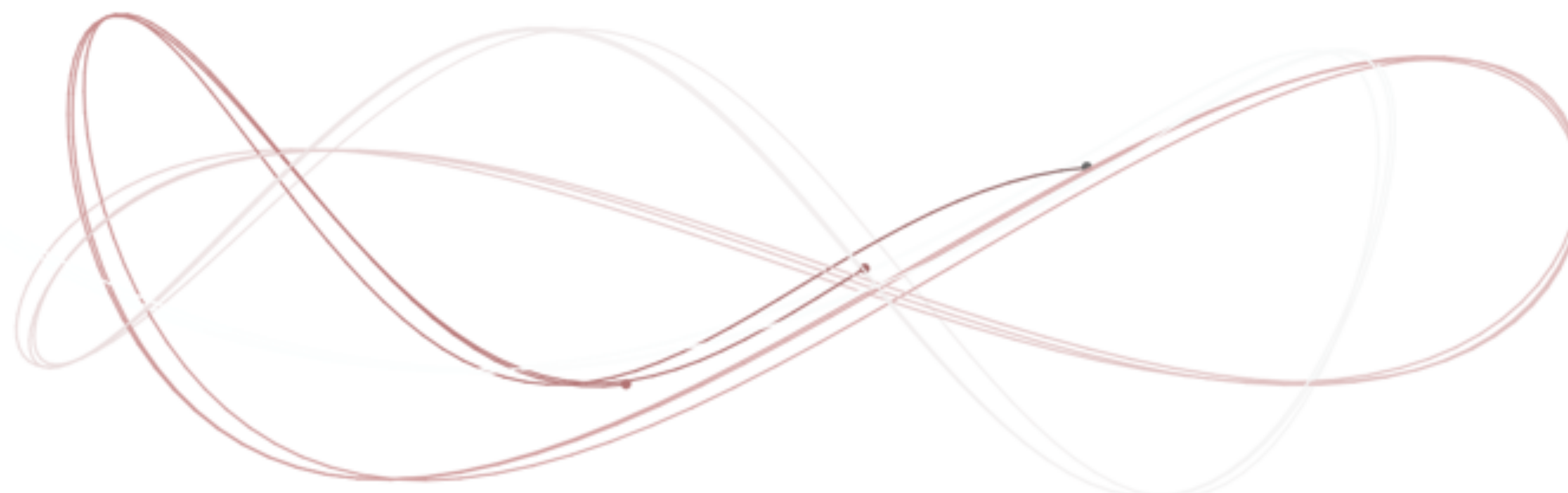
Truncated Random Measures

Jonathan Huggins
MIT CSAIL and Dept. of EECS

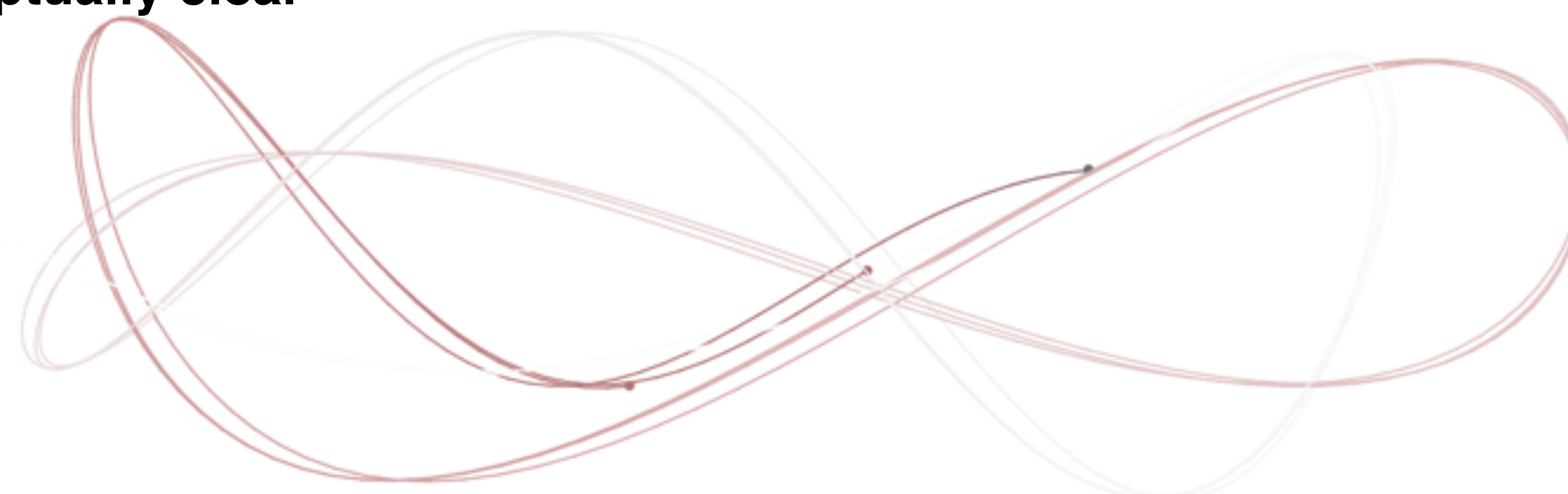
with: T. Campbell, J. How, T. Broderick

What leads to a statistical method being used for science?



What leads to a statistical method being used for science?

1. **Conceptually clear**



What leads to a statistical method being used for science?

1. **Conceptually clear**

- Bayesian methods are **conceptually clear**…

What leads to a statistical method being used for science?

1. **Conceptually clear**

- Bayesian methods are **conceptually clear**…

2. **Easy to use**

What leads to a statistical method being used for science?

1. Conceptually clear

- Bayesian methods are **conceptually clear**…

2. Easy to use

- …but often **not easy to use**…

What leads to a statistical method being used for science?

1. **Conceptually clear**

- Bayesian methods are **conceptually clear**…

2. **Easy to use**

- …but often **not easy to use**…

3. **Reliable**

What leads to a statistical method being used for science?

1. **Conceptually clear**

- Bayesian methods are **conceptually clear**…

2. **Easy to use**

- …but often **not easy to use**…

3. **Reliable**

- …which makes them **less reliable**

What leads to a statistical method being used for science?

1. **Conceptually clear**

- Bayesian methods are **conceptually clear**…

2. **Easy to use**

- …but often **not easy to use**…

3. **Reliable**

- …which makes them **less reliable**
- How to fix this? **probabilistic programming**

What leads to a statistical method being used for science?

1. **Conceptually clear**

- Bayesian methods are **conceptually clear**…

2. **Easy to use**

- …but often **not easy to use**…

3. **Reliable**

- …which makes them **less reliable**
- How to fix this? **probabilistic programming**
  - Write down the model, but don't worry about inference

What leads to a statistical method being used for science?

1. **Conceptually clear**

- Bayesian methods are **conceptually clear**…

2. **Easy to use**

- …but often **not easy to use**…

3. **Reliable**

- …which makes them **less reliable**
- How to fix this? **probabilistic programming**
  - Write down the model, but don't worry about inference
  - v1.0: **BUGS/JAGS** (Gibbs sampling)

BUGS

What leads to a statistical method being used for science?

1. **Conceptually clear**

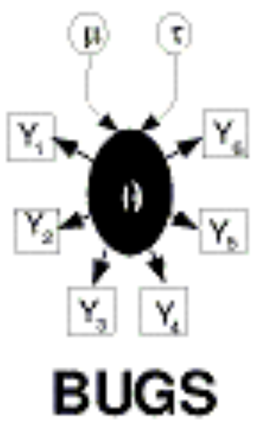
- Bayesian methods are **conceptually clear**…

2. **Easy to use**

- …but often **not easy to use**…

3. **Reliable**

- …which makes them **less reliable**
- How to fix this? **probabilistic programming**
  - Write down the model, but don't worry about inference
  - v1.0: **BUGS/JAGS** (Gibbs sampling)
  - v2.0: **Stan** (HMC or variational inference or MAP estimation)



What leads to a statistical method being used for science?

1. **Conceptually clear**

- Bayesian methods are **conceptually clear**…

2. **Easy to use**

- …but often **not easy to use**…

3. **Reliable**

- …which makes them **less reliable**
- How to fix this? **probabilistic programming**
    - Write down the model, but don't worry about inference
    - v1.0: **BUGS/JAGS** (Gibbs sampling)
    - v2.0: **Stan** (HMC or variational inference or MAP estimation)
- **Goal:** integrate BNP priors into PPLs like Stan

BNP: awesome, but challenging to use

BNP: awesome, but challenging to use

Need models that can extract new, useful information from infinite streams of data

BNP: awesome, but challenging to use

Need models that can extract new, useful information from infinite streams of data

e.g. keep learning new topics from a stream of documents

BNP: awesome, but challenging to use

Need models that can extract new, useful information from infinite streams of data

e.g. keep learning new topics from a stream of documents

Bayesian nonparametrics: achieves growing model size via infinite parameters

BNP: awesome, but challenging to use

Need models that can extract new, useful information from infinite streams of data

e.g. keep learning new topics from a stream of documents

Bayesian nonparametrics: achieves growing model size via infinite parameters

[Gopalan 2014] [Teh 2006] [Huang 2014] [Michini 2015] [Lennox 2010] [Prunster 2014] [Yang 2015] [Yu 2012] [Ozaki 2008] [Kottas 2008]

BNP: awesome, but challenging to use

Need models that can extract new, useful information from infinite streams of data

e.g. keep learning new topics from a stream of documents

movie text medicine robotics genetics

finance astronomy traffic agriculture pathology

hard work!

Bayesian nonparametrics: achieves growing model size via infinite parameters

[Gopalan 2014] [Teh 2006] [Huang 2014] [Michini 2015] [Lennox 2010] [Prunster 2014] [Yang 2015] [Yu 2012] [Ozaki 2008] [Kottas 2008]

BNP: awesome, but challenging to use

Need models that can extract new, useful information from infinite streams of data

e.g. keep learning new topics from a stream of documents

Bayesian nonparametrics: achieves growing model size via infinite parameters

hard work!

automate inference with probabilistic programming

[Gopalan 2014] [Teh 2006] [Huang 2014] [Michini 2015] [Lennox 2010] [Prunster 2014] [Yang 2015] [Yu 2012] [Ozaki 2008] [Kottas 2008]

Inference in BNP models

Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation

Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Problem:

Wide variety of priors in BNP with **no finite approximation**

Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Problem:

Wide variety of priors in BNP with **no finite approximation**

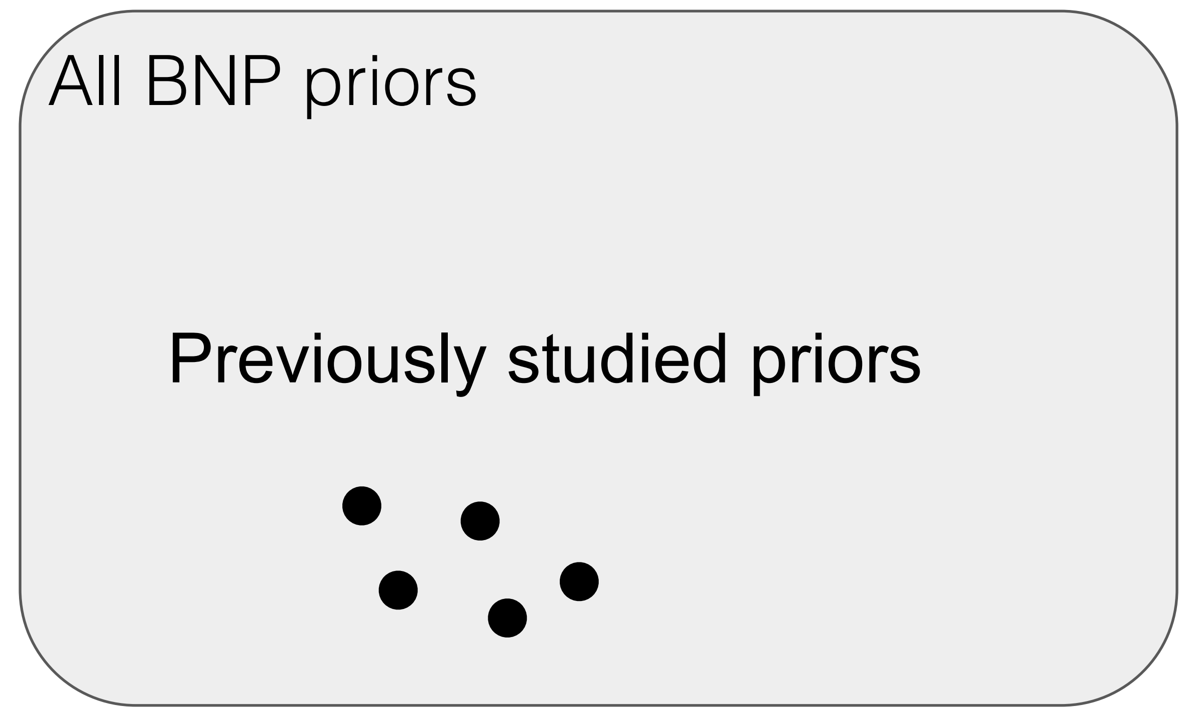
All BNP priors

Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Problem:

Wide variety of priors in BNP with **no finite approximation**

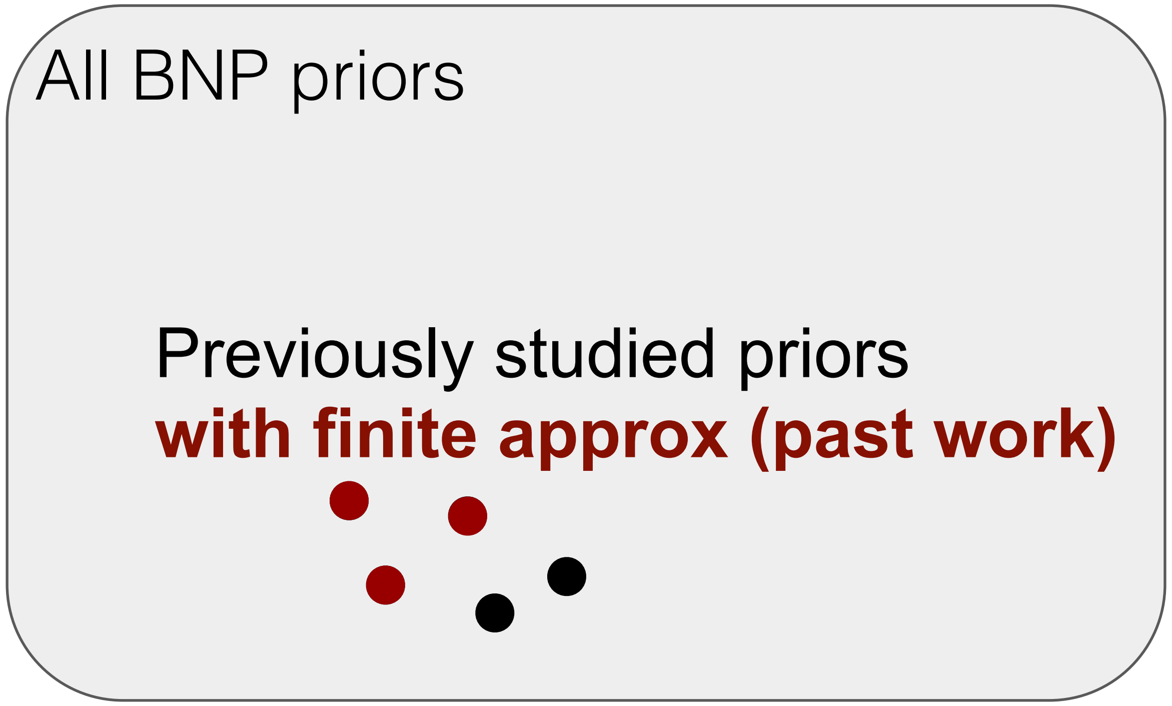


Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Problem:

Wide variety of priors in BNP with **no finite approximation**

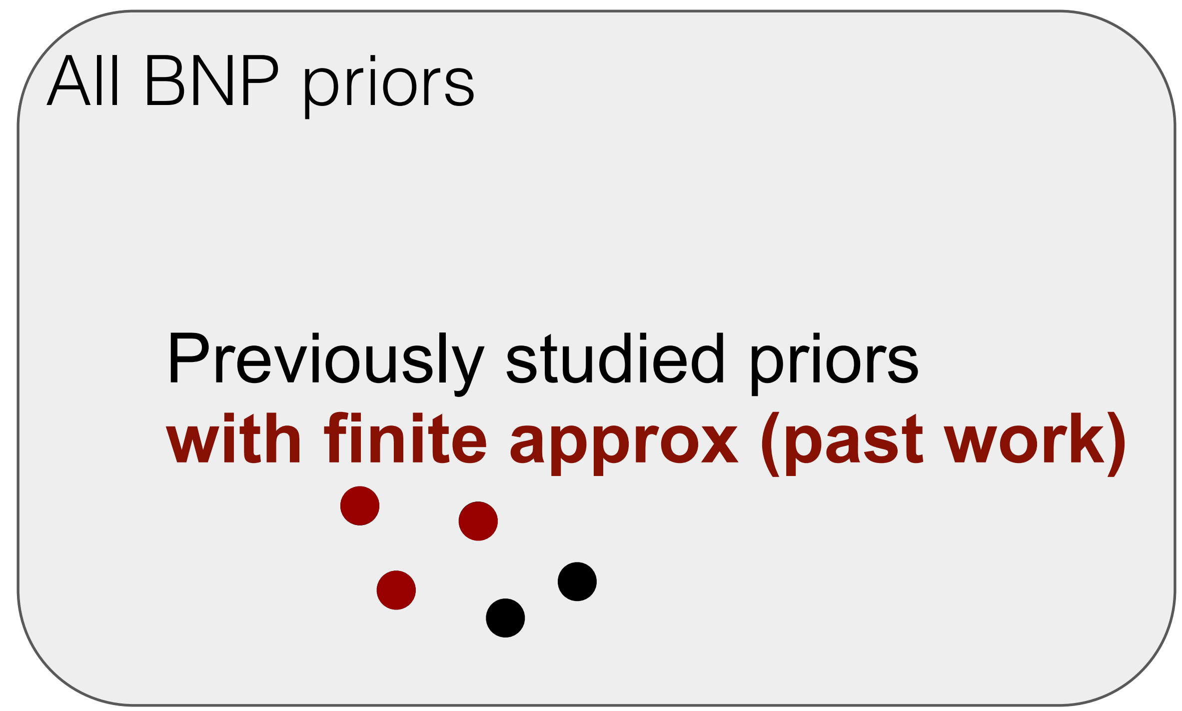


Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Problem:

Wide variety of priors in BNP with **no finite approximation**



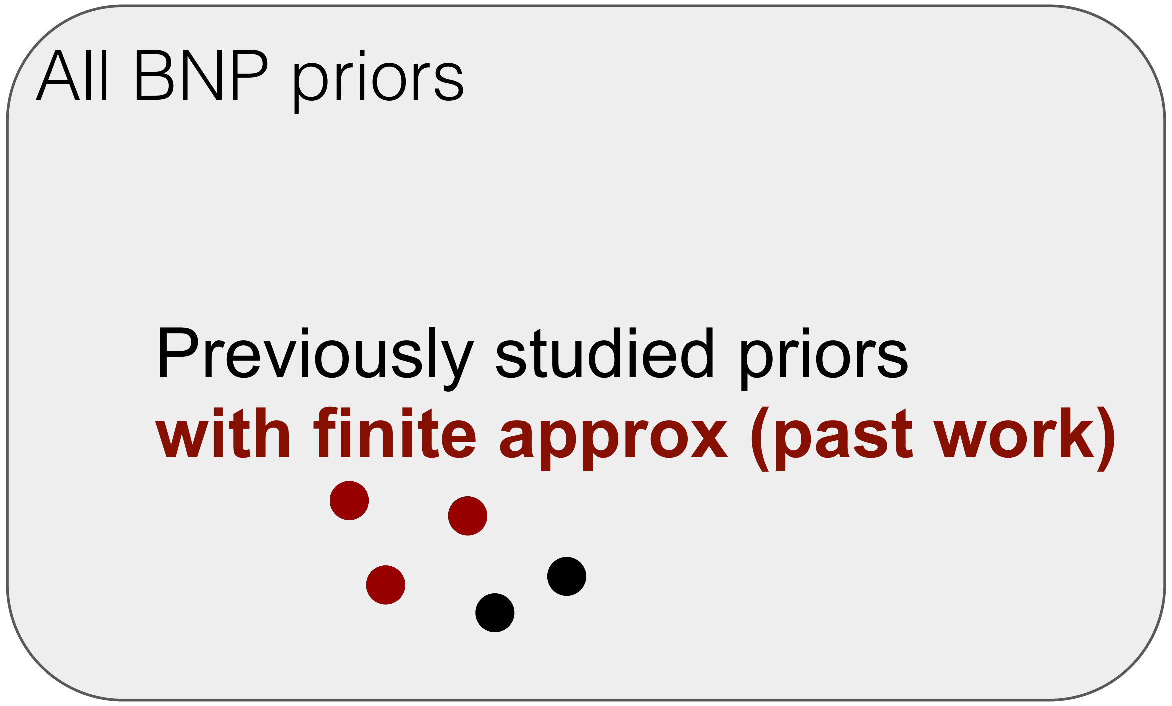
Contributions:

Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Problem:

Wide variety of priors in BNP with **no finite approximation**



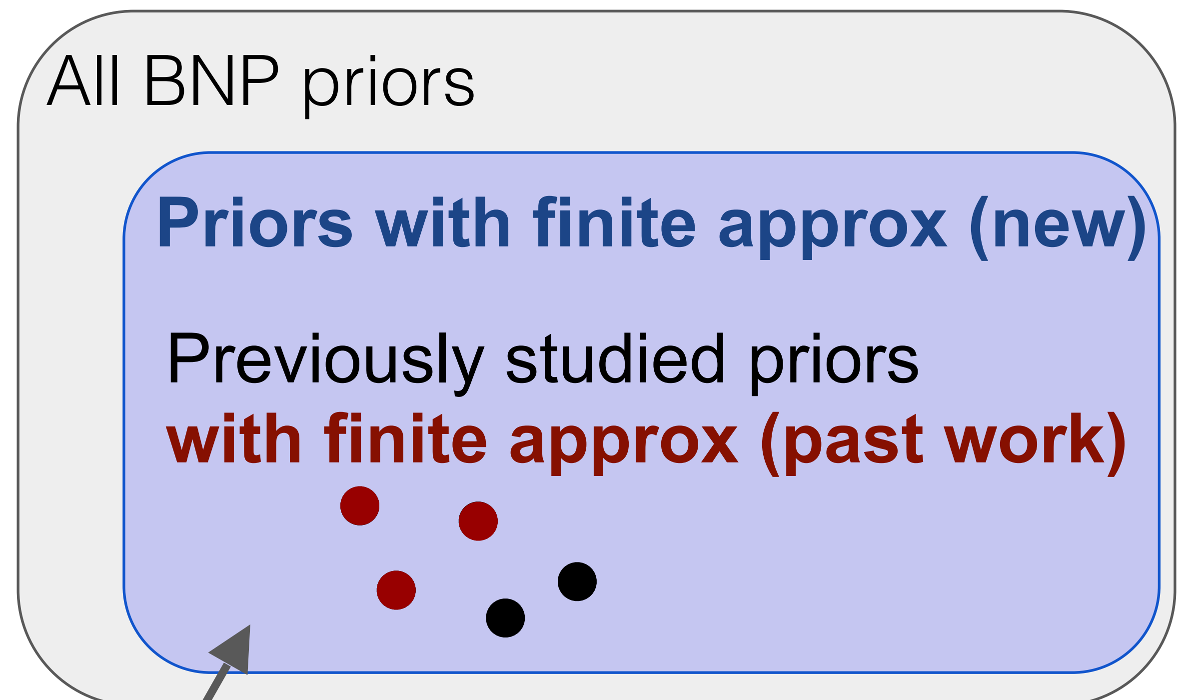
Contributions:

- 2 representation forms (7 reps total) that allow finite approximation of ***(normalized) completely random measures ((N)CRMs)***

Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Problem:
Wide variety of priors in BNP with **no finite approximation**



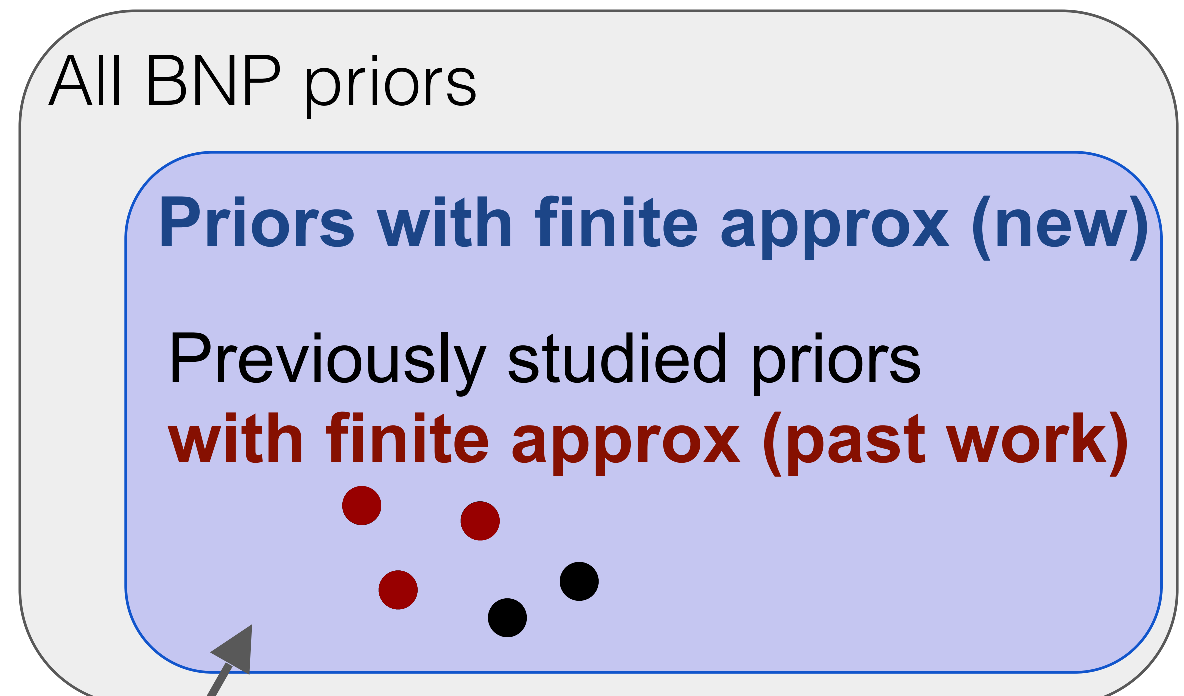
Contributions:

- 2 representation forms (7 reps total) that allow finite approximation of ***(normalized) completely random measures ((N)CRMs)***

Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Problem:
Wide variety of priors in BNP with **no finite approximation**



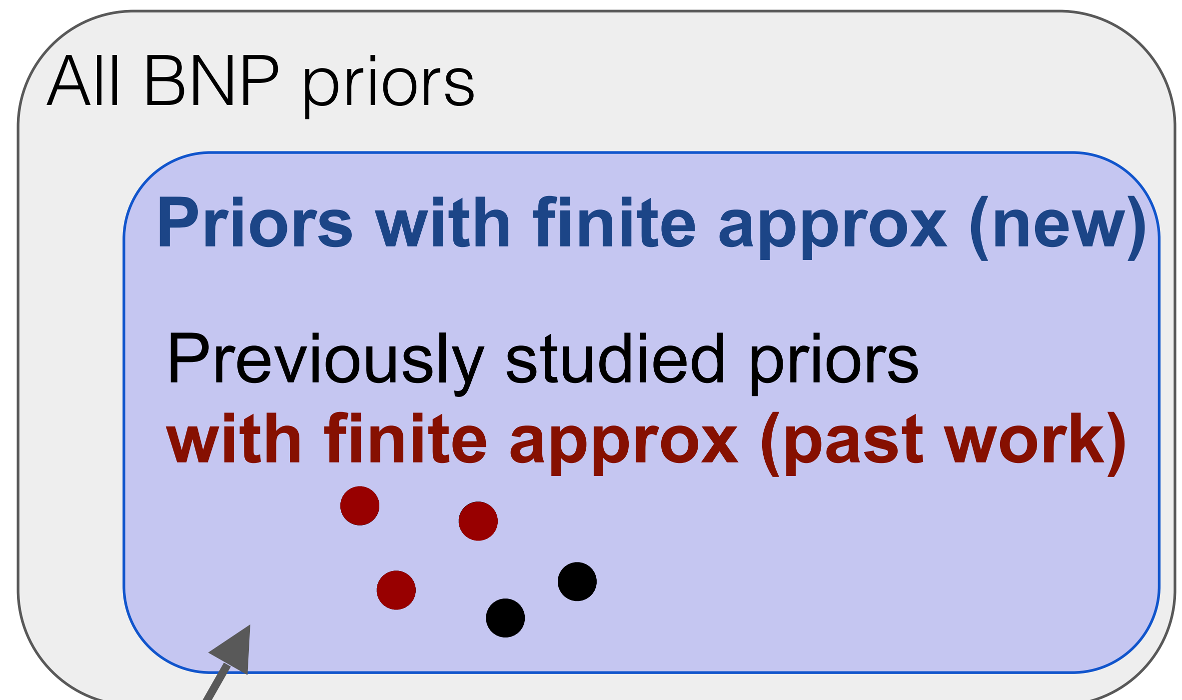
Contributions:

- 2 representation forms (7 reps total) that allow finite approximation of ***(normalized) completely random measures ((N)CRMs)***
- Approximation error analysis

Inference in BNP models

- Option #1: Integrate out the parameter (CRP, IBP, etc.)
  issues: care about the parameters, using approximations (HMC/VB), distributed computation
- Option #2: use a **finite approximation...**
  with e.g. variational inference, HMC
  [Blei 06; Neal 10]

Problem:
Wide variety of priors in BNP with **no finite approximation**



Contributions:

- 2 representation forms (7 reps total) that allow finite approximation of ***(normalized) completely random measures ((N)CRMs)***
- Approximation error analysis
- Computational complexity analysis (not in this talk)

Past work: finite approximations to BNP priors

	Finite Approximation	Approximation Error Bounds	Computational Complexity
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓		
ΓP	✓	✓	✓
(N)CRM	✓		

Past work: finite approximations to BNP priors

	Finite Approximation	Approximation Error Bounds	Computational Complexity
DP	✓ [Sethuraman 94] [Roychowdhury 15]	✓ [Ishwaran 01]	✓
BP	✓ [Teh 07] [Paisley 12] [Thibaux 07]	✓ [Doshi-Velez 09] [Paisley 12]	✓
BPP	✓ [Broderick 14]		
ΓP	✓ [Bondesson 82] [Roychowdhury 15]	✓ [Roychowdhury 15]	✓
(N)CRM	✓ [Broderick 14]		

Past work: finite approximations to BNP priors

	Finite Approximation	Approximation Error Bounds	Computational Complexity
DP	✓ [Sethuraman 94] [Roychowdhury 15]	✓ [Ishwaran 01]	✓
BP	✓ [Teh 07] [Paisley 12] [Thibaux 07]	✓ [Doshi-Velez 09] [Paisley 11]	✓
BPP	✓ [Broderick 14]		
ΓP	✓ [Bondesson 82] [Roychowdhury 15]	✓ [Roychowdhury 15]	✓
(N)CRM	✓ [Broderick 14]		

Sparse results for a few priors in BNP

Past work: finite approximations to BNP priors

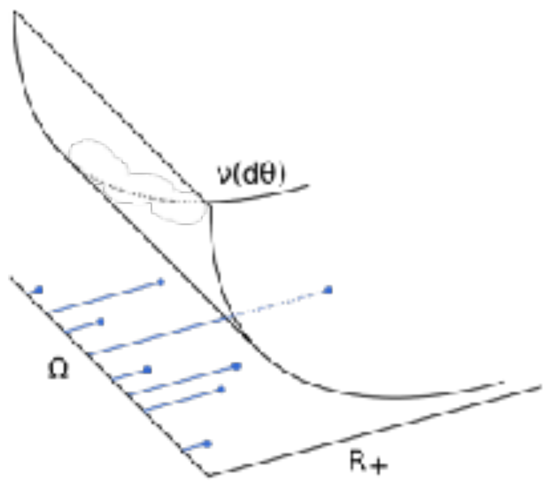
	Finite Approximation	Approximation Error Bounds	Computational Complexity
DP	✓ [Sethuraman 94] [Roychowdhury 15]	✓ [Ishwaran 01]	✓
BP	✓ [Teh 07] [Paisley 12] [Thibaux 07]	✓ [Doshi-Velez 09] [Paisley 1[illegible]]	✓
BPP	✓ [Broderick 1[illegible]]		
ΓP	✓ [Bondesson 82] [Roychowdhury 15]	✓ [Roychowdhury 15]	✓
(N)CRM	✓ [Broderick 1[illegible]]		

Sparse results for a few priors in BNP

No general theory

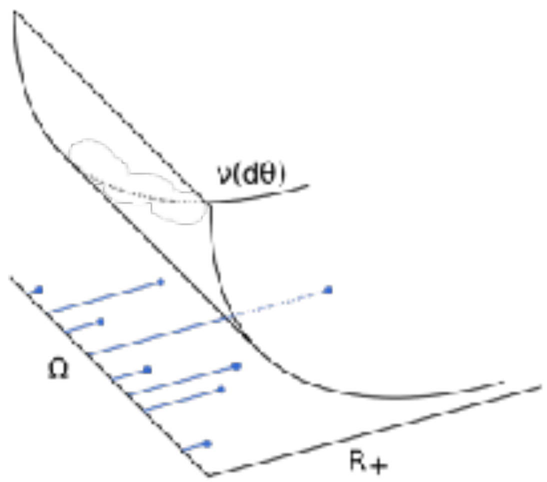
Truncation Roadmap

Truncation Roadmap



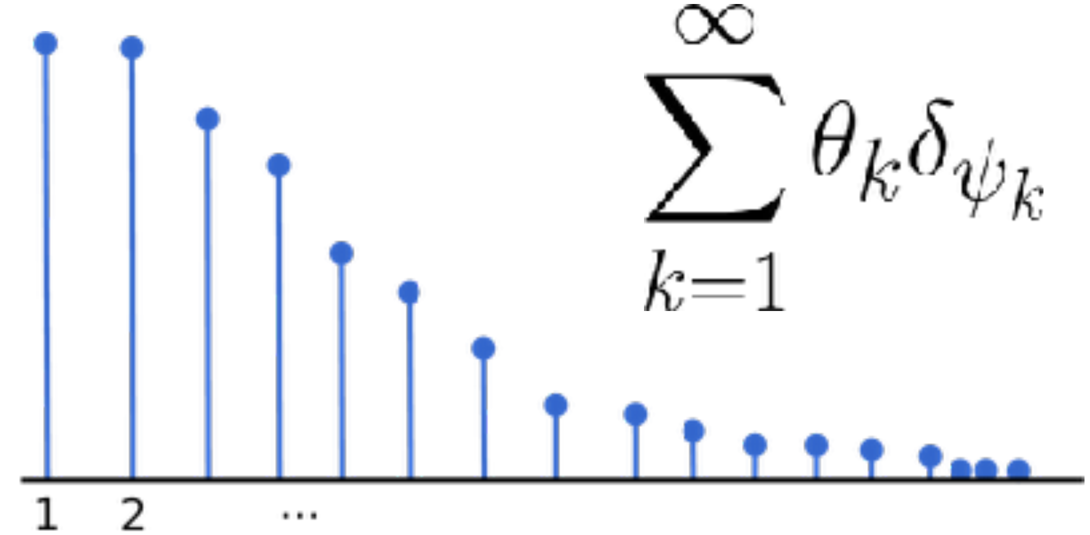
Tractable models
in BNP

Truncation Roadmap



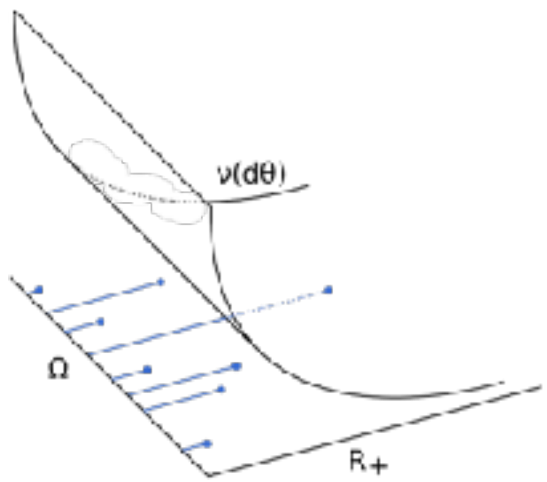
Tractable models
in BNP

two forms for sequential
representations



$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

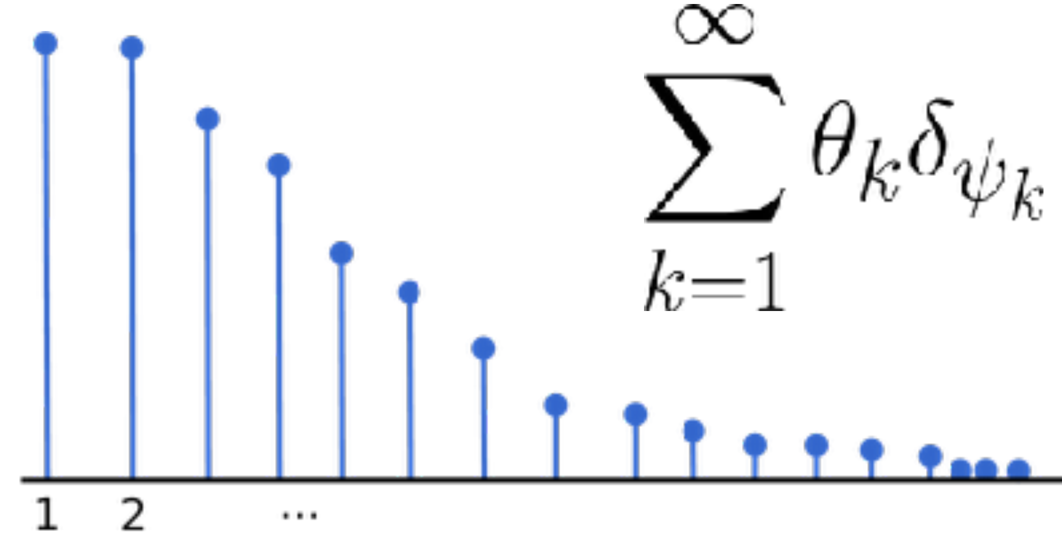
Truncation Roadmap



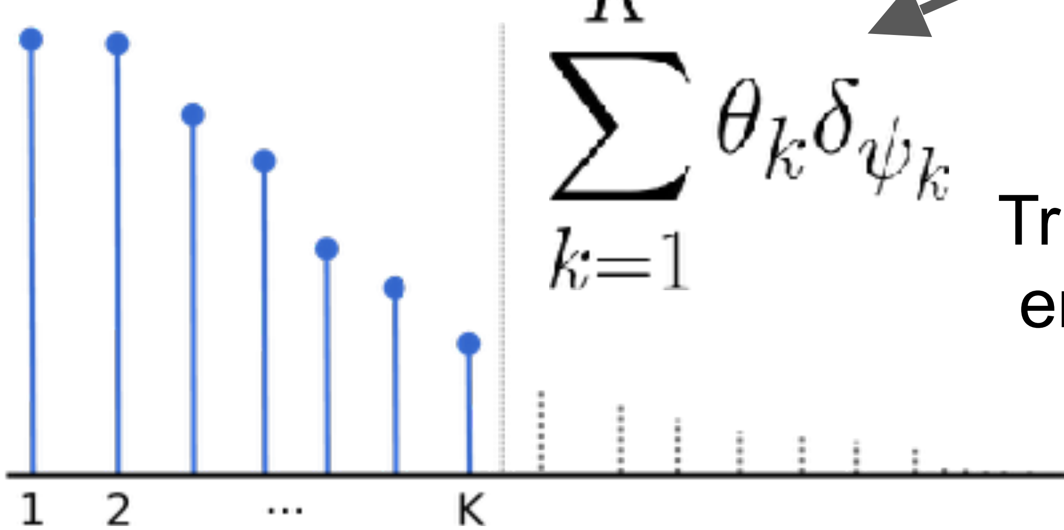
Tractable models
in BNP

two forms for sequential
representations

$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

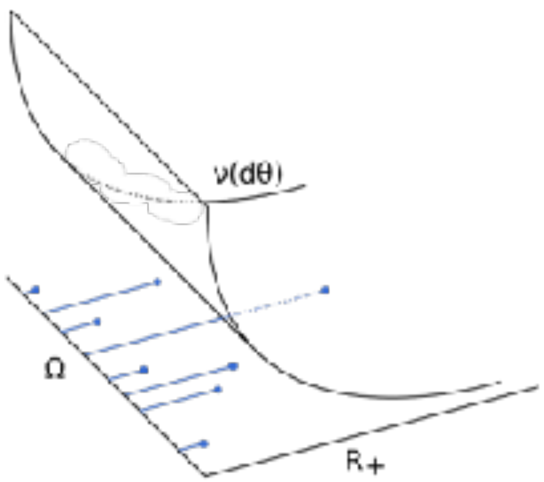


$$\sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$



Truncation and
error analysis

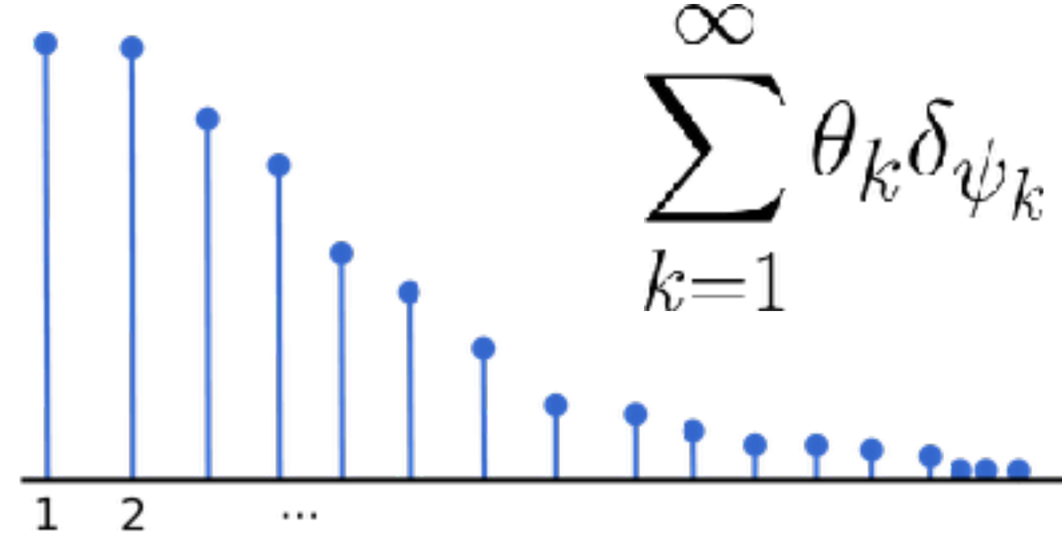
Truncation Roadmap



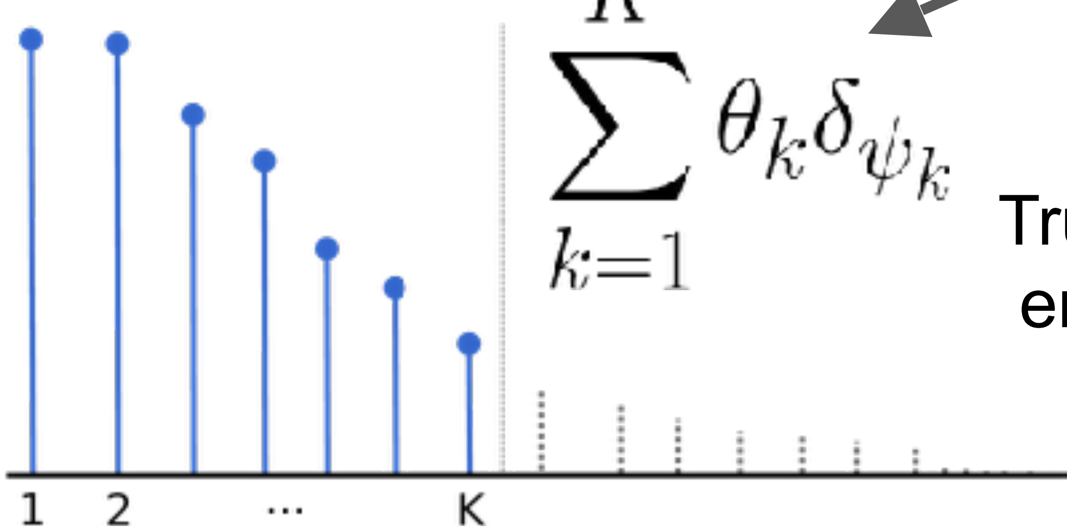
Tractable models in BNP

two forms for sequential representations

$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$



$$\sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$



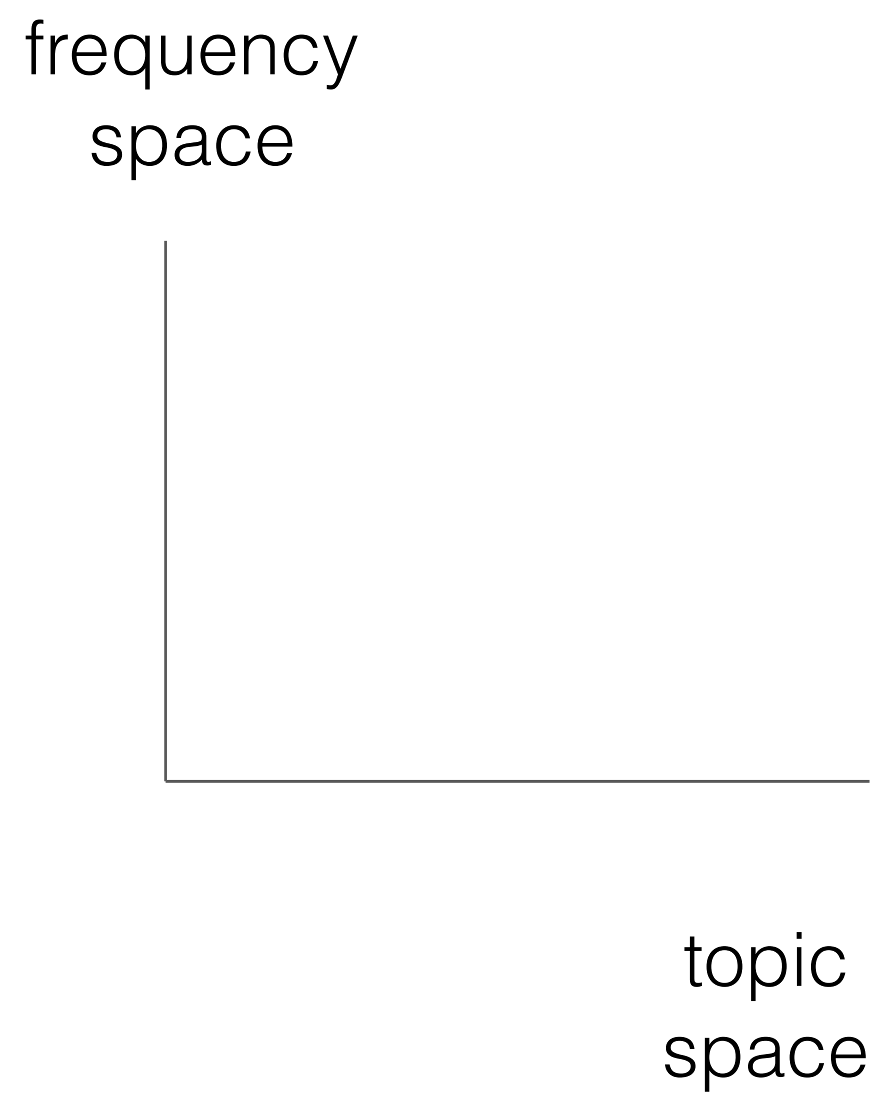
Truncation and error analysis

The Standard Model in BNP (By Example)

	sports	politics	food	…
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	

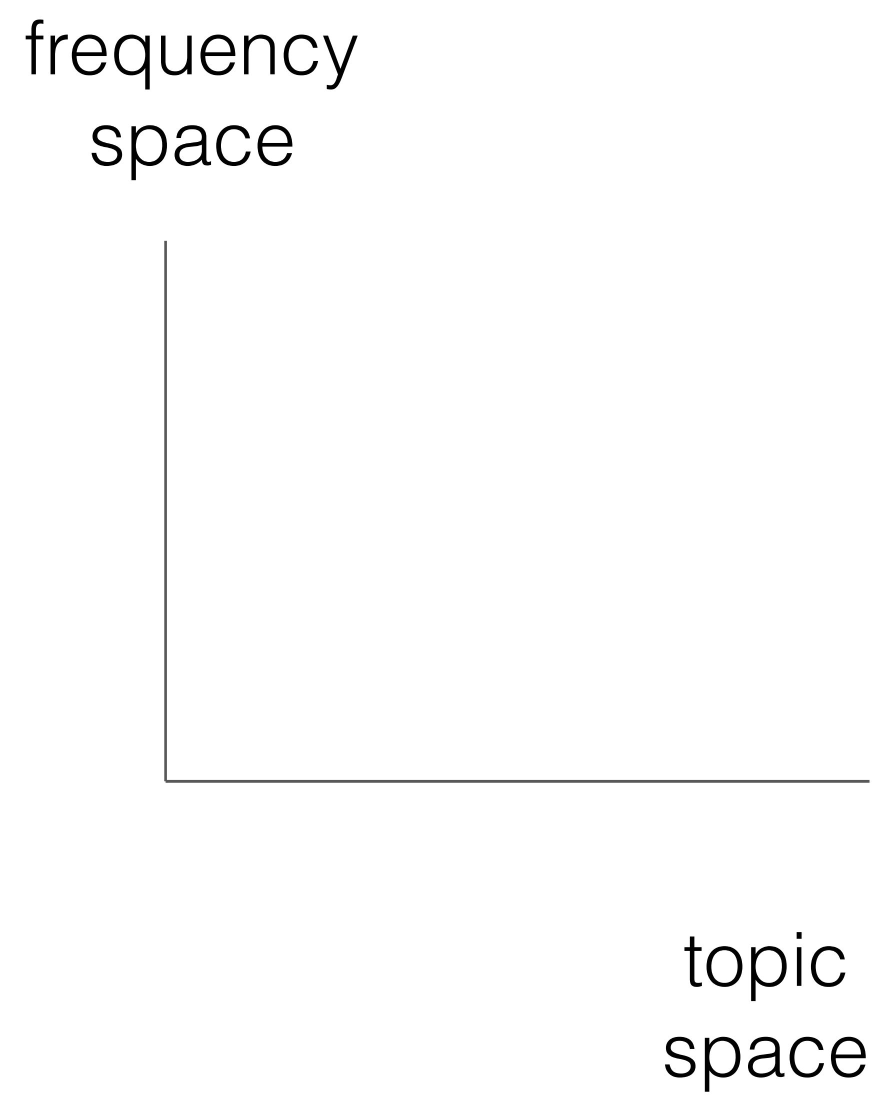
The Standard Model in BNP (By Example The New York Times)

	sports	politics	food	...
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	

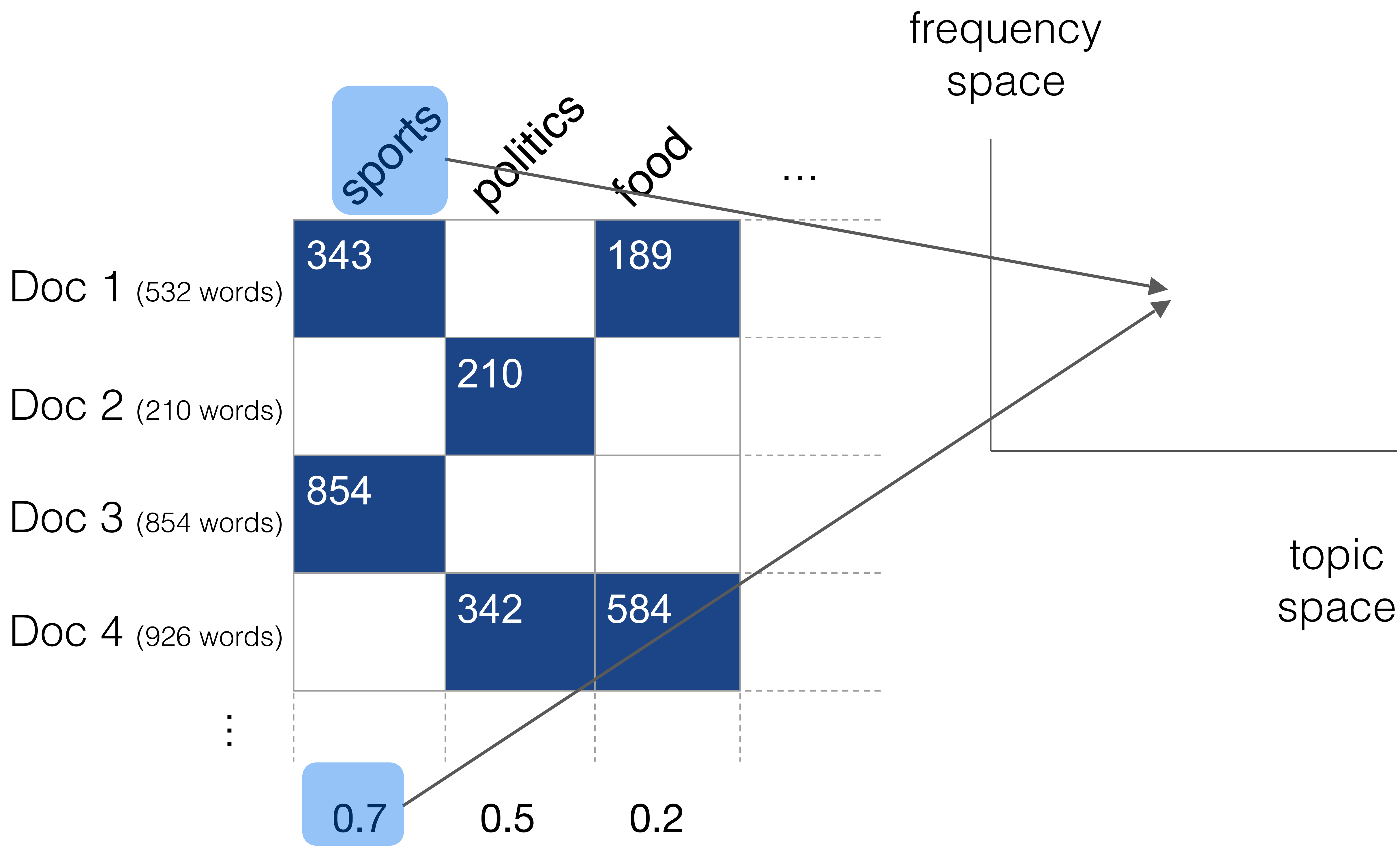


The Standard Model in BNP (By Example The New York Times)

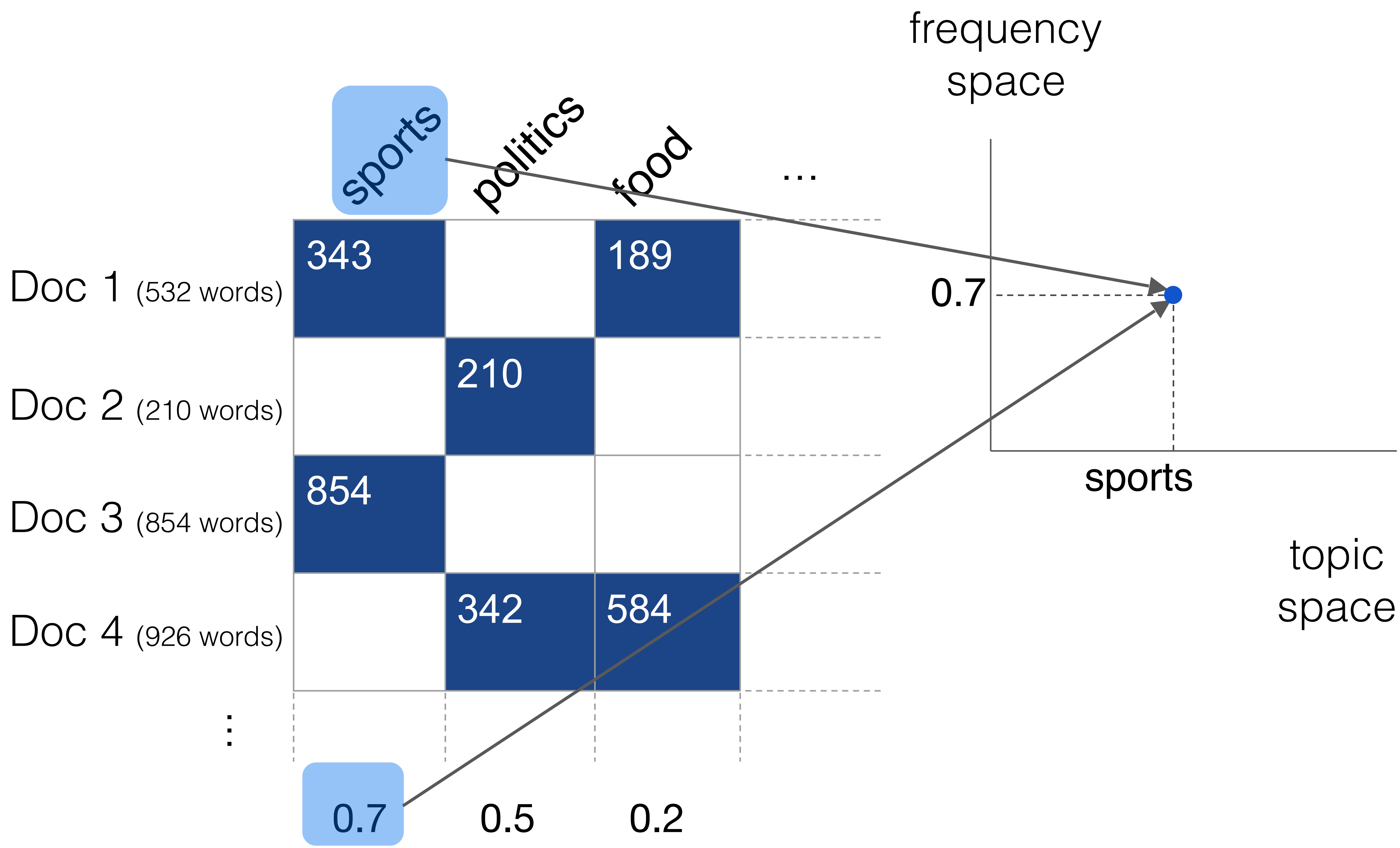
	sports	politics	food	...
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



The Standard Model in BNP (By Example)

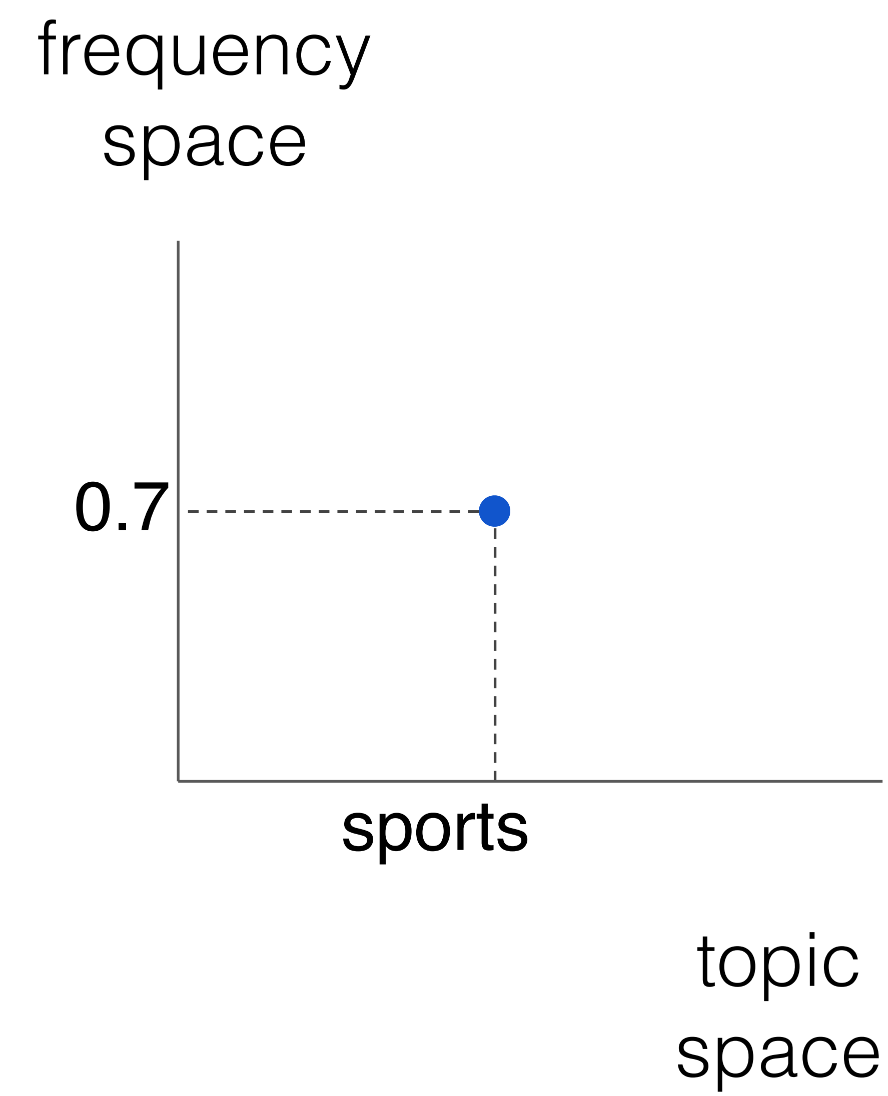


The Standard Model in BNP (By Example The New York Times)



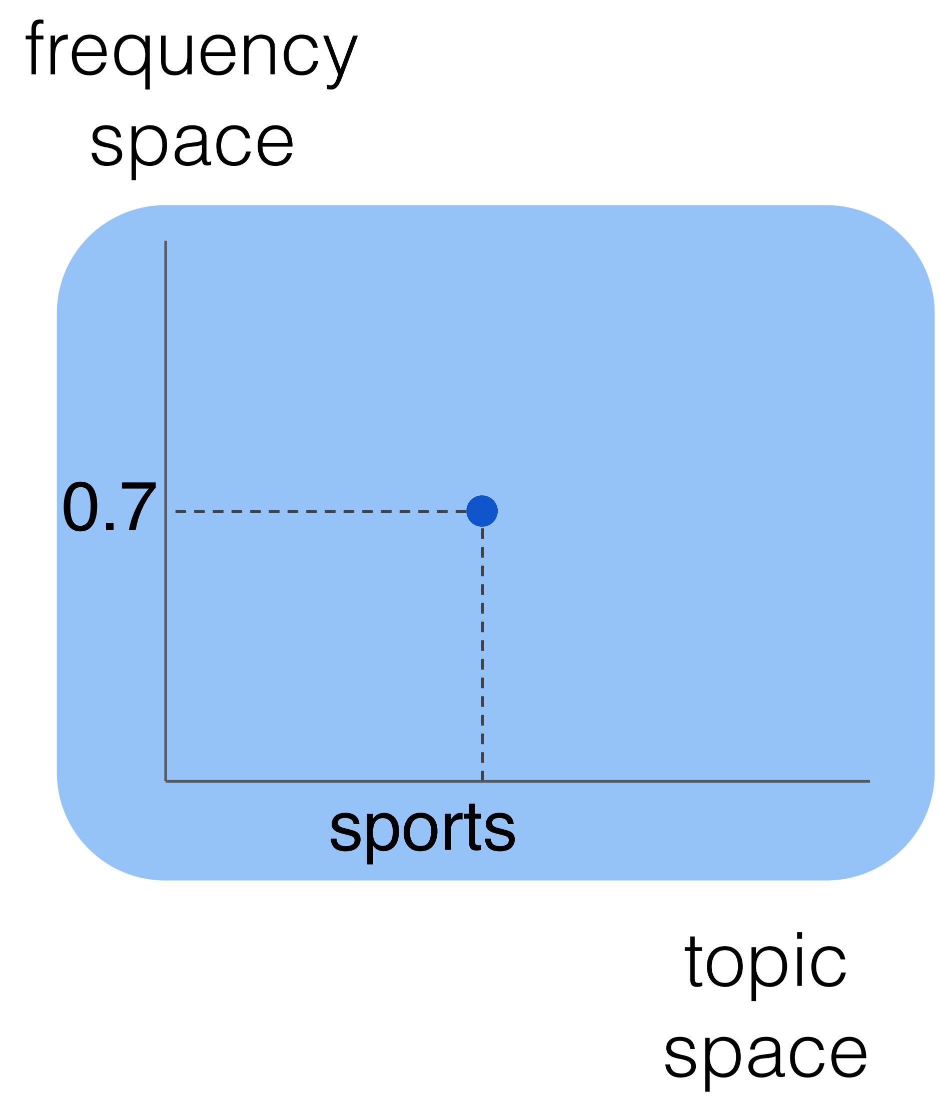
The Standard Model in BNP (By Example)

	sports	politics	food	…
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



The Standard Model in BNP (By Example The New York Times)

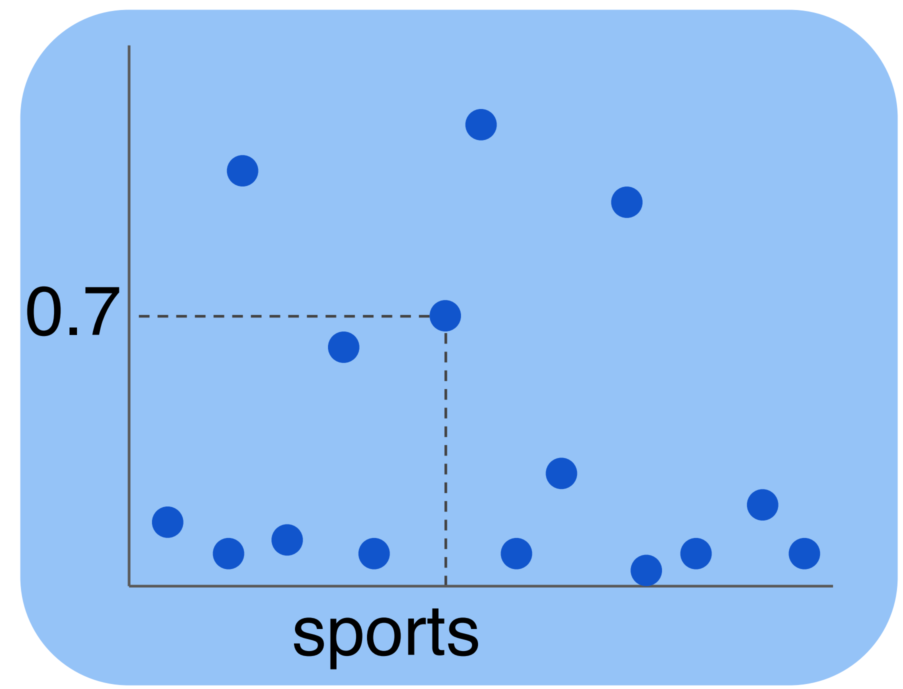
	sports	politics	food	...
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



The Standard Model in BNP (By Example)

	sports	politics	food	…
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	

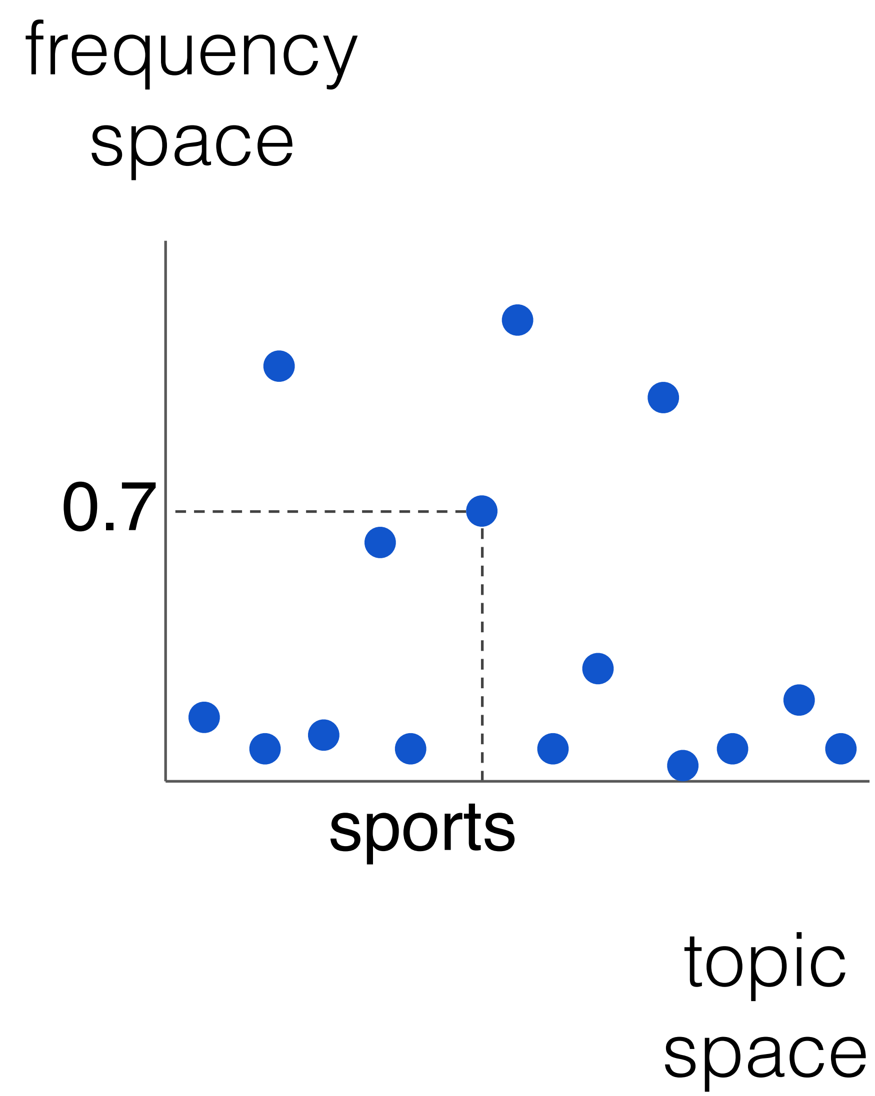
frequency space



topic space

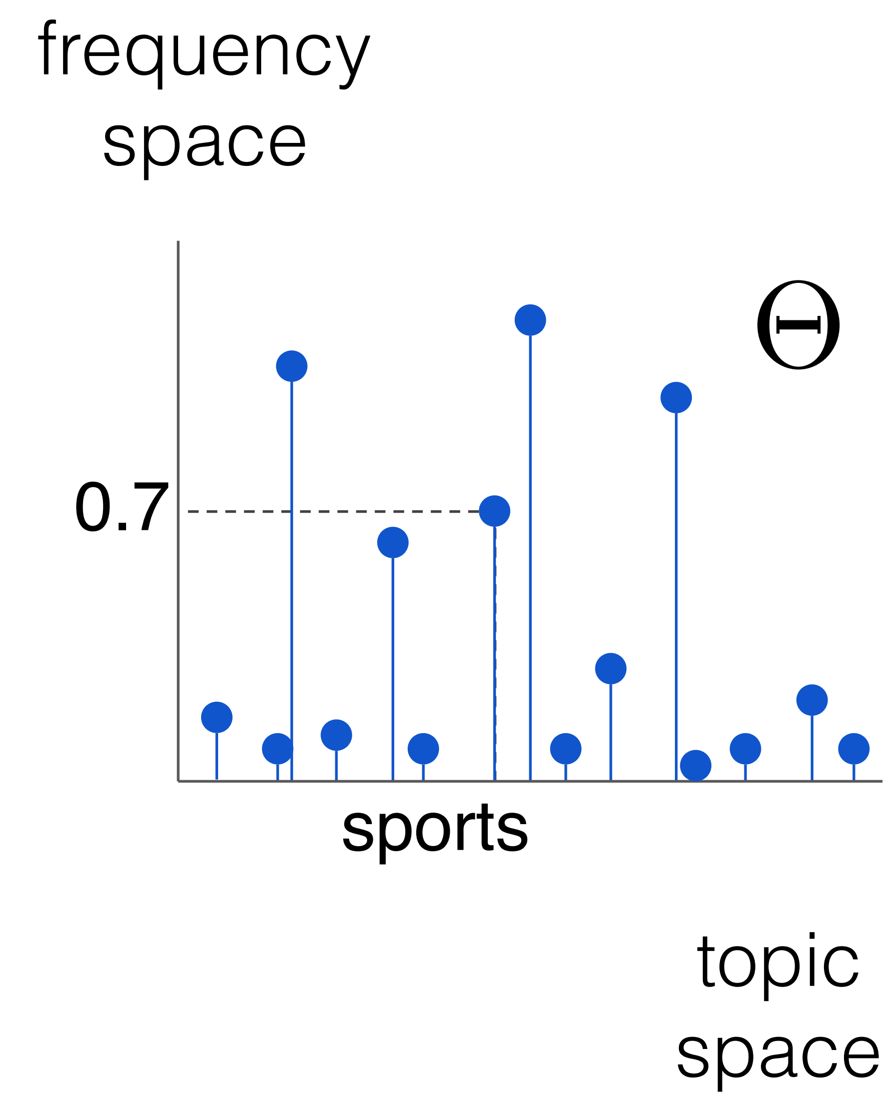
The Standard Model in BNP (By Example)

	sports	politics	food	…
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



The Standard Model in BNP (By Example The New York Times)

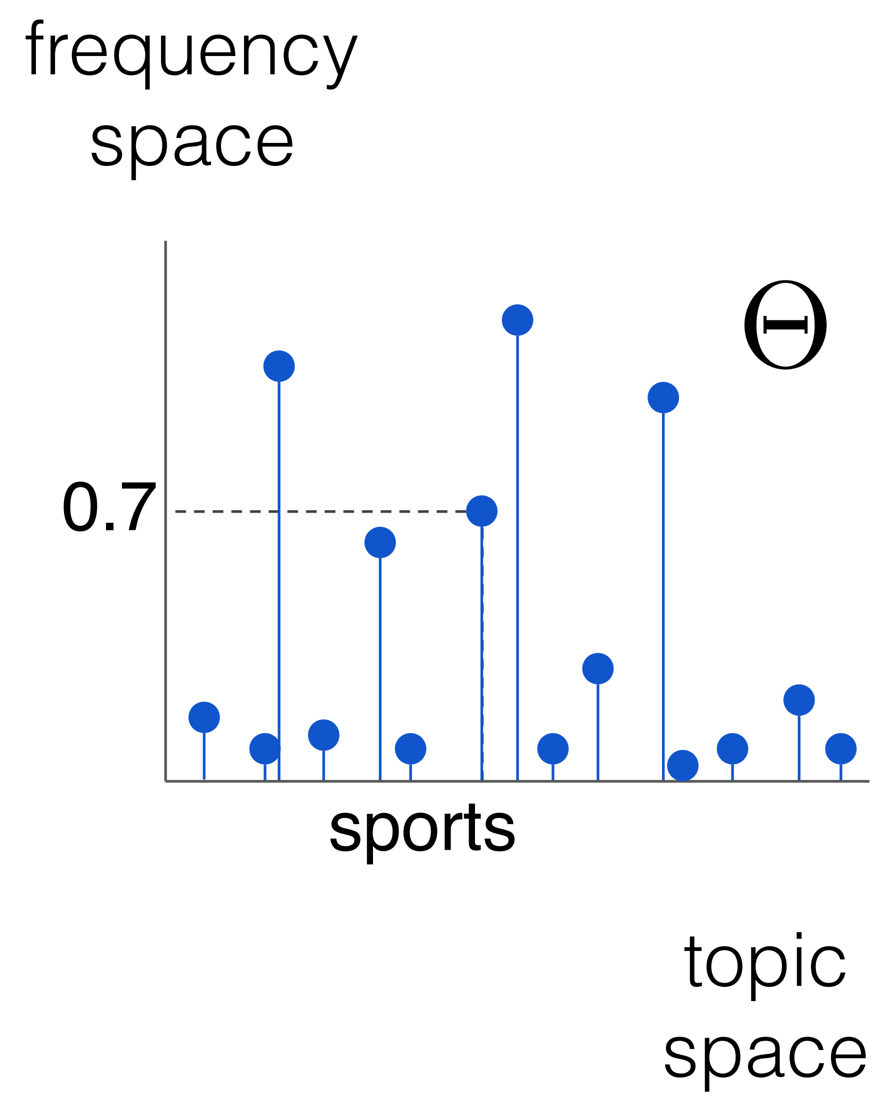
	sports	politics	food	...
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



Θ is a ***random discrete measure on the topics***

The Standard Model in BNP (By Example The New York Times)

	sports	politics	food	...
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



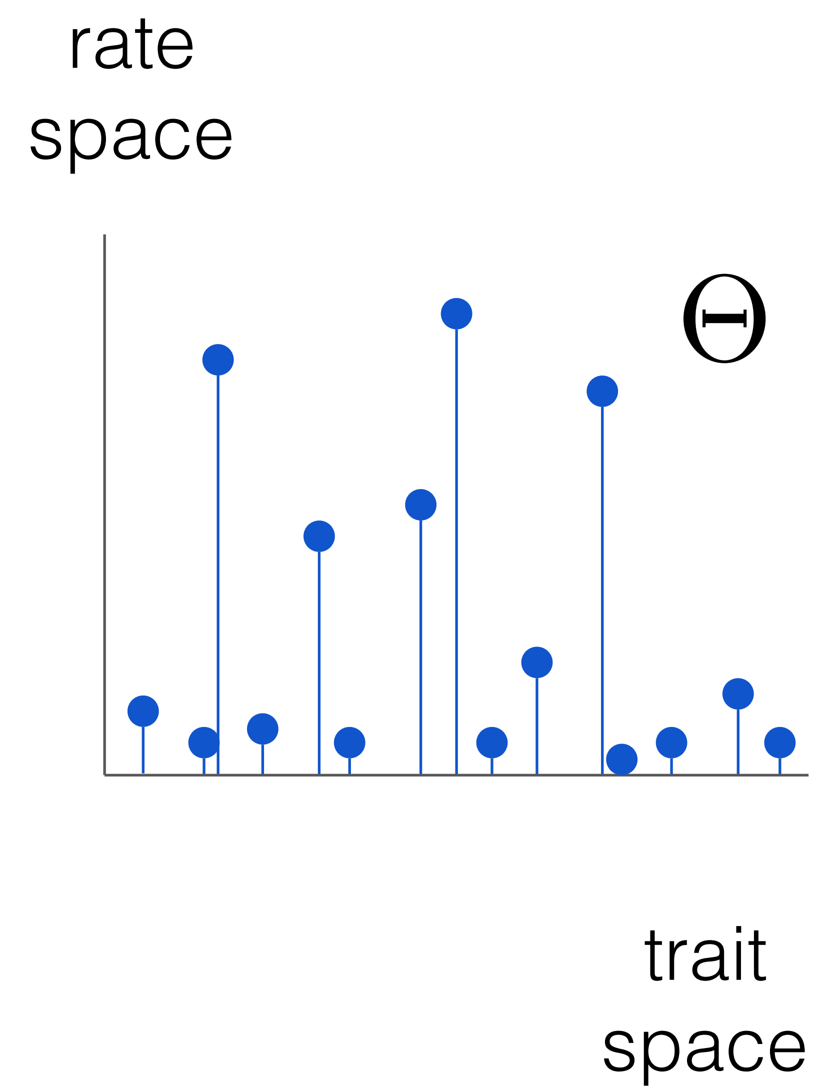
Θ is a ***random discrete measure on the topics***

The Standard Model in BNP (By Example)

“traits”

	ψ_1	ψ_2	ψ_3	…
Obs 1	343		189	
Obs 2		210		
Obs 3	854			
Obs 4		342	584	
⋮				
	θ_1	θ_2	θ_3	

“rates”

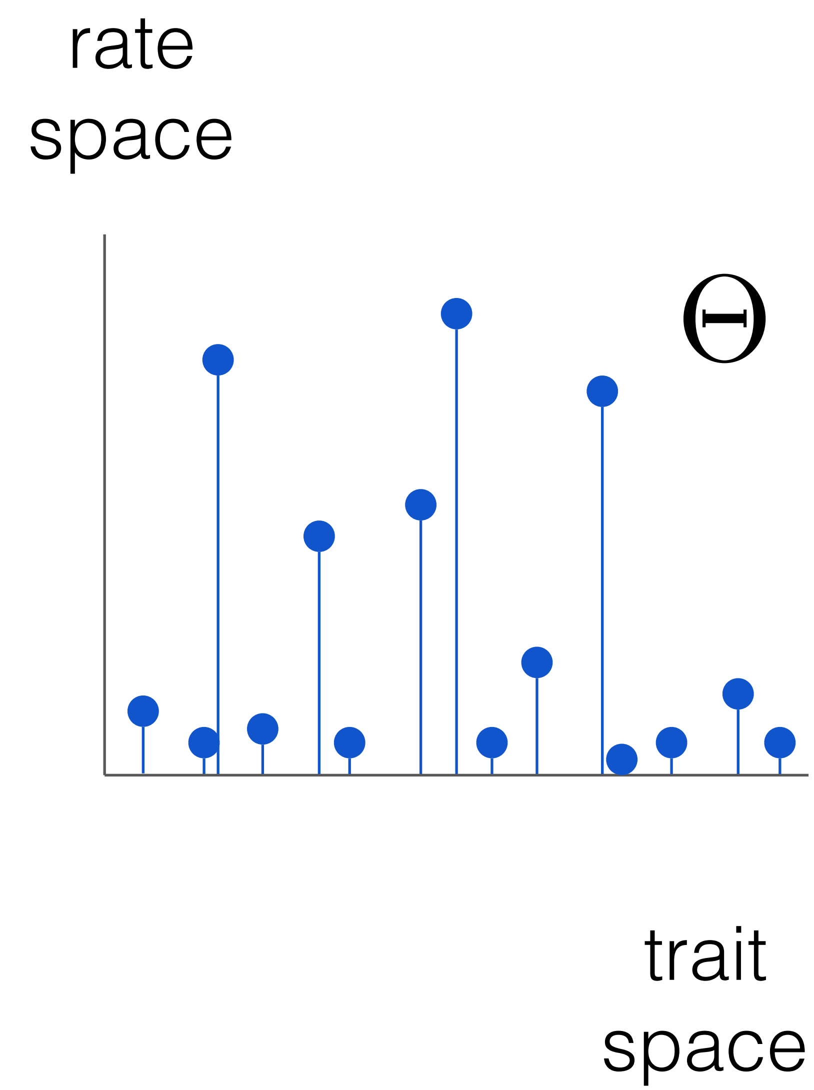


Θ is a ***random discrete measure on the ~~topics~~*** traits

The Standard Model in BNP (By Example The New York Times)

	ψ_1	ψ_2	ψ_3	*"traits"* …
Obs 1	343		189	
Obs 2		210		
Obs 3	854			
Obs 4		342	584	
⋮				
	θ_1	θ_2	θ_3	

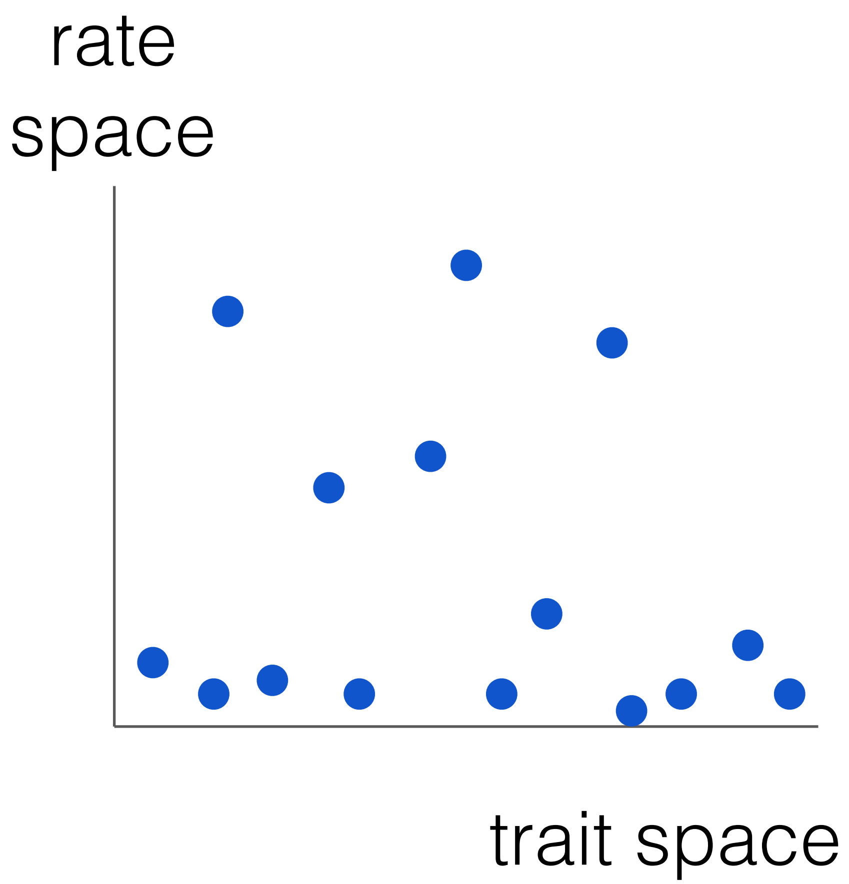
"rates"



Θ is a ***random discrete measure on the ~~topics~~*** traits

Poisson processes and (N)CRMs

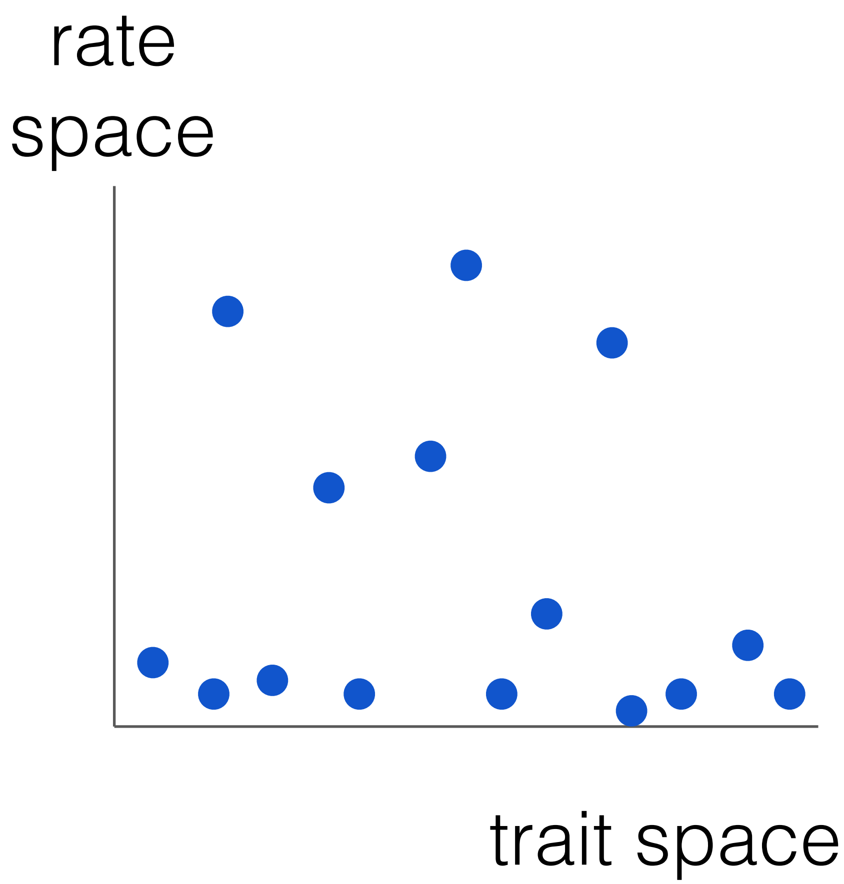
How do we generate infinitely many trait/rate points (ψ, θ)?



Poisson processes and (N)CRMs

How do we generate infinitely many trait/rate points (ψ, θ)?

Poisson point process with measure $\nu(d\theta \times d\psi)$:

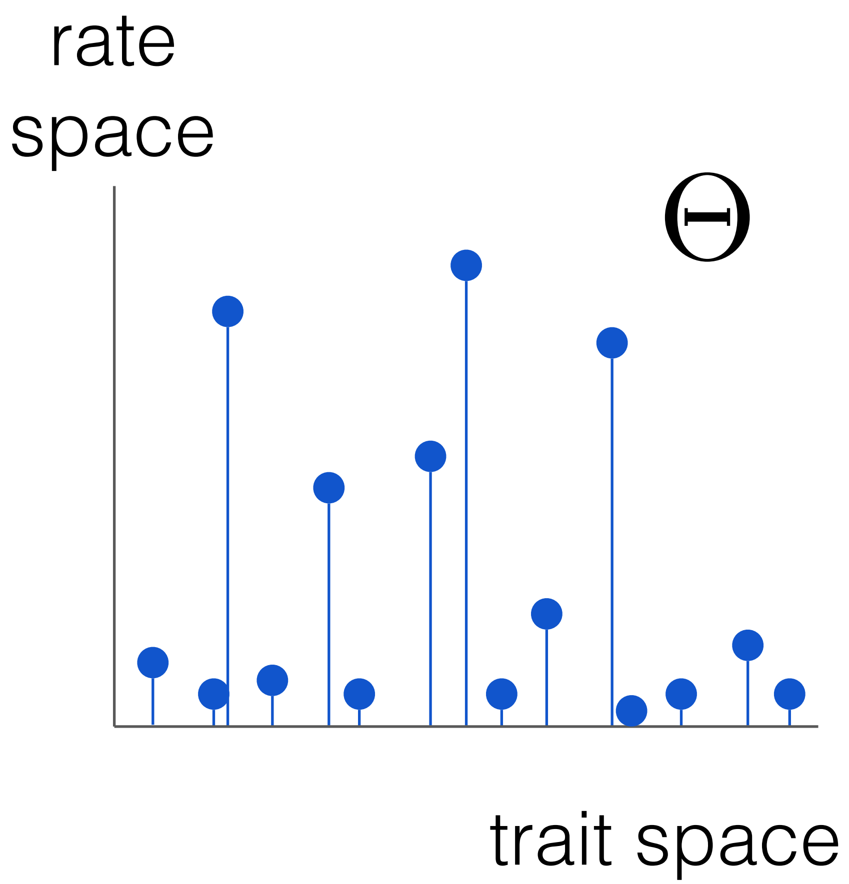


[Kingman 93]

Poisson processes and (N)CRMs

How do we generate infinitely many trait/rate points (ψ, θ)?

Poisson point process with measure $\nu(d\theta \times d\psi)$:

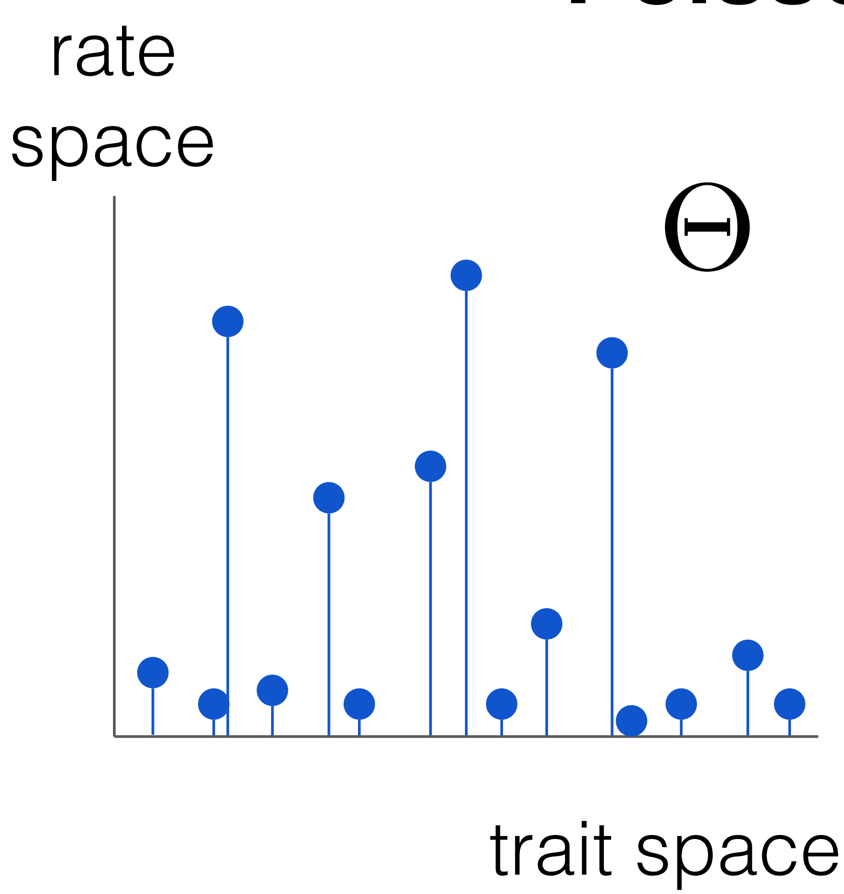


completely random measure (CRM)
(e.g. BP, ΓP)

[Kingman 93]

Poisson processes and (N)CRMs

How do we generate infinitely many trait/rate points (ψ, θ)?



Poisson point process with measure $\nu(d\theta \times d\psi)$:

completely random measure (CRM)
(e.g. BP, ΓP)

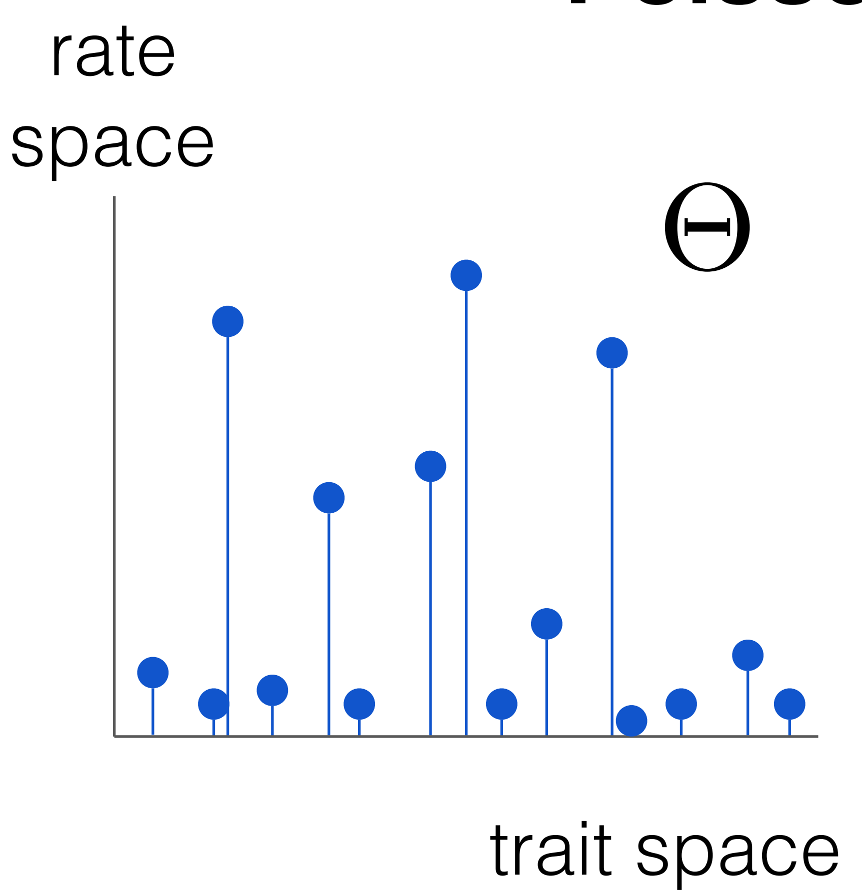
Normalize rates: **normalized CRM**
(NCRM) (e.g. DP)

[Kingman 93]

Poisson processes and (N)CRMs

How do we generate infinitely many trait/rate points (ψ, θ)?

Poisson point process with measure $\nu(d\theta \times d\psi)$:



completely random measure (CRM) (e.g. BP, ΓP)

Normalize rates: **normalized CRM** (NCRM) (e.g. DP)

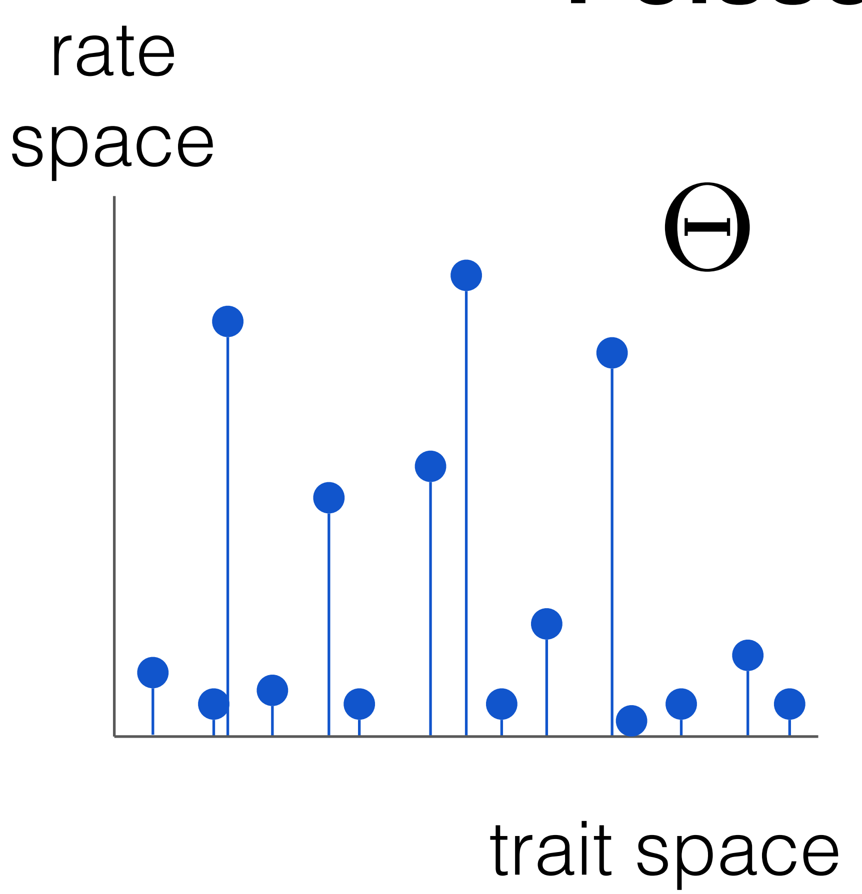
Captures a large class of useful priors in BNP

[Kingman 93]

Poisson processes and (N)CRMs

How do we generate infinitely many trait/rate points (ψ, θ)?

Poisson point process with measure $\nu(d\theta \times d\psi)$:



completely random measure (CRM) (e.g. BP, ΓP)

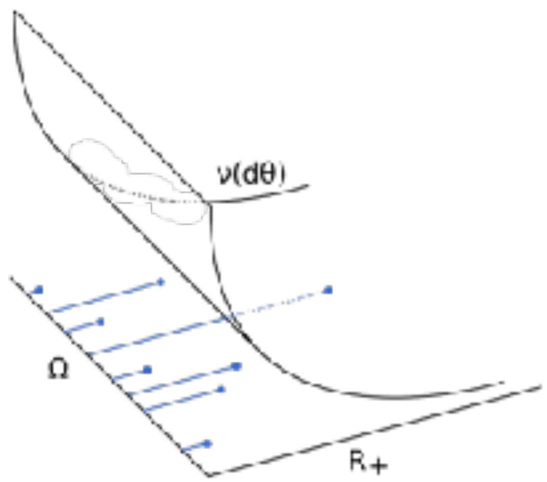
Normalize rates: **normalized CRM** (NCRM) (e.g. DP)

Captures a large class of useful priors in BNP

How do we pick a finite subset of the points?

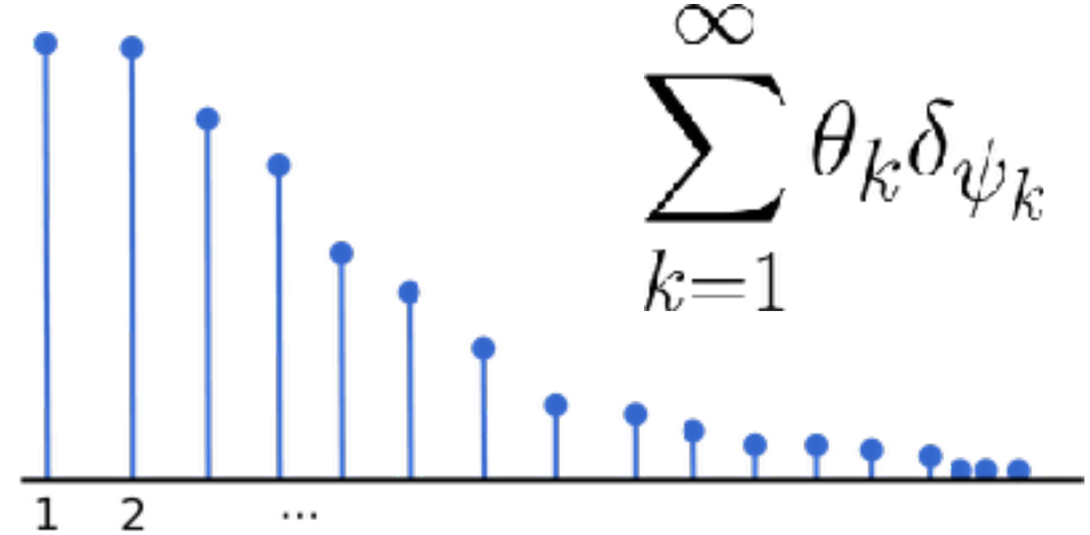
[Kingman 93]

Truncation Roadmap

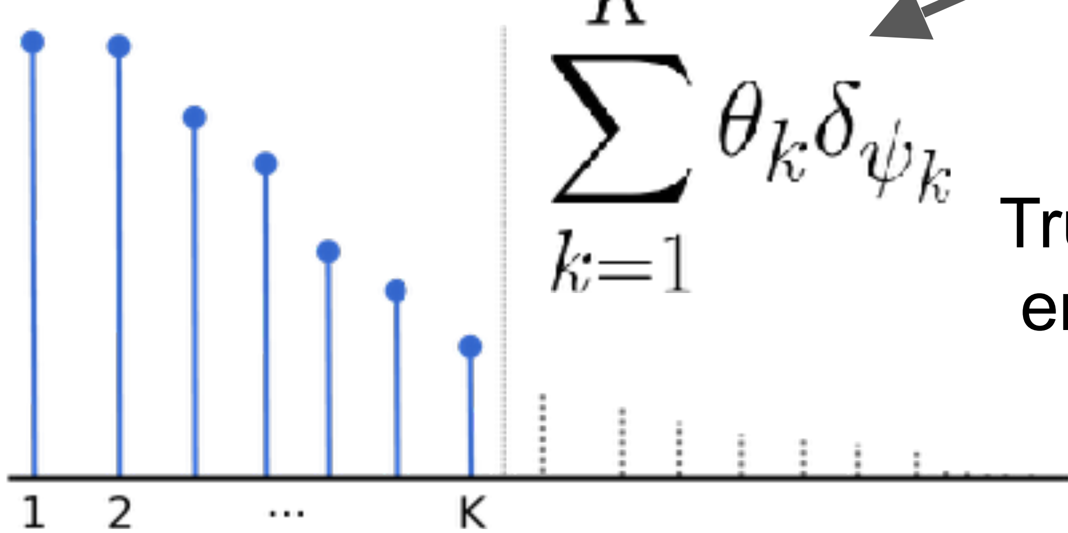


Tractable models
in BNP

two forms for sequential
representations



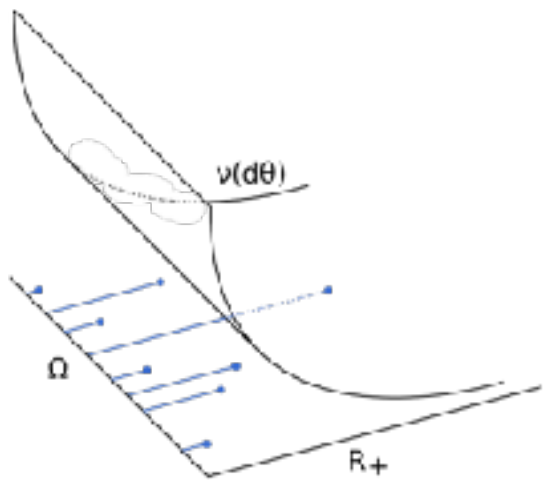
$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$



$$\sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

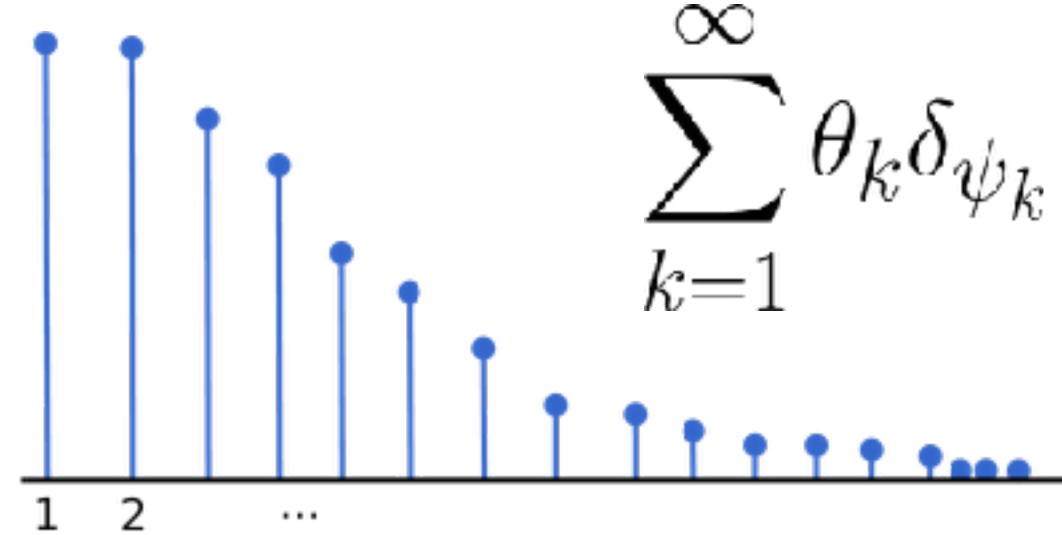
Truncation and
error analysis

Truncation Roadmap

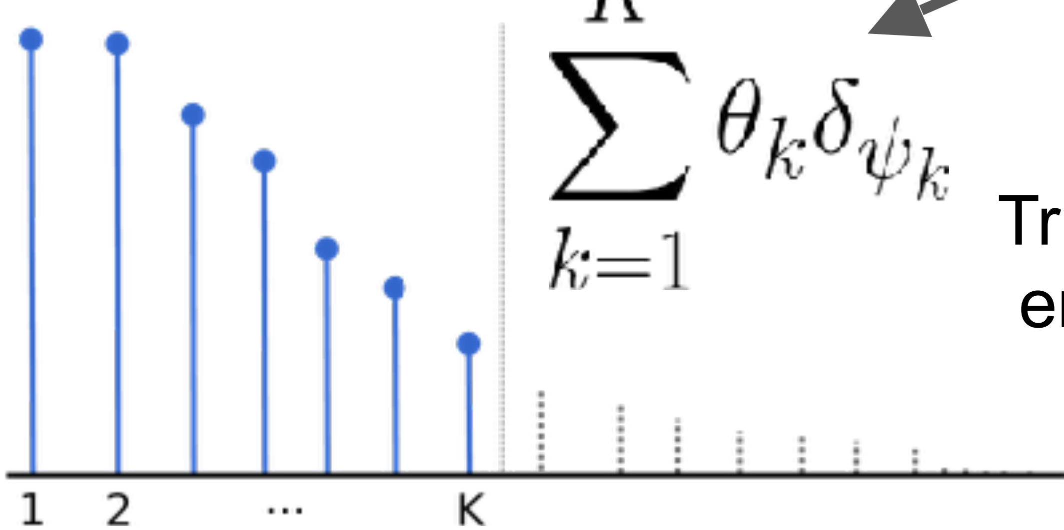


Tractable models in BNP

two forms for sequential representations



$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

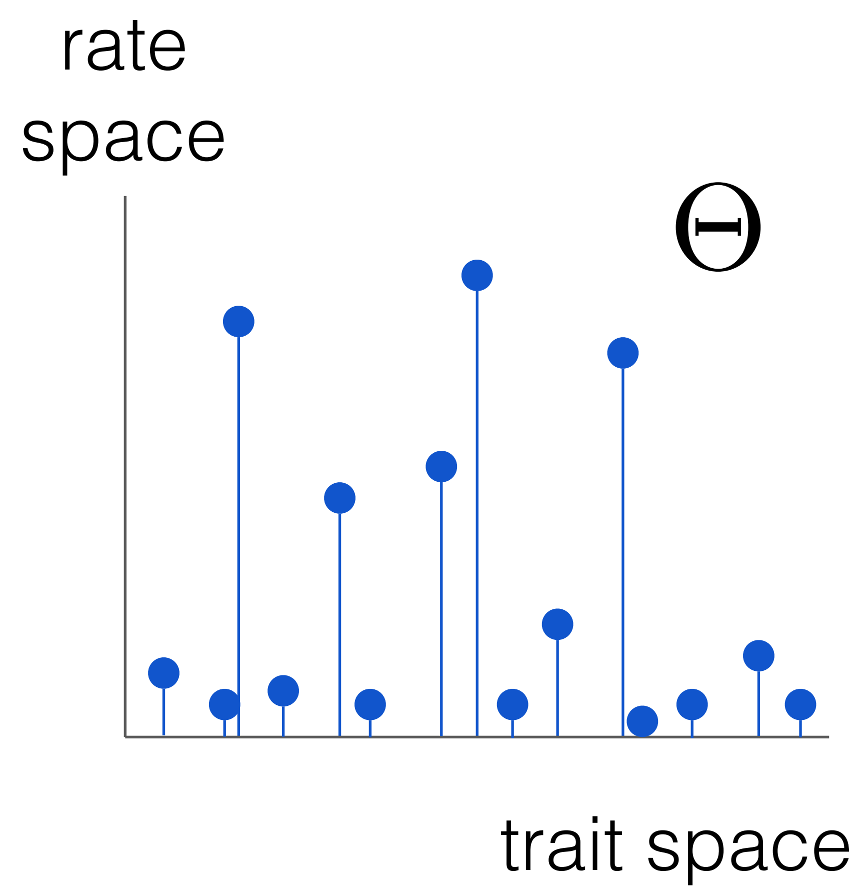


$$\sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

Truncation and error analysis

Sequential representation & truncation

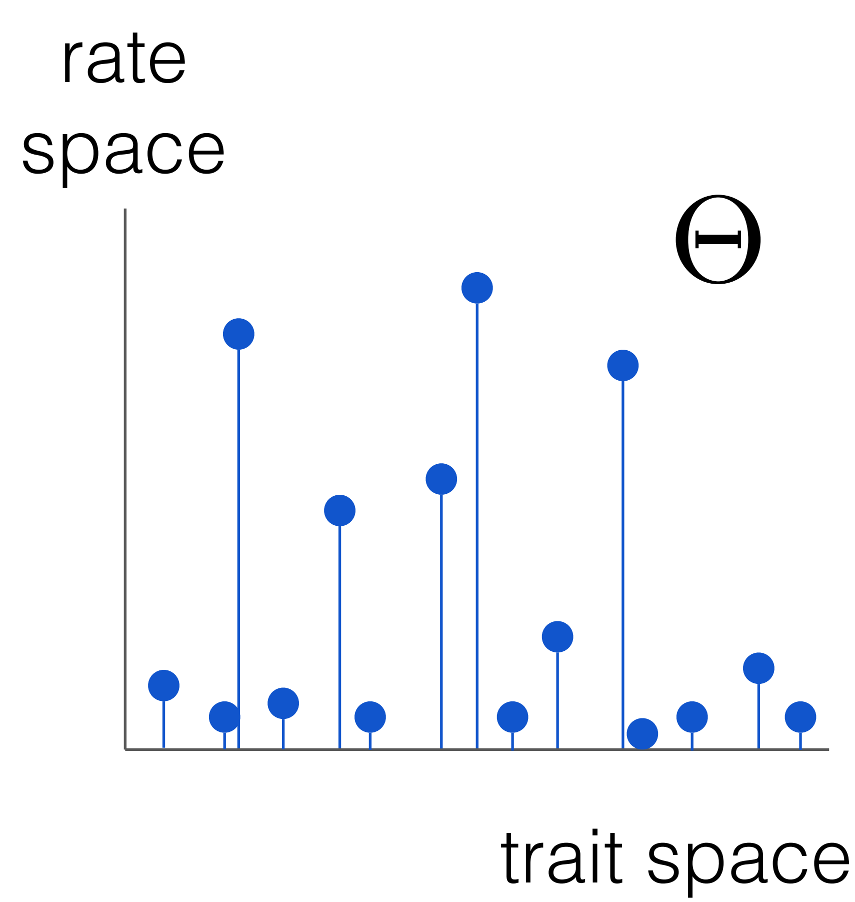
We pick a finite subset of atoms (ψ, θ) by:



Sequential representation & truncation

We pick a finite subset of atoms (ψ, θ) by:

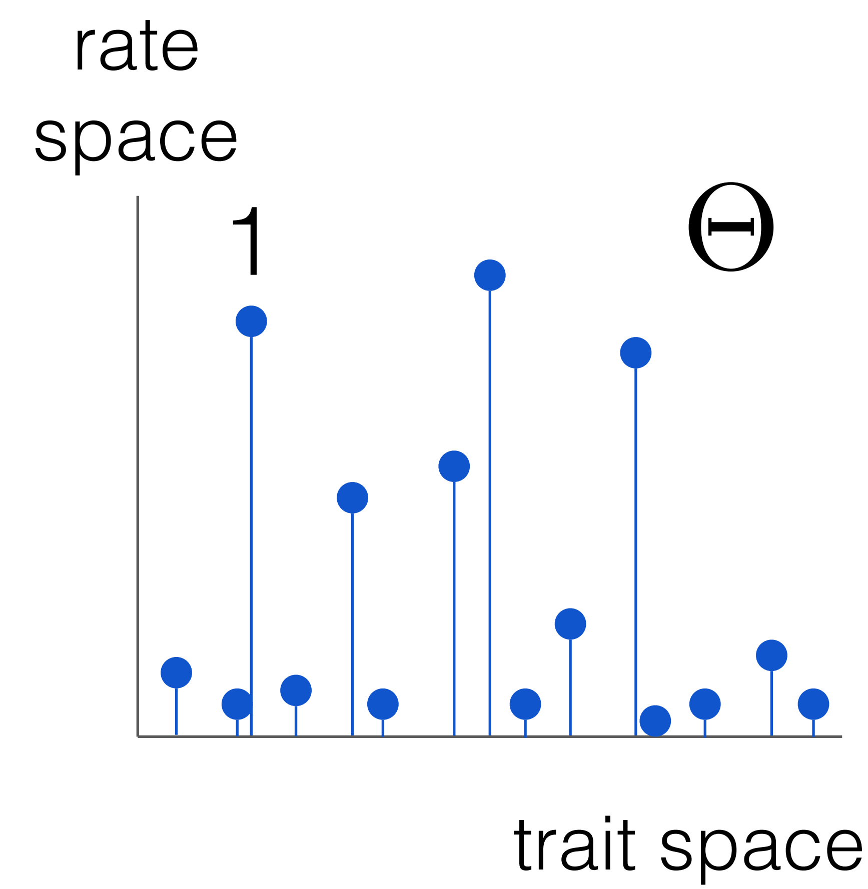
1) ordering the atoms **(sequential representation)**



Sequential representation & truncation

We pick a finite subset of atoms (ψ,θ) by:

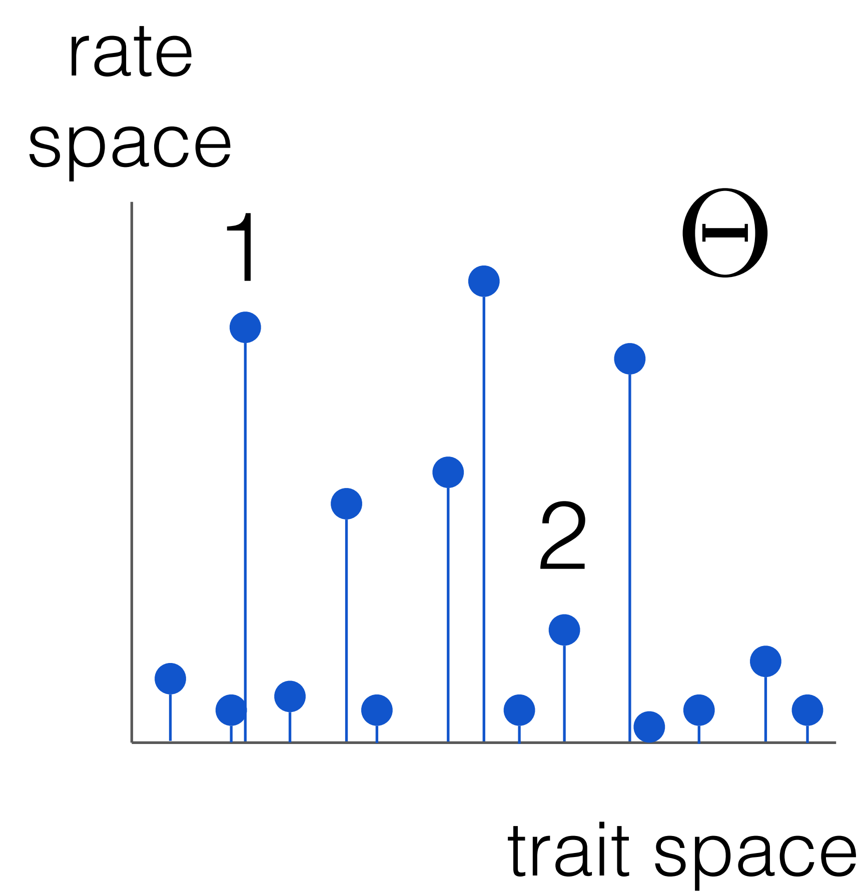
1) ordering the atoms **(sequential representation)**



Sequential representation & truncation

We pick a finite subset of atoms (ψ, θ) by:

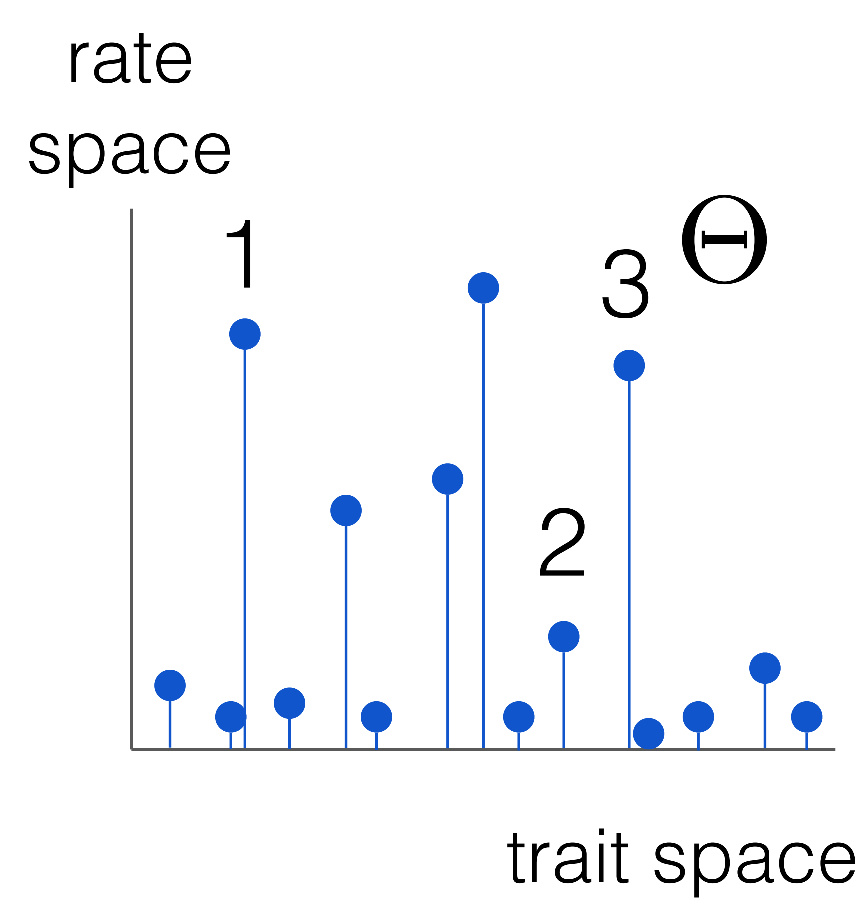
1) ordering the atoms **(sequential representation)**



Sequential representation & truncation

We pick a finite subset of atoms (ψ, θ) by:

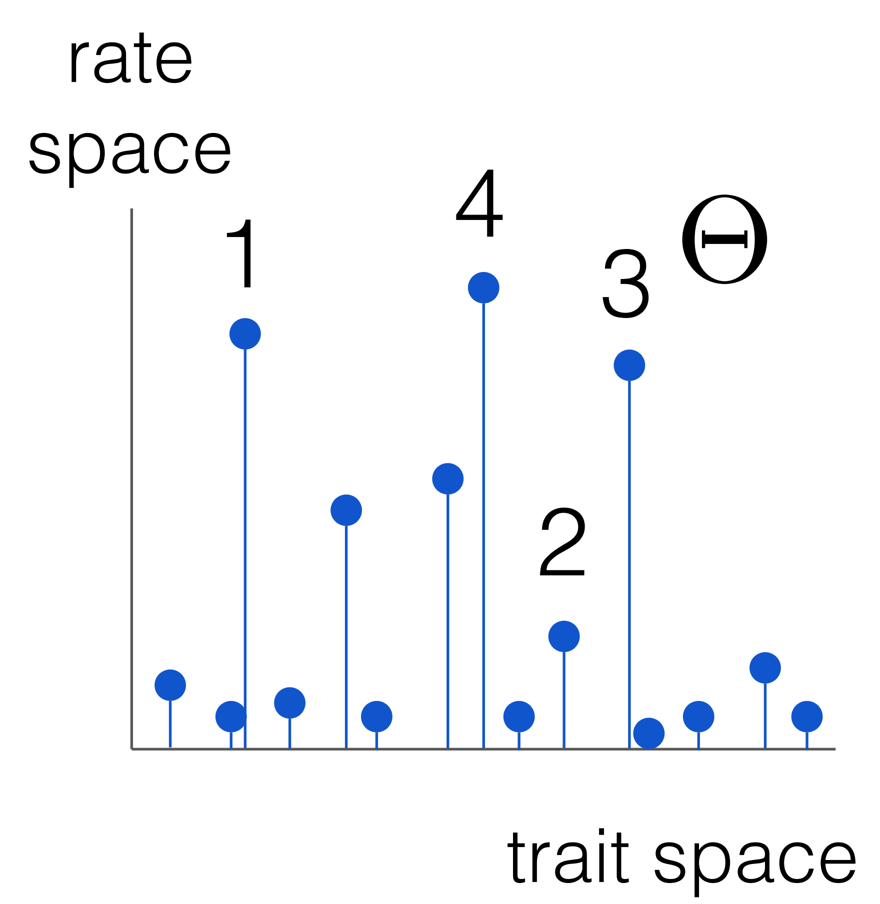
1) ordering the atoms **(sequential representation)**



Sequential representation & truncation

We pick a finite subset of atoms (ψ, θ) by:

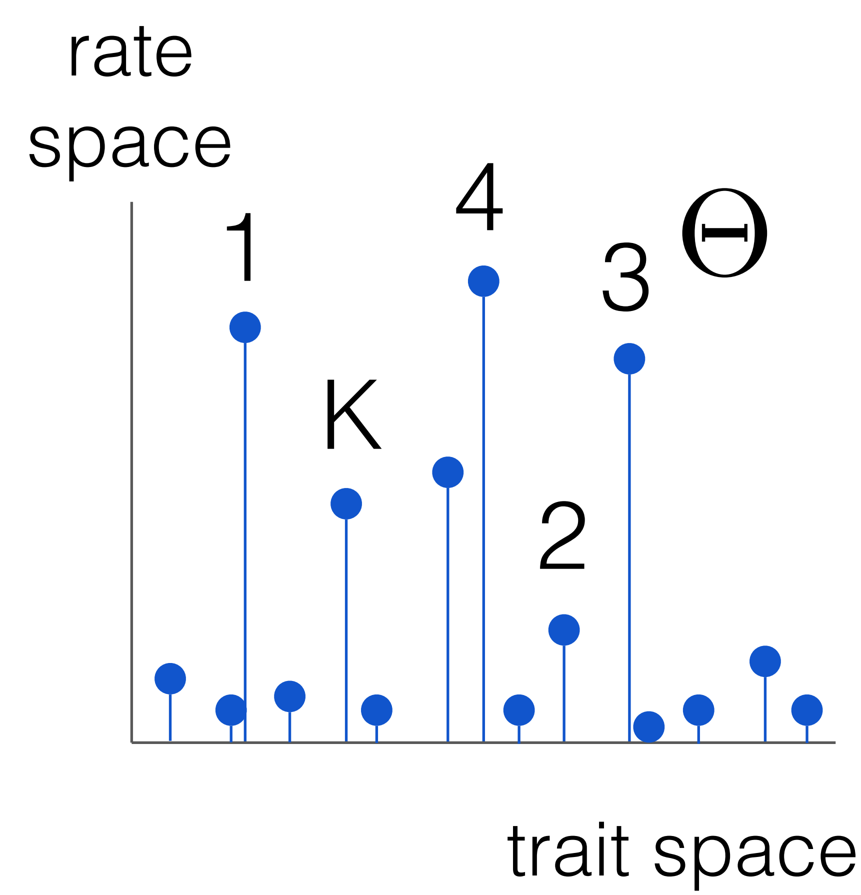
1) ordering the atoms **(sequential representation)**



Sequential representation & truncation

We pick a finite subset of atoms (ψ,θ) by:

1) ordering the atoms **(sequential representation)**

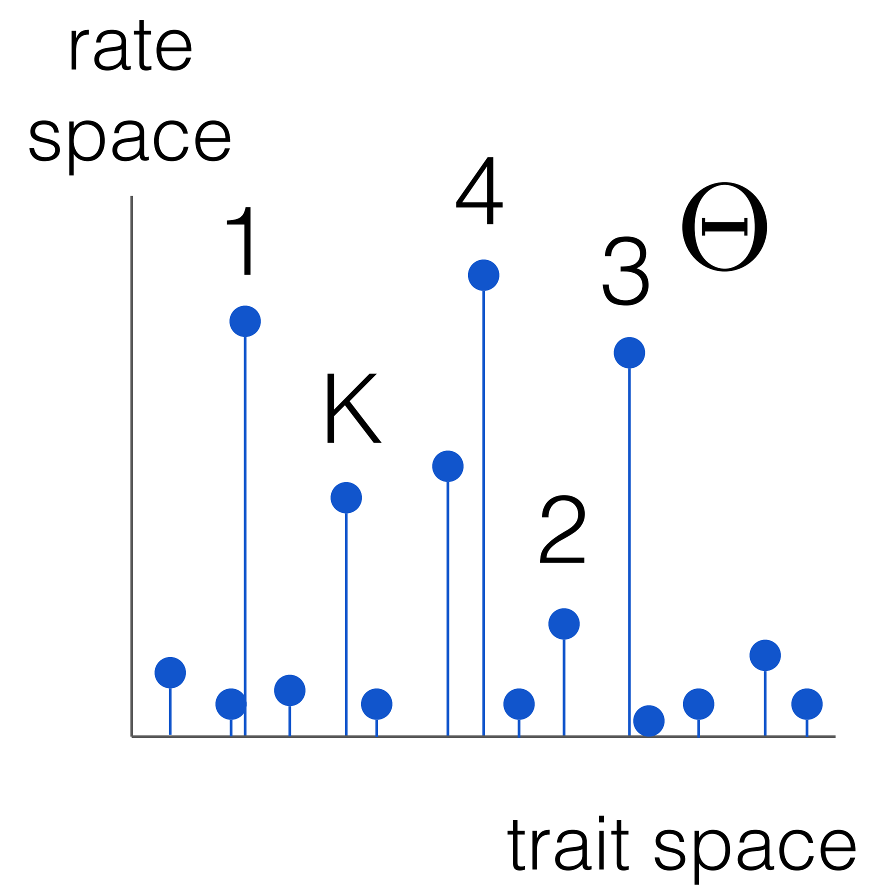


Sequential representation & truncation

We pick a finite subset of atoms (ψ,θ) by:

1) ordering the atoms **(sequential representation)**

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

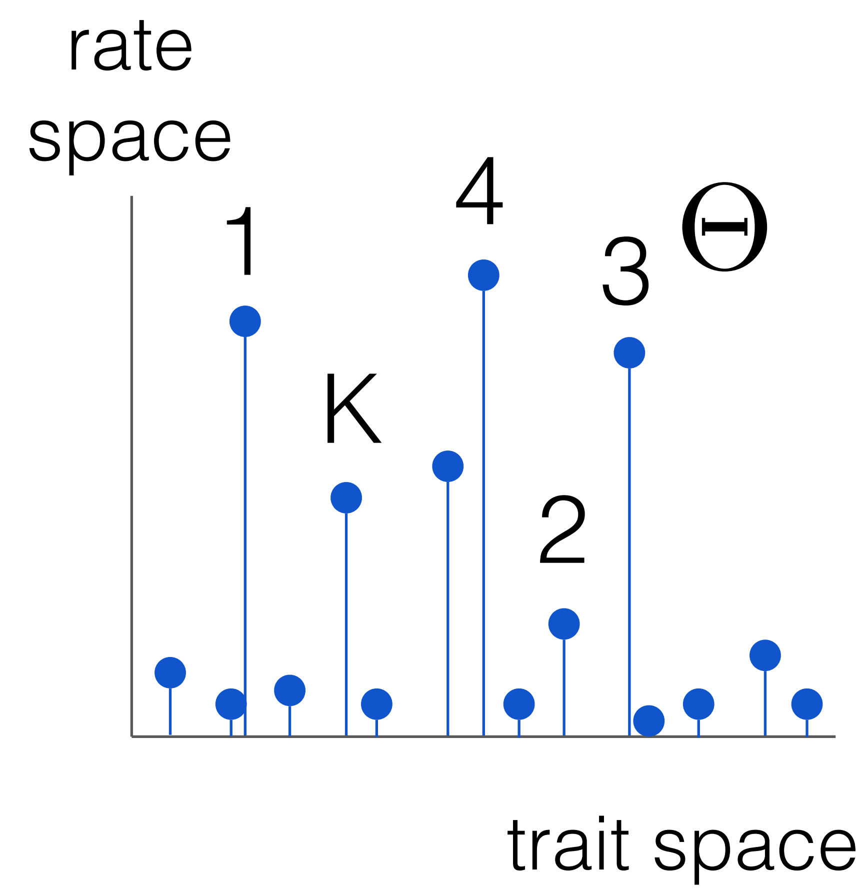


Sequential representation & truncation

We pick a finite subset of atoms (ψ,θ) by:

1) ordering the atoms **(sequential representation)**
2) removing any atoms beyond the K-th **(truncation)**

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

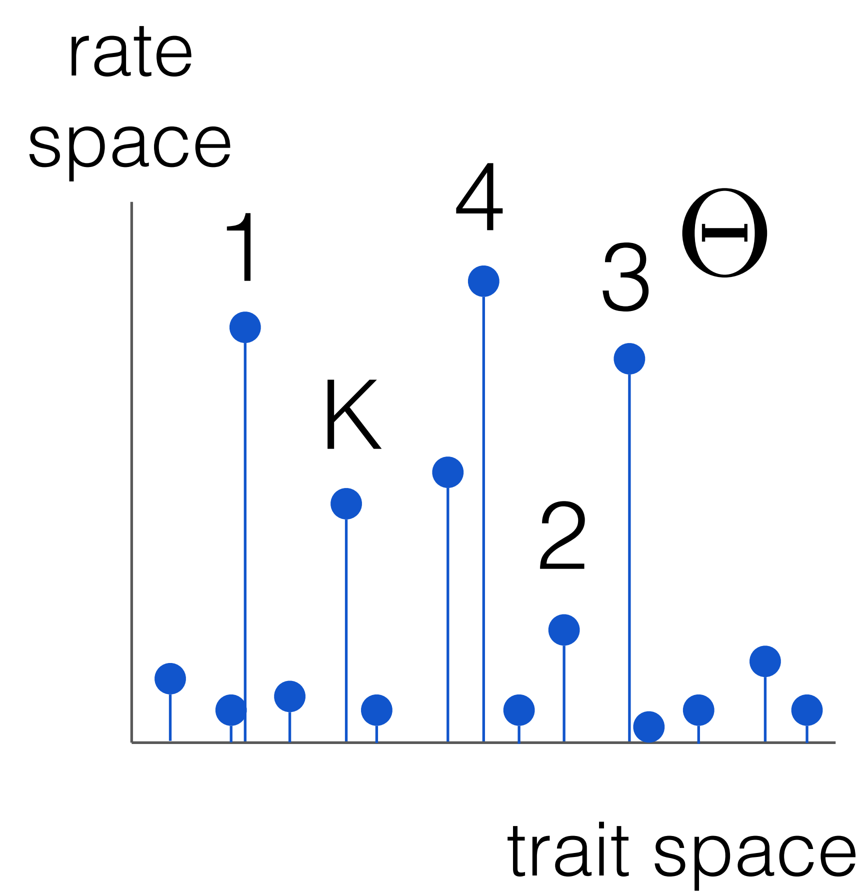


Sequential representation & truncation

We pick a finite subset of atoms (ψ,θ) by:

1) ordering the atoms **(sequential representation)**
2) removing any atoms beyond the K-th **(truncation)**

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

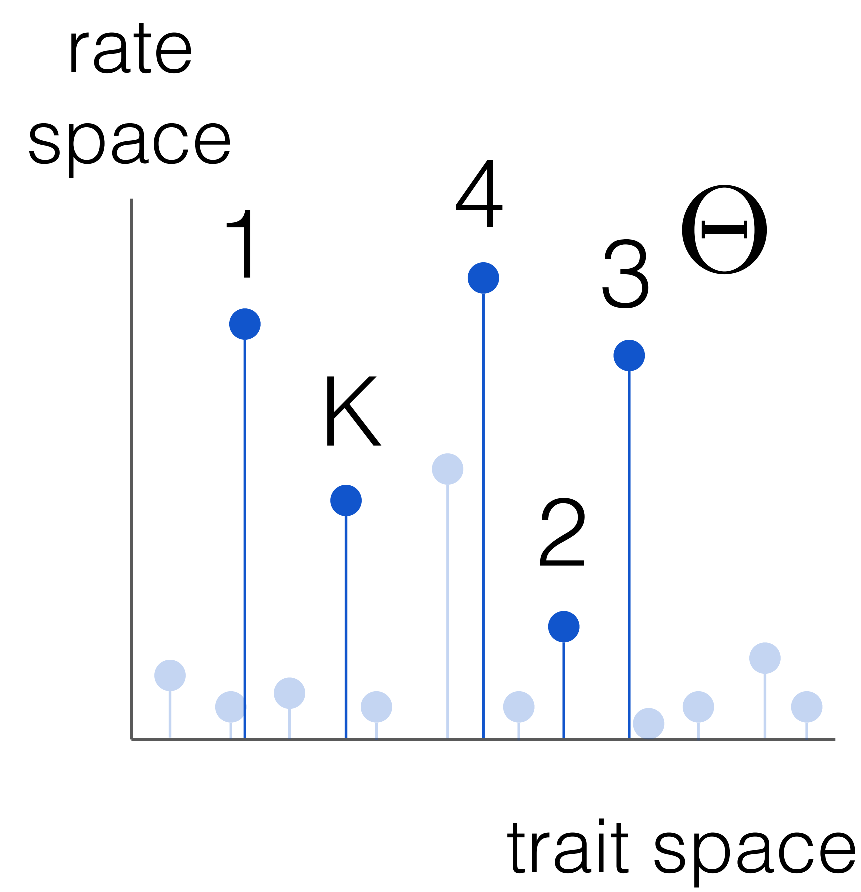


Sequential representation & truncation

We pick a finite subset of atoms (ψ,θ) by:

1) ordering the atoms **(sequential representation)**
2) removing any atoms beyond the K-th **(truncation)**

$$\Theta = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

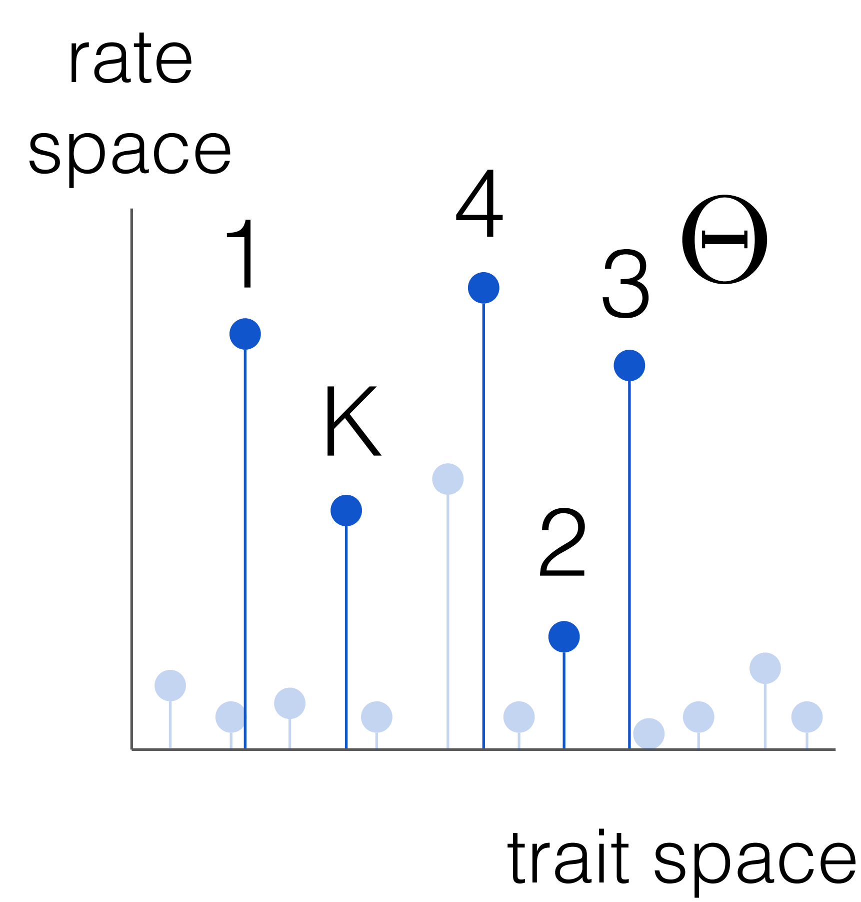


Sequential representation & truncation

We pick a finite subset of atoms (ψ,θ) by:

1) ordering the atoms **(sequential representation)**
2) removing any atoms beyond the K-th **(truncation)**

$$\Theta = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$



Ordering of (N)CRM atoms

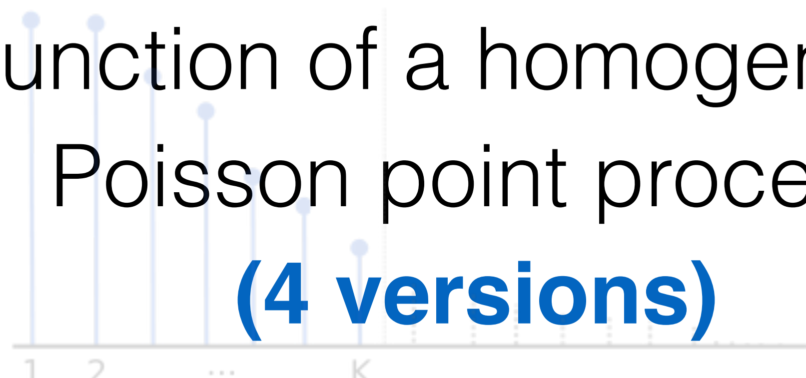
We describe 2 forms for sequential representations

Ordering of (N)CRM atoms

We describe 2 forms for sequential representations

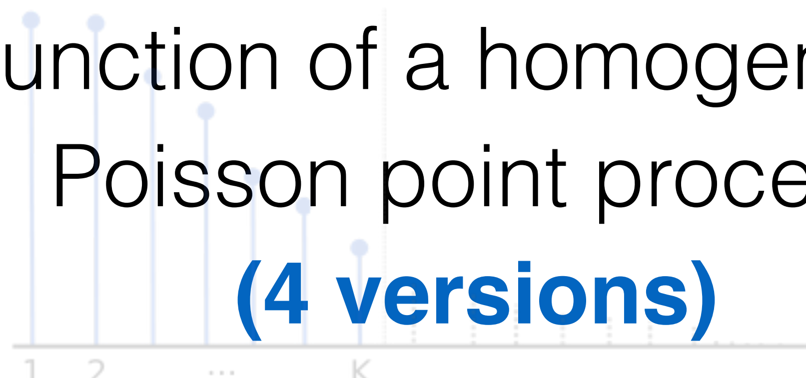
Series representation

function of a homogenous
Poisson point process
(4 versions)



Ordering of (N)CRM atoms

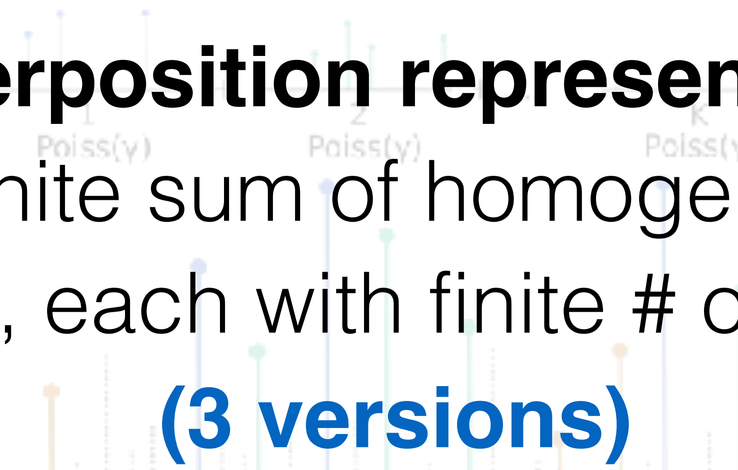
We describe 2 forms for sequential representations



Series representation

function of a homogenous Poisson point process

(4 versions)



Superposition representation

infinite sum of homogenous CRMs, each with finite # of atoms

(3 versions)

Ordering of (N)CRM atoms

We describe 2 forms for sequential representations

Series representation
function of a homogenous Poisson point process
(4 versions)

Superposition representation
infinite sum of homogenous CRMs, each with finite # of atoms
(3 versions)

Theorem (H., Campbell, How, Broderick).
Can generate (N)CRMs using all 7 sequential representations

Sequential representation comparison

Why so many representations?

Sequential representation comparison

Why so many representations?

They're all useful in different circumstances

Sequential representation comparison

Why so many representations?

They're all useful in different circumstances

	Series Reps				**Superposition Reps**		
	B-Rep	IL-Rep	R-Rep	T-Rep	DB-Rep	PL-Rep	SB-Rep
Error Bound Decay	✓ (exp)	✓ (exp)	✓/✗	✗	✓ (exp)	✓ (exp)	✗
Ease of Analysis	✗	✗✗	✗	✗	✓	✓	✓
Generality	✓	✓	✓	✓	✓	✓	✓
Known # Atoms	✓	✓	✗	✗	✗	✗	✗

Sequential representation example

Given Gamma process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$

Sequential representation example

Given Gamma process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$

Step 1: compute $c := \lim_{\theta \to 0} \theta\nu(\theta)$

Sequential representation example

Given Gamma process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$

Step 1: compute $c := \lim_{\theta \to 0} \theta\nu(\theta) = \gamma\lambda$

Sequential representation example

Given Gamma process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$

Step 1: compute $c := \lim_{\theta \to 0} \theta\nu(\theta) = \gamma\lambda$

Step 2: compute $f(\theta) := -c^{-1}\frac{d}{d\theta}\left[\theta\nu(\theta)\right]$

Sequential representation example

Given Gamma process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$

Step 1: compute $c := \lim_{\theta\to 0} \theta\nu(\theta) = \gamma\lambda$

Step 2: compute $f(\theta) := -c^{-1}\frac{d}{d\theta}\left[\theta\nu(\theta)\right] = \lambda e^{-\lambda\theta}$

Sequential representation example

Given Gamma process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$

Step 1: compute $c := \lim_{\theta \to 0} \theta\nu(\theta) = \gamma\lambda$

Step 2: compute $f(\theta) := -c^{-1}\frac{d}{d\theta}\left[\theta\nu(\theta)\right] = \lambda e^{-\lambda\theta}$

Exponential(λ)
density!

Sequential representation example

Given Gamma process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$

Step 1: compute $c := \lim_{\theta \to 0} \theta\nu(\theta) = \gamma\lambda$

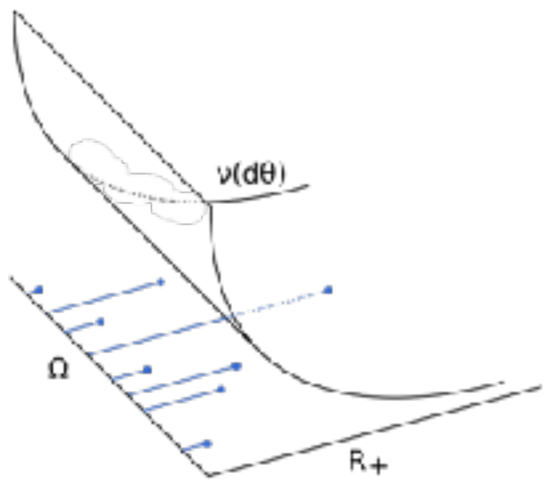
Step 2: compute $f(\theta) := -c^{-1}\frac{d}{d\theta}[\theta\nu(\theta)] = \lambda e^{-\lambda\theta}$

Exponential(λ) density!

Step 3: plug in!

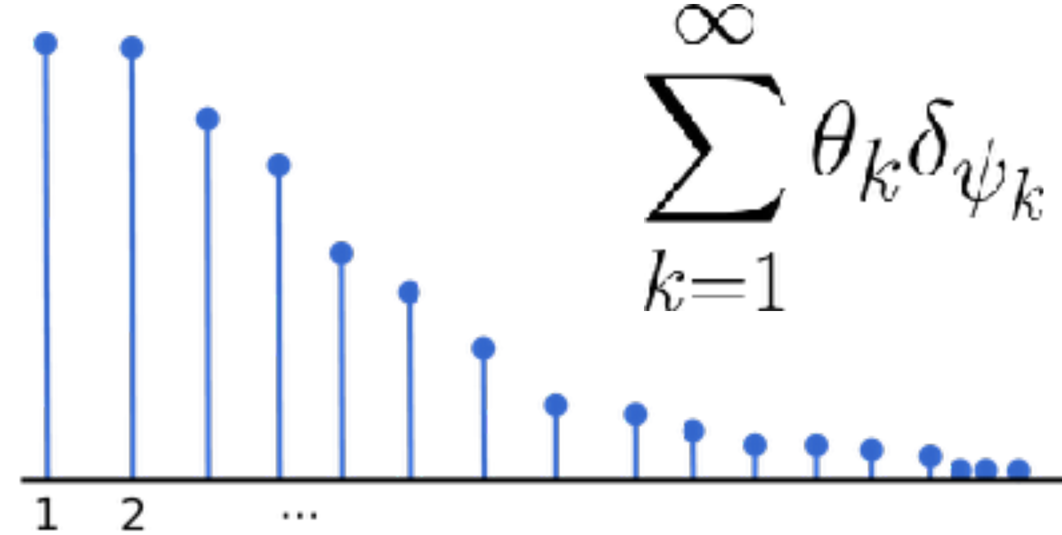
$$\Theta = \sum_{k=1}^{\infty} V_k e^{-\Gamma_k}\delta_{\psi_k}, \quad V_k \overset{\text{iid}}{\sim} f, \quad \Gamma \sim \text{PoissonP}(c)$$

Truncation Roadmap



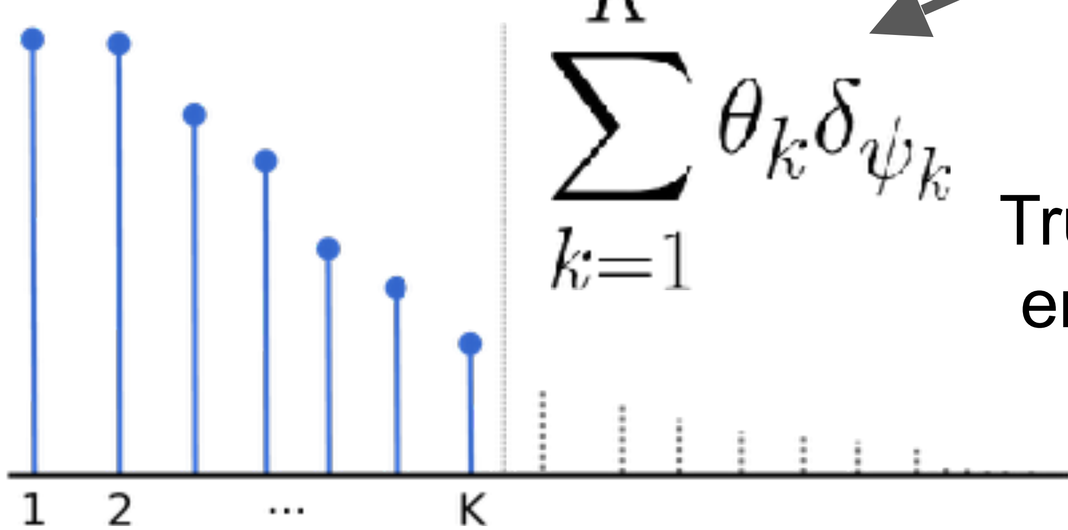
Tractable models in BNP

two forms for sequential representations



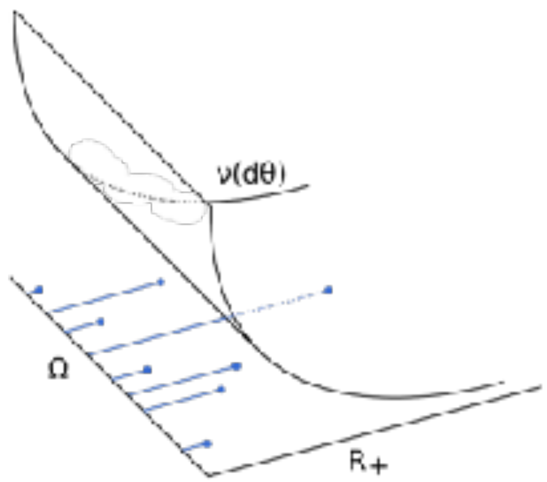
$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

$$\sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

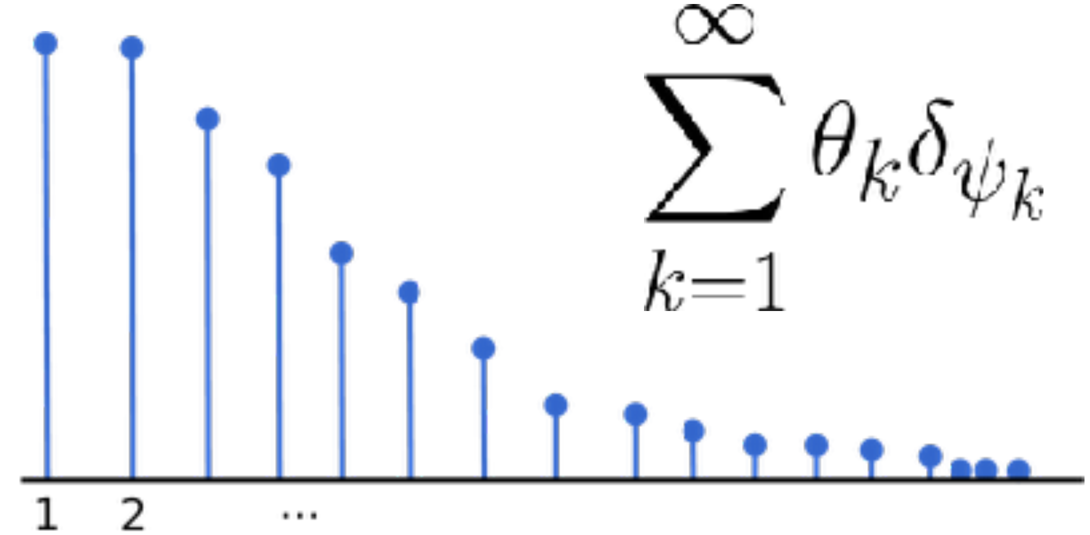


Truncation and error analysis

Truncation Roadmap

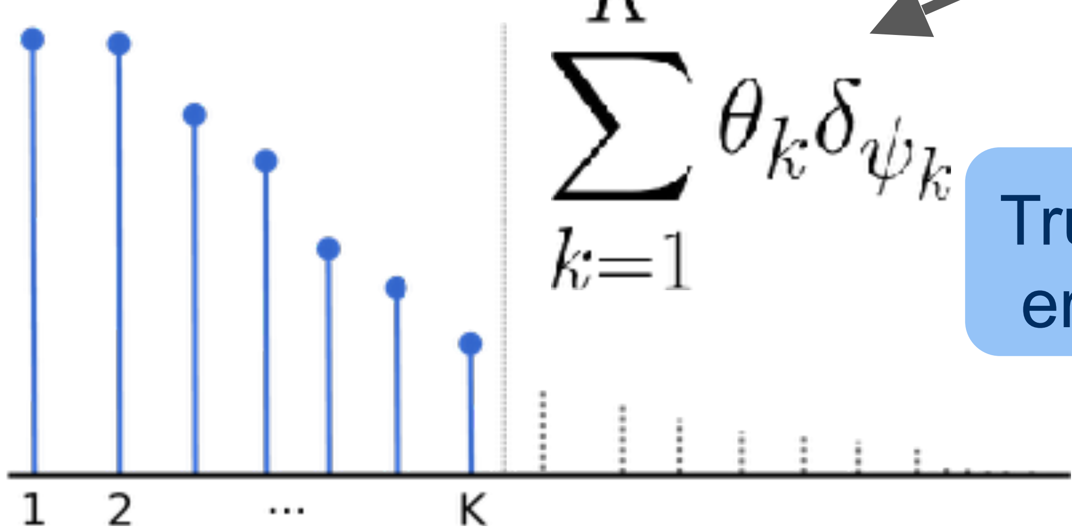


Tractable models
in BNP



$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

two forms for sequential
representations



$$\sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

Truncation and
error analysis

Choosing between the seven representations

How close is our finite approximation?

Choosing between the seven representations

How close is our finite approximation?

Truncation error: $\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int \left| p_{N,\infty}(y) - p_{N,K}(y) \right| \mathrm{d}y$

Choosing between the seven representations

How close is our finite approximation?

Truncation error: $\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2}\int \left|p_{N,\infty}(y) - p_{N,K}(y)\right| \mathrm{d}y$

full infinite
Θ

truncated
$Θ_K$

Choosing between the seven representations

How close is our finite approximation?

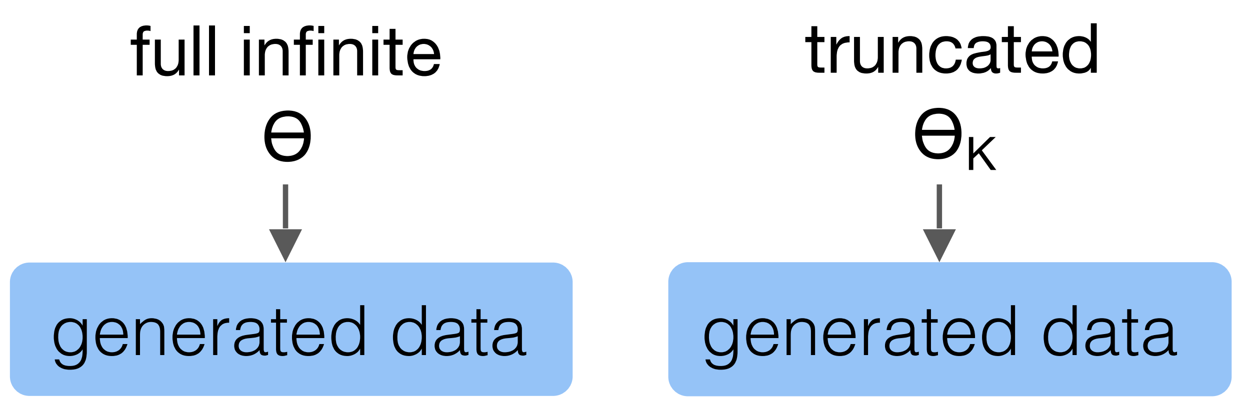
Truncation error: $\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2}\int \left|p_{N,\infty}(y) - p_{N,K}(y)\right| \mathrm{d}y$

full infinite	truncated
θ	θ_K
$\downarrow$	$\downarrow$
generated data	generated data

Choosing between the seven representations

How close is our finite approximation?

Truncation error: $\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(y) - p_{N,K}(y)| \, \mathrm{d}y$



Compare the distribution of the data under full vs. truncated

Choosing between the seven representations

How close is our finite approximation?

Truncation error: $\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2}\int \left|p_{N,\infty}(y) - p_{N,K}(y)\right| \mathrm{d}y$

Depends on **number of observations N** and **truncation level K**

Choosing between the seven representations

How close is our finite approximation?

Truncation error: $\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2}\int \left|p_{N,\infty}(y) - p_{N,K}(y)\right| \mathrm{d}y$

Depends on **number of observations N** and **truncation level K**

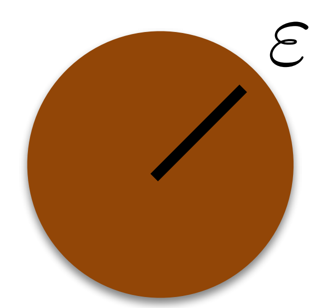
As N gets larger, error increases

Choosing between the seven representations

How close is our finite approximation?

Truncation error: $\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int \left| p_{N,\infty}(y) - p_{N,K}(y) \right| \mathrm{d}y$

Depends on **number of observations N** and **truncation level K**



As N gets larger, error increases

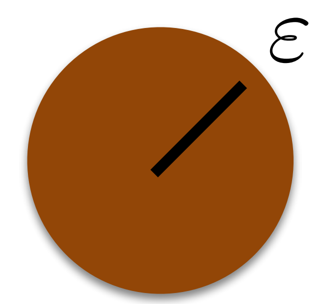
As K gets larger, error decreases

Choosing between the seven representations

How close is our finite approximation?

Truncation error: $\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2}\int |p_{N,\infty}(y) - p_{N,K}(y)|\,\mathrm{d}y$

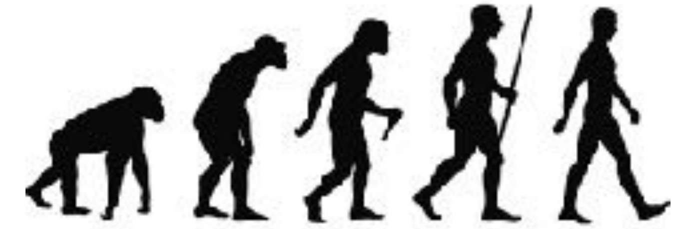
Depends on **number of observations N** and **truncation level K**



As N gets larger, error increases

As K gets larger, error decreases

Cannot evaluate exactly, so we develop **new upper bounds**

Protobound 

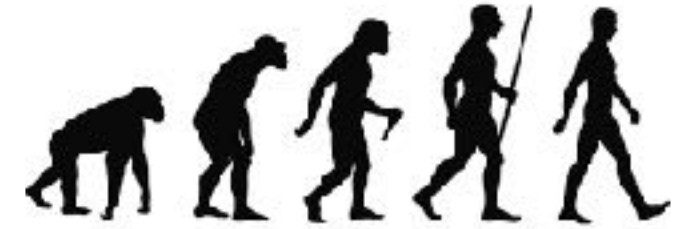
Leads to all the other truncation error bounds in this work

Lemma (H., Campbell, How, Broderick).

$$\|p_{N,\infty} - p_{N,K}\|_1 \leq \mathbb{P}\left(\text{any datum selects a removed trait}\right)$$

The truncation error

i.e. P(whoops!)

Protobound 

Leads to all the other truncation error bounds in this work

Lemma (H., Campbell, How, Broderick).

$$\|p_{N,\infty} - p_{N,K}\|_1 \le \mathbb{P}\left(\text{any datum selects a removed trait}\right)$$

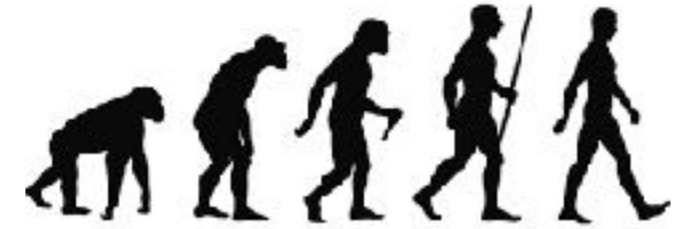
The truncation error

i.e. P(whoops!)

Theorem (HCHB). The series rep error is bounded by

$$\|p_{N,\infty} - p_{N,K}\|_1$$

$$\le 1 - e^{-\int_0^\infty \mathbb{E}[\bar{\pi}(\tau(V,u+G_K))^N]\mathrm{d}u}$$

Protobound 

Leads to all the other truncation error bounds in this work

Lemma (H., Campbell, How, Broderick).

$$\|p_{N,\infty} - p_{N,K}\|_1 \leq \mathbb{P}\left(\text{any datum selects a removed trait}\right)$$

The truncation error

i.e. P(whoops!)

Theorem (HCHB). The series rep error is bounded by

$$\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - e^{-\int_0^\infty \mathbb{E}[\bar{\pi}(\tau(V, u + G_K))^N] \mathrm{d}u}$$

Theorem (HCHB). The superposition rep error is bounded by

$$\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - e^{-\int_0^\infty \bar{\pi}(\theta)^N \nu_K^+(\mathrm{d}\theta)}$$

Error bound example

Given Gamma-Poisson process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$ $\pi(\theta) = e^{-\theta}$

Error bound example

Given Gamma-Poisson process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$ $\quad \pi(\theta) = e^{-\theta}$

Step 1: bound the integral, where $G_K \sim \mathrm{Gamma}(K, c)$:

$$\int_0^\infty (1 - \mathbb{E}\left[\pi(\theta e^{-G_K})\right]\nu(\mathrm{d}\theta)$$

Error bound example

Given Gamma-Poisson process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta}$ $\quad \pi(\theta) = e^{-\theta}$

Step 1: bound the integral, where $G_K \sim \mathrm{Gamma}(K, c)$:

$$\int_0^\infty \left(1 - \mathbb{E}\left[\pi(\theta e^{-G_K})\right]\right] \nu(\mathrm{d}\theta) = \gamma\lambda\mathbb{E}\left[\log(1 + e^{-G_K}/\lambda)\right]$$ Integration by parts

Error bound example

Given Gamma-Poisson process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta} \qquad \pi(\theta) = e^{-\theta}$

Step 1: bound the integral, where $G_K \sim \mathrm{Gamma}(K, c)$:

$$\int_0^\infty \left(1 - \mathbb{E}\left[\pi(\theta e^{-G_K})\right]\right] \nu(\mathrm{d}\theta) = \gamma\lambda\mathbb{E}\left[\log(1 + e^{-G_K}/\lambda)\right] \quad \text{Integration by parts}$$

$$\leq \gamma\mathbb{E}\left[e^{-G_K}\right] \quad \log(1+x) \leq x$$

Error bound example

Given Gamma-Poisson process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta} \qquad \pi(\theta) = e^{-\theta}$

Step 1: bound the integral, where $G_K \sim \mathrm{Gamma}(K, c)$:

$$
\begin{aligned}
\int_0^\infty (1 - \mathbb{E}\left[\pi(\theta e^{-G_K})\right] \nu(\mathrm{d}\theta) &= \gamma\lambda\mathbb{E}\left[\log(1 + e^{-G_K}/\lambda)\right] && \text{Integration by parts} \\
&\leq \gamma\mathbb{E}\left[e^{-G_K}\right] && \log(1+x) \leq x \\
&= \gamma\left(\frac{\gamma\lambda}{1+\gamma\lambda}\right)^K && \text{Gamma expectation}
\end{aligned}
$$

Error bound example

Given Gamma-Poisson process: $\nu(\theta) = \gamma\lambda\theta^{-1}e^{-\lambda\theta} \qquad \pi(\theta) = e^{-\theta}$

Step 1: bound the integral, where $G_K \sim \mathrm{Gamma}(K, c)$:

$$\int_0^\infty \left(1 - \mathbb{E}\left[\pi(\theta e^{-G_K})\right]\right) \nu(\mathrm{d}\theta) = \gamma\lambda\mathbb{E}\left[\log(1 + e^{-G_K}/\lambda)\right]$$ Integration by parts

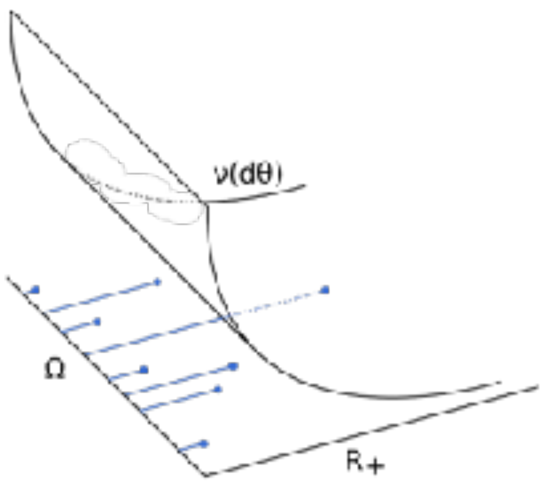
$$\leq \gamma\mathbb{E}\left[e^{-G_K}\right]$$ $\log(1+x) \leq x$

$$= \gamma\left(\frac{\gamma\lambda}{1+\gamma\lambda}\right)^K$$ Gamma expectation

Step 2: plug in!

$$\frac{1}{2}\|p_{N,\infty} - p_{N,K}\|_1 \leq 1 - \exp\left\{-N\gamma\left(\frac{\gamma\lambda}{1+\gamma\lambda}\right)^K\right\}$$

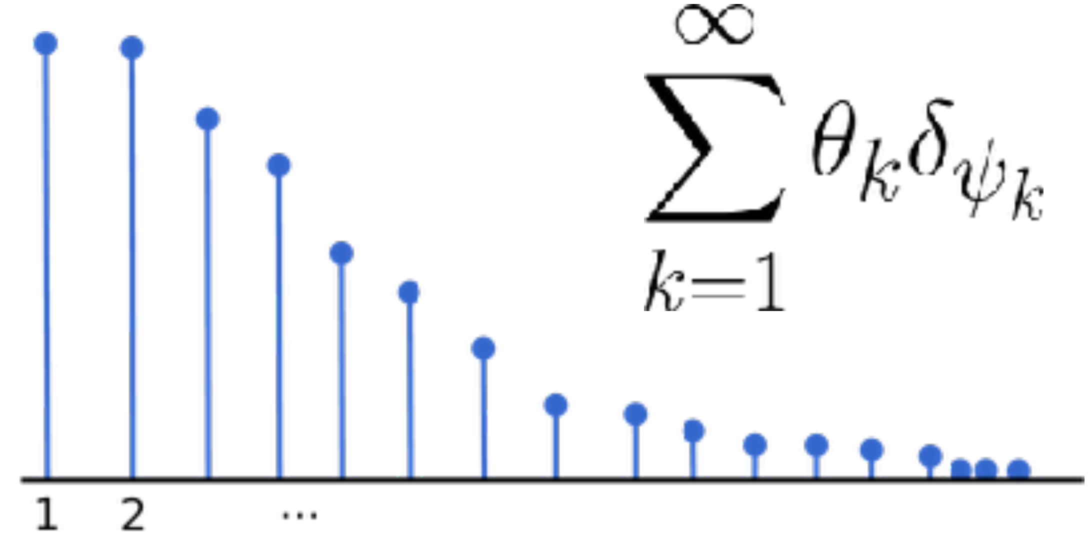
Truncation Roadmap



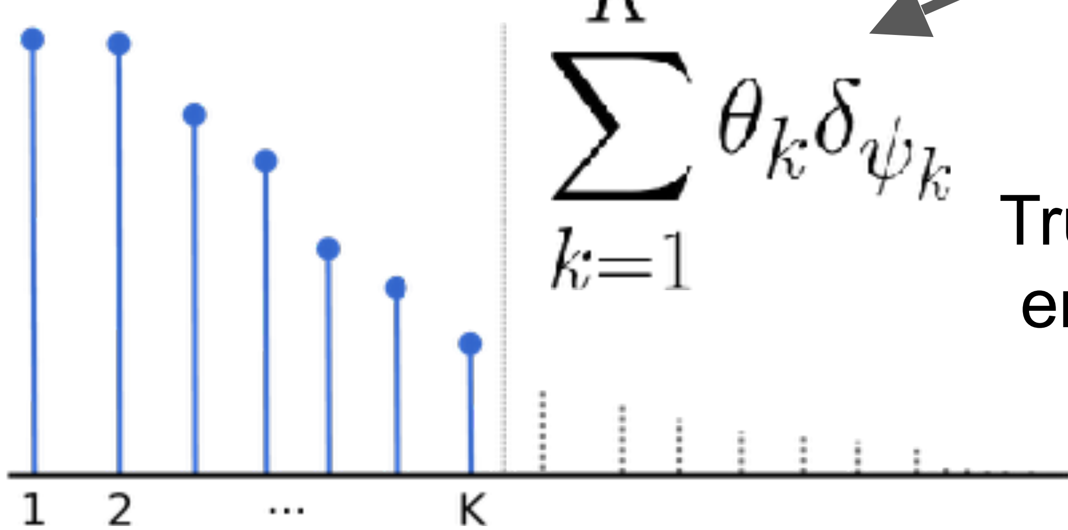
Tractable models
in BNP

two forms for sequential
representations

$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

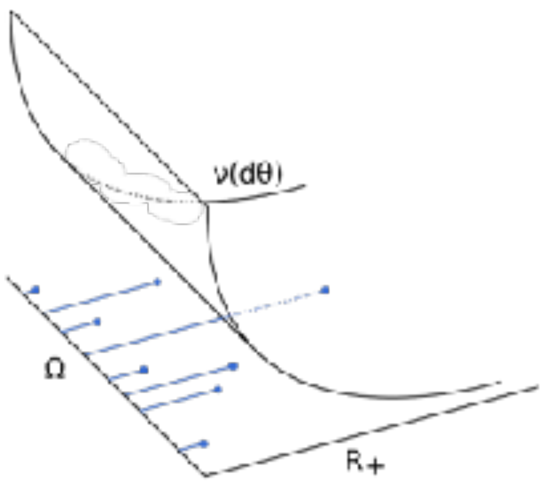


$$\sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$



Truncation and
error analysis

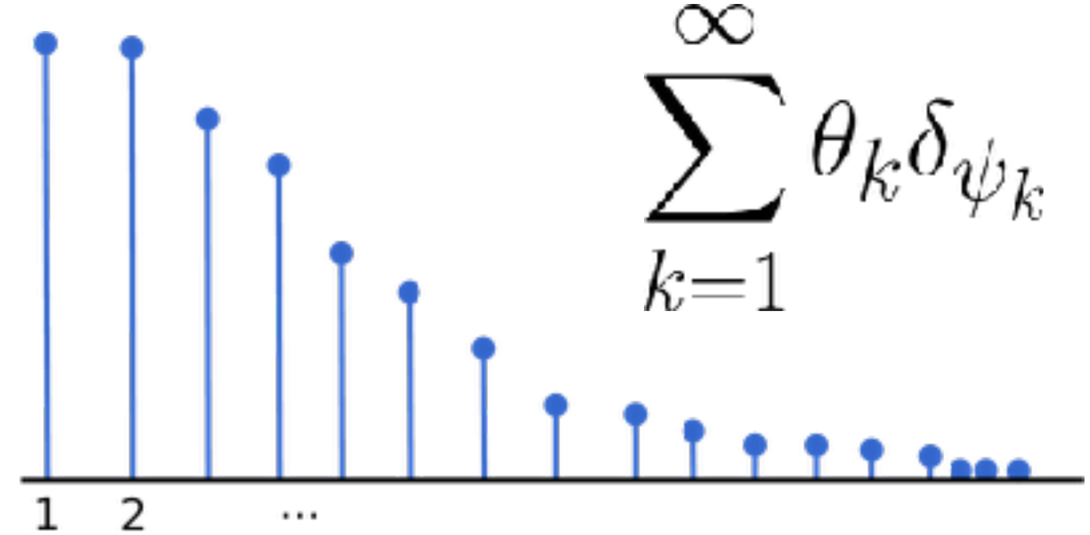
Truncation Roadmap



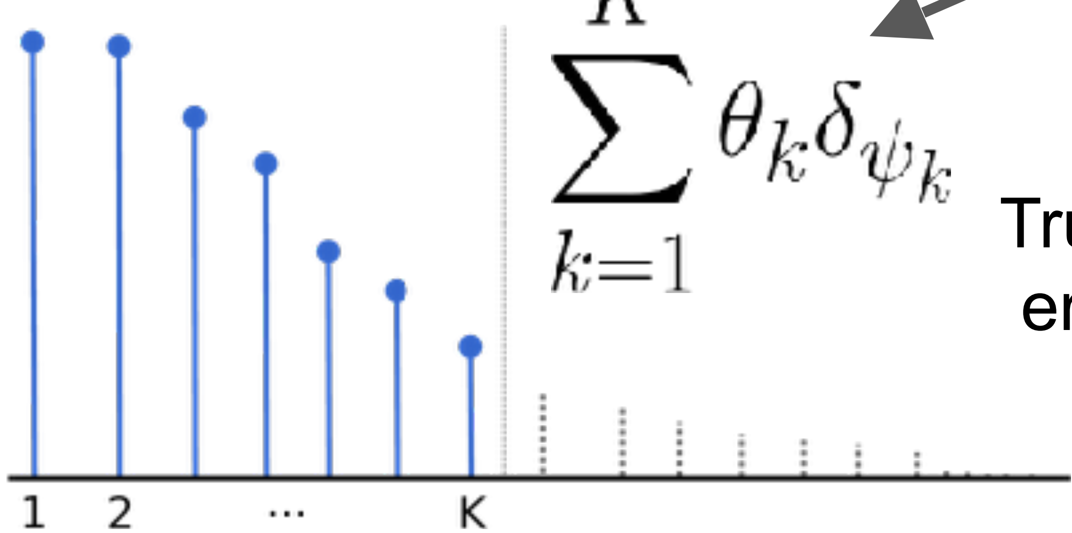
Tractable models
in BNP

two forms for sequential
representations

$$\sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$



$$\sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$



Truncation and
error analysis

Previous Work

	Finite Approximation	Approximation Error Bounds	Computational Complexity
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓		
ΓP	✓	✓	✓
(N)CRM	✓		

Our Work

	Finite Approximation	Approximation Error Bounds	Computational Complexity
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓		
ΓP	✓	✓	✓
(N)CRM	✓	✓	✓

Our Work

	Finite Approximation	Approximation Error Bounds	Computational Complexity
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓	✓	✓
ΓP	✓	✓	✓
(N)CRM	✓	✓	✓

Conclusions

Conclusions

The **sequential representations** and **truncation error bounds** we develop…

Conclusions

The **sequential representations** and
truncation error bounds we develop…

- Expand the class of BNP priors that admit efficient inference

Conclusions

The **sequential representations** and
truncation error bounds we develop…

- Expand the class of BNP priors that admit efficient inference
- Help automate the use of BNP models (e.g. in PPLs)

Conclusions

The **sequential representations** and **truncation error bounds** we develop…

- Expand the class of BNP priors that admit efficient inference
- Help automate the use of BNP models (e.g. in PPLs)
- Facilitates the use of “modern” inference methods (e.g. HMC and VB) with BNP models

Conclusions

The **sequential representations** and
truncation error bounds we develop…

- Expand the class of BNP priors that admit efficient inference
- Help automate the use of BNP models (e.g. in PPLs)
- Facilitates the use of “modern” inference methods (e.g. HMC and VB) with BNP models
- Trade off computational efficiency and statistical accuracy of truncated model

Conclusions

The **sequential representations** and
truncation error bounds we develop…

- Expand the class of BNP priors that admit efficient inference
- Help automate the use of BNP models (e.g. in PPLs)
- Facilitates the use of “modern” inference methods (e.g. HMC and VB) with BNP models
- Trade off computational efficiency and statistical accuracy of truncated model

J. Huggins*, T. Campbell*, J. How, T. Broderick.
Truncated Random Measures. Submitted, 2016.
Available online: **https://arxiv.org/abs/1603.00861**