An Analysis of Activity Changes in Muscular Sclerosis Patients

A Case Study in the Use of Bayesian Nonparametrics

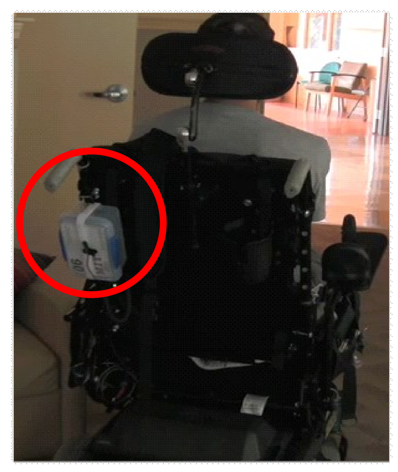
Finale Doshi-Velez and Nicholas Roy

Massachusetts Institute of Technology

Setting and Data

The Boston Home is an eldercare facility in Dorchester, Massachusetts that specializes in the long-term care of residents with muscular sclerosis (MS). All of the residents are in wheelchairs. The facility has dedicated staff to keep residents mobile; however, as the disease progresses residents find it motor skills more and more challenging. At the same time, social programs at the facility are designed to keep the residents engaged and manage cognitive aspects of the disease. Overall, MS progresses in relatively discrete stages: residents will be at one stage for several months before experiencing a sharp change (usually over the period of 2-4 weeks).

The Boston Home Team at MIT mounted wifi-enabled Nokia tablets to 20 resident wheelchairs in December 2009. Since then (with some system outages), the team has monitored which of the facility's 52 access points are visible, and their associated signal strengths, every 30 seconds between 6am and 11pm.

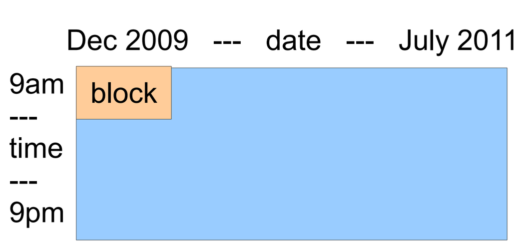


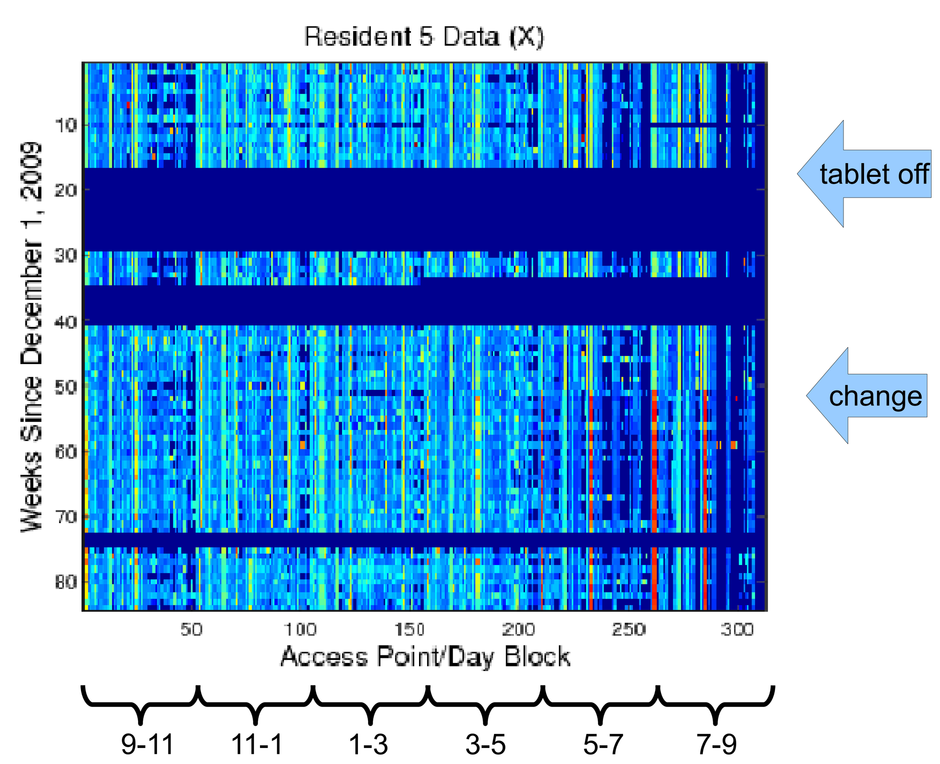
Problem Statement

While the data was originally intended for a project involving resident localization, we are curious to what extent this data can also be used to detect and characterize long-term changes in resident activity. The Boston Home provides an ideal setting for a preliminary study of this form because detected changes can be verified with the staff (with certain privacy constraints). For this study, we chose 5 of the 20 residents.

Data Pre-Processing

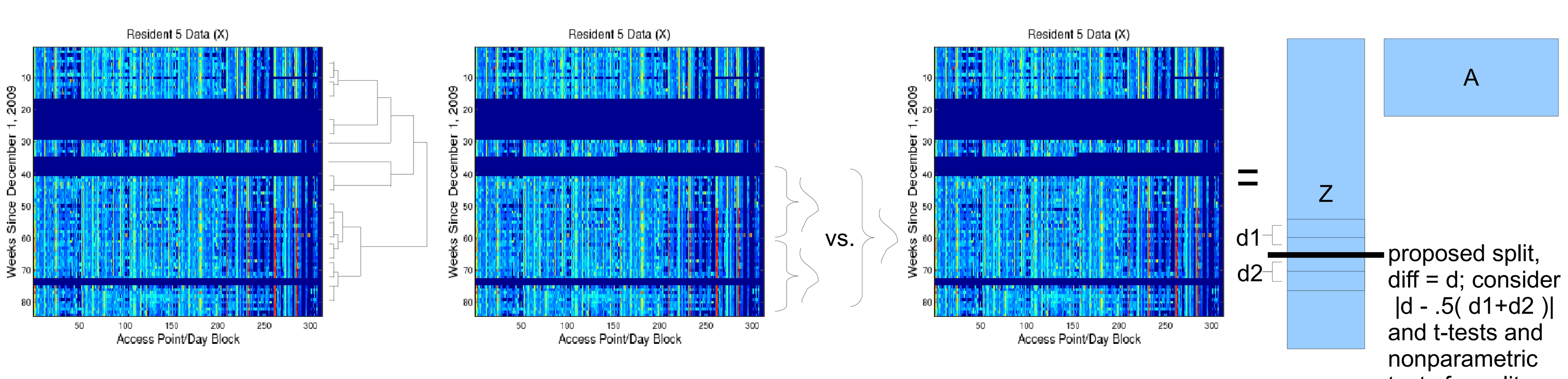
The data is recorded in dB. Knowing that changes take place on the order of weeks, and resident have similar activity schedules for any particular day, we compressed the data by summing the signal strengths for each access point (AP) for each week starting from December 1, 2009 to July 15, 2011 and for each 2-hour block from 9am-11am, 11am-1pm, 1pm-3pm, 3pm-5pm, 5pm-7pm, and 7pm-9pm. These blocks coincided with common blocks in resident activity schedules.





Baselines

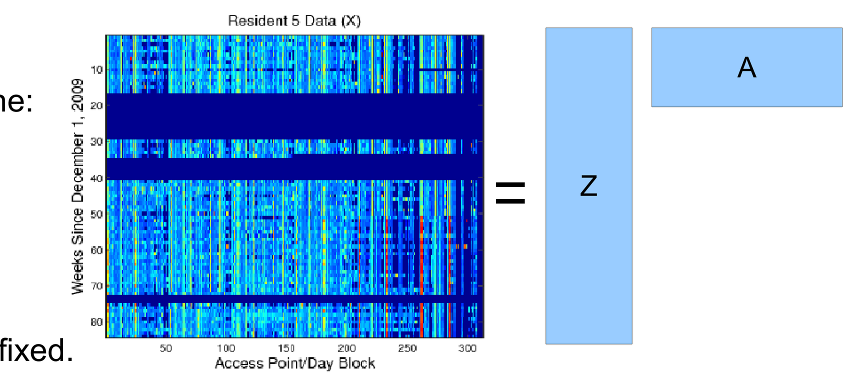
- **Hierarchical clustering** (HC): Greedily combine adjacent clusters with the closest means, look at how often data points end up in the same cluster to define a change probability.
- **Bayesian clustering** (BC): apply Gaussian priors N(0 , S) for the cluster centers, apply collapsed Gibbs to separate clusters with DP rich-get-richer prior.
- **Principle Components Analysis** (PCA): apply PCA to reduce dimensionality by factoring the data X = ZA; look for differences in the reduced matrix Z and interpret activities in the matrix A.
- **Non-negative Matrix Factorization** (NMF): same as PCA, except use multiplicative reweighting to factor X.



Our BNP approach

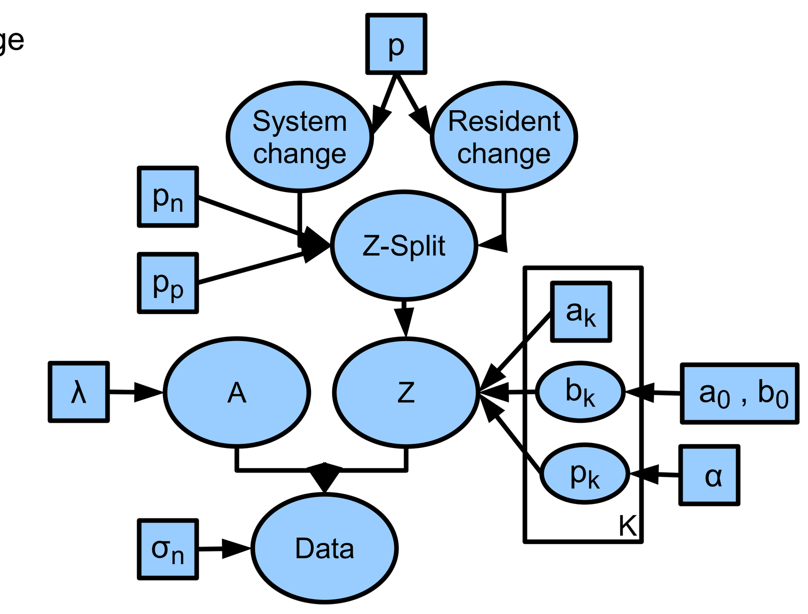
Also a matrix factorization X = ZA with the following properties:

- Elements of the matrix Z are non-negative, and each row sums to one: interpret as proportions of activities.
- Elements of the matrix A are non-negative, represent "activities" that the resident might have been doing.
- The number of activities is learned.



Generative process:

- Elements of A are chosen with an exponential prior: $A_{kd} \sim Exp(\lambda)$, λ fixed.
- System and resident changes s_t and r_t are chosen independently with Bern(p); changes are seen in Z-Split with Bern(p_n) if there is no change and Bern(p_p) if there is a change; p, p_n and p_p are fixed.
- Within each block, $\dot{Z}_{tk} \sim Bern(p_k)Gamma(a_k , b_k)$ and $Z_{tk} = \dot{Z}_{tk} / (sum_j \dot{Z}_{tk})$, where
  - $a_k$is fixed
  - b_k is drawn from a prior Gamma(a_0 , b_0)
  - p_k is drawn from a Beta(α/K , 1) – weak limit IBP approximation.
- $X = ZA + E$, where each element $E_{td} \sim N(0 , \sigma_n)$.



Inference

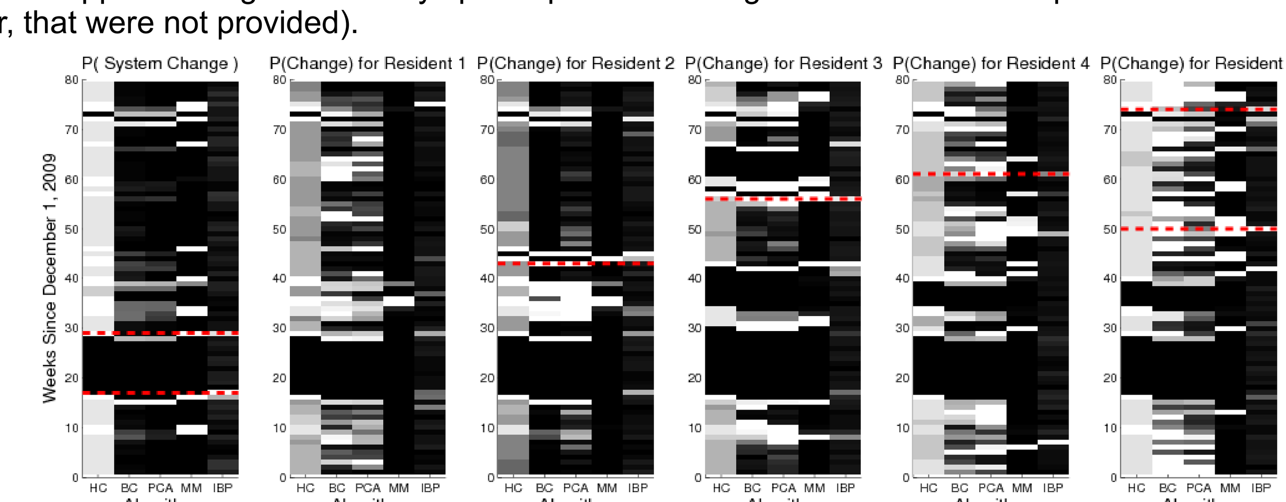
- Blocked sampler for sampling split points, A, and Z.
- Split points, A are relatively straightforward to implement.
- For Z, use an MH kernel that perturbs each $\dot{Z}_{tk}$ independently.
- Marginalize p_k and b_k.

Main Results

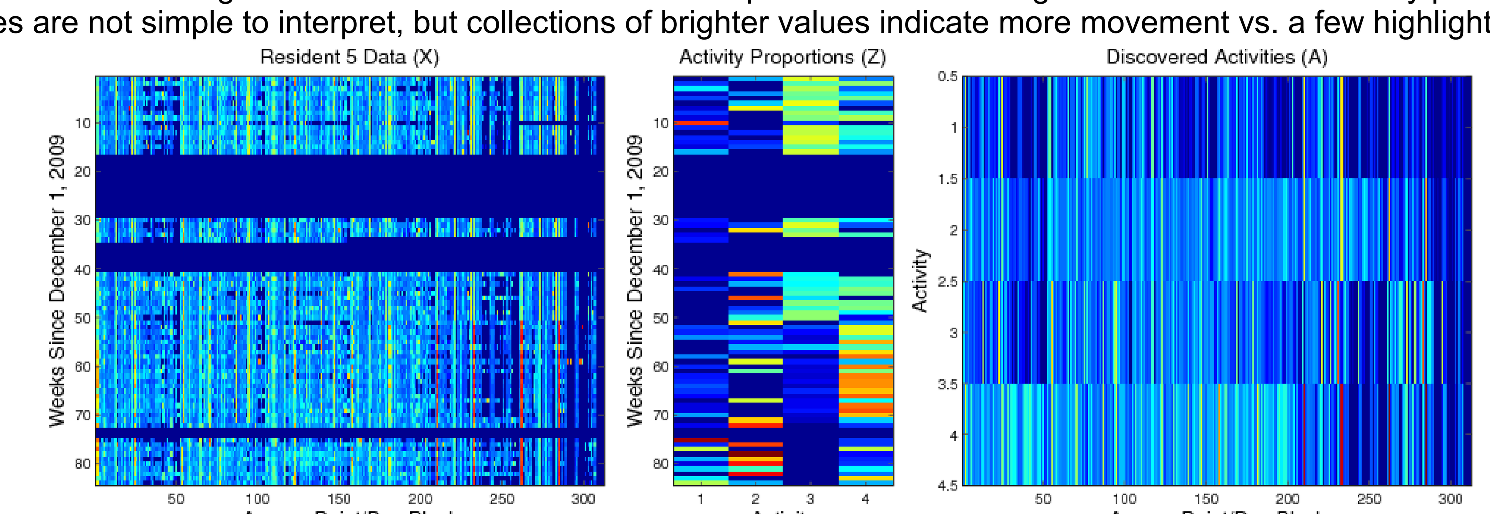
Computational time: Running in Matlab, with fairly optimized code:

- Hierarchical clustering took 9.3 seconds, and Bayesian clustering took 262 seconds (for 1000 MCMC iterations)
- The PCA-based factorization took 4.1 seconds, and the NMF took 5.6 seconds.
- Our BNP approach took 4900 seconds for 100 MCMC.

Comparision of Changepoints: For reasonably well-chosen parameters; lighter means that a change is more likely. Both the NMF and the BNP approaches give relatively sparse possible changes. Red lines correspond to known changes (there may be more, however, that were not provided).



Example case: the changes and the activities detected correspond to known changes in the resident's activity patterns. (The AP signatures are not simple to interpret, but collections of brighter values indicate more movement vs. a few highlights).



Questions

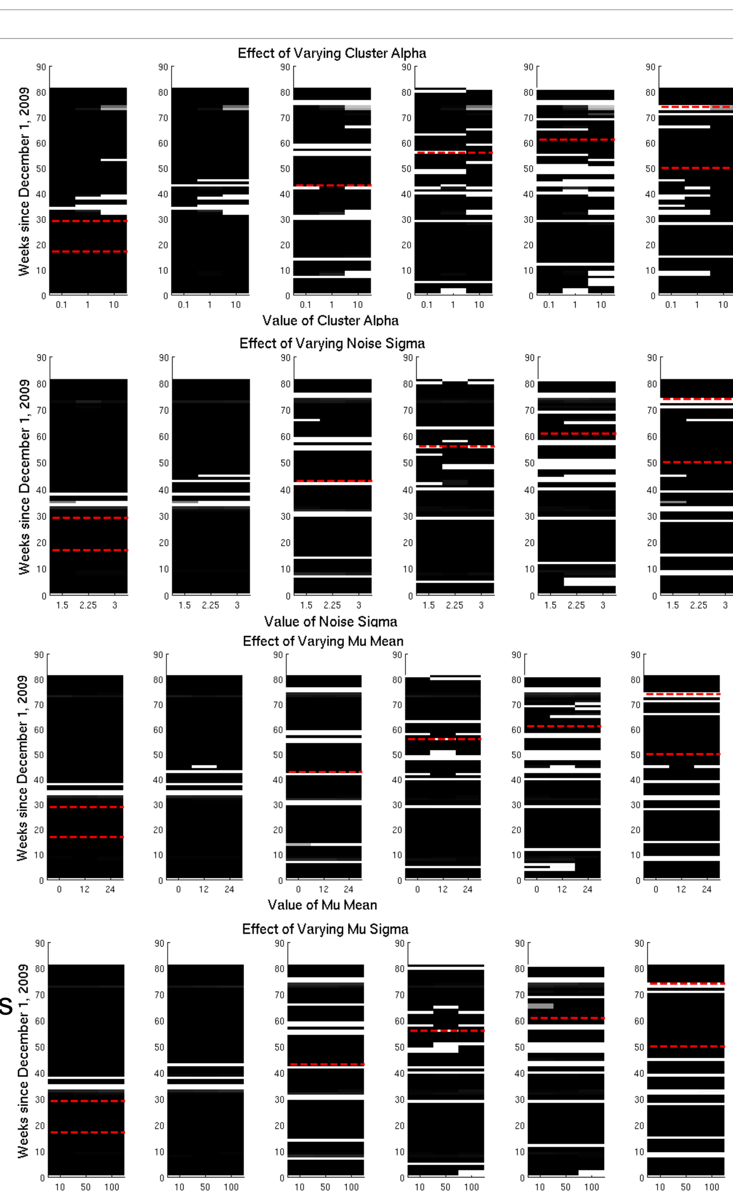
- What are appropriate models for this kind of data? (There are obvious extensions to more sophisticated change point detection problems, but is that what we really need? Likewise, these mixtures bear a close resemblance to topic models.)
- Is adjusting the priors the best way to address questions of interpretablility?
- What is the best way to validate these kinds of results?

Sensitivity Investigation: Bayesian Clustering

We tested a few parameter settings for alpha (for the DP clustering), noise, and the prior placed on the mean.

To the right are the differences in split points as the parameters are varied. Most of the lines go straight across, indicating relatively little sensitivity.

The most sensitive parameter is the alpha; in general larger alpha (as expected) encourages more change points. More splits in the resident changes (first five) results in fewer splits being predicted for the system changes.



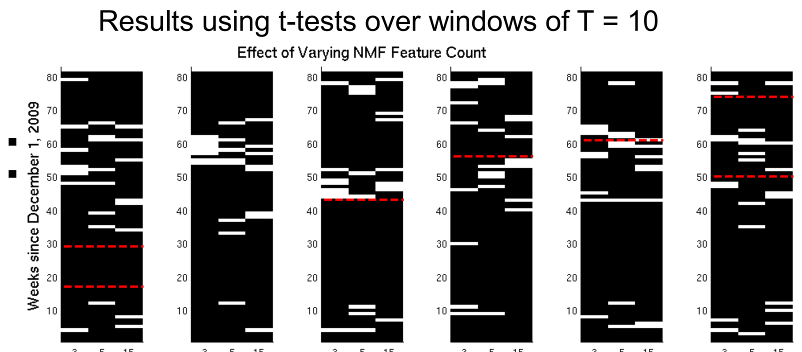
Sensitivity Investigation: NMF/Split Heuristics

We tested three different ways of defining a split for three different settings of K (the number of activities in the NMF).

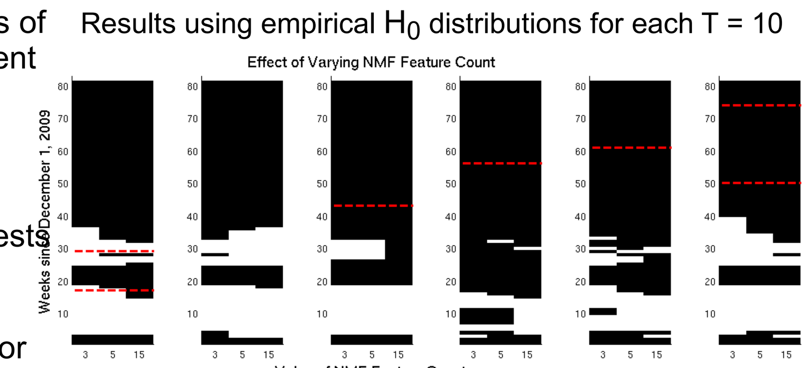
There seems to be some sensitivity to K when using t-tests and empirical H_0 distributions (both with a BH correction for multiple comparisons), none for the difference of difference heuristic from the main results.

The t-test and the empirical H_0 approaches also seem not to be finding particularly relevant split points.

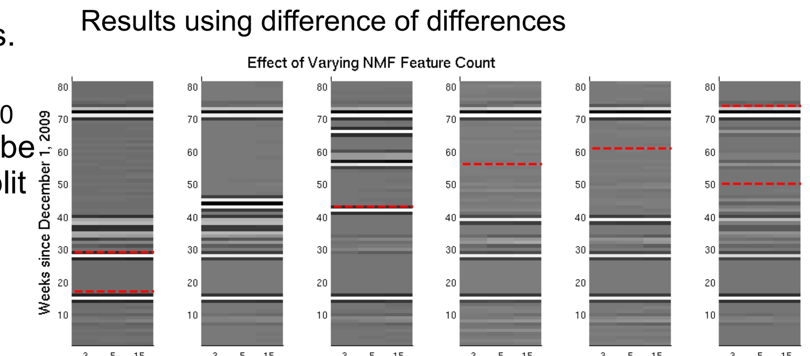
Results using t-tests over windows of T = 10



Results using empirical H_0 distributions for each T = 10



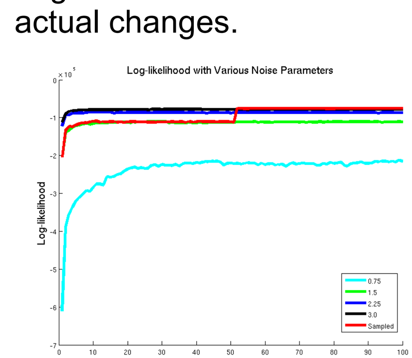
Results using difference of differences

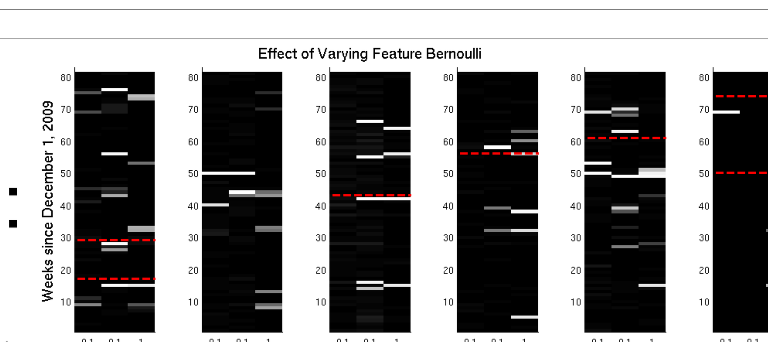


Sensitivity Investigation: Our BNP

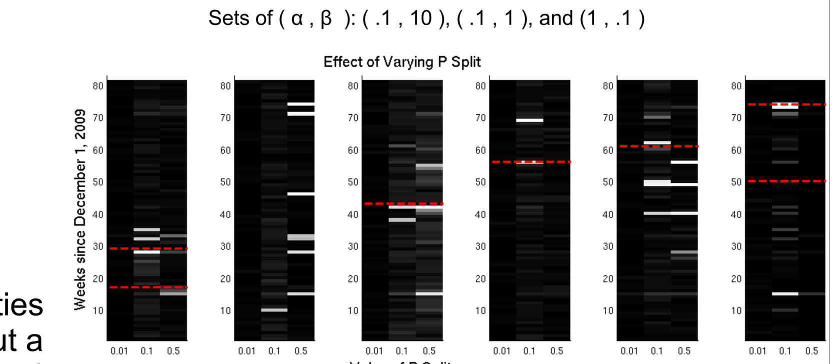
We varied the noise parameter (as well as tried learning it), combinations of the change probabilities, and the parameters for the beta distribution on p_k.

As expected, large beta parameters induces fewer changes, as do lower penalties when Z-split changes without a resident or system change and larger (and learned) noise... but larger noise finds few of the actual changes.





Sets of (α , β): (.1 , 10), (.1 , 1), and (1 , .1)



Sets of (p, p_n ,p_p): (0.01 , 0.01 , 0.01), (0.1, 0.001, 0.1), (0.5, 0.00001, 0.5)