a civil rights bill. Schickler explained that signing a discharge petition is costly for members of Congress, as it is a sign of violating "congressional norms" and intruding on "committee authority". Only members who "cared enough" about a civil rights issue would sign a petition (p183). Hence, discharge petition signatures serve as a good measure of the *intensity* of a member's preference.

Floor Speeches. We use a dummy variable to indicate whether a House member delivered at least one speech on the floor to support civil rights in a certain Congress. Floor speeches serve as an important way for House members to signal commitments to their constituents. Competition for speech time is fierce on the floor. Using precious time to deliver a speech to support civil rights indicates that a civil rights issue is one of the member's top priorities. Table 1 shows that the speech variable varies a lot over time. For instance, while in the 73rd Congress (1933-35), only about 2.5% members spoke about civil rights, in the 89th Congress (1965-67), about 22.5% Democrats and 16.1% Republicans delivered speeches supporting civil rights issues.

Bill Sponsorship. Bill Sponsorship is a count variable measuring how many civil rights bills a member initialized in a certain Congress. On average, one member sponsored less than 1 civil rights bill. The variation is large, as some members initialized many bills while most other members didn't initialize any civil rights bills.

Table 1: Summary Statistics of Four Measures of House Members' Positions on Civil Rights.

Con.	Year	Roll Calls			Petitions			Speeches		Bill Sponsorship	
		No.	Dem.	Rep.	No.	Dem.	Rep.	Dem.	Rep.	Dem.	Rep.
73	1933-35	0			1	0.156 (0.363)	0.717 (0.453)	0.025 (0.156)	0.025 (0.157)		
74	1935-37	0			1	0.392 (0.489)	0.771 (0.422)	0.037 (0.189)	0.029 (0.167)		
75	1937-39	2	0.507 (0.5)	0.521 (0.501)	2	0.224 (0.417)	0.618 (0.487)	0.059 (0.236)	0.097 (0.297)		
76	1939-41	2	0.476 (0.5)	0.964 (0.186)	3	0.161 (0.368)	0.354 (0.479)	0.052 (0.222)	0.096 (0.295)		
77	1941-43	3	0.589 (0.492)	0.975 (0.157)	4	0.16 (0.366)	0.222 (0.416)	0.029 (0.169)	0.018 (0.132)		
78	1943-45	2	0.478 (0.5)	0.924 (0.265)	4	0.178 (0.383)	0.209 (0.407)	0.022 (0.147)	0.005 (0.068)		
79	1945-47	3	0.531 (0.499)	0.895 (0.307)	4	0.305 (0.46)	0.312 (0.464)	0.036 (0.188)	0.015 (0.123)		
80	1947-49	1	0.422 (0.495)	0.941 (0.235)	2	0.19 (0.393)	0.081 (0.273)	0.051 (0.22)	0.004 (0.063)	0.102 (0.606)	0.087 (0.399)
81	1949-51	4	0.567 (0.496)	0.83 (0.376)	3	0.234 (0.424)	0.076 (0.265)	0.053 (0.224)	0.023 (0.149)	0.132 (0.768)	0.08 (0.292)
82	1951-53	0			1	0.05 (0.218)	0.019 (0.138)	0.074 (0.263)	0.024 (0.154)	0.107 (0.536)	0.053 (0.33)
83	1953-55	0			2	0.272 (0.445)	0.045 (0.208)	0.073 (0.261)	0.041 (0.198)	0.178 (0.79)	0.032 (0.199)
84	1955-57	1	0.522 (0.501)	0.876 (0.33)	1	0.403 (0.491)	0.232 (0.423)	0.14 (0.348)	0.039 (0.195)	0.43 (1.625)	0.074 (0.358)
85	1957-59	2	0.55 (0.498)	0.896 (0.306)	2	0.154 (0.361)	0.071 (0.257)	0.142 (0.349)	0.078 (0.27)	0.446 (1.494)	0.118 (0.428)
86	1959-61	5	0.591 (0.492)	0.87 (0.337)	1	0.564 (0.497)	0.289 (0.455)	0.143 (0.351)	0.075 (0.265)	0.254 (1.022)	0.176 (0.792)
87	1961-63	1	0.655 (0.476)	0.855 (0.353)	0			0.095 (0.294)	0.023 (0.149)	0.374 (1.453)	0.158 (1.004)
88	1963-65	2	0.635 (0.482)	0.861 (0.347)	2	0.283 (0.451)	0.066 (0.249)	0.182 (0.386)	0.143 (0.351)	0.517 (1.642)	0.522 (1.169)
89	1965-67	2	0.761 (0.427)	0.793 (0.406)	0			0.225 (0.418)	0.161 (0.369)	0.225 (0.684)	0.629 (0.845)
90	1967-69	0			0			0.23 (0.422)	0.053 (0.224)	0.254 (0.953)	0.026 (0.191)
91	1969-71	0			1	0.46 (0.499)	0.11 (0.314)			0.352 (0.951)	0.14 (0.481)
92	1971-73	0			1	0.318 (0.467)	0.406 (0.492)			0.512 (1.813)	0.299 (0.993)

Note: This table presents the distributions of four measures of civil rights positions of House members by parties and by year. Roll calls represent a dummy variable indicating whether a member voted "yes" or not for a civil rights bill; petitions represent a dummy variable for signing a discharge petition for advancing a civil right bill; speeches represent a dummy variable indicating whether a member delivered at least one pro-civil right speech during a certain Congress; bill sponsorship is a count variable measuring how many civil rights bills a member initialized during a certain Congress.

5.2 Statistical Model for Analyzing Racial Realignment

Here we introduce the specific statistical model we use to analyze the data of congressional action. We regard a party's House members in one state as the basic unit of analysis, assuming members from the same state and in the same party share the same constraint imposed by their constituencies. We use the IgCRP model along with the voting block model of Spirling and Quinn (2010) as the data model, to group party-state units and to model the evolvement of group memberships over the years. Party-state units in the same group share the group-specific parameters that we use to model the four types of legislative behaviors. In this way, we analyze the four types of data measuring House members' civil rights positions in a unified framework.

Specifically, let $t = 73, 74, ..., 92$ index the 73rd Congress (1933-35) to the 92nd Congress (1971-73), $s = 1, 2, ..., 50$ index 50 states, and D and R represent the Democratic Party and the Republican Party respectively. Let $i = 1, 2, ..., I^D_{st}$(or I^R_{st}) represent a Democratic/Republican House member from state s in the tth Congress, and $j = 1, 2, ..., J^V_t/J^P_t$ index a civil rights vote or petition in the tth Congress.

Let $g^D[st]$ represent the group the Democratic Party of state s in the tth Congress belongs to. Similarly, $g^R[st]$ represents the group of the Republican party. We assume $g^D[st]$ and $g^R[st]$ share the same IgCRP prior:

$$g^D, g^R \sim \mathsf{IgCRP}(\gamma, \alpha_p, \beta_p)$$

Roll-call Votes. Let V^D_{istj} represent the outcome of roll-call vote j in the tth Congress for Democratic member i from state s. Similarly, V^R_{istj} represents the roll-call vote of a Republican member. We assume V^D_{istj} and V^R_{istj} follow the following distributions:

$$V^D_{istj} \overset{ind.}{\sim} \mathsf{Bernoulli}(\theta_{g^D[st]})$$

$$V^R_{istj} \overset{ind.}{\sim} \mathsf{Bernoulli}(\theta_{g^R[st]})$$

For $g^D[st] = 1, 2, ..., k...$ and $g^R[st] = 1, 2, ..., k...$,

$$\theta_k \sim \mathsf{Beta}(\alpha_\theta, \beta_\theta)$$

Discharge Petition Signatures. Let P^D_{istj} and P^R_{istj} represent whether member i from state s signed petition j in the tth Congress. D and R index Democrats and Republicans. We assume P^D_{istj} and P^R_{istj} follow the following distributions:

$$P^D_{istj} \overset{ind.}{\sim} \mathsf{Bernoulli}(\eta_{g^D[st]})$$

$$P^R_{istj} \overset{ind.}{\sim} \mathsf{Bernoulli}(\eta_{g^R[st]})$$

For $g^D[st] = 1, 2, ..., k...$ and $g^R[st] = 1, 2, ..., k...$,

$$\eta_k \sim \mathsf{Beta}(\alpha_\eta, \beta_\eta)$$

Floor Speeches. Let S^D_{ist} and S^R_{ist} represent whether Democratic or Republican member i from state s delivered at least one floor speech to support civil rights. Their distributions are:

$$S^D_{ist} \overset{ind.}{\sim} \mathsf{Bernoulli}(\omega_{g^D[st]})$$

$$S_{ist}^{R} \overset{ind.}{\sim} \mathsf{Bernoulli}(\omega_{g^R[st]})$$

For $g^D[st] = 1, 2, ..., k...$ and $g^R[st] = 1, 2, ..., k...$,

$$\omega_k \sim \mathsf{Beta}(\alpha_\omega, \beta_\omega)$$

Bill Sponsorship. Finally, let B_{ist}^{D} and B_{ist}^{R} represent the number of civil rights bills member i from state s sponsored in the $t$the Congress. B_{ist}^{D} and B_{ist}^{R} follow the following distributions:

$$B_{ist}^{D} \overset{ind.}{\sim} \mathsf{Poisson}(\lambda_{g^D[st]})$$

$$B_{ist}^{R} \overset{ind.}{\sim} \mathsf{Poisson}(\lambda_{g^R[st]})$$

For $g^D[st] = 1, 2, ..., k...$ and $g^R[st] = 1, 2, ..., k...$,

$$\lambda_k \sim \mathsf{Gamma}(a_\lambda, b_\lambda)$$

5.3 MCMC Algorithm

As introduced in Section 3.3, the MCMC algorithm includes two major steps. One step is to sample the group-specific parameters, here θ_k, η_k, ω_k, and λ_k, for $k = 1, 2, ...$, conditioned on group assignments $g^D[st]$ and $g^R[st]$ for $s = 1, 2, ..., 50$ and $t = 73, 74, ..., 92$. We set priors for the group-specific parameters conjugate priors. Conditioned on group assignments, posterior distributions of θ_k, η_k, ω_k are Beta distributions, and the posterior distribution of λ_k is a Gamma distribution. The second step is to sample group assignments $g^D[st]$ and $g^R[st]$ for $s = 1, 2, ..., 50$ and $t = 73, 74, ..., 92$, conditioned on the group-specific parameters. We combine the forward-backward approach for change-point models and the truncation approximation approach for stick-breaking priors to sample group assignments. Details of the algorithm are shown in Appendix A.

5.4 Empirical Results

In this part, we first check whether our results are consistent with the original findings of Schickler (2016). Beyond this, we present the changes of the two parties' positions in individual states, cross-party coalition within a state, as well as cross-state coalition within a party.

Differences Across Parties. We use posterior means of outcome variables, $\{\theta_{g^D[st]}, \theta_{g^R[st]}\}$, $\{\eta_{g^D[st]}, \eta_{g^R[st]}\}$, $\{\omega_{g^D[st]}, \omega_{g^R[st]}\}$, $\{\lambda_{g^D[st]}, \lambda_{g^R[st]}\}$ to measure a party's position in state s and the tth Congress. The posterior mean carries easily-interpreted empirical meaning. For instance, $\theta_{g^D[st]}$ means the probability that in the tth Congress a Democratic member from state s would vote "yes" for a civil rights roll call; $\lambda_{g^D[st]}$ means the average number of civil rights bill Democratic members in state s sponsored in the tth Congress.

Then we investigate how the difference between the two parties' positions changed over time.

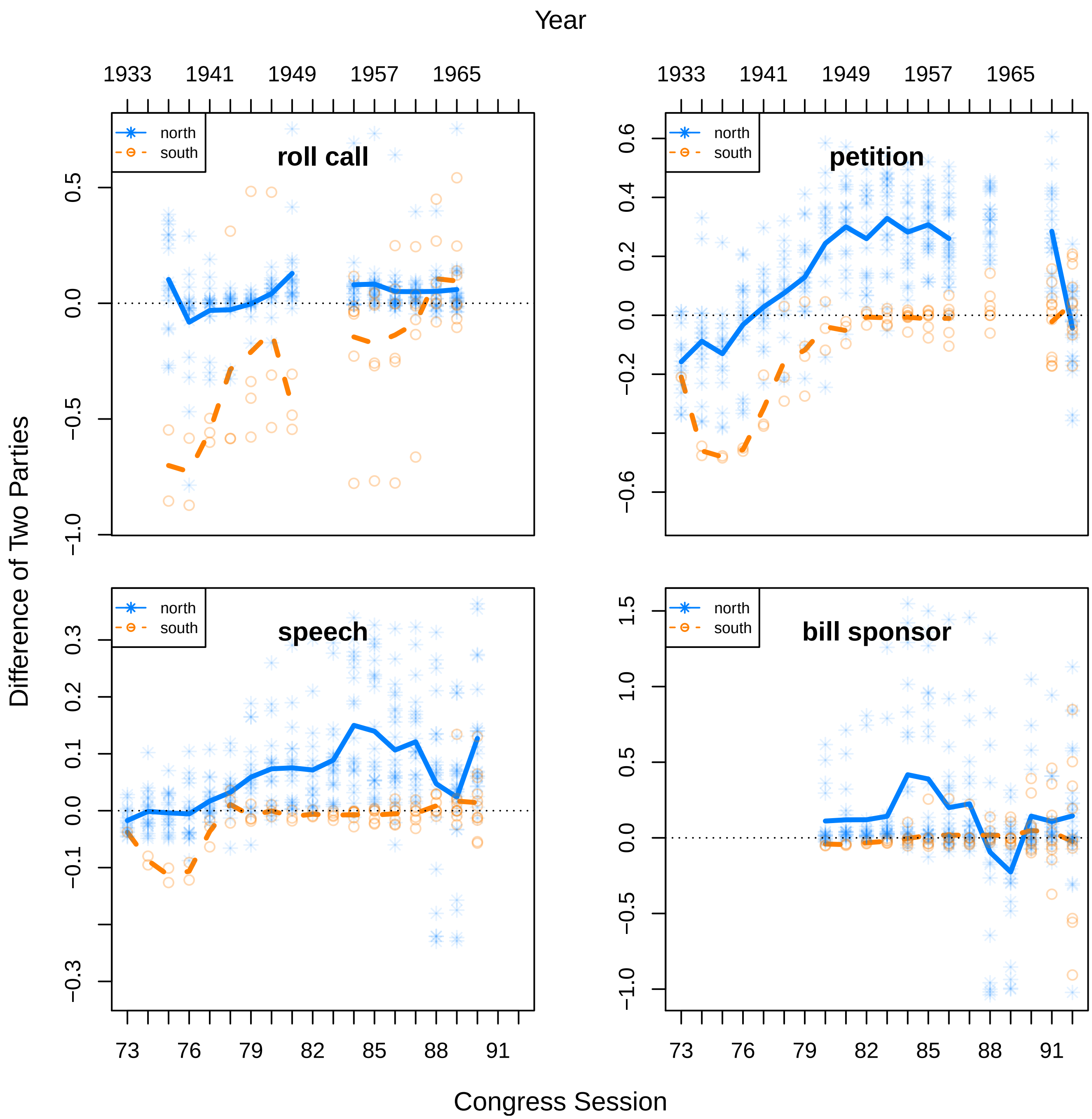


Figure 5: Differences between the Democratic Party and the Republican Party on Civil Rights Position in Each Congress. A positive value in the y-axis indicates that Democrats are more supportive of civil rights. Each point represents the posterior mean of an outcome variable in the corresponding state and Congress. The solid and dotted lines represent averages of northern and southern states respectively.

To compare our results with the original findings, We take the average of northern states and southern states separately to check whether the two parties' positions had started to diverge longer before the 1960s in the North.

We present the results in Figure 5. Consistent with the original findings, we find that ever since the 1940s, on average, northern Democrats were more supportive of civil rights than northern

Republicans, and the gap of the two parties in the North kept increasing until the 1960s. Blue solid lines in each panel take positive values on the y-axis, which indicates that northern Democrats are more supportive of civil rights than Republicans from the same states. While the Democratic Party was only moderately different from the Republican Party in terms of roll call votes, for all the other three measures, the gaps were substantially large.

The results also automatically capture the well-known situation in the South that southern Democrats stood firmly against civil rights legislation. In the 1930s, the average differences between Democrats and Republicans in the South were negative. However, since the 1940s, the partisan gap decreased gradually, and by the end of the 1950s, the differences became near 0. This change was partly driven by the shift of southern Republicans' positions and partly driven by the gradually dissolved Democratic coalition in the South. We will further explore this change in the following analysis.

Beyond the average differences between Democrats and Republicans, our model also reveals how the partisan differences in individual states evolved over time. Figure 5 shows that even viewed from the perspective of roll-call votes, a measure only moderately captures the two parties' difference, the pattern of partisan differences started to change in northern states. In the 1930s, there were many northern states where Republicans were more supportive of civil rights than Democrats. Representative states were Missouri, Nebraska, and Oregon.[4] In the 1940s, the number of such states decreased gradually. By the end of the 1940s, in few states the partisan differences were negative. The proposed methodology also helps identify states where Democrats pioneered in advancing civil rights. New York, California, and Illinois are such states.

Cross-party Coalitions within a State. Besides posterior means of outcome variables, we also use the probability that the Democratic Party and the Republican Party are in the same group to measure how the two parties differed in each state and each Congress. In addition, we calculate the average probabilities of northern and southern states separately.

As shown in Figure 6, the average probability of northern states declines fast in the 1940s. Before, in some northern states such as Connecticut and New Jersey, the probabilities are as high as 0.7 to 0.8, meaning the Democrats and Republicans took very similar positions on civil rights issues. Since the 1950s, in nearly all northern states the probabilities are below 0.5. In many states, the probabilities are near zero, meaning Democrats and Republicans took very different positions on civil rights issues.

In the South, the Democratic Party and the Republican Party used to hold sharply different attitudes toward civil rights. The probabilities of southern states are near zero in most Congresses before 1950. In the 1950s, the probabilities increase fast. During some Congresses of the 1950s and the 1960s, in states such as Virginia and North Carolina, the probabilities approach 1, meaning that in these states southern Democrats and Republicans hardly disagreed with each other on civil rights issues.

[4]Note that the "North" is defined all the non-South states. See footnote 1

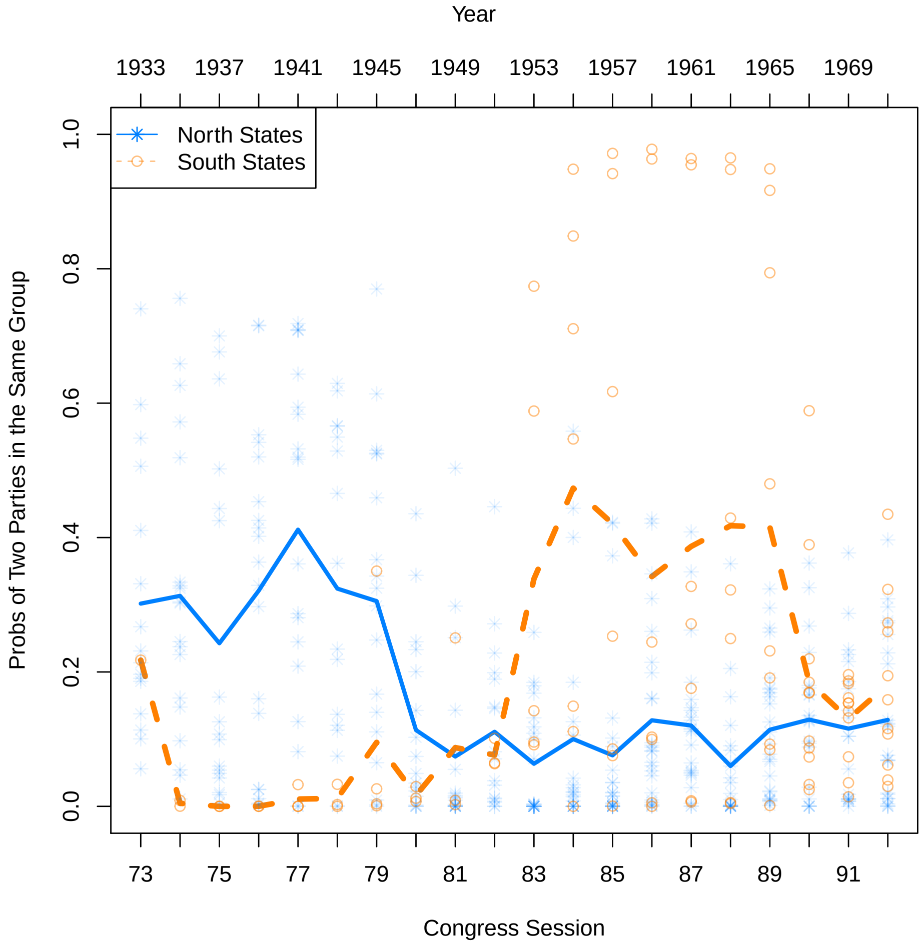


Figure 6: Probabilities that the Democratic Party and the Republican Party Are in the Same Group in Supporting Civil Rights in Each Congress. Each point represents the probability of an individual state. The solid and dotted lines represent averages of northern and southern states respectively.

Cross-state Coalitions within a Party. Besides coalitions across parties, we also illustrate within-party coalitions across different states and how the coalitions changed over time. In Figure 7, we show the probabilities that Democrats/Republicans in different states are in the same group. We also show the average probabilities of northern states and southern states separately. The probabilities measure the degree of party solidarity.

In the 1930s, northern Republicans were more unified than northern Democrats. Starting from the end of the 1930s, northern Democrats became more likely to act together. While Democratic solidarity decreased slightly in the first half part of the 1940s, after the 80th Congress (1947-49), the degree of solidarity restarted to increase slowly again. At the same time, Republican solidarity decreased gradually over time. By the end of the 1950s, northern Republicans had become less united than northern Democrats.

In contrast, in the South, Democratic used to be firmly unified. As northern states started to act together to support civil rights, southern Democrats became more united in opposing civil rights legislation. However, since the 1950s, the southern solidarity of the Democratic Party

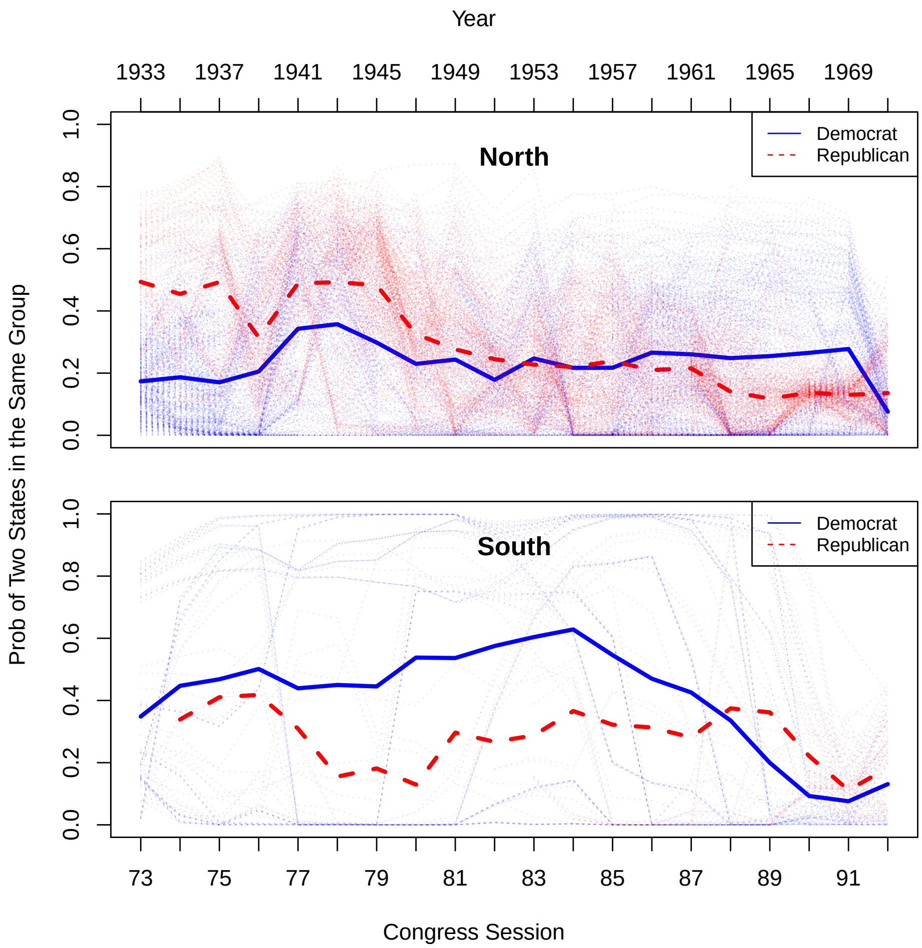


Figure 7: Probabilities that a Party in Two Different States are in the Same Groups in Each Congress. Light lines represent the probabilities of two specific states. Thick lines represent average probabilities of the North and the South.

started to dissolve gradually. By the end of the 1960s, the average probability that Democrats in two southern states are in the same group has declined to around 0.1.

6 Concluding Remarks

Most societal changes are continuous processes rather than revolutions. As our empirical analysis indicates, the Democratic Party's shift to the pro-civil rights position began in the 1930s and continued gradually until the 1960s. Before the landmark presidential election of 1964, the Southern Democrats' unity against civil rights started eroding in the 1950s. To properly analyze such changes, one needs to model a continuing process of gradual shifts of data generating processes.

This paper presents a novel nonparametric Bayesian model, called the intergenerational Chinese restaurant process, to achieve that goal. In IgCRP, units are clustered into latent groups, and the group assignment is assumed to be a Markov process. In each time period, a unit remains in the same group as the previous period with a probability controlled by a stickiness parameter, or moves to another group with a probability determined by the Chinese restaurant process conditional on the group assignments in the previous period. Due to this group assignment process, IgCRP can properly model a gradual structural change.

IgCRP is widely applicable to any data set with the repeated measurement of multiple units, and thus the use of IgCRP would benefit many social scientists who are interested in temporal dynamics of unobserved heterogeneity. To facilitate this, we are developing free, open-source, and easy-to-use software for implementing IgCRP. We intend to create a package in the `R` system consisting of pre-written functions for estimation algorithms of the IgCRP model so that applied quantitative social scientists can simply use the package to conduct Bayesian analysis using the model.

References

Blackwell, D., and MacQueen, J. B. (1973), "Ferguson distributions via Pólya urn schemes," *The annals of statistics*, 1(2), 353–355.

Blackwell, M. (2018), "Game Changers: Detecting Shifts in Overdispersed Count Data," *Political Analysis*, 26(2), 230239.

Chib, S. (1998), "Estimation and Comparison of Multiple Change-point Models," *Journal of Econometrics*, 86(2), 221–241.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1(2), 209–230.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011), "A Sticky HDP-HMM with Application to Speaker Diarization," *The Annals of Applied Statistics*, 5(2A), 1020–1056.

Garmines, E. G., and Stimson, J. A. (1989), *Issue Evolution: Race and the Transformation of American Politics*, Princeton, NJ: Princeton University Press.

Gill, J., and Casella, G. (2009), "Nonparametric Priors for Ordinal Bayesian Social Science Models: Specification and Estimation," *Journal of the American Statistical Association*, 104(486), 453–454.
URL: *https://doi.org/10.1198/jasa.2009.0039*

Imai, K., and Tingley, D. (2012), "A Statistical Method for Empirical Testing of Competing Theories," *American Journal of Political Science*, 56(1), 218–236.

Ishwaran, H., and James, L. F. (2001), "Gibbs Sampling Methods for Stick-Breaking Priors," *Journal of the American Statistical Association*, 96(453), 161–172.

Kim, I. S., Liao, S., and Imai, K. (2020), "Measuring Trade Profile with Granular Product-Level Data," *American Journal of Political Science*, 64(1), 102–117.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12473*

Kyung, M., Gill, J., and Casella, G. (2011), "New findings from terrorism data: Dirichlet process random-effects models for latent groups," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 60(5), 701–721.
URL: *https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9876.2011.00768.x*

Pang, X., Friedman, B., Martin, A. D., and Quinn, K. M. (2012), "Endogenous Jurisprudential Regimes," *Political Analysis*, 20(4), 417–436.

Park, J. H. (2011), "Changepoint Analysis of Binary and Ordinal Probit Models: An Application to Bank Rate Policy Under the Interwar Gold Standard," *Political Analysis*, 19(2), 188–204.
URL: *http://www.jstor.org/stable/23011263*

Park, J. H. (2012), "A Unified Method for Dynamic and Cross-Sectional Heterogeneity: Introducing Hidden Markov Panel Models," *American Journal of Political Science*, 56(4), 1040–1054.
URL: *https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5907.2012.00590.x*

Schickler, E. (2016), *Racial Realignment: The Transformation of American Liberalism, 1932-1965*, Princeton, NJ: Princeton University Press.

Schickler, E., Pearson, K., and Feinstein, B. D. (2010), "Congressional Parties and Civil Rights Politics from 1933 to 1972," *The Journal of Politics*, 72(3), 672–689.

Sethuraman, J. (1994), "A Constructive Definition of Dirichlet Priors," *Statistica Sinica*, pp. 639–650.

Spirling, A., and Quinn, K. (2010), "Identifying Intraparty Voting Blocs in the UK House of Commons," *Journal of the American Statistical Association*, 105(490), 447–457.

Teh, Y. W. (2010), "Dirichlet Process," in *Encyclopedia of Machine Learning*, New York: Springer, pp. 280–287.

Traunmller, R., Murr, A., and Gill, J. (2015), "Modeling Latent Information in Voting Data with Dirichlet Process Priors," *Political Analysis*, 23(1), 1–20.
URL: *http://www.jstor.org/stable/24572838*

A MCMC Algorithm for the Empirical Application

In Section 3.3, we introduce the MCMC algorithm for a general IgCRP process. In this section, we present a specific algorithm for the empirical application. Most parts of the algorithm are the same as the corresponding parts of the general algorithm. Here we explain the parts specific to the application example.

As in Section 3.3, at the beginning of the algorithm, set an arbitrarily large number K to truncate the number of clusters. Then initialize the starting values of $g^D[st]$ and $g^R[st]$ for $s = 1, 2, ..., 50$ and $t = 73, 74, ..., 92$. After initialization, each iteration of the Gibbs sampler proceeds as follows:

1. Update $\theta_k, \eta_k, \omega_k, \lambda_k$ for $k = 1, 2, ..., K$

   (a) The Posterior Distribution of θ_k

$$\theta_k \sim \mathsf{Beta}(\alpha_\theta + N_\theta^1, \beta_\theta + N_\theta^0)$$

$$N_\theta^1 = \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^D}\sum_{j=1}^{J_t^V} V_{istj}^D \mathbb{I}(g^D[st] = 1) + \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^R}\sum_{j=1}^{J_t^V} V_{istj}^R \mathbb{I}(g^R[st] = 1)$$

$$N_\theta^0 = \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^D}\sum_{j=1}^{J_t^V}(1 - V_{istj}^D) \mathbb{I}(g^D[st] = 1) + \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^R}\sum_{j=1}^{J_t^V}(1 - V_{istj}^R) \mathbb{I}(g^R[st] = 1)$$

   (b) The Posterior Distribution of η_k

$$\eta_k \sim \mathsf{Beta}(\alpha_\eta + N_\eta^1, \beta_\eta + N_\eta^0)$$

$$N_\eta^1 = \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^D}\sum_{j=1}^{J_t^P} P_{istj}^D \mathbb{I}(g^D[st] = 1) + \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^R}\sum_{j=1}^{J_t^P} P_{istj}^R \mathbb{I}(g^R[st] = 1)$$

$$N_\theta^0 = \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^D}\sum_{j=1}^{J_t^P}(1 - P_{istj}^D) \mathbb{I}(g^D[st] = 1) + \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^R}\sum_{j=1}^{J_t^P}(1 - P_{istj}^R) \mathbb{I}(g^R[st] = 1)$$

   (c) The Posterior Distribution of ω_k

$$\omega_k \sim \mathsf{Beta}(\alpha_\omega + N_\omega^1, \beta_\omega + N_\omega^0)$$

$$N_\omega^1 = \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^D} S_{ist}^D \mathbb{I}(g^D[st] = 1) + \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^R} S_{ist}^R \mathbb{I}(g^R[st] = 1)$$

$$N_\omega^0 = \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^D}(1 - S_{ist}^D) \mathbb{I}(g^D[st] = 1) + \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^R}(1 - S_{ist}^R) \mathbb{I}(g^R[st] = 1)$$

(d) The Posterior Distribution of λ_k

$$\lambda_k \sim \mathsf{Gamma}(\alpha_\omega + C_\lambda, \beta_\omega + N_\lambda)$$

$$C_\lambda = \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^D} B_{ist}^D \mathbb{I}(g^D[st]=1) + \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^R} B_{ist}^R \mathbb{I}(g^R[st]=1)$$

$$N_\lambda = \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^D} \mathbb{I}(g^D[st]=1) + \sum_{t=73}^{92}\sum_{s=1}^{50}\sum_{i=1}^{I_{st}^R} \mathbb{I}(g^R[st]=1)$$

2. Sample the Transition Probability p

   To sample p, we first introduce a series of dummy variables d_{st}^D and d_{st}^R for $t = 2, 3, ...T$ and $i = 1, 2, ..., N$ to indicate self-transitions.

   (a) Sample d_{st}^D and d_{st}^D

$$p(d_{st}^D = 1) = \begin{cases} 0 & \text{if } g^D[st] \neq g^D[s, t-1]; \\ \frac{p}{p+(1-p)q_k^t} & \text{if } g^D[st] = g^D[s, t-1] = k. \end{cases}$$

$$p(d_{st}^R = 1) = \begin{cases} 0 & \text{if } g^R[st] \neq g^R[s, t-1]; \\ \frac{p}{p+(1-p)q_k^t} & \text{if } g^R[st] = g^R[s, t-1] = k. \end{cases}$$

   (b) Sample p

$$p \sim \mathsf{Beta}(\alpha_p + N_1, \beta_p + N_2)$$

$$N_1 = \sum_{s=1}^{50}\sum_{t=72}^{93} d_{st}^D + \sum_{s=1}^{50}\sum_{t=72}^{93} d_{st}^R$$

$$N_2 = \sum_{s=1}^{50}\sum_{t=72}^{93} (1 - d_{st}^D) + \sum_{s=1}^{50}\sum_{t=72}^{93} (1 - d_{st}^R)$$

3. Update the Stick-breaking Weight π_k^t and q_k^t:

$$\pi_k^t \sim Beta(1 + n_k^{t-1} + n_k^t, \gamma + \sum_{l=k+1}^{K} n_l^{t-1} + \sum_{l=k+1}^{K} n_l^t)$$

$$n_k^{t-1} = \sum_{s=1}^{50} \mathbb{I}(g^D[s, t-1] = k) + \sum_{s=1}^{50} \mathbb{I}(g^R[s, t-1] = k)$$

$$n_k^t = \sum_{i=1}^{N} (1 - d_{st}^D)\mathbb{I}(g^D[st] = k) + \sum_{i=1}^{N} (1 - d_{st}^R)\mathbb{I}(g^R[st] = k)$$

$$q_k^t = \pi_k^t \prod_{l=1}^{k-1}(1 - \pi_l^t)$$

4. Update $g^D[st]$ and $g^R[st]$

Here we introduce the sampling algorithm for $g^D[st]$. The algorithm for $g^R[st]$ follows the same pattern.

Let define $\mathbf{g}^D[t] \equiv (g^D[1t], g^D[2t], ..., g^D[50t])'$, $\mathbf{q}^t \equiv (q_1^t, q_2^t, ..., q_K^t)'$. Let Y_{st}^D represent the collection of $(V_{istj}^D, P_{istj}^D, S_{ist}^D, B_{ist}^D)'$ for all $i = 1, 2, ..., I_{st}^D$ and $j = 1, 2, ..., J_t^V/J_t^P$; $\mathbf{Y}_t^D \equiv (Y_{1t}^D, Y_{2t}^D, ..., Y_{50t}^D)'$. Finally, let define $\Theta_k \equiv (\theta_k, \eta_k, \omega_k, \lambda_k)'$.

We sample $\mathbf{g}^D[92]$,...,$\mathbf{g}^D[t]$, ..., $\mathbf{g}^D[73]$in turn.

$$Pr(g^D[st] = k|\mathbf{g}^D[92], ..., \mathbf{g}^D[t+1], p, \mathbf{q}^{92}, ..., \mathbf{q}^{73}, \mathbf{Y}_{92}^D, ..., \mathbf{Y}_{73}^D, \Theta_1, ..., \Theta_K)$$

$$\propto \underbrace{Pr(g^D[s, t+1]|p, g^D[st] = k, \mathbf{q}^{t+1})}_{\text{part 1}} \underbrace{Pr(g^D[st] = k|p, \mathbf{q}^t, ..., \mathbf{q}^{73}, \mathbf{Y}_t^D, ..., \mathbf{Y}_{73}^D, \Theta_1, ..., \Theta_K)}_{\text{part 2}}$$

part 1:

$$Pr(g^D[s, t+1] = l|p, g^D[st] = k, q_l^{t+1}) = (1-p)q_l^{t+1} + p\mathbb{I}(l = k)$$

part 2:

$$Pr(g^D[st] = k|p, \mathbf{q}^t, ..., \mathbf{q}^{73}, \mathbf{Y}_t^D, ..., \mathbf{Y}_{73}^D, \Theta_1, ..., \Theta_K))$$

$$\propto \underbrace{f(Y_{st}|g^D[st] = k, \mathbf{Y}_{t-1}^D, ..., \mathbf{Y}_1^D, \Theta_1, ..., \Theta_K)}_{\text{part } a} \underbrace{Pr(g^D[st] = k|p, \mathbf{q}^t, ..., \mathbf{q}^{73}, \mathbf{Y}_{t-1}^D, ..., \mathbf{Y}_{73}^D, \Theta_1, ..., \Theta_K)}_{\text{part } b}$$

part a:

$$f(Y_{st}^D|g^D[st] = k, \mathbf{Y}_{t-1}^D, ..., \mathbf{Y}_{73}^D, \Theta_1, ..., \Theta_K) = f(Y_{st}^D|\Theta_{g^D[st]})$$

$$f(Y_{st}^D|\Theta_{g^D[st]}) = f_V f_P f_S f_B$$

$$f_V = \prod_{i=1}^{I_{st}^D} \prod_{j=1}^{J_t^V} (\theta_{g^D[st]}^{V_{istj}^D} (1 - \theta_{g^D[st]})^{(1-V_{istj}^D)})$$

$$f_P = \prod_{i=1}^{I_{st}^D} \prod_{j=1}^{J_t^P} (\eta_{g^D[st]}^{p_{istj}^D} (1 - \eta_{g^D[st]})^{(1-P_{istj}^D)})$$

$$f_S = \prod_{i=1}^{I_{st}^D} (\omega_{g^D[st]}^{S_{ist}^D} (1 - \omega_{g^D[st]})^{(1-S_{ist}^D)})$$

$$f_S = \prod_{i=1}^{I_{st}^D} \frac{\lambda_{g^D[st]}^{B_{ist}^D} e^{-\lambda_{g^D[st]}}}{B_{ist}^D!}$$

part b:

$$Pr(g^D[st] = k|p, \mathbf{q}^t, ..., \mathbf{q}^{73}, \mathbf{Y}_{t-1}^D, ..., \mathbf{Y}_{73}^D, \Theta_1, ..., \Theta_K)$$

$$= \sum_l Pr(g^D[st] = k|p, q_k^t, g^D[s, t-1] = l) Pr(g^D[s, t-1] = l|p, \mathbf{q}^{t-1}, ..., \mathbf{q}^{73}, \mathbf{Y}_{t-1}^D, ..., \mathbf{Y}_{73}^D, \Theta_1, ..., \Theta_K)$$