

# Latent Dirichlet Allocation: Mixed Membership Modeling



Emily Fox & Carlos Guestrin

Machine Learning Specialization

University of Washington

# Mixed membership models for documents



# So far, clustered articles into groups



Doc labeled  
with a topic  
assignment



**Clustering goal:** discover groups of related docs

# Are documents about just one thing?



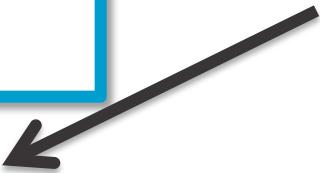
Is this article  
just about  
**science?**



# Soft assignments capture uncertainty



Soft assignment  $r_{ik}$  tells us this doc could be about **world news or science**



**But,** clustering model still specifies each doc belongs to **1 topic**



# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

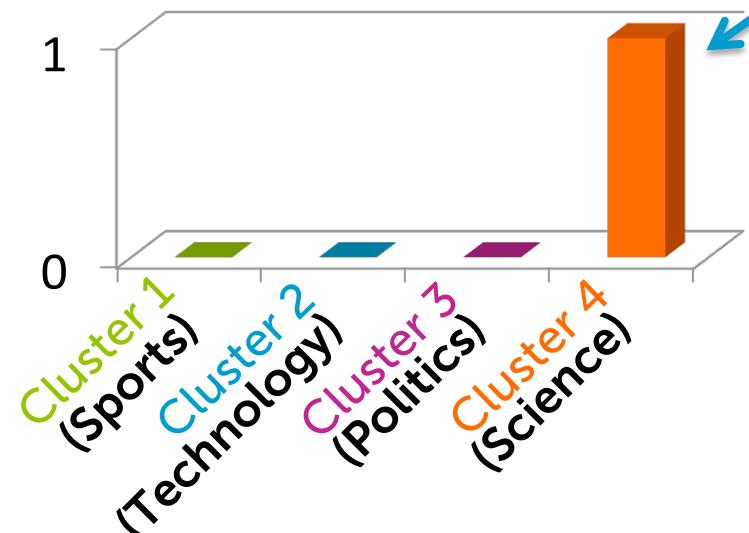
**Keywords:** Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

Encoding of cluster membership  $z_i = 4$

Based on science related words, maybe doc in cluster 4



# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full clinical seizures. We propose a Bayesian nonparametric graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

Soft assignments capture uncertainty in  $z_i = 2$  or 4

graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

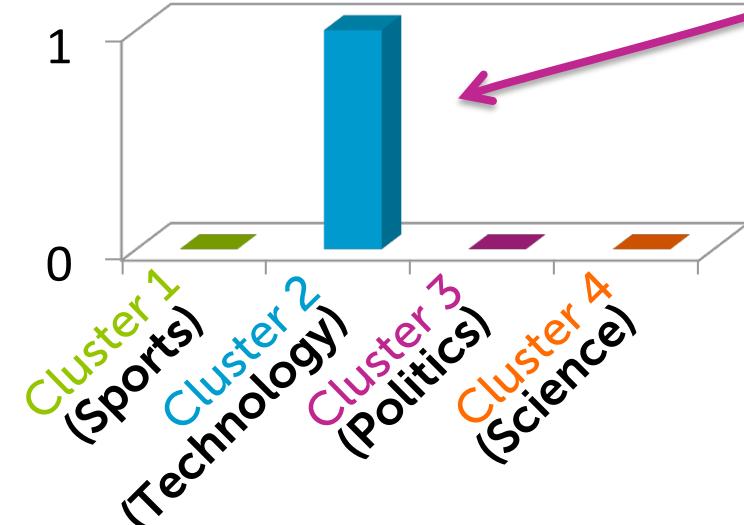
**Keywords:** Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

Encoding of cluster membership  $z_i = 2$

Or maybe cluster 2 (technology) is a better fit



# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A key challenge is how to model the complex interactions between patients into (iEEG) data. We show that our model can switch between different hidden variable states, which can be used to predict unknown future states. This allows us to set out a graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

“ $z_i$ ” is both 2 and 4

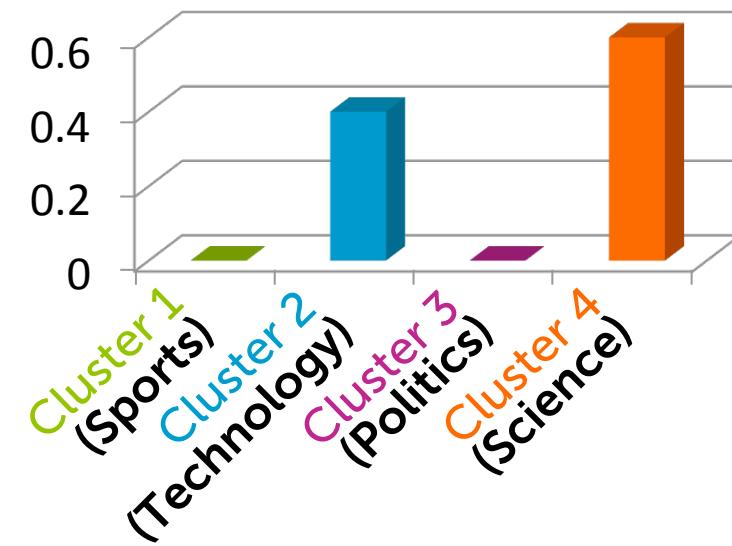


**Keywords:** Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

Really, it's about science  
and technology



# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

*Keywords:* Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

# Mixed membership models

Want to discover a set of memberships

(In contrast, cluster models aim at discovering a single membership)

# Building up to document mixed membership models



# An alternative document clustering model



(Back to  
clustering,  
not mixed  
membership  
modeling)



# So far, we have considered...

Modeling the Complex Dynamics and Changing  
Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces informative state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

---

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



# Bag-of-words representation

Modeling the Complex Dynamics and Changing  
Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

## Bayesian

*Keywords:* Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

---

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



# Bag-of-words representation

Modeling the Complex Dynamics and Changing  
Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

*Keywords:* Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

---

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

$\mathbf{x}_i = \{ \text{modeling, complex, epilepsy, modeling, Bayesian, clinical, epilepsy, EEG, data, dynamic...} \}$

multiset

= unordered set of words with  
duplication of unique elements  
mattering



# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

*Keywords:* Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

# A model for bag-of-words representation

As before, the “prior”  probability that doc i is from topic k is:

$$p(z_i = k) = \pi_k$$

$\boldsymbol{\pi} = [\pi_1 \ \pi_2 \dots \pi_K]$   
represents corpus-wide  
topic prevalence

# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

# A model for bag-of-words representation

Assuming doc  $i$  is from topic  $k$ , words occur with probabilities:

SCIENCE	
patients	0.05
clinical	0.01
epilepsy	0.002
seizures	0.0015
EEG	0.001
...	...

words in vocab

# Topic-specific word probabilities

Distribution on words in vocab for **each topic**

SCIENCE	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TECH	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

SPORTS	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

...



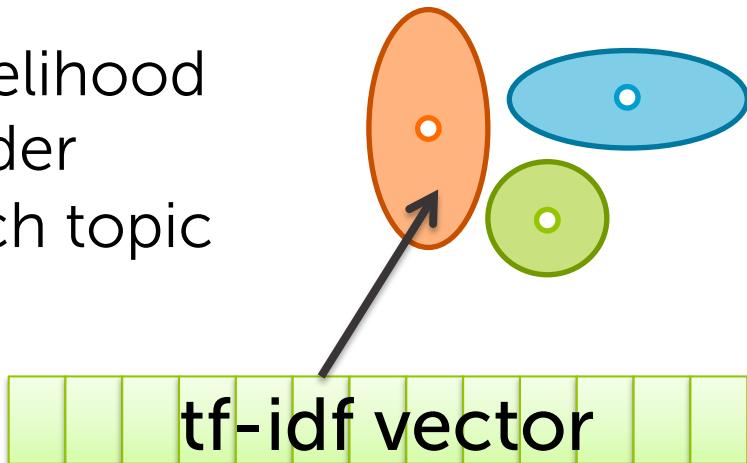
(table now organized by decreasing probabilities  
showing top words in each category)

# Comparing and contrasting

Prior topic probabilities

$$p(z_i = k) = \pi_k$$

Likelihood under each topic



compute likelihood of **tf-idf** vector under each **Gaussian**

## Now

$$p(z_i = k) = \pi_k$$

SCIENCE	TECH	SPORTS
experiment 0.1	develop 0.18	player 0.15
test 0.08	computer 0.09	score 0.07
discover 0.05	processor 0.032	team 0.06
hypothesize 0.03	user 0.027	goal 0.03
climate 0.01	internet 0.02	injury 0.01
...	...	...

...

{modeling, complex, epilepsy,  
modeling, Bayesian, clinical,  
epilepsy, EEG, data, dynamic...}

compute likelihood of the  
**collection of words** in doc  
under each **topic distribution**

# Latent Dirichlet allocation (LDA)

# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

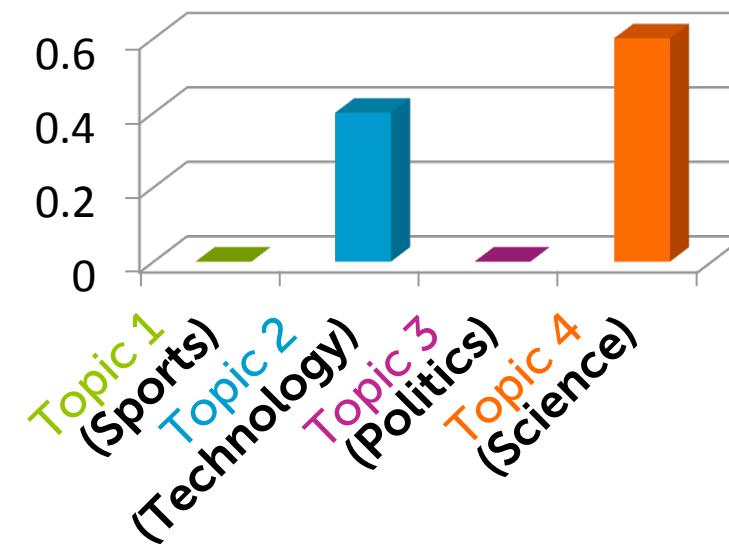
**Keywords:** Bayesian, nonparametric, EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

# LDA is a mixed membership model

Want to discover a  set of topics



## Topic vocab distributions:

SCIENCE	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TECH	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

SPORTS	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (IEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of IEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

*Keywords:* Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



# Clustering:

One topic indicator  
 $z_i$  per document



All words come from  
(get scored under)  
same topic  $z_i$



Distribution on  
prevalence of  
topics in corpus

$$\pi = [\pi_1 \ \pi_2 \ \dots \ \pi_K]$$



## Same topic distributions:

SCIENCE	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TECH	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

SPORTS	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian, nonparametric, EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

## In LDA:

One topic indicator  
 $z_{iw}$  per word in doc i

Each word gets scored under its topic  $z_{iw}$

Distribution on prevalence of topics in document

$$\pi_i = [\pi_{i1} \ \pi_{i2} \ \dots \ \pi_{iK}]$$

## Topic vocab distributions:

SCIENCE	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TECH	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

SPORTS	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

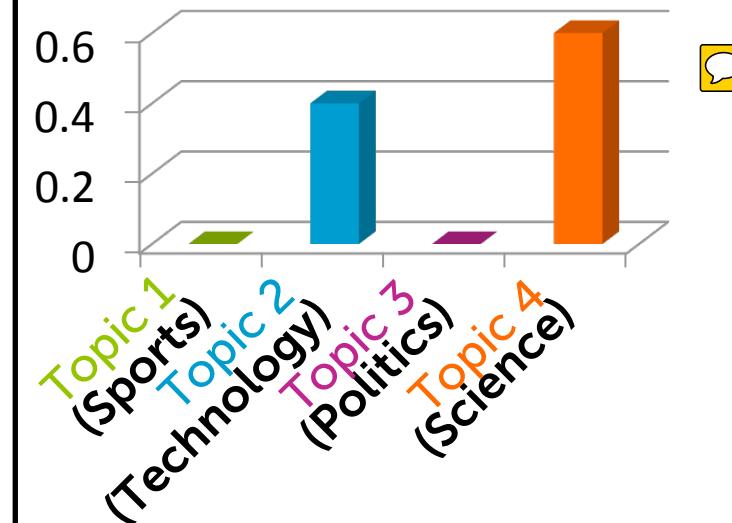
*Keywords:* Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

## Document topic proportions:

$$\pi_i = [\pi_{i1} \ \pi_{i2} \dots \ \pi_{iK}]$$



# Inference in LDA models

## Topic vocab distributions:

SCIENCE	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TECH	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

SPORTS	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

# Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

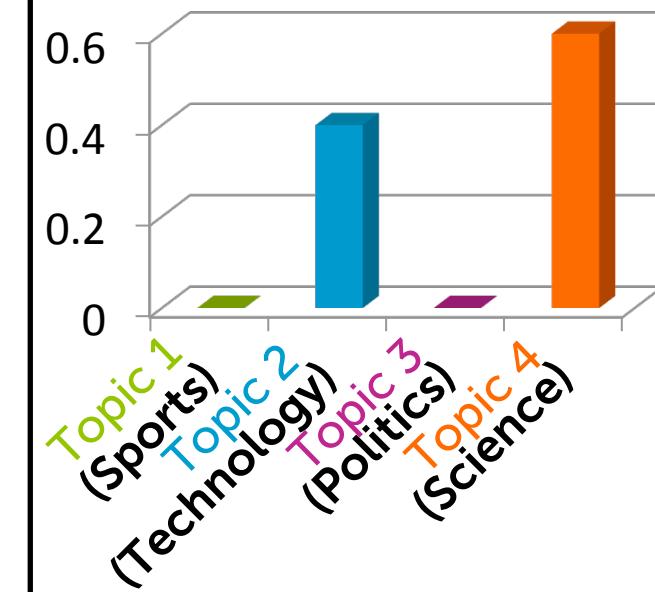
*Keywords:* Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

## Document topic proportions:

$$\pi_i = [\pi_{i1} \ \pi_{i2} \dots \ \pi_{iK}]$$



# Topic vocab distributions:

TOPIC 1	
Word 1	?
Word 2	?
Word 3	?
Word 4	?
Word 5	?
...	...

TOPIC 2	
Word 1	?
Word 2	?
Word 3	?
Word 4	?
Word 5	?
...	...

TOPIC 3	
Word 1	?
Word 2	?
Word 3	?
Word 4	?
Word 5	?
...	...

## Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

### Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

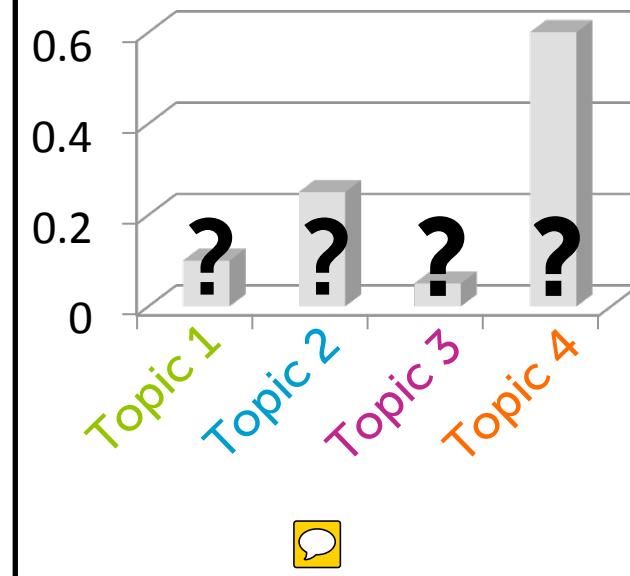
**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

### 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

## Document topic proportions:

$$\pi_i = [\pi_{i1} \ \pi_{i2} \dots \ \pi_{iK}]$$



## Topic vocab distributions:

TOPIC 1	
Word 1	?
Word 2	?
Word 3	?
Word 4	?
Word 5	?
...	...

TOPIC 2	
Word 1	?
Word 2	?
Word 3	?
Word 4	?
Word 5	?
...	...

TOPIC 3	
Word 1	?
Word 2	?
Word 3	?
Word 4	?
Word 5	?
...	...

## Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

### LDA inputs:

- Set of words per doc for each doc in corpus

### LDA outputs:

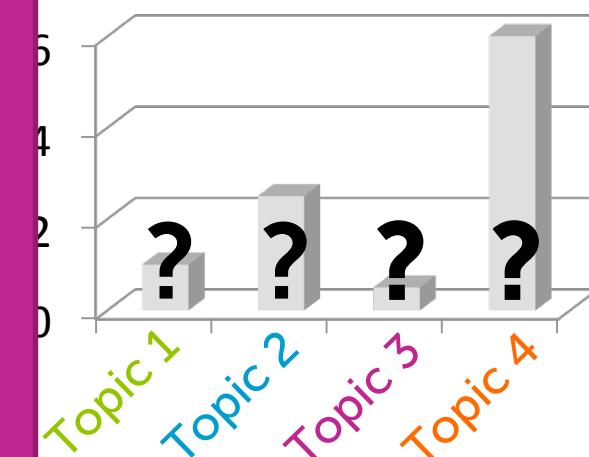
- Corpus-wide topic vocab distributions
- Topic assignments per word
- Topic proportions per doc

#### 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

## Document topic proportions:

$$\pi_i = [\pi_{i1} \ \pi_{i2} \dots \ \pi_{iK}]$$



# Interpreting LDA outputs



## TOPIC 1

experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

## TOPIC 2

develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

## TOPIC 3

player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

## Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

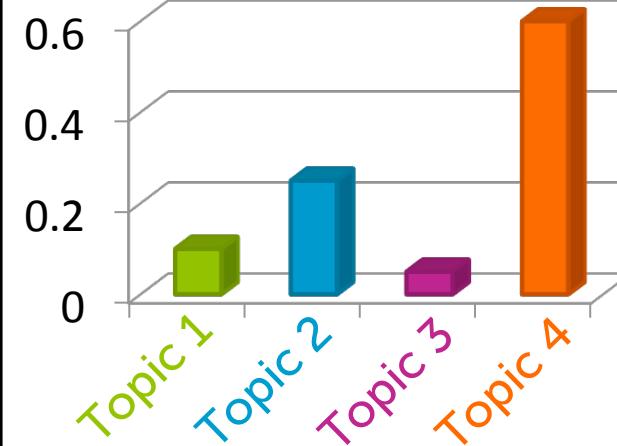
### Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

### 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



# Interpreting LDA outputs

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

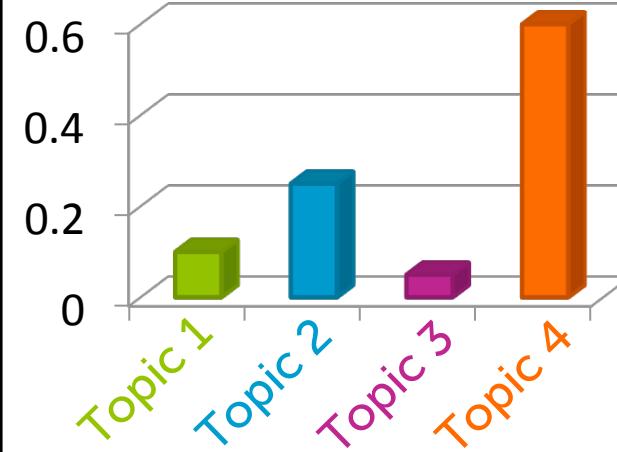
**Abstract**

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamics across seizures between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the change dynamics and demonstrate the importance of this model in learning and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



Examine **coherence** of learned topics

- What are top words per topic?
- Do they form meaningful groups?
- Use to post-facto label topics (e.g., science, tech, sports,...)

# Interpreting LDA outputs

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

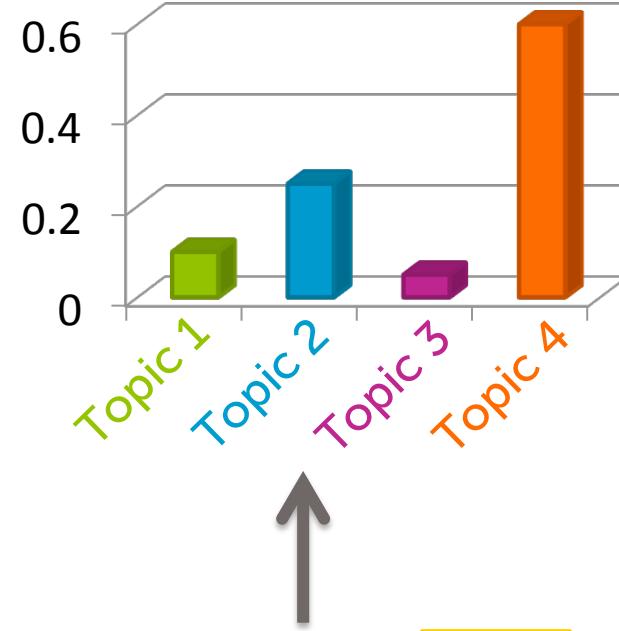
Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



Doc-specific topic proportions can be used to:

- Relate documents
- Study user topic preferences
- Assign docs to multiple categories

# Interpreting LDA outputs

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

## Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

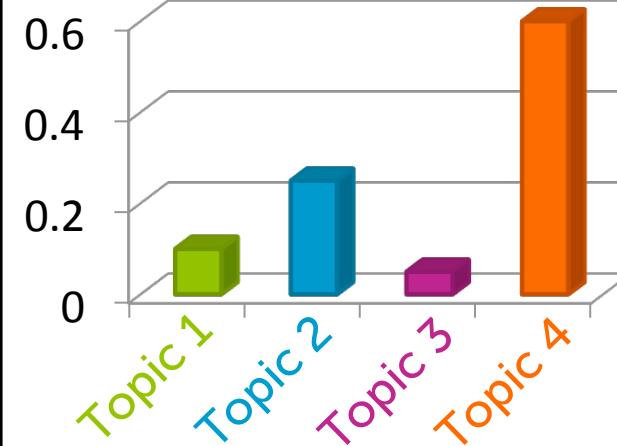
### Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

### 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



Typically **not** interested  
in word assignments



# An inference algorithm for LDA: Gibbs sampling



# Clustering algorithms **so far**

## k-means

Assign observations to  
closest cluster center

$$z_i \leftarrow \arg \min_j \|\mu_j - \mathbf{x}_i\|_2^2$$

Revise cluster centers

$$\mu_j \leftarrow \arg \min_{\mu} \sum_{i:z_i=j} \|\mu - \mathbf{x}_i\|_2^2$$

Iterative **hard** assignment  
to max objective

## EM for MoG

**E-step:** estimate cluster  
responsibilities

$$\hat{r}_{ik} = \frac{\hat{\pi}_k N(x_i | \hat{\mu}_k, \hat{\Sigma}_k)}{\sum_{j=1}^K \hat{\pi}_j N(x_i | \hat{\mu}_j, \hat{\Sigma}_j)}$$

**M-step:** maximize likelihood  
over parameters

$$\hat{\pi}_k, \hat{\mu}_k, \hat{\Sigma}_k | \{\hat{r}_{ik}, x_i\}$$

Iterative **soft** assignment  
to max objective

# What can we do for our bag-of-words models?

## Part 1: Clustering model

SCIENCE	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TECH	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

SPORTS	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

:

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

*Keywords:* Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

One topic indicator  
 $z_i$  per document  $i$

All words come from  
(get scored under)  
same topic  $z_i$

Distribution on  
prevalence of  
topics in corpus

$$\boldsymbol{\pi} = [\pi_1 \ \pi_2 \ \dots \ \pi_K]$$

# What can we do for our bag-of-words models?

## Part 1: Clustering model

SCIENCE	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TECH	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

SPORTS	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

:

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

*Keywords:* Bayesian nonparametric, EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible

Can derive  
EM algorithm:



- Gaussian likelihood of tf-idf vector



multinomial likelihood  
of word counts

( $m_w$  successes of word  $w$ )

- Result: mixture of multinomial model

# What can we do for our bag-of-words models?

## Part 2: LDA model

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

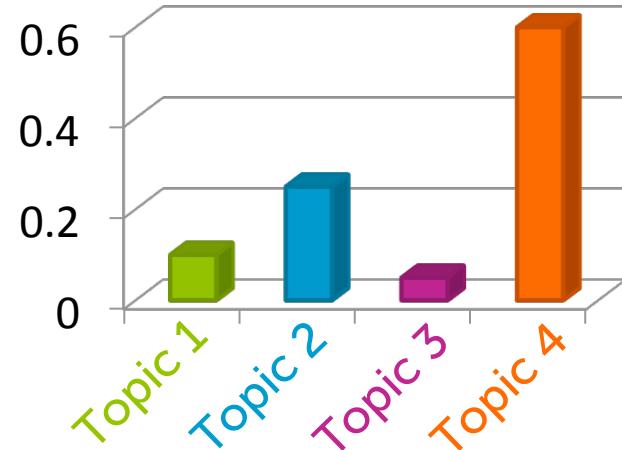
Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



Can derive  
EM algorithm,  
but not common  
(performs poorly)



# Typical LDA implementations

Normally LDA is specified as a **Bayesian model**  
(otherwise, “probabilistic latent semantic analysis/indexing”)

- Account for uncertainty in parameters 
- when making predictions
- Naturally regularizes parameter estimates  
in contrast to MLE

EM-like algorithms (e.g., “variational EM”), or...

# Gibbs sampling for Bayesian inference

# Gibbs sampling

Iterative **random** hard assignment!



Benefits:

- Typically intuitive updates
- Very straightforward to implement

# Random sample #10000



## TOPIC 1

experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

## TOPIC 2

develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

## TOPIC 3

player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

## Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

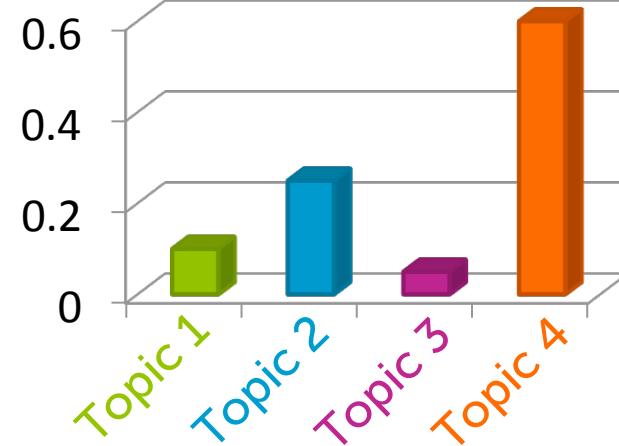
### Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

### 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



## Current set of assignments

# Random sample #10001

## TOPIC 1

experiment	0.12
test	0.06
hypothesize	0.042
discover	0.04
climate	0.011
...	...

## TOPIC 2

develop	0.16
computer	0.11
user	0.03
processor	0.029
internet	0.023
...	...

## TOPIC 3

player	0.15
score	0.07
team	0.06
offense	0.02
defense	0.018
...	...

## Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

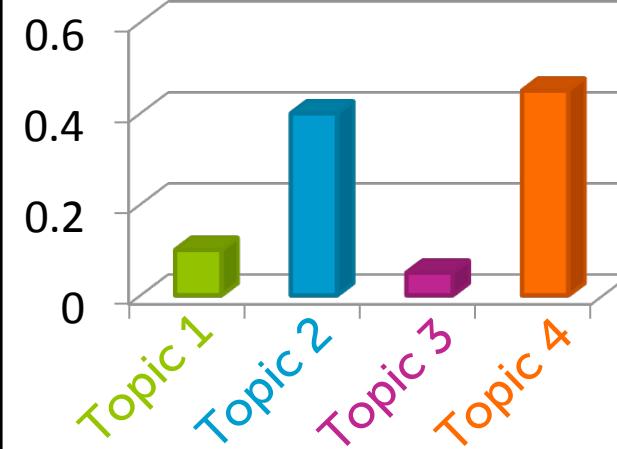
### Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

### 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



## Current set of assignments

# Random sample #10002

## TOPIC 1

experiment	0.10
discover	0.055
hypothesize	0.043
test	0.042
examine	0.015
...	...

## TOPIC 2

computer	0.12
develop	0.115
user	0.031
device	0.022
cloud	0.018
...	...

## TOPIC 3

player	0.17
score	0.09
game	0.062
team	0.043
win	0.03
...	...

## Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

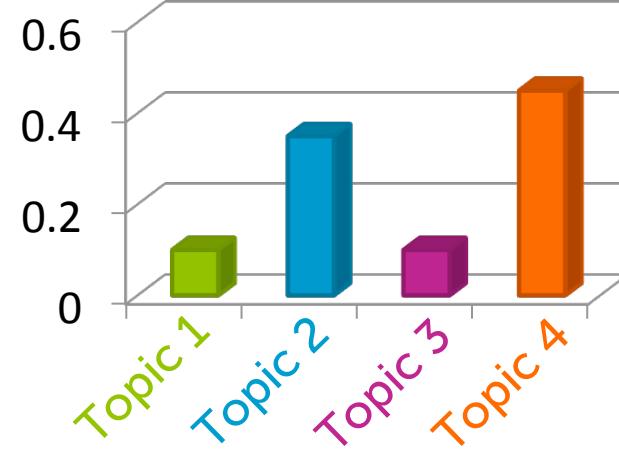
### Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

### 1. Introduction

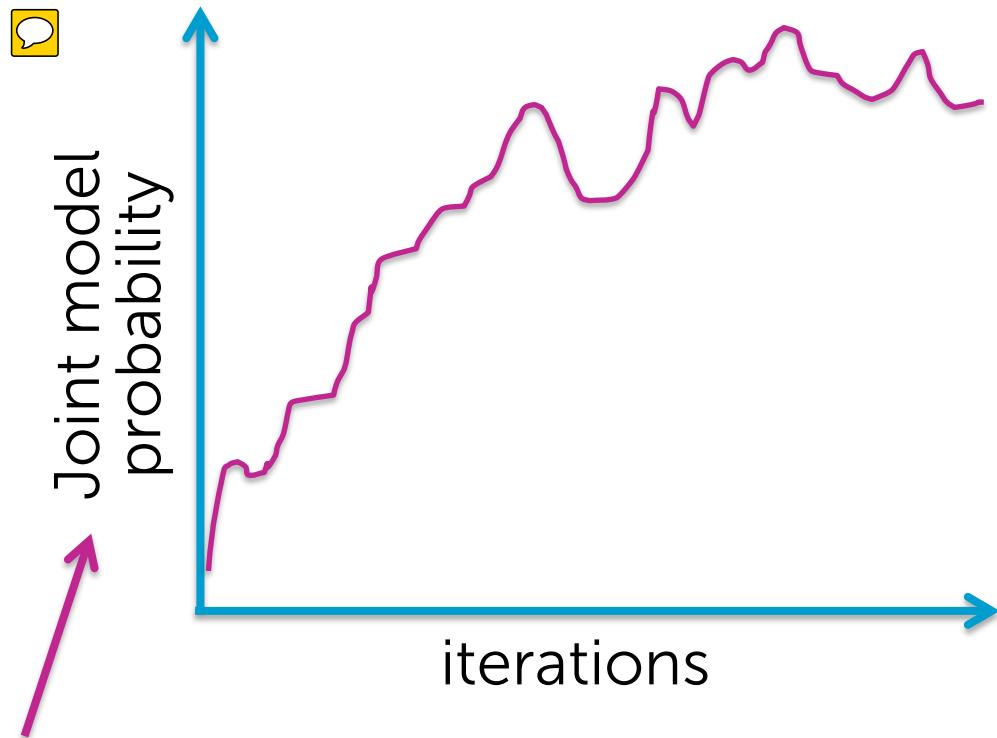
Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



## Current set of assignments

# What do we know about this process?

Not an optimization algorithm



probability of observations given variables/parameters  
and probability of variables/parameters themselves

Eventually  
provides  
“correct”  
Bayesian  
estimates...

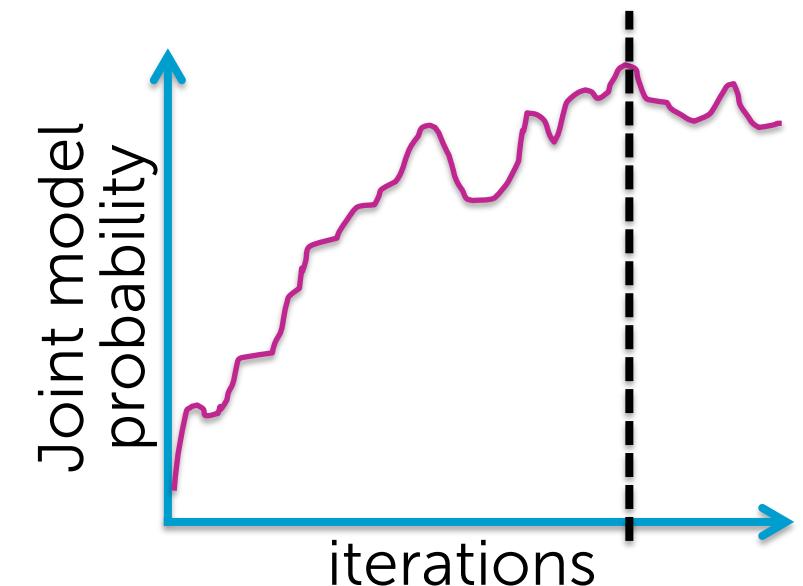
# What to do with sampling output?

## Predictions:

1. Make prediction for each snapshot of randomly assigned variables/parameters (full iteration)
2. Average predictions for final result

## Parameter or assignment estimate:

- Look at snapshot of randomly assigned variables/parameters that maximizes "joint model probability"





# Standard Gibbs sampling steps

# Gibbs sampling algorithm outline

Iterative **random** hard assignment!

Assignment variables and model parameters  
treated similarly

Iteratively **draw variable/parameter from  
conditional distribution** having fixed:

- all other variables/parameters
  - values randomly selected in previous rounds
  - changes from iter to iter
- observations
  - always the same values

# Gibbs sampling for LDA

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

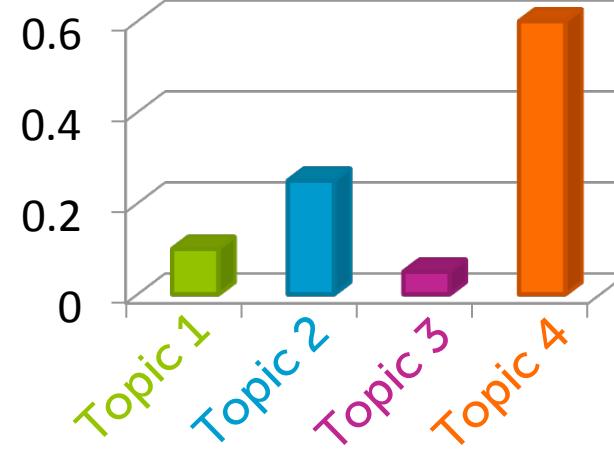
Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



## Current set of assignments

# Gibbs sampling for LDA

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

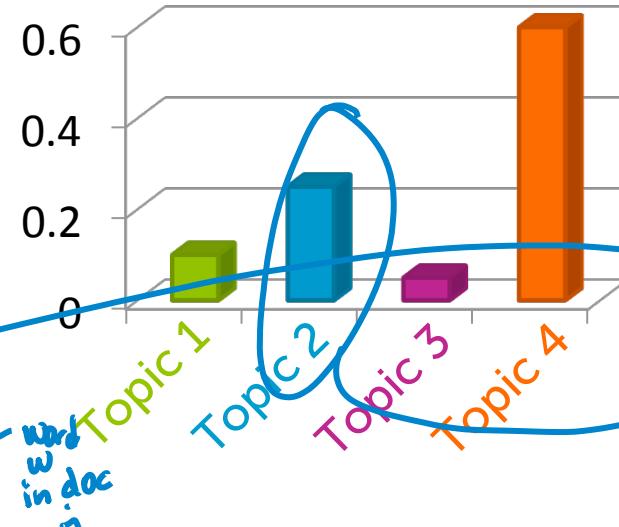
**Abstract**

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian, nonparametric, EEG, factorial hidden Markov model, graphical model, time series

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



**Step 1: Randomly reassigned all  $z_{iw}$  based on**

- doc topic proportions
- topic vocab distributions

Draw randomly from responsibility vector  
 $[r_{iw1} \ r_{iw2} \ \dots \ r_{iwK}]$

$$r_{iw2} = \frac{\text{prior prob. } z_{iw=2} \cdot P(\text{"EEG"} | z_{iw=2})}{\sum_{j=1}^K \text{prior prob. } z_{iw=j} \cdot P(\text{"EEG"} | z_{iw=j})}$$

prob. of assigning  $z_{iw}=2$



# Gibbs sampling for LDA

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



**Step 2: Randomly reassign doc topic proportions based on assignments  $z_{iw}$  in current doc**



# Gibbs sampling for LDA

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

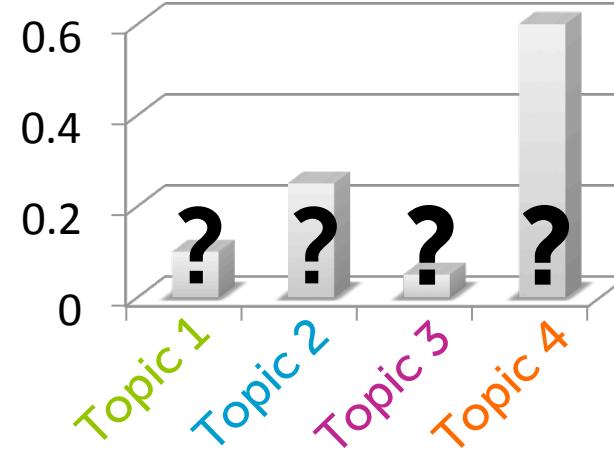
Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



**Step 3: Repeat for all docs**

# Gibbs sampling for LDA

TOPIC 1	
Word 1	?
Word 2	?
Word 3	?
Word 4	?
Word 5	?
...	...

TOPIC 2	
Word 1	?
Word 2	?
Word 3	?
Word 4	?
Word 5	?
...	...

TOPIC 3	
Word 1	?
Word 2	?
Word 3	?
Word 4	?
Word 5	?
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

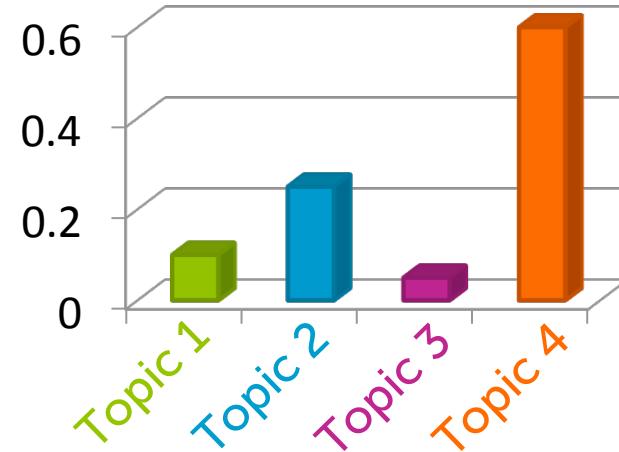
Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

---

1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



**Step 4: Randomly  
reassign topic vocab  
distributions based on  
assignments  $z_{iw}$  in  
entire corpus**

# Gibbs sampling for LDA

## TOPIC 1

experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

## TOPIC 2

develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

## TOPIC 3

player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

## Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

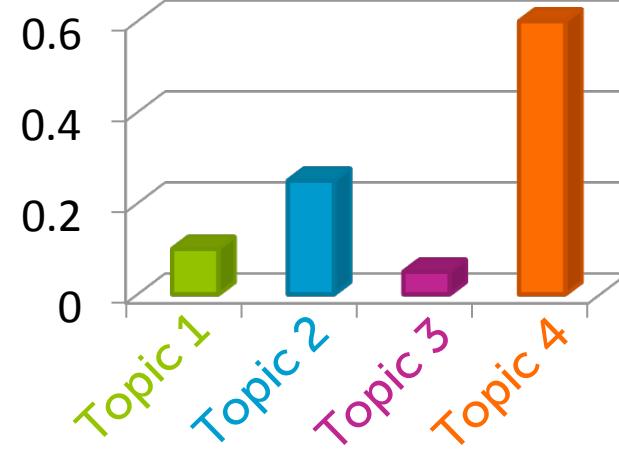
### Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

### 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



**Repeat Steps 1-4 until  
max iter reached**

# Collapsed Gibbs sampling in LDA



# “Collapsed” Gibbs sampling for LDA

Based on special structure of LDA model, can sample **just** indicator variables  $z_{iw}$

- No need to sample other parameters
  - corpus-wide topic vocab distributions
  - per-doc topic proportions

Often leads to much better performance because examining uncertainty in smaller space

# Collapsed Gibbs sampling for LDA

TOPIC 1	
experiment	0.
test	0.
discover	0.
hypothesize	0.
climate	0.
...	...

Never draw topic vocab distributions or doc topic proportions

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

## Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

## 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



Randomly reassign  $z_{iw}$  based on current assignments  $z_{jv}$  of all other words **in document and corpus**



# Select a document

epilepsy	dynamic	Bayesian	EEG	model



5 word document

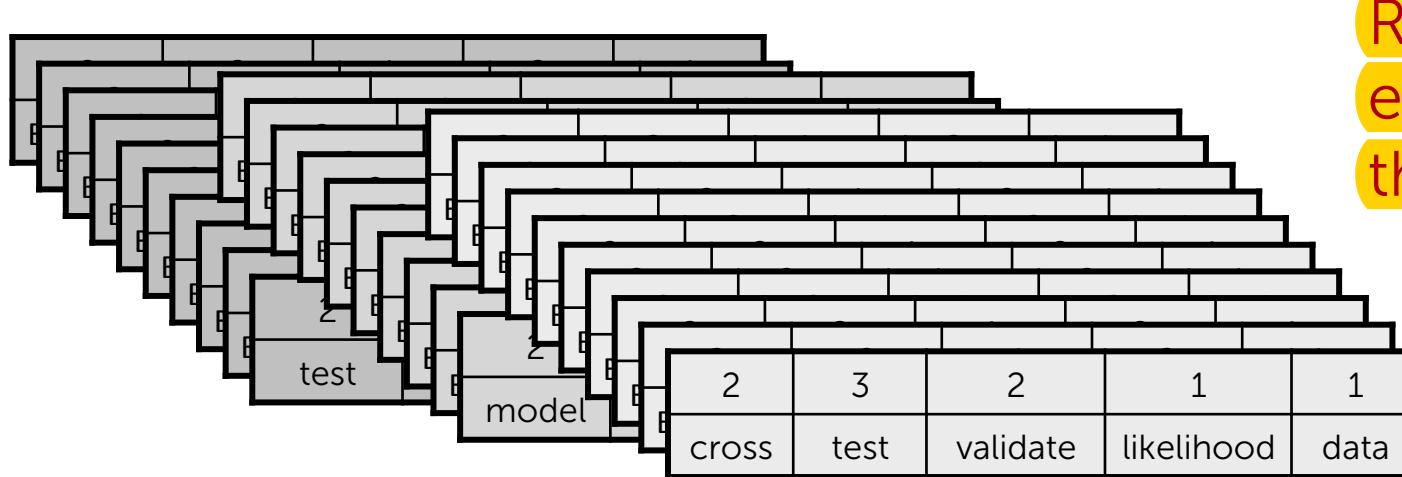
# Randomly assign topics

3	2	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

(one possible approach) 

# Randomly assign topics

3	2	1	3	1
epilepsy	dynamic	Bayesian	EEG	model



Repeat for  
each doc in  
the corpus



# Maintain local statistics

3	2	1	3	1
epilepsy	dynamic	Bayesian	EEG	model



	Topic 1	Topic 2	Topic 3
Doc i	2	1	2



# Maintain global statistics

3	2	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

	Topic 1	Topic 2	Topic 3
epilepsy	1	0	35
Bayesian	50	0	1
model	42	1	0
EEG	0	0	20
dynamic	10	8	1
...			

	Topic 1	Topic 2	Topic 3
Doc i	2	1	2

 Total counts from **all** docs

# Randomly reassign topics



3	<del>2</del>	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

	Topic 1	Topic 2	Topic 3
epilepsy	1	0	35
Bayesian	50	0	1
model	42	1	0
EEG	0	0	20
dynamic	10	<del>7</del> <sup>8</sup>	1
...			

	Topic 1	Topic 2	Topic 3
Doc i	2	<del>0</del> <sup>1</sup>	2

decrementing  
counts  
after removing  
current assignment  
 $z_{iw} = 2$

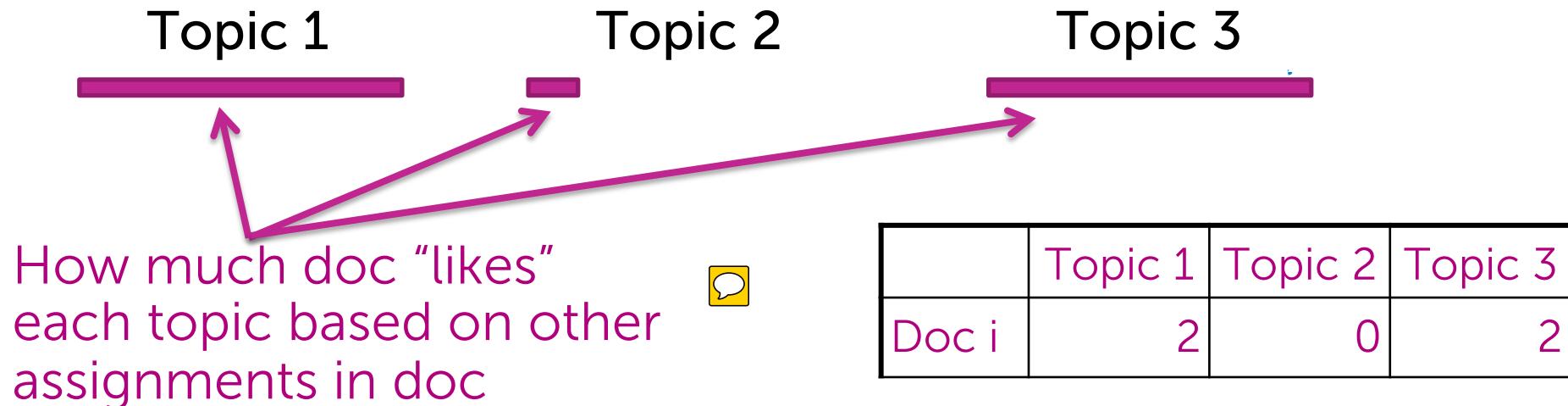
# Probability of new assignment

3	?	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

Reassign with probability  
 $p(z_{iw} | \text{every other } z_{jv} \text{ in corpus},$   
 $\text{words in corpus})$  

# Probability of new assignment

3	?	1	3	1
epilepsy	dynamic	Bayesian	EEG	model



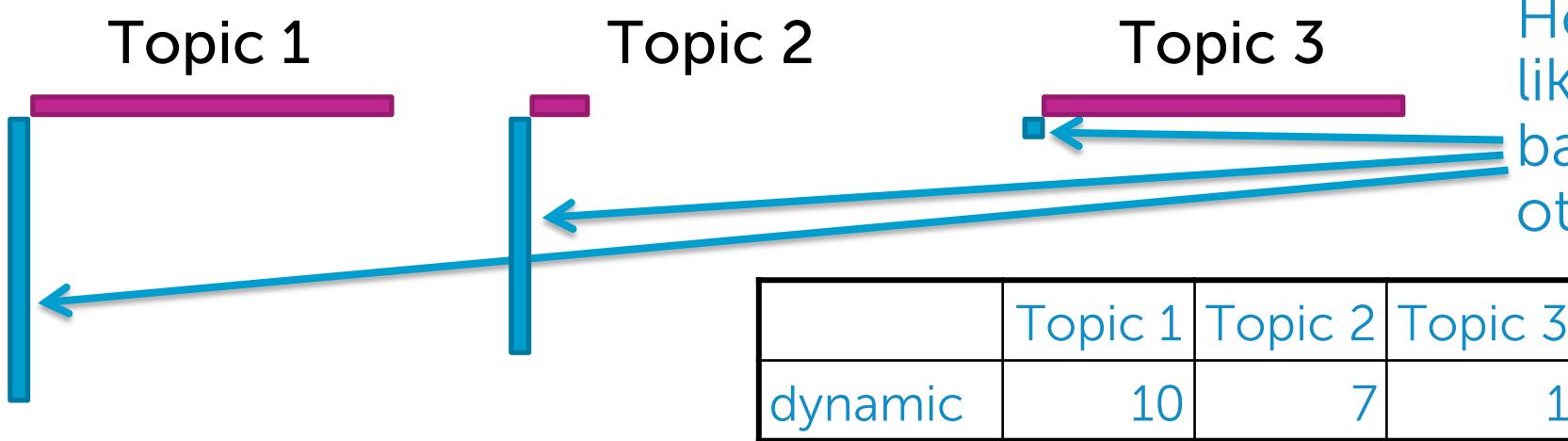
$$\text{# current assignments to topic } k \text{ in doc } i \rightarrow \frac{n_{ik} + \alpha}{N_i - 1 + K\alpha}$$

smoothing param *from Bayes prior*  
ignore current word

# words in doc i →  $N_i - 1 + K\alpha$

# Probability of new assignment

3	?	1	3	1
epilepsy	dynamic	Bayesian	EEG	model



How much each topic likes the word "dynamic" based on assignments in other docs in corpus

# assignments **corpus-wide** of word "dynamic" to topic k



$$\frac{m_{\text{dynamic},k} + \gamma}{\sum_{w \in V} m_{w,k} + V\gamma}$$

smoothing param *from Bayes prior*  
size of vocab

# Probability of new assignment

3	?	1	3	1
epilepsy	dynamic	Bayesian	EEG	model



Topic 2 also really likes “dynamic”,  
but in a different context...  
e.g., a topic on fluid dynamics

# Probability of new assignment

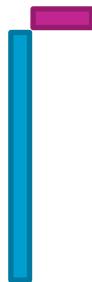
3	?	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

Topic 1



Topic fits word  
and document

Topic 2



Topic fits word,  
but not doc

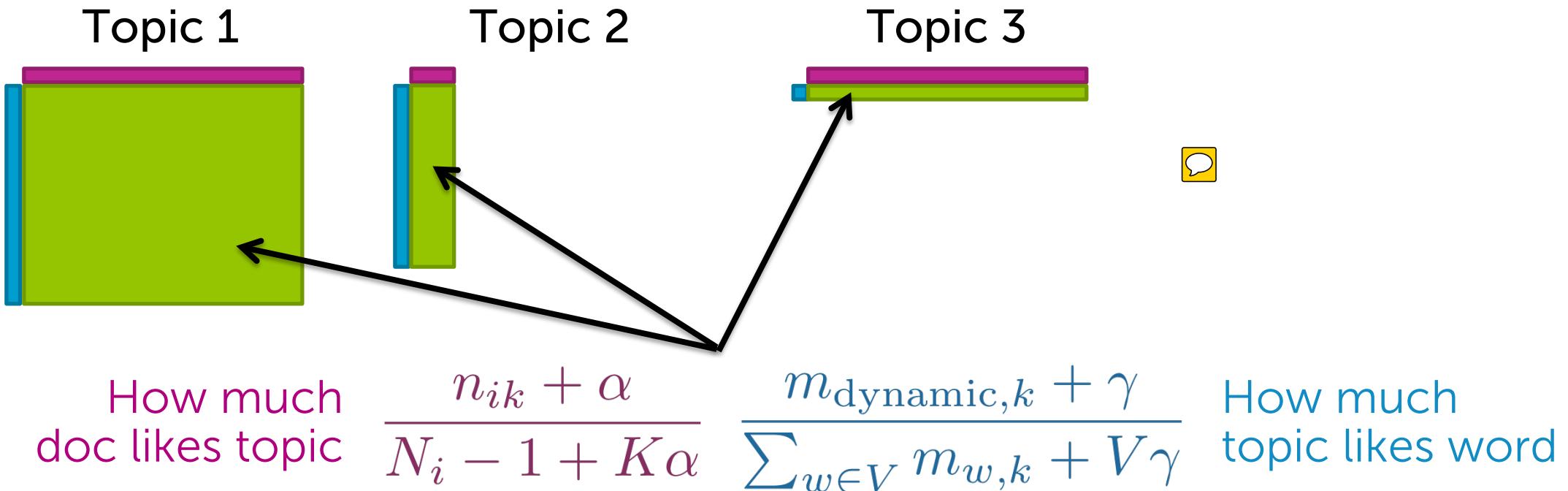
Topic 3



Topic fits doc,  
but not word

# Probability of new assignment

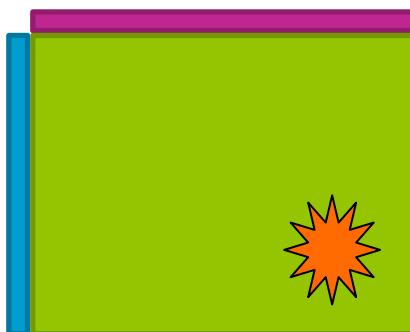
3	?	1	3	1
epilepsy	dynamic	Bayesian	EEG	model



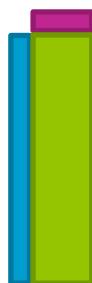
# Randomly draw a new topic indicator

3	?	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

Topic 1



Topic 2



Topic 3



To draw new topic assignment (equivalently):

- roll K-sided die with these probabilities
- throw dart at these regions

Normalize this product of terms over K possible topics!

How much doc likes topic

$$\frac{n_{ik} + \alpha}{N_i - 1 + K\alpha} \quad \frac{m_{\text{dynamic},k} + \gamma}{\sum_{w \in V} m_{w,k} + V\gamma}$$

How much topic likes word

# Update counts

3	1	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

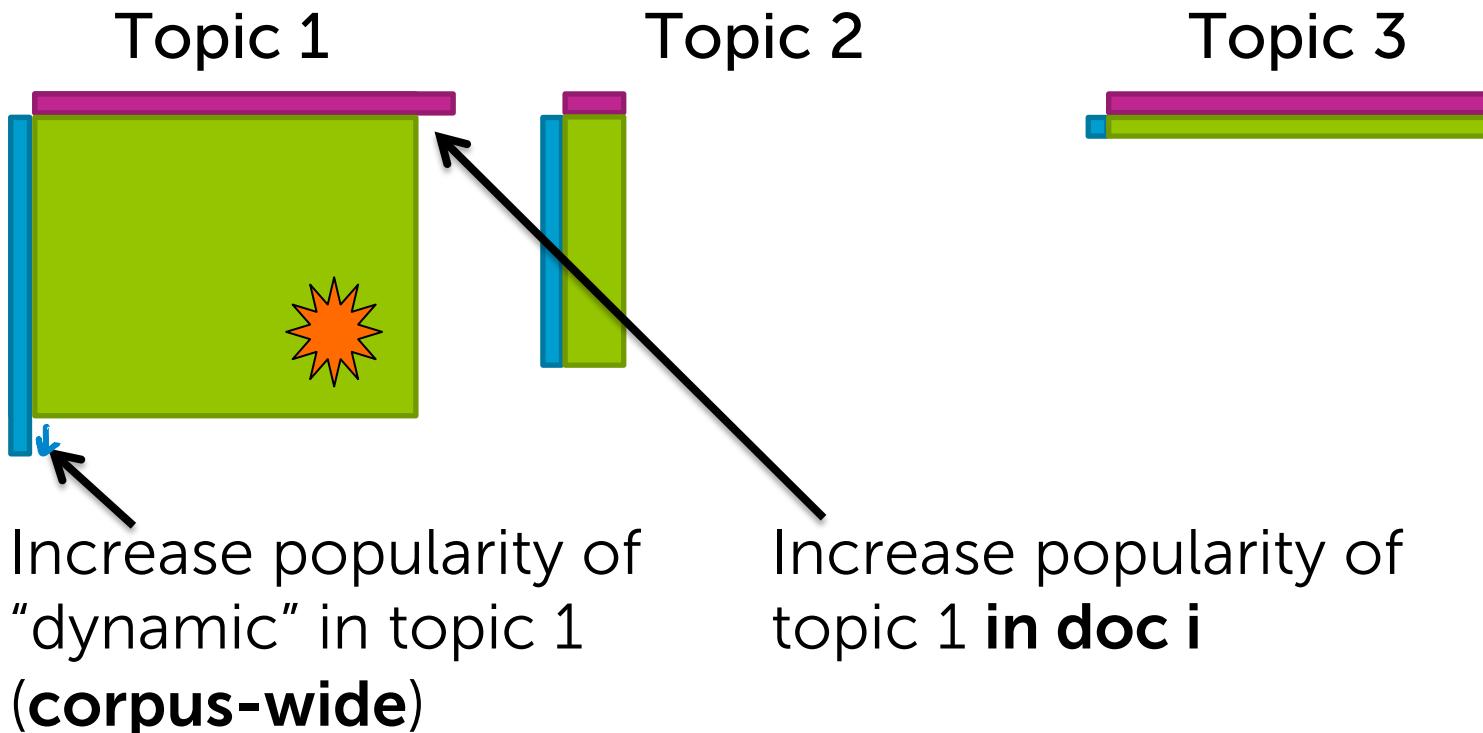
	Topic 1	Topic 2	Topic 3
epilepsy	1	0	35
Bayesian	50	0	1
model	42	1	0
EEG	0	0	20
dynamic	11	7	1
...			

	Topic 1	Topic 2	Topic 3
Doc i	3	2	0

increment counts  
based on new  
assignment of  
 $z_{iw}=1$

# Geometrically...

3	1	1	3	1
epilepsy	dynamic	Bayesian	EEG	model



# Iterate through all words/docs

3	1	1	3	1
epilepsy	dynamic	Bayesian	EEG	model

	Topic 1	Topic 2	Topic 3
epilepsy	1	0	35
Bayesian	50	0	1
model	42	1	0
EEG	0	0	20
dynamic	10	7	1
...			

	Topic 1	Topic 2	Topic 3
Doc i	2	0	2



# Using samples from collapsed Gibbs



# What to do with the collapsed samples?

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

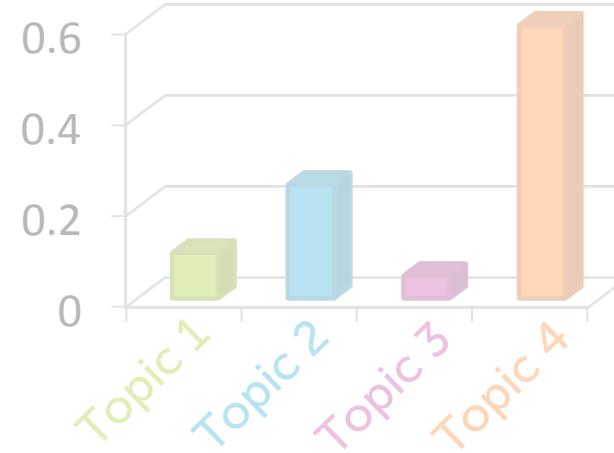
Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



From “best” sample of  $\{z_{iw}\}$ ,  
can infer:

# What to do with the collapsed samples?

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

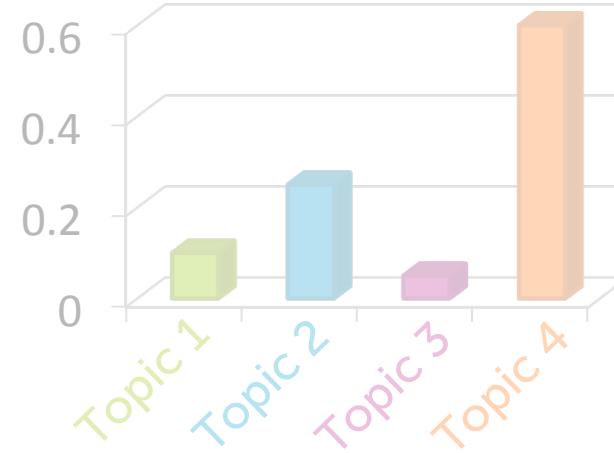
Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



From “best” sample of  $\{z_{iw}\}$ , can infer:

1. Topics from conditional distribution... need corpus-wide info

# What to do with the collapsed samples?

## TOPIC 1

experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

## TOPIC 2

develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	...

## TOPIC 3

player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

## Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA

<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA

<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

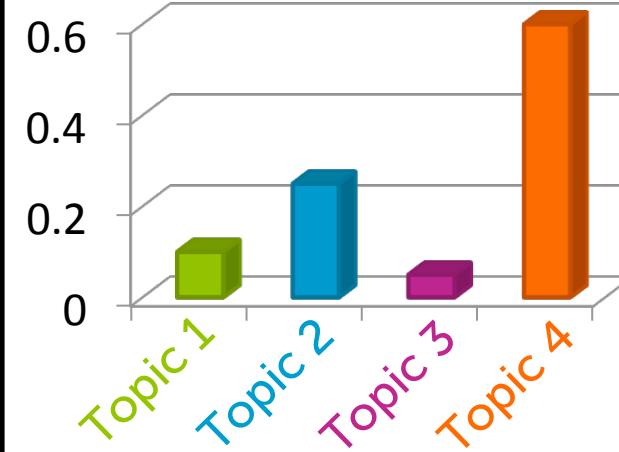
### Abstract

Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

### 1. Introduction

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



From “best” sample of  $\{z_{iw}\}$ ,  
can infer:

1. Topics from conditional distribution...  
need corpus-wide info
2. Document “embedding”...  
need doc info only

# Embedding new documents

TOPIC 1	
experiment	0.1
test	0.08
discover	0.05
hypothesize	0.03
climate	0.01
...	...

TOPIC 2	
develop	0.18
computer	0.09
processor	0.032
user	0.027
internet	0.02
...	

TOPIC 3	
player	0.15
score	0.07
team	0.06
goal	0.03
injury	0.01
...	...

⋮

Modeling the Complex Dynamics and Changing Correlations of Epileptic Events

Drausin F. Wulsin<sup>a</sup>, Emily B. Fox<sup>c</sup>, Brian Litt<sup>a,b</sup>

<sup>a</sup>Department of Bioengineering, University of Pennsylvania, Philadelphia, PA  
<sup>b</sup>Department of Neurology, University of Pennsylvania, Philadelphia, PA  
<sup>c</sup>Department of Statistics, University of Washington, Seattle, WA

---

**Abstract**

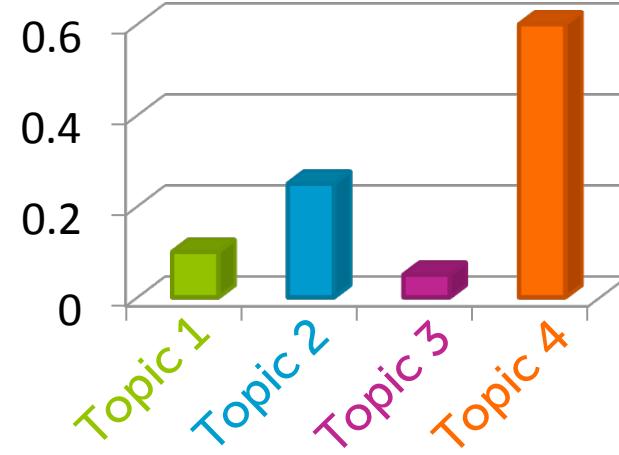
Patients with epilepsy can manifest short, sub-clinical epileptic “bursts” in addition to full-blown clinical seizures. We believe the relationship between these two classes of events—something not previously studied quantitatively—could yield important insights into the nature and intrinsic dynamics of seizures. A goal of our work is to parse these complex epileptic events into distinct dynamic regimes. A challenge posed by the intracranial EEG (iEEG) data we study is the fact that the number and placement of electrodes can vary between patients. We develop a Bayesian nonparametric Markov switching process that allows for (i) shared dynamic regimes between a variable number of channels, (ii) asynchronous regime-switching, and (iii) an unknown dictionary of dynamic regimes. We encode a sparse and changing set of dependencies between the channels using a Markov-switching Gaussian graphical model for the innovations process driving the channel dynamics and demonstrate the importance of this model in parsing and out-of-sample predictions of iEEG data. We show that our model produces intuitive state assignments that can help automate clinical analysis of seizures and enable the comparison of sub-clinical bursts and full clinical seizures.

**Keywords:** Bayesian nonparametric EEG, factorial hidden Markov model, graphical model, time series

---

**1. Introduction**

Despite over three decades of research, we still have very little idea of what defines a seizure. This ignorance stems both from the complexity of epilepsy as a disease and a paucity of quantitative tools that are flexible



## Simple approach:

1. Fix topics based on training set collapsed sampling
2. Run uncollapsed sampler on new doc(s) only

# Summary for LDA and Gibbs sampling

# What you can do now...

- Compare and contrast clustering and mixed membership models
- Describe a document clustering model for the bag-of-words doc representation
- Interpret the components of the LDA mixed membership model
- Analyze a learned LDA model
  - Topics in the corpus
  - Topics per document
- Describe Gibbs sampling steps at a high level
- Utilize Gibbs sampling output to form predictions or estimate model parameters
- Implement collapsed Gibbs sampling for LDA

# Acknowledgements

Thanks to David Mimno  
(<http://mimno.infosci.cornell.edu>)  
for the collapsed LDA example outline

