Homework1
Khanh-Tung Nguyen-Ba (kn2465)
October 15, 2019

1 Problem 1

1.1 The generative model

The generative model we use for the senate dataset is:

Draw proportion $\theta \sim Dir_K(\alpha = 1)$

For each component $k \in [1..K]$:

1. draw $\beta_k \sim Beta(\beta; a = 1, b = 1)$

For each datapoint $i \in [1..n]$,

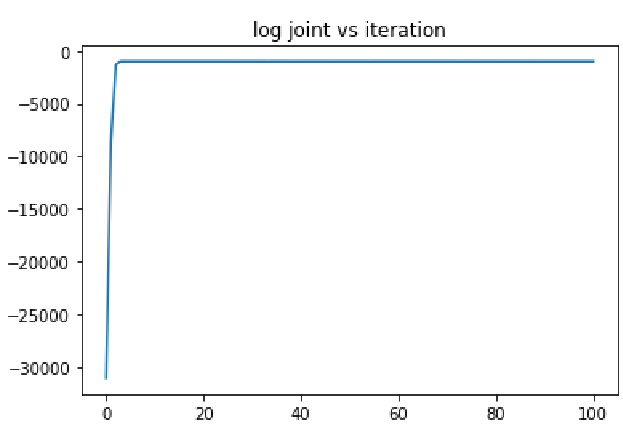
1. draw assignment $z_i|\theta \sim Cat(\theta)$

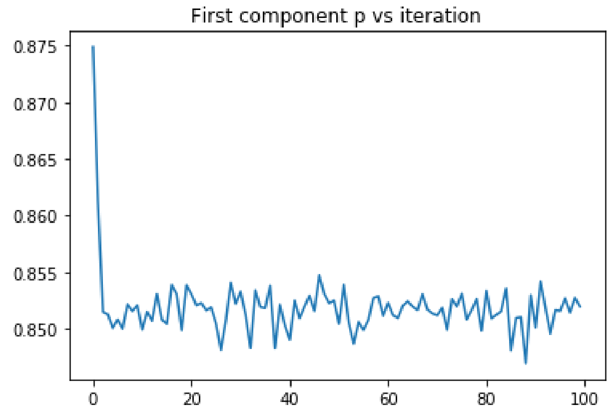
2. draw datapoint $x_i|\beta, z_i \sim Binomial(x; v, \beta_{z_i})$ where v is the number of bills to vote

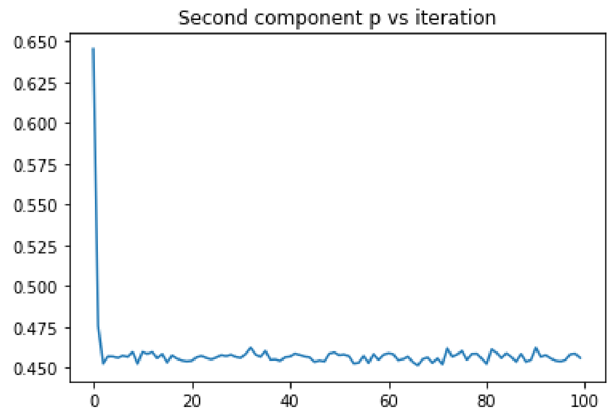
1.2 Code and Discussion

Details of the Gibbs sampling implementation can be found in the accompanying Homework1-Problem1 Jupyter notebook and the gibbs.py module.

The algorithm quickly converges after a few iterations. Below are the charts that show the logjoint and the component p against the iterations.







The model captures the partisanship well using K = 2. Not much information was given about the senators other than party and state, so there is little value to run a higher number of clusters.

z	party	state
0	D	54
	I	2
	R	2
1	D	1
	R	44

A mixed-membership model in which we add a latent variable for attributes of the bills would add more depth to the analysis. The mixture model only reflects the partisanship of the senate, the mixed-membership model could potentially let us know more about the cluster of bills that the senate tends to be more (or less) partisan on.

2 Problem2

One important topic in quantitative finance is volatility modeling where quantitative researchers try to find the right model to explain certain stylized facts of an implied volatility surface's behaviors. Classical dimension reduction models like PCA are often employed to learn the lower dimension representation that can capture the shift, twist, and curvature of a volatility surface over time. Those models often lack in interpretability and flexibility (fail to capture regime switching).

My dataset is 10 years of daily volatility surfaces for SPX (2500 data points), each has 9 tenors (when the option expires) and 9 strikes (where the option kicks in).

If we let $X = (x_1, ..., x_t, ..., x_{2500})$, $x_t \in R^{81}$, we want to find the latent variable $Z = (z_1, ..., z_t, ..., z_{2500})$ where $z_t \in R^m$ that $m << 81$ for the state-space model.

Several questions that I am investigating:

1. X is consist of volatilities that cannot be negative, what is the likelihood to use?

2. How to capture the stylized facts (shift, twist, curvature) by the latent variables?

3. How to capture the regime switch(where is a structural change in the volatility level that breaks away from the prior correlation)?

4. How to adequately captured the temporal correlation structure by the posterior $p(Z|X)$ by an approximate posteriors?