An Introduction to the Dirichlet Process and Nonparametric Bayesian Models

Frank Wood

Columbia University

26 Apr 2010

Spike Sorting

Spike Sorting Schematic

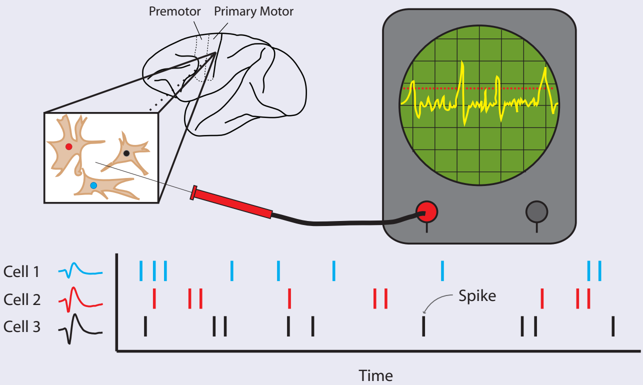


Figure: Illustration of spike train acquisition

Spike Sorting

Steps

1. Eliminate noise
2. Detect action potentials
3. Deconvolve overlapping action potentials
4. Identify the number of neurons in the recording
5. Attribute spikes to neurons
6. Track changes in action potential waveshape
7. Detect appearance and disappearance of neurons

Spike Sorting

Steps

1. Eliminate noise
2. Detect action potentials
3. Deconvolve overlapping action potentials
4. Identify the number of neurons in the recording
5. Attribute spikes to neurons
6. Track changes in action potential waveshape
7. Detect appearance and disappearance of neurons

Spike Sorting

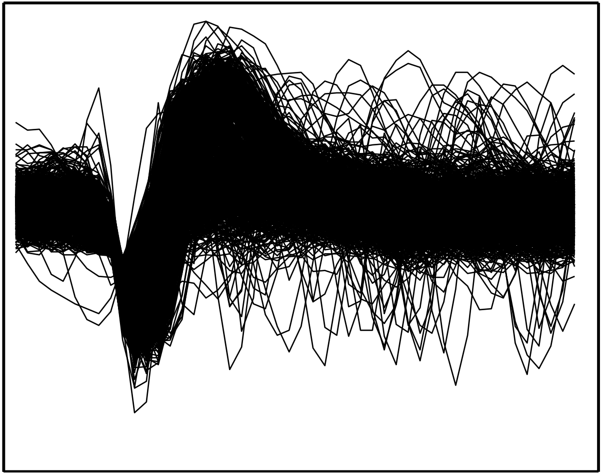


Figure: Single channel, all detected action potentials.

Spike Sorting

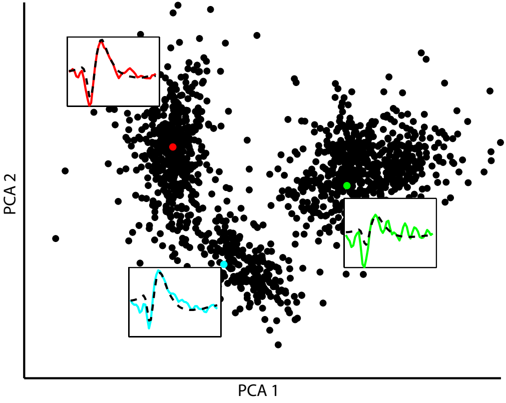


Figure: Projection of waveforms onto first 2 PCA basis vectors.

Spike Sorting

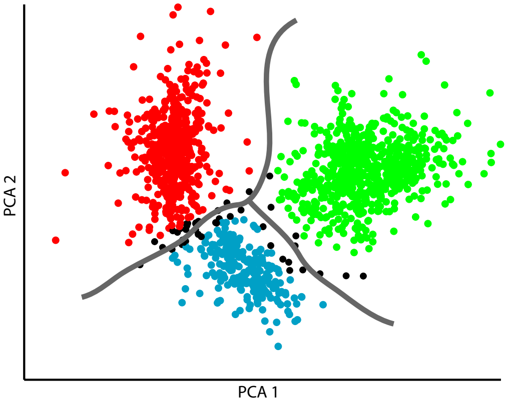
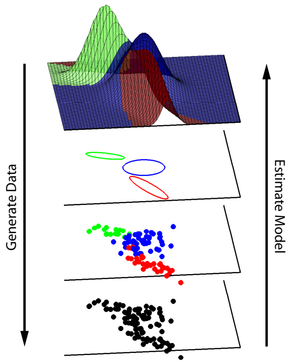


Figure: Spike train variability arising from clustering ambiguity.

Gaussian Mixture Model (GMM) Spike Sorting [Lewicki, 1994]



GMM

$$\begin{aligned} \theta_k &= \{\vec{\mu}_k, \Sigma_k\} \\ c_i \mid \vec{\pi} &\sim \text{Discrete}(\pi_1, \ldots, \pi_K) \\ \vec{y}_i \mid c_i = k, \Theta &\sim \text{Gaussian}(\theta_k) \end{aligned}$$

Finite Gaussian mixture model estimation

Estimation

- Expectation Maximization (EM)
- Variational inference
- Markov chain Monte Carlo (MCMC)

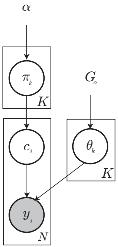
A challenge to pick the “best” model

- Complexity $\leftrightarrows$ Model selection $\leftrightarrows$ Neuron cardinality
- Clustering $\leftrightarrows$ Attributing spikes to neurons

Approaches

- Reversible jump MCMC
- Penalized likelihood (Bayesian information criteria)
- Cross validation on held out data
- or...

Bayesian GMM $\rightarrow$ IGMM as $K \rightarrow \infty$ [Rasmussen, 2000]



$$\Sigma_k \sim \text{Inverse-Wishart}_{\upsilon_0}(\Lambda_0^{-1})$$
$$\vec{\mu}_k \sim \text{Gaussian}(\vec{\mu}_0, \Sigma_k/\kappa_0)$$

$$\theta_k = \{\vec{\mu}_k, \Sigma_k\}$$

$$\begin{aligned}
\pi_1, \ldots, \pi_K | \alpha &\sim \text{Dirichlet}(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K}) \\
c_i | \vec{\pi} &\sim \text{Discrete}(\pi_1, \ldots, \pi_K) \\
\vec{y}_i | c_i = k, \Theta &\sim \text{Gaussian}(\theta_k) \\
\Theta &\sim \mathcal{G}_0
\end{aligned}$$

Key insight

IGMM posterior distribution consists of infinite mixture models that vary in realized complexity

Key insight

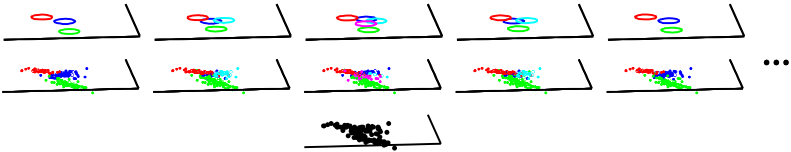


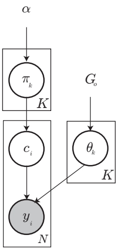
Figure: Multiple generative models for same data.

IGMM Posterior Estimation

Many approaches from DP mixture model literature

- Batch:
  - Markov chain Monte Carlo (MCMC)
    - Gibbs [Neal, 2000, MacEachern and Muller, 1998]
    - Split merge [Jain and Neal, 2004]
  - Variational [Blei and Jordan, 2005]
- Sequential: Particle filter [MacEachern et al., 1999, Fearnhead, 2004]

IGMM $\approx$ Bayesian GMM in $K \to \infty$ limit



$$\begin{aligned}
\Sigma_k &\sim \text{Inverse-Wishart}_{\upsilon_0}(\Lambda_0^{-1}) \\
\vec{\mu}_k &\sim \text{Gaussian}(\vec{\mu}_0, \Sigma_k/\kappa_0)
\end{aligned}$$

$$\theta_k = \{\vec{\mu}_k, \Sigma_k\}$$

$$\begin{aligned}
\pi_1, \dots, \pi_K | \alpha &\sim \text{Dirichlet}(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}) \\
c_i | \vec{\pi} &\sim \text{Discrete}(\pi_1, \dots, \pi_K) \\
\vec{y}_i | c_i = k, \Theta &\sim \text{Gaussian}(\theta_k) \\
\Theta &\sim \mathcal{G}_0
\end{aligned}$$

Dirichlet Multinomial Conjugacy

$$
\begin{aligned}
\vec{\pi} &\sim \text{Dir}\left(\frac{\alpha}{K}, \ldots, \frac{\alpha}{K}\right) \\
P(c_{i+1} = k | c_1, \ldots, c_i, \alpha) &= \int P(c_{i+1} = k | \vec{\pi}) p(\vec{\pi} | c_1, \ldots, c_i, \alpha) d\vec{\pi} \\
= \frac{\Gamma(\alpha + i)}{\prod_{j=1}^{K} \Gamma(\frac{\alpha}{K} + n_j)} & \int \pi_1^{\frac{\alpha}{K} + n_1 - 1} \ldots \pi_k^{\frac{\alpha}{K} + n_k} \ldots \pi_K^{\frac{\alpha}{K} + n_K - 1} d\vec{\pi} \\
&= \frac{n_k + \frac{\alpha}{K}}{\alpha + i}
\end{aligned}
$$

Where n_k is the number of c_j, $j = 1, \ldots, i$ such that $c_j = k$.

Note that this is true regardless of the "order" of the observations (c_i's form an exchangeable sequence).

Chinese Restaurant Process

Order the clusters so $n_k > 0$ if $k \leq K_+$ and $n_k = 0$ if $k > K_+$. Then as $K \to \infty$

$$P(c_{i+1} = k | c_1, \ldots, c_i, \alpha) = \begin{cases} \frac{n_k}{\alpha+i} & k \leq K_+ \\ \frac{\alpha}{\alpha+i} & k > K_+ \end{cases} .$$

This is the *Chinese Restaurant Process*, $CRP(\alpha)$.

Fully Conjugate Collapsed IGMM Gibbs Sampler

One can make a "collapsed" Gibbs sampler for the IGMM by analytically integrating out the π's and θ's leaving on the c's to sample. The Gibbs sampler will be constructed by finding the posterior distribution of a single c conditioned on the state of all other variables, then sampling each of these variables repeatedly.

Gibbs sampler for the IGMM

$$
\begin{aligned}
P(c_i = j | \mathcal{C}_{-i}, \mathcal{Y}, \alpha; \mathcal{H}) &\propto P(\mathcal{Y}|\mathcal{C}; \mathcal{H}) P(\mathcal{C}|\alpha) \\
&\propto P(y_i | \mathcal{Y}^{(j)} \backslash y_i; \mathcal{H}) P(c_i = j | \mathcal{C}_{-i}, \alpha)
\end{aligned}
$$

Fully Conjugate Collapsed IGMM Gibbs Sampler

Per class posterior predictive distribution

$$y_i|\mathcal{Y}^{(j)}\backslash y_i;\mathcal{H} \sim t_{\nu_n-D+1}(\vec{\mu}_n, \mathbf{\Lambda}_n(\kappa_n+1)/(\kappa_n(\nu_n-D+1))) \quad (1)$$

where

$$\begin{aligned}
\vec{\mu}_n &= \frac{\kappa_0}{\kappa_0+N}\vec{\mu}_0 + \frac{N}{\kappa_0+N}\bar{y} \\
\kappa_n &= \kappa_0 + N \\
\nu_n &= \nu_0 + N \\
\mathbf{\Lambda}_n &= \mathbf{\Lambda}_0 + \mathbf{S} + \frac{\kappa_0 n}{\kappa_0+N}(\bar{y}-\vec{\mu}_0)(\bar{y}-\vec{\mu}_0)^T
\end{aligned}$$

Microelectrode Array Data la040325MI

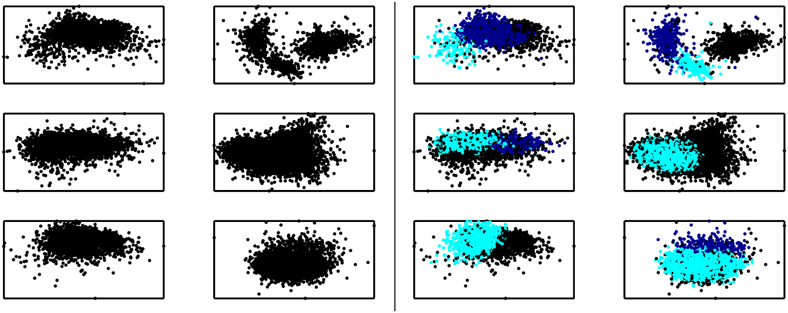


Figure: Unsorted & Human Sorted

Microelectrode Array Data la040325MI

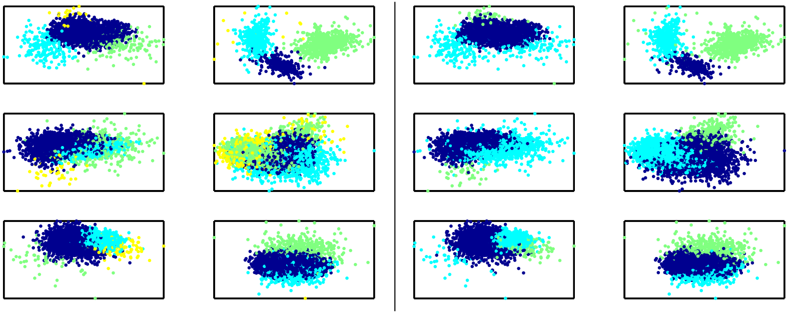
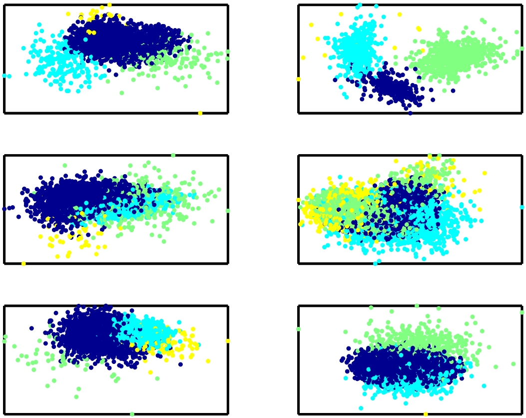
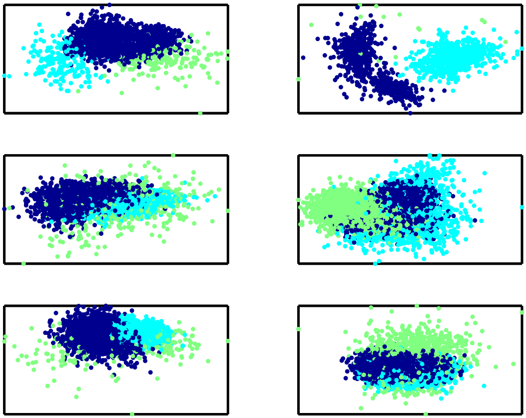


Figure: IGMM

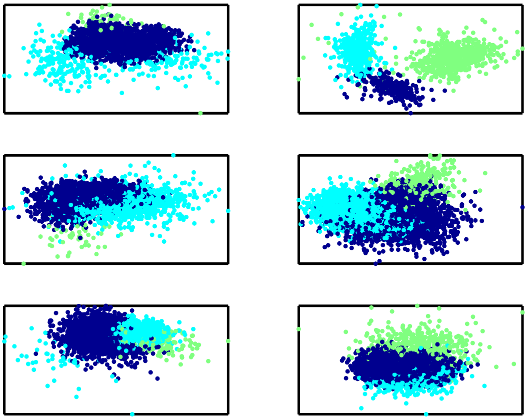
Microelectrode Array Data la040325MI



Microelectrode Array Data la040325MI



Microelectrode Array Data la040325MI



Microelectrode Array Data la040325MI

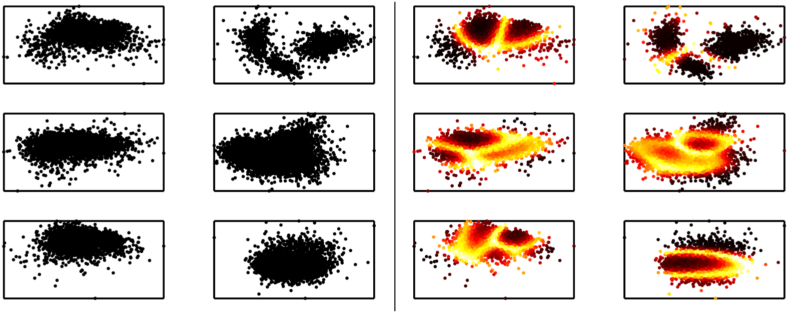


Figure: Unsorted & Ambiguity

Bibliography I

D. Blei and M. Jordan. Variational inference for Dirichlet process mixtures. *Journal of Bayesian Analysis*, 1(1):121–144, 2005.

P. Fearnhead. Particle filters for mixture models with an unknown number of components. *Journal of Statistics and Computing*, 14:11–21, 2004.

S. Jain and R. M. Neal. A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *Journal of Computational and Graphical Statistics*, 13(1):158–182, March 2004.

M. S. Lewicki. Bayesian modeling and classification of neural signals. *Neural Computation*, 6:1005–1030, 1994.

S. MacEachern and P. Muller. Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics*, 7:223–238, 1998.

Bibliography II

S. N. MacEachern, M. Clyde, and J. Liu. Sequential importance sampling for nonparametric Bayes models: the next generation. *The Canadian Journal of Statistics*, 27:251–267, 1999.

R. M. Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265, 2000.

C. Rasmussen. The infinite Gaussian mixture model. In *Advances in Neural Information Processing Systems 12*, pages 554–560. MIT Press, Cambridge, MA, 2000.