An Analysis of Activity Changes in Muscular Sclerosis Patients
A Case Study in the Use of Bayesian Nonparametrics
Finale Doshi-Velez and Nicholas Roy

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 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).

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 straight across, indicating system changes.
- **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: \( \text{A} \sim \text{Exp}(\lambda) \), \( \lambda \) fixed.
- System and resident changes are independent with Bernoulli priors; changes are seen in Z-Split with Bernoulli priors. If there is no change and Bernoulli priors are fixed, within each block, \( Z_{k} \sim \text{Bernoulli}(\rho_{k}, 0.5) \) and \( Z_{k} = Z_{k} + \text{sum}(Z_{k}) \), where
- \( \lambda \) fixed
- \( \rho_{k} \) drawn from a prior Gamma(0.1, 1.0)
- \( \text{t-test} \) weak informative prior approximation.
- \( \text{t-test} \) weak informative prior approximation.

Inference
- **Blocked sampler for sampling split points, A, and Z.**
- **For Z, use an MH kernel that perturbs each block independently.**
- **Marginalize \( \rho_{k} \) 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 4000 seconds for 100 MCMC runs.**

Comparison 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, if we have other data).

Example case: the changes and the activities detected correspond to known changes in the resident's activity patterns. (The AP changes and the activities detected correspond to known changes in the resident's activity patterns. (The AP changes and the activities detected correspond to known changes in the resident's activity patterns. (The AP changes and the activities detected correspond to known changes in the resident's activity patterns. (The AP changes and the activities detected correspond to known changes in the resident's activity patterns. (The AP changes and the activities detected correspon..."

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 interpretability?**
- **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 means.

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 H0 distributions (both with a BH correction for multiple comparisons), none for the difference of difference heuristics from the main results.

The t-test and the empirical H0 approaches also seem not to be finding particularly relevant split points.

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 \( \rho_{k} \).

As expected, large beta parameters induce 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 five of the actual changes.