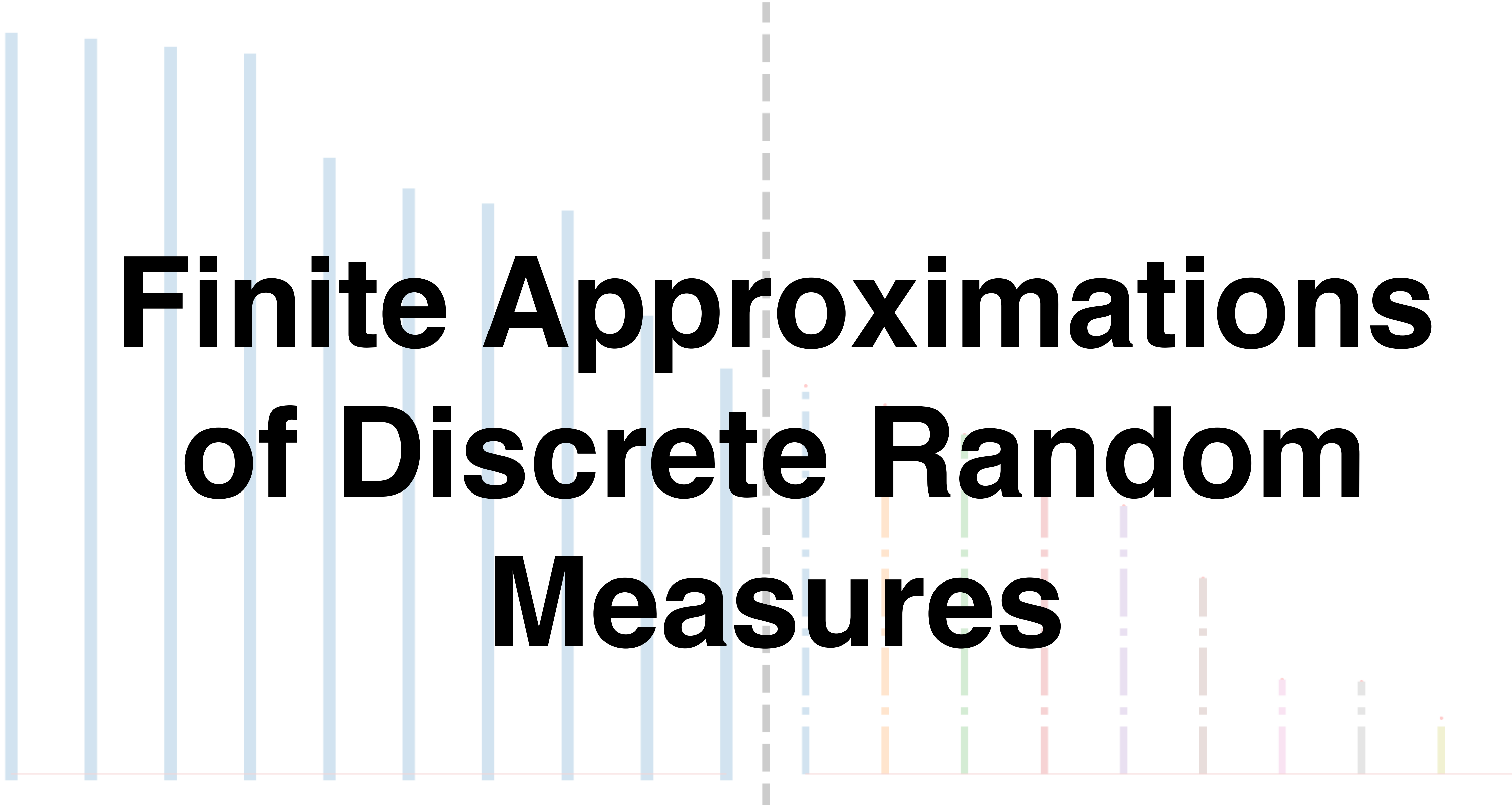


Finite Approximations of Discrete Random Measures

Jonathan H. Huggins
Postdoctoral Research Fellow
Department of Biostatistics, Harvard

with: Trevor Campbell, Jonathan How, Lorenzo Masoero, Lester Mackey, Tamara Broderick

Bayesian nonparametrics

Bayesian nonparametrics

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

Bayesian nonparametrics

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

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

Bayesian nonparametrics

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

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

Bayesian nonparametrics: achieves growing model size via infinite parameters

Bayesian nonparametrics

Need models that can extract new, useful information from unbounded 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]

Bayesian nonparametrics

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

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

Bayesian nonparametrics: achieves growing model size via infinite parameters

movies | text | medicine | robotics | genetics

finance | astronomy | traffic | agriculture | pathology

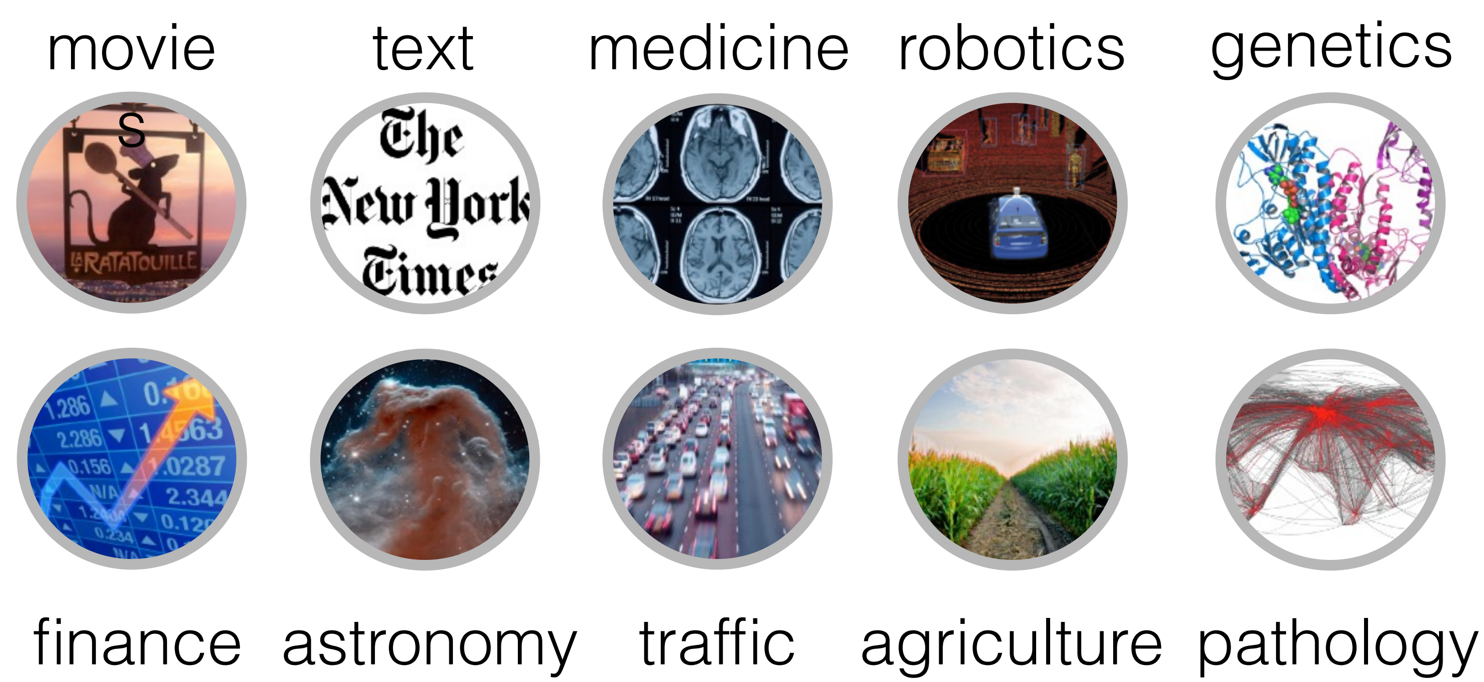
$$\Pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\Pi_0(\mathrm{d}\Theta)$$

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

Bayesian nonparametrics

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

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



Bayesian nonparametrics: achieves growing model size via infinite parameters

parameter

$$\Pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\Pi_0(\mathrm{d}\Theta)$$

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

Bayesian nonparametrics

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

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

movie text medicine robotics genetics

s

finance astronomy traffic agriculture pathology

Bayesian nonparametrics: achieves growing model size via infinite parameters

parameter

$$\Pi(\mathrm{d}\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\Pi_0(\mathrm{d}\Theta)$$

likelihood

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

Bayesian nonparametrics

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

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

movie
s text medicine robotics genetics

finance astronomy traffic agriculture pathology

Bayesian nonparametrics: achieves growing model size via infinite parameters

parameter

$$\Pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\Pi_0(\mathrm{d}\Theta)$$

likelihood **data**

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

Bayesian nonparametrics

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

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

Bayesian nonparametrics: achieves growing model size via infinite parameters

movie
s

text

medicine

robotics

genetics

finance

astronomy

traffic

agriculture

pathology

parameter **prior**

$$\Pi(\mathrm{d}\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\Pi_0(\mathrm{d}\Theta)$$

likelihood **data**

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

Bayesian nonparametrics

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

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

movie text medicine robotics genetics

finance astronomy traffic agriculture pathology

Bayesian nonparametrics: achieves growing model size via infinite parameters

posterior **parameter** **prior**

$$\Pi(\mathrm{d}\Theta \mid X) \propto_\Theta f(X \mid \Theta)\Pi_0(\mathrm{d}\Theta)$$

likelihood **data**

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

Inference in BNP models

$$\pi(\mathrm{d}\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\pi_0(\mathrm{d}\Theta)$$

Inference in BNP models

$$\pi(\mathrm{d}\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\pi_0(\mathrm{d}\Theta)$$

- Option #1: Integrate out the parameter Θ (CRP, IBP, etc.)
  issues: care about the parameters, using certain inference algs. (HMC/VB), distributed computation, discrete latent variables instead

Inference in BNP models

$$\pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\pi_0(\mathrm{d}\Theta)$$

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

Inference in BNP models

$$\pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\pi_0(\mathrm{d}\Theta)$$

- Option #1: Integrate out the parameter Θ (CRP, IBP, etc.)
  issues: care about the parameters, using certain inference algs. (HMC/VB), distributed computation, discrete latent variables instead
- 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 or poorly understood finite approximation

Inference in BNP models

$$\pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\pi_0(\mathrm{d}\Theta)$$

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

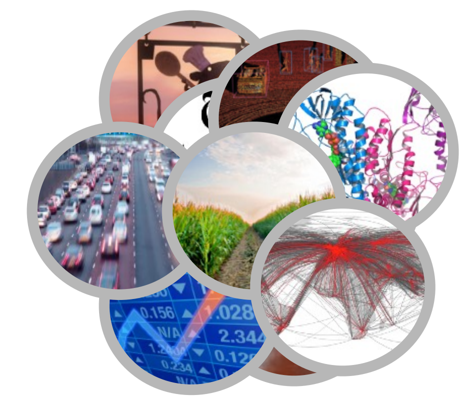
Problem: Wide variety of priors in BNP with
no or poorly understood finite approximation

In this talk:

Inference in BNP models

$$\pi(\mathrm{d}\Theta \,|\, X) \propto_\Theta f(X \,|\, \Theta)\pi_0(\mathrm{d}\Theta)$$

- Option #1: Integrate out the parameter Θ (CRP, IBP, etc.)
  issues: care about the parameters, using certain inference algs. (HMC/VB), distributed computation, discrete latent variables instead
- 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 or poorly understood finite approximation

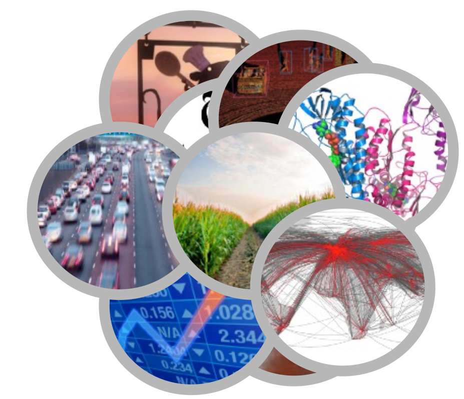
In this talk:

1) Two finite approximation types: **truncated** and **non-nested**

Inference in BNP models

$$\pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\pi_0(\mathrm{d}\Theta)$$

- Option #1: Integrate out the parameter Θ (CRP, IBP, etc.)
  issues: care about the parameters, using certain inference algs. (HMC/VB), distributed computation, discrete latent variables instead
- 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 or poorly understood finite approximation**

In this talk:

1) Two finite approximation types: **truncated** and **non-nested**
2) Two truncated forms (7 reps total) that allow finite approximation of ***(normalized) completely random measures*** *[(N)CRMs]*

Inference in BNP models

$$\pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\pi_0(\mathrm{d}\Theta)$$

- Option #1: Integrate out the parameter Θ (CRP, IBP, etc.)
  issues: care about the parameters, using certain inference algs. (HMC/VB), distributed computation, discrete latent variables instead
- 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 or poorly understood finite approximation**

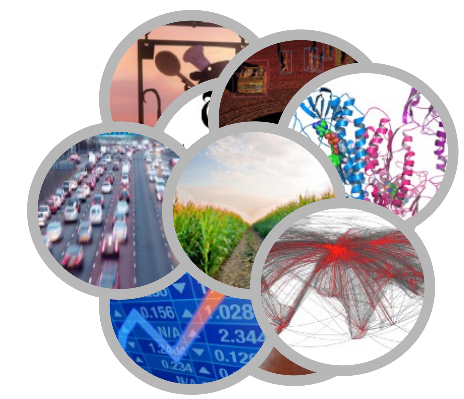
In this talk:

1) Two finite approximation types: **truncated** and **non-nested**
2) Two truncated forms (7 reps total) that allow finite approximation of ***(normalized) completely random measures*** *[(N)CRMs]*
3) Truncation approximation error analysis

Inference in BNP models

$$\pi(\mathrm{d}\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\pi_0(\mathrm{d}\Theta)$$

- Option #1: Integrate out the parameter Θ (CRP, IBP, etc.)
  issues: care about the parameters, using certain inference algs. (HMC/VB), distributed computation, discrete latent variables instead
- 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 or poorly understood finite approximation**

In this talk:

1) Two finite approximation types: **truncated** and **non-nested**
2) Two truncated forms (7 reps total) that allow finite approximation of ***(normalized) completely random measures*** *[(N)CRMs]*
3) Truncation approximation error analysis
4) One non-nested form for (N)CRMs

Outline

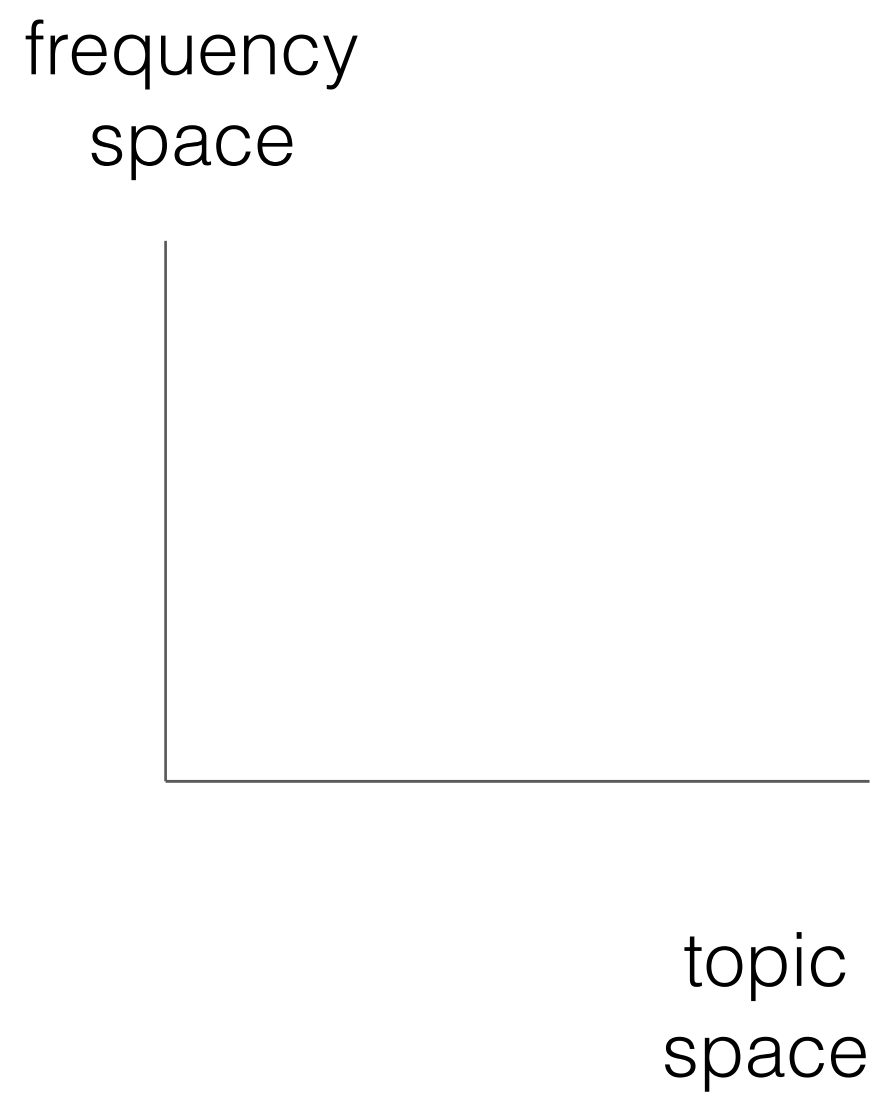
- ➡ **Tractable priors in BNP**
- Truncated approximations
    - Two forms for sequential representations
    - Truncation and error analysis
- Non-nested approximations

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	

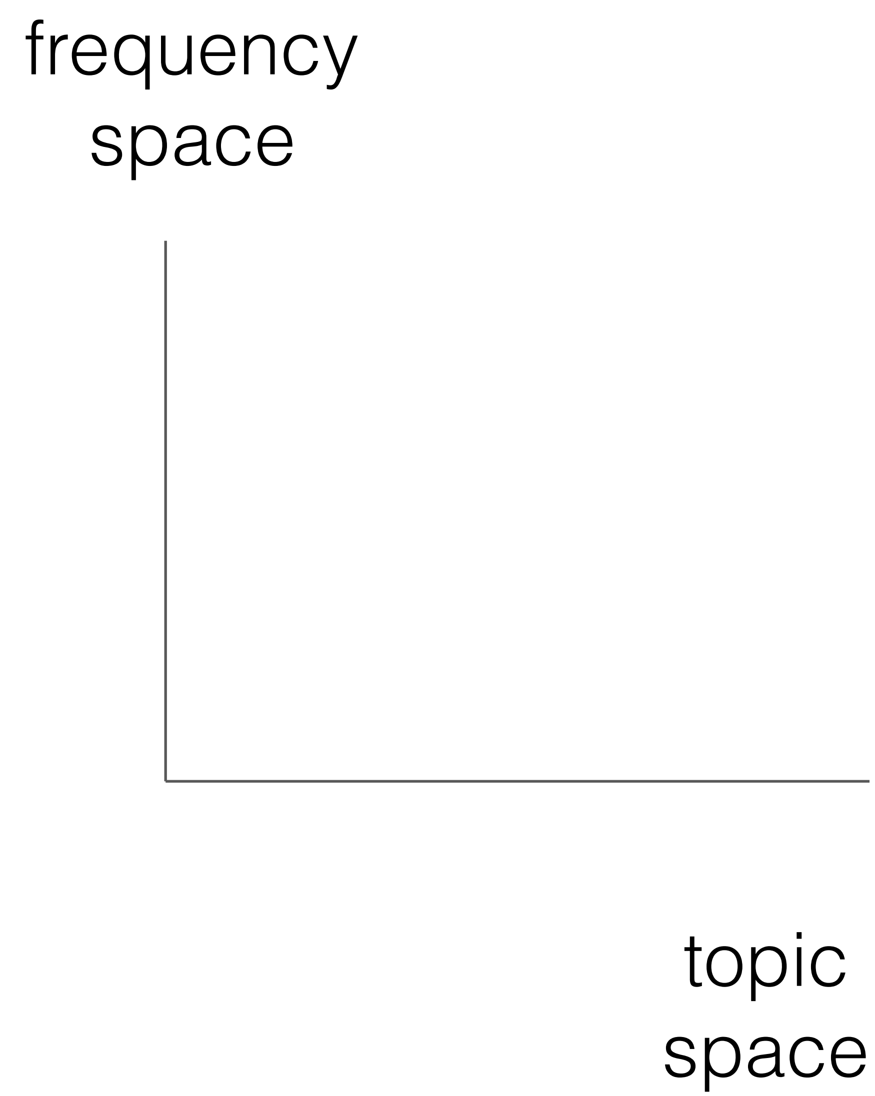
The Standard Model in BNP (By Example)

	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	

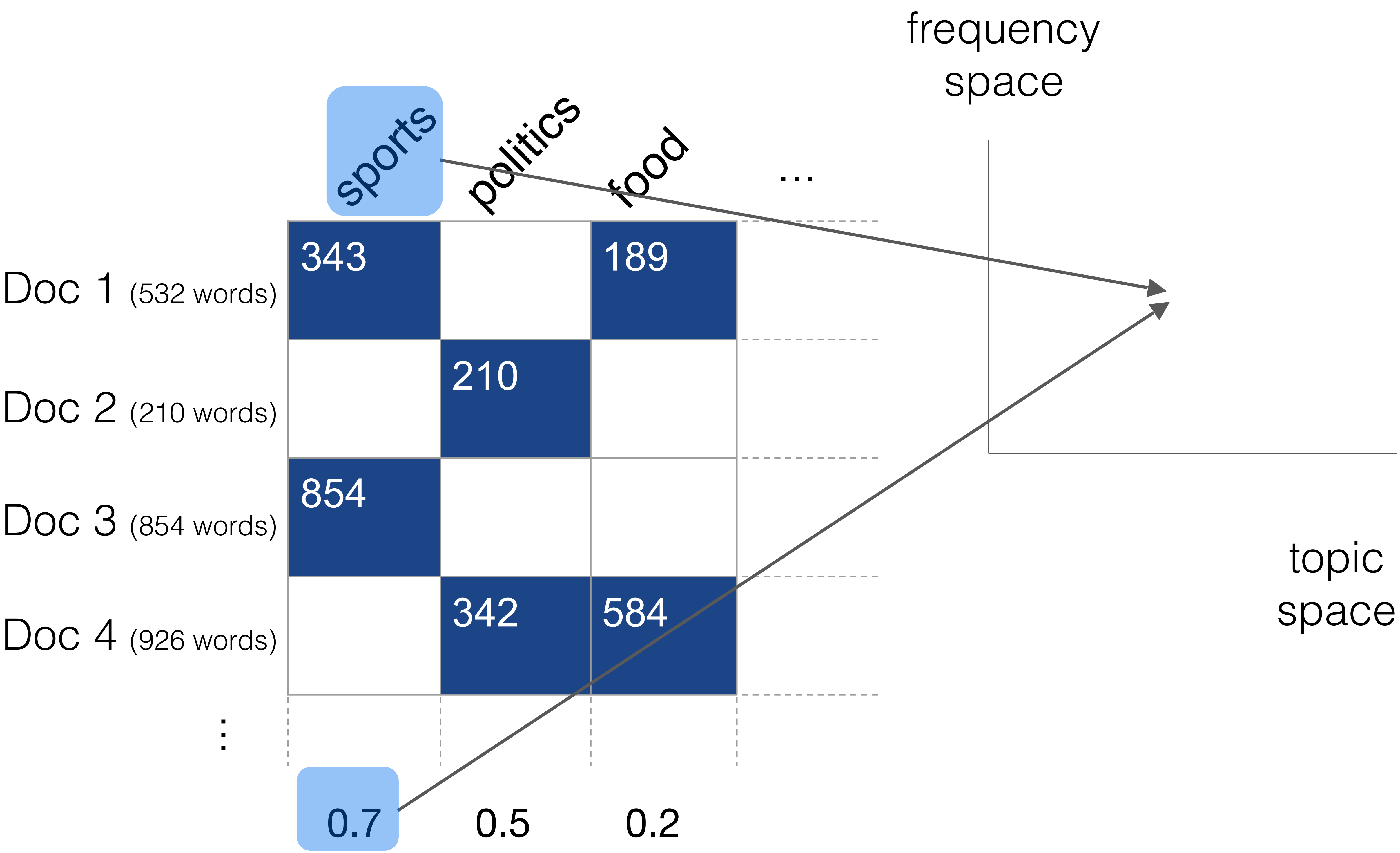


The Standard Model in BNP (By Example)

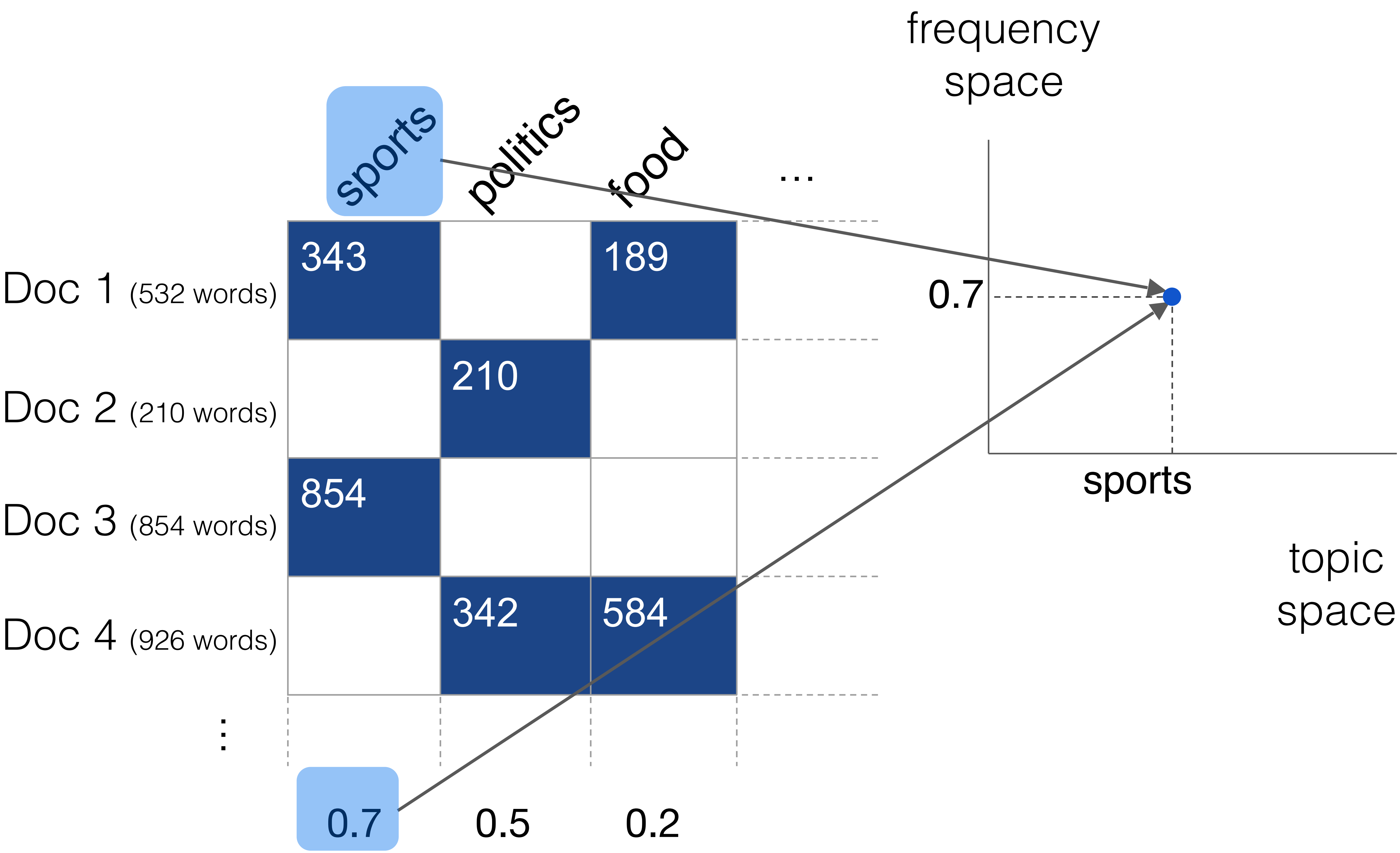
	sports	politics	food	...
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



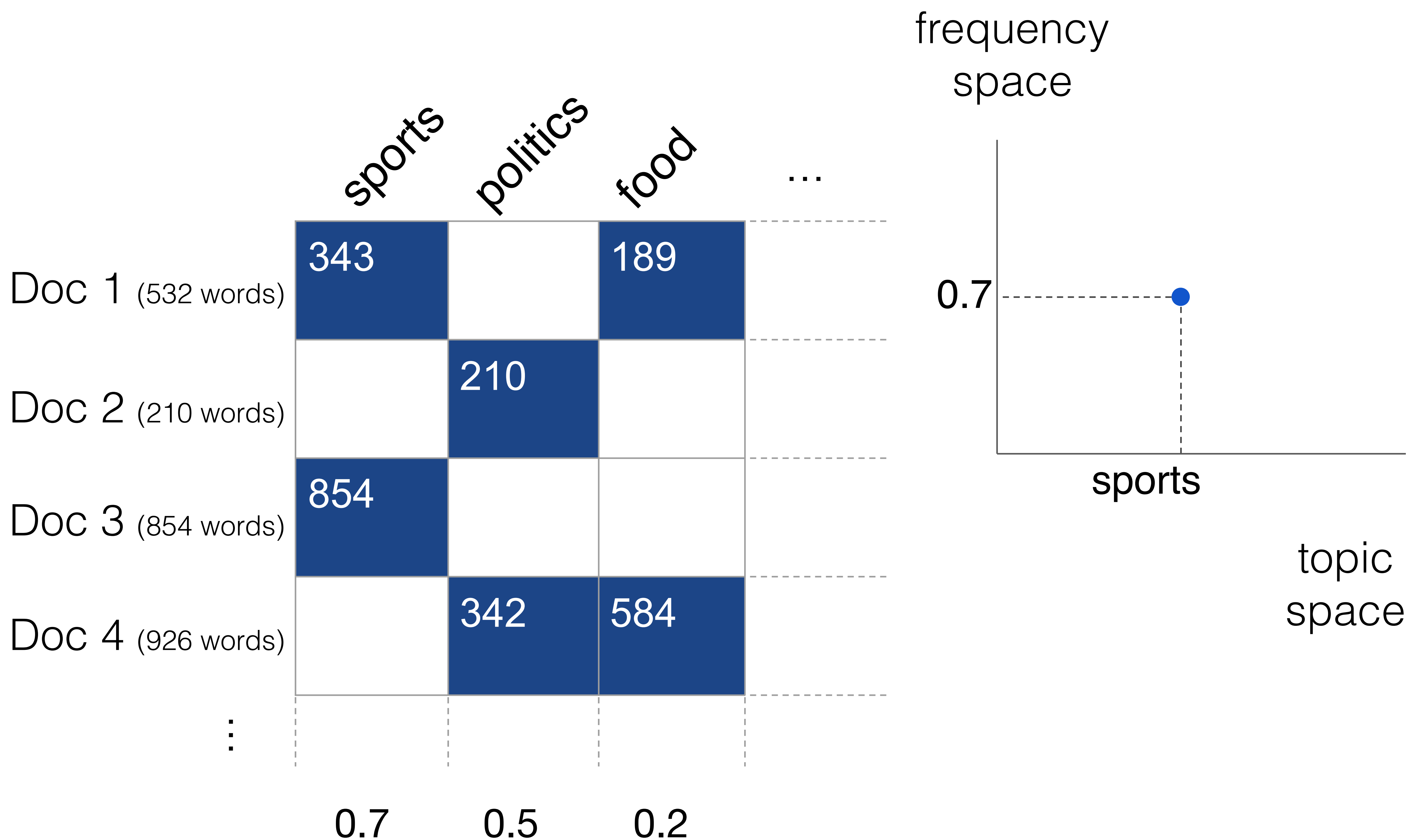
The Standard Model in BNP (By Example)



The Standard Model in BNP (By Example)

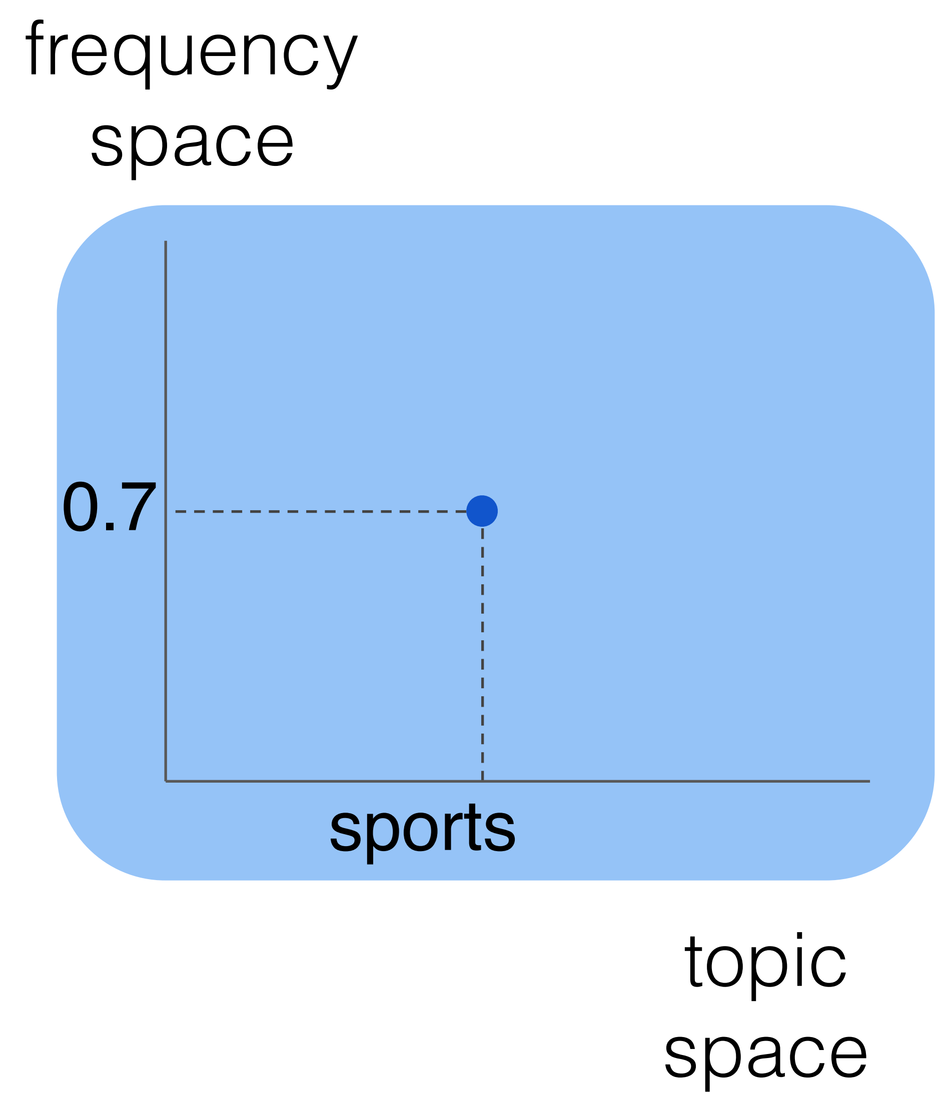


The Standard Model in BNP (By Example)



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

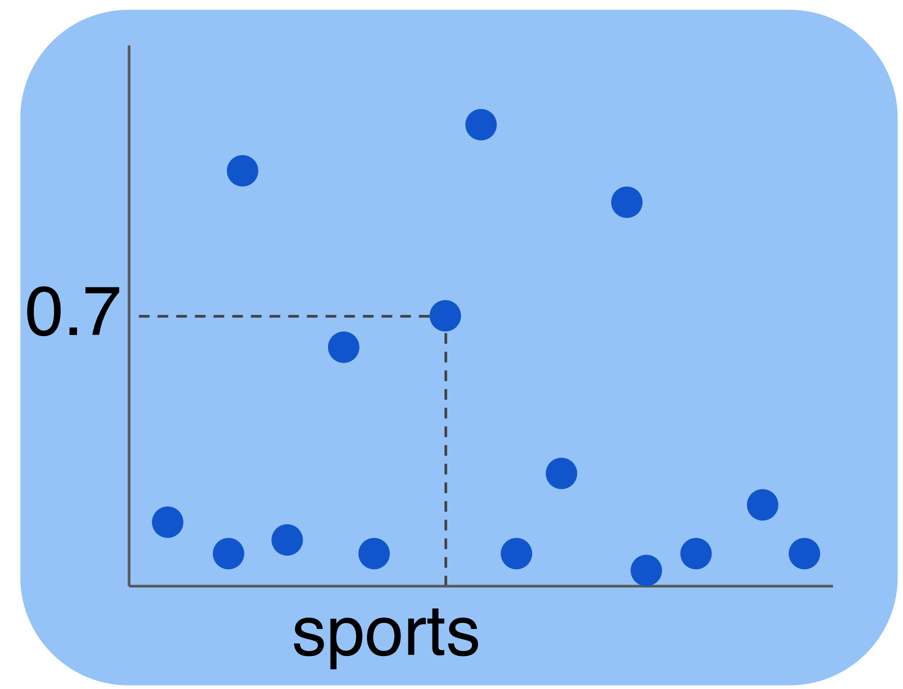
	sports	politics	food	…
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



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	

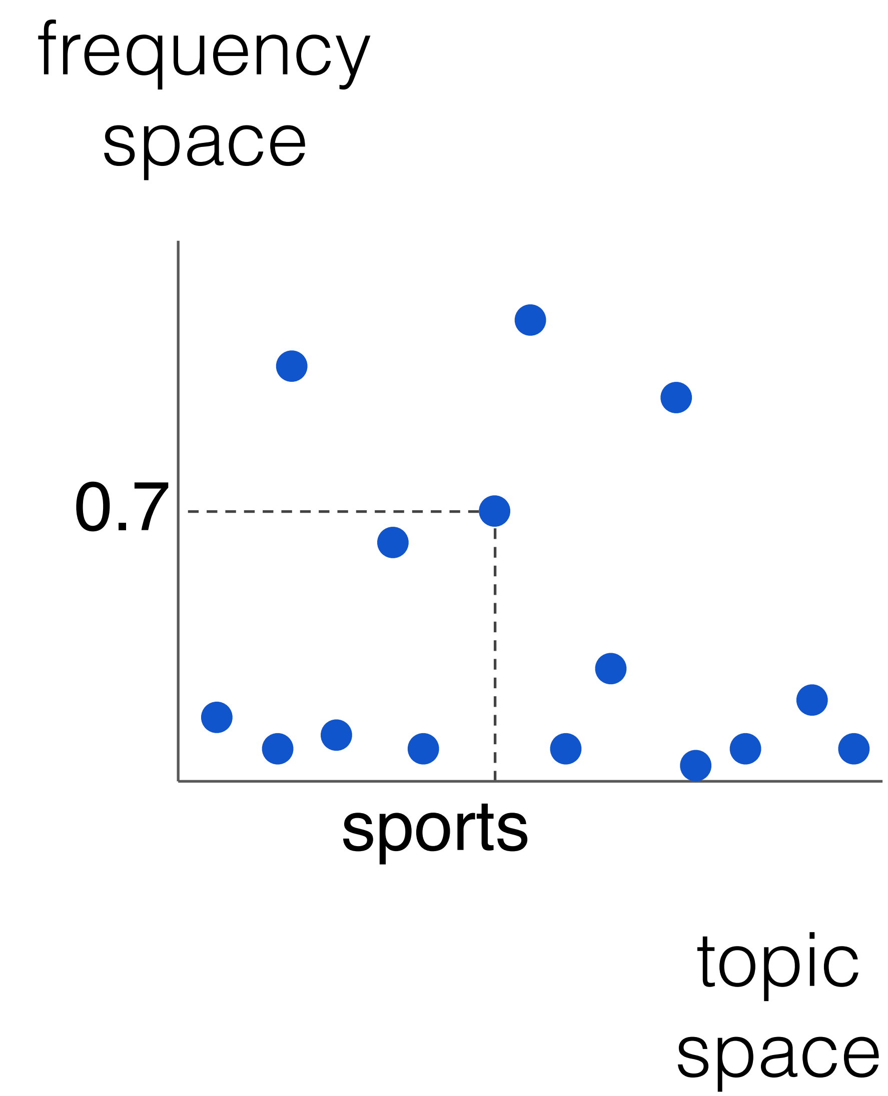
frequency space



topic space

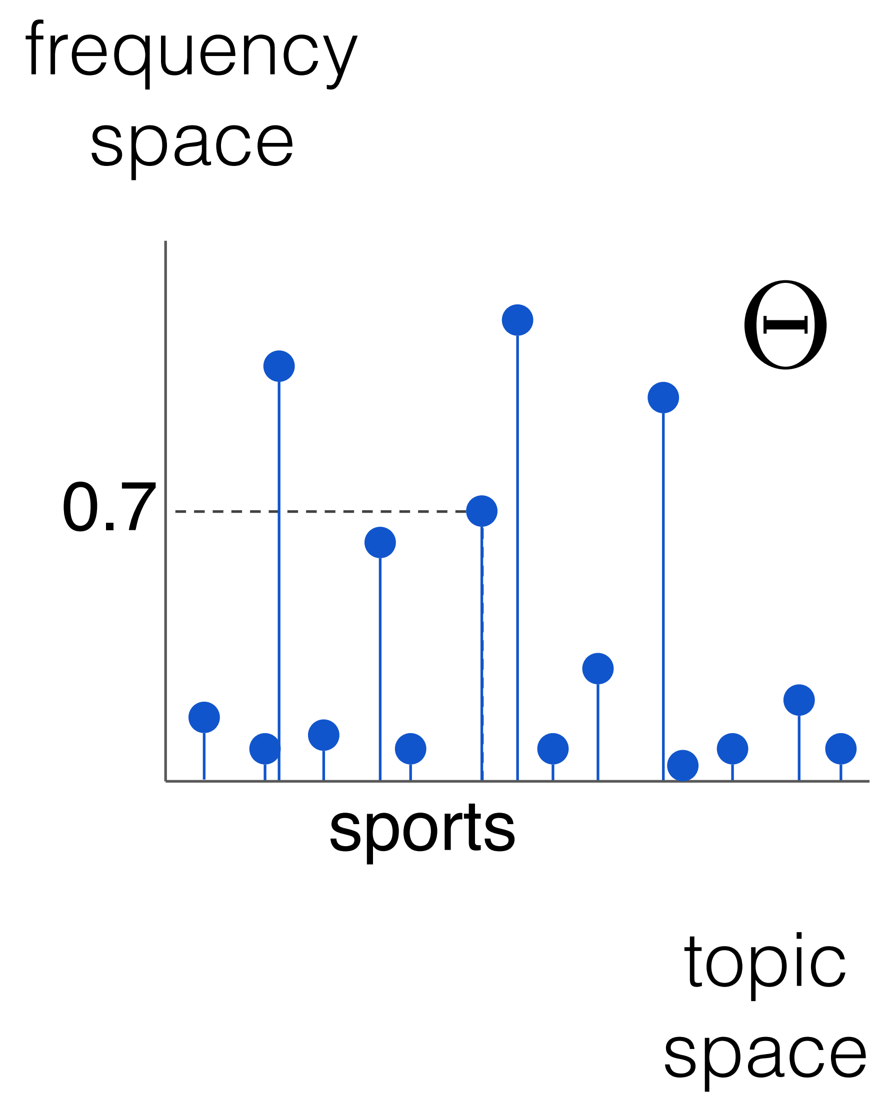
The Standard Model in BNP (By Example)

	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)

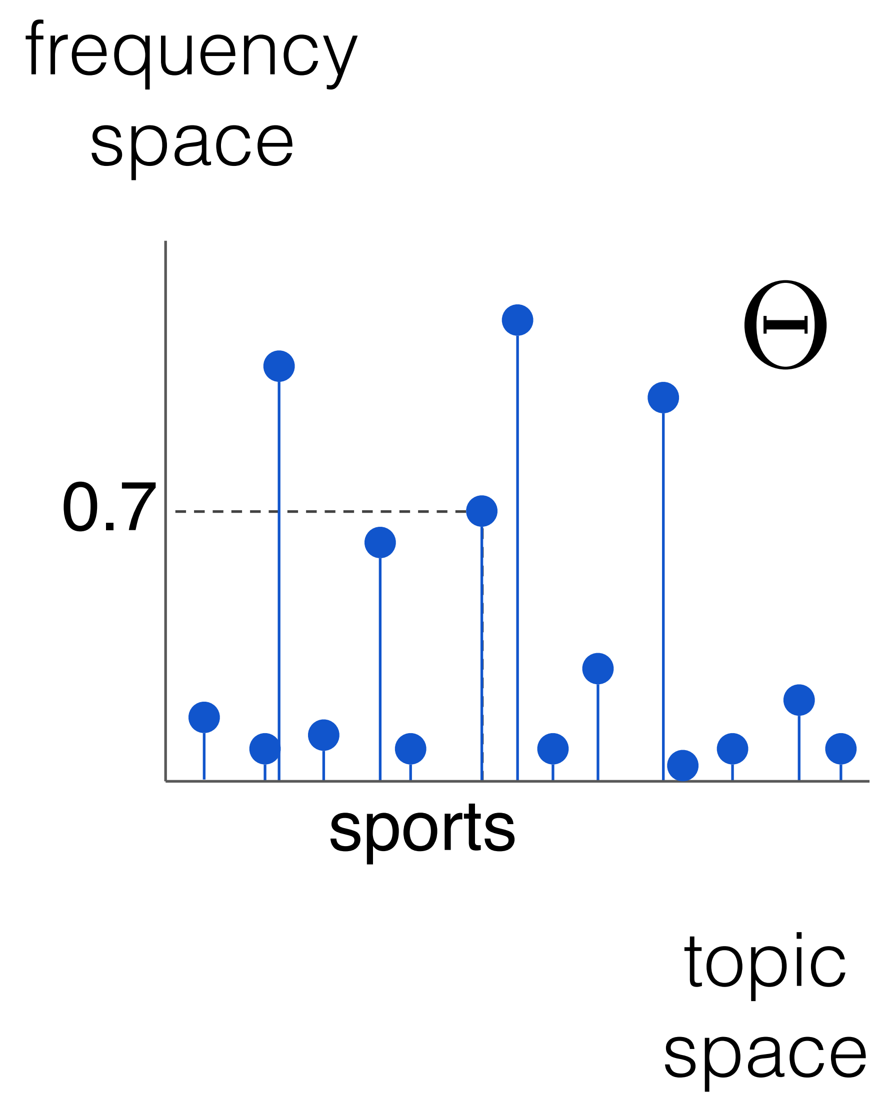
	sports	politics	food	…
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



Θ is a *random discrete measure on the topics*

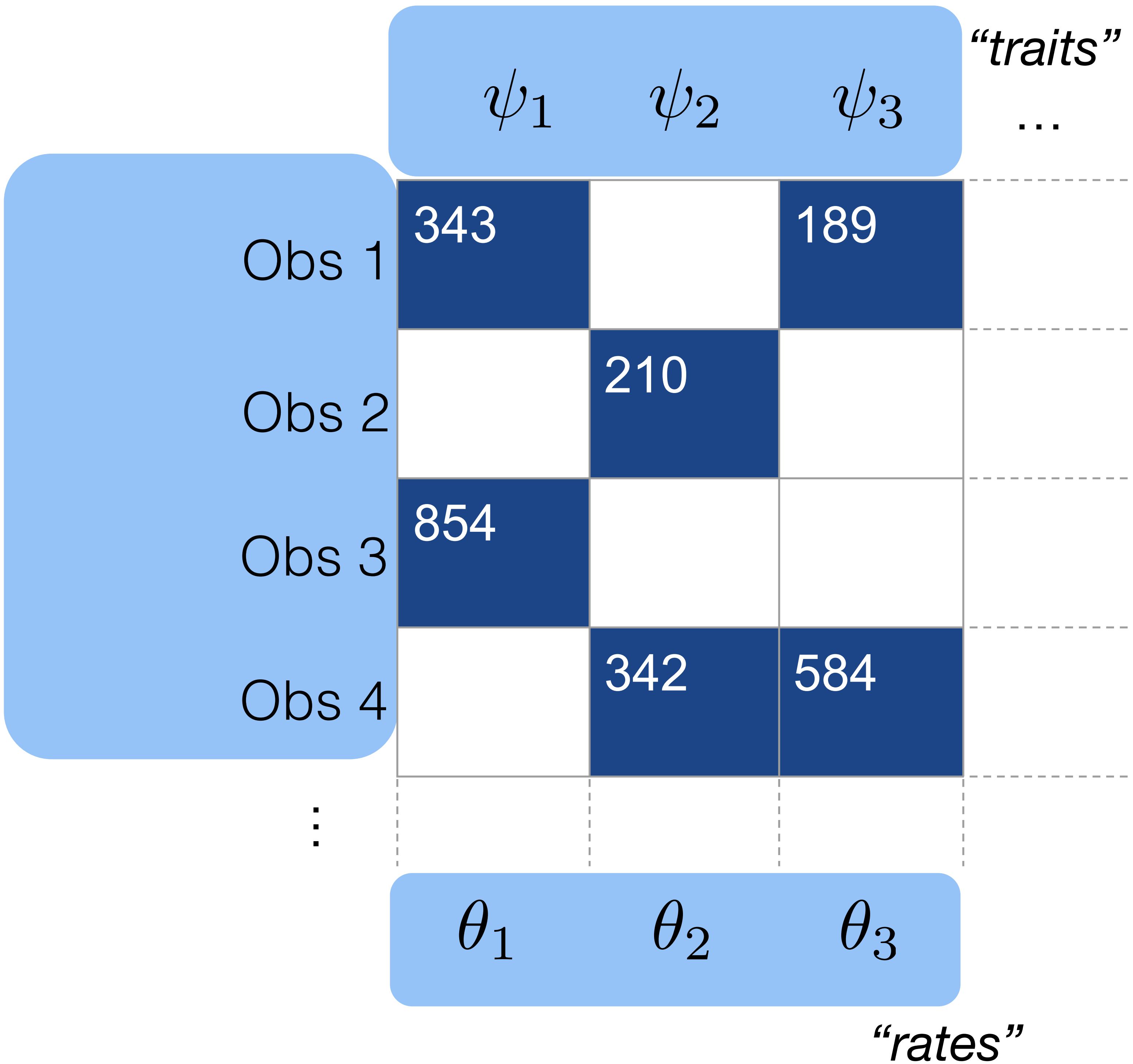
The Standard Model in BNP (By Example)

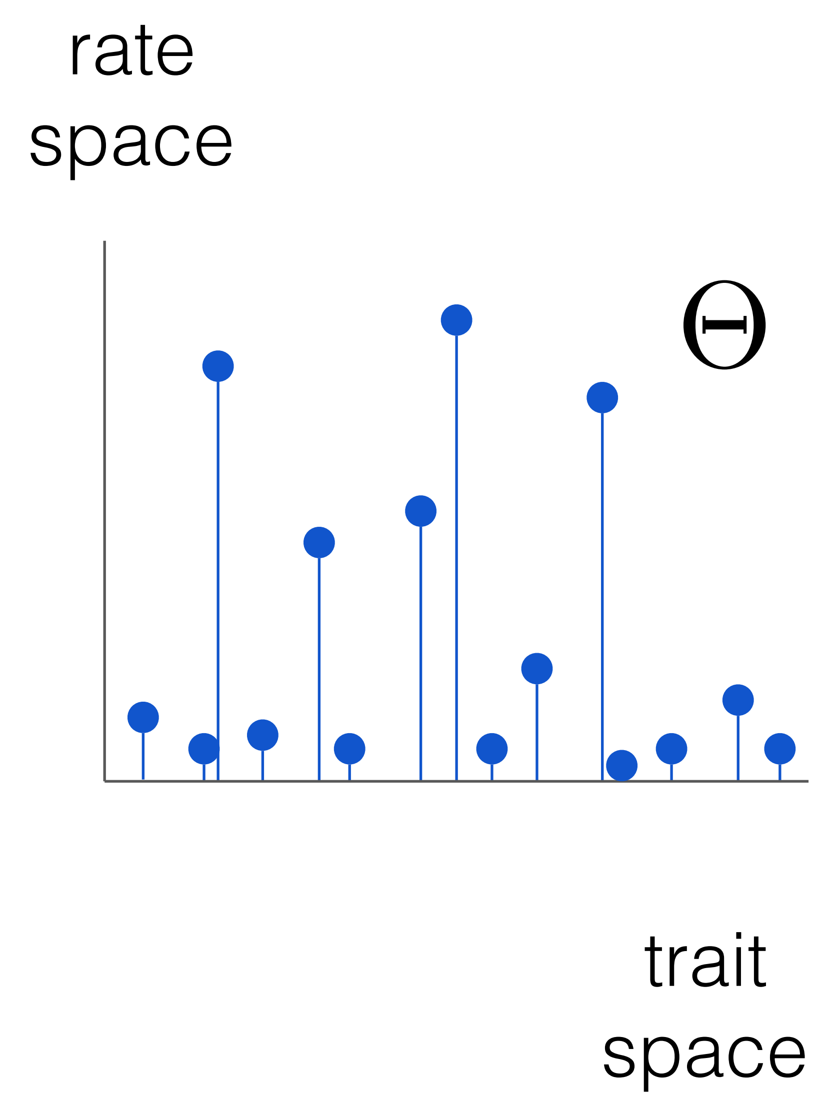
	sports	politics	food	…
Doc 1 (532 words)	343		189	
Doc 2 (210 words)		210		
Doc 3 (854 words)	854			
Doc 4 (926 words)		342	584	
⋮				
	0.7	0.5	0.2	



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

The Standard Model in BNP (By Example)



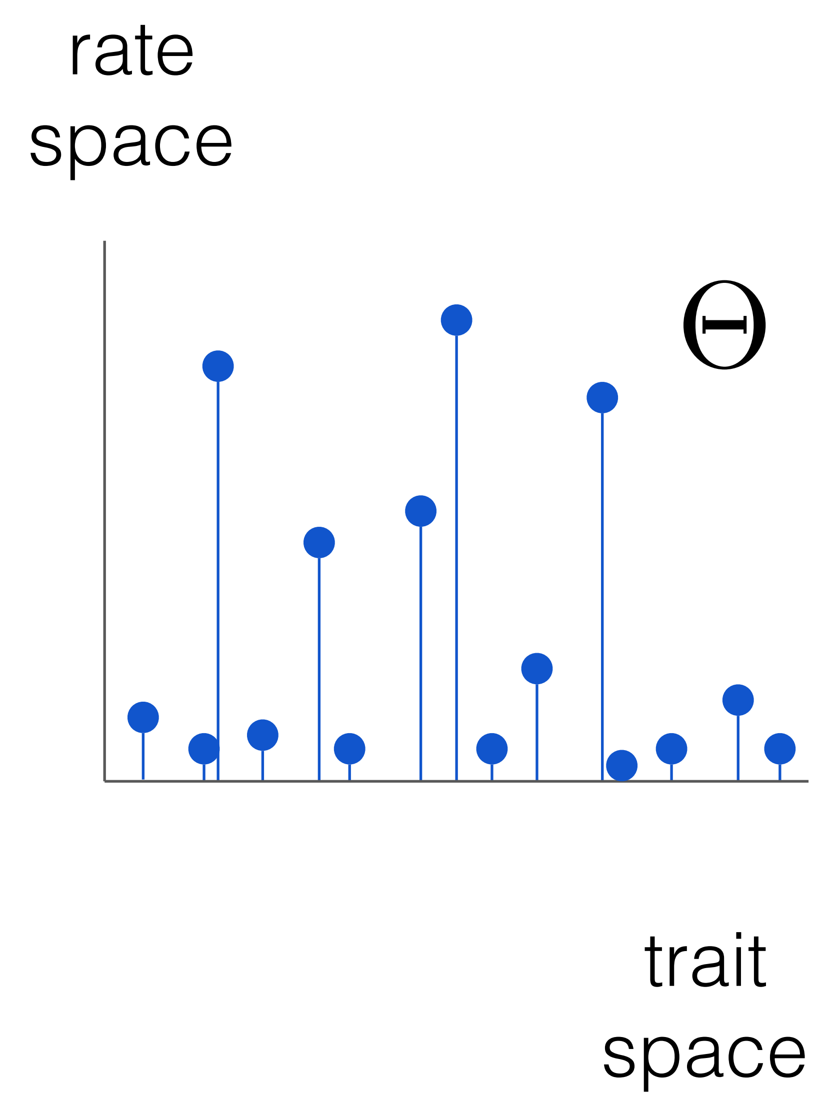


Θ is a ***random discrete measure on the ~~topics~~*** traits

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	

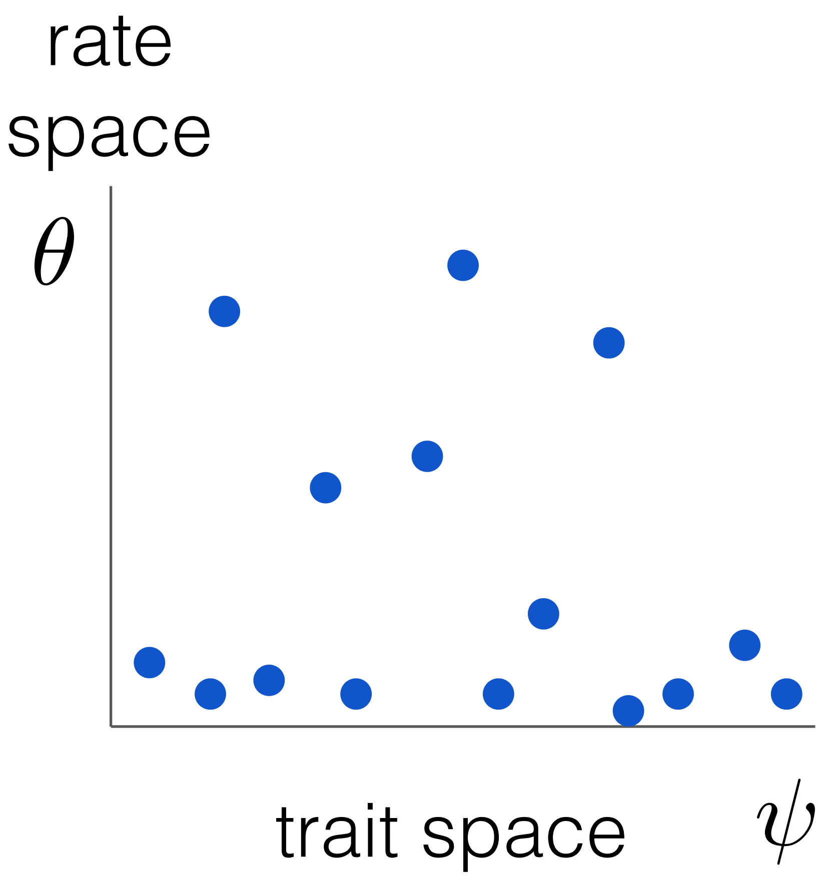
"rates"



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

Poisson processes and (N)CRMs

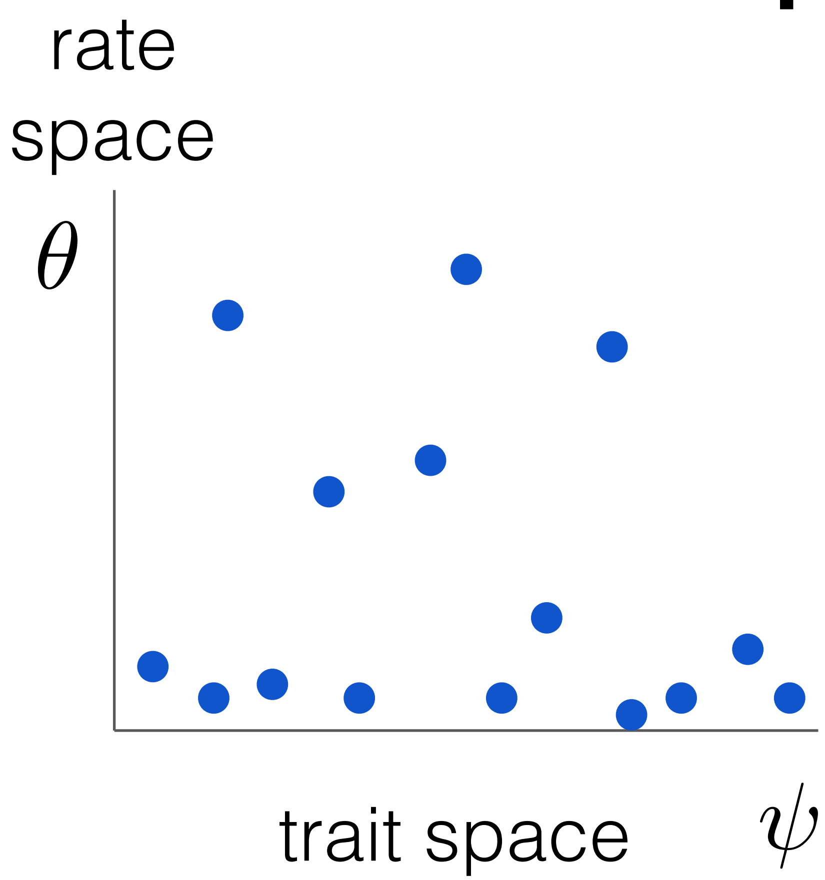
How do we generate infinitely many trait/rate points (ψ, θ)?



Poisson processes and (N)CRMs

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

Poisson process with intensity measure $\mu(\mathrm{d}\theta \times \mathrm{d}\psi)$

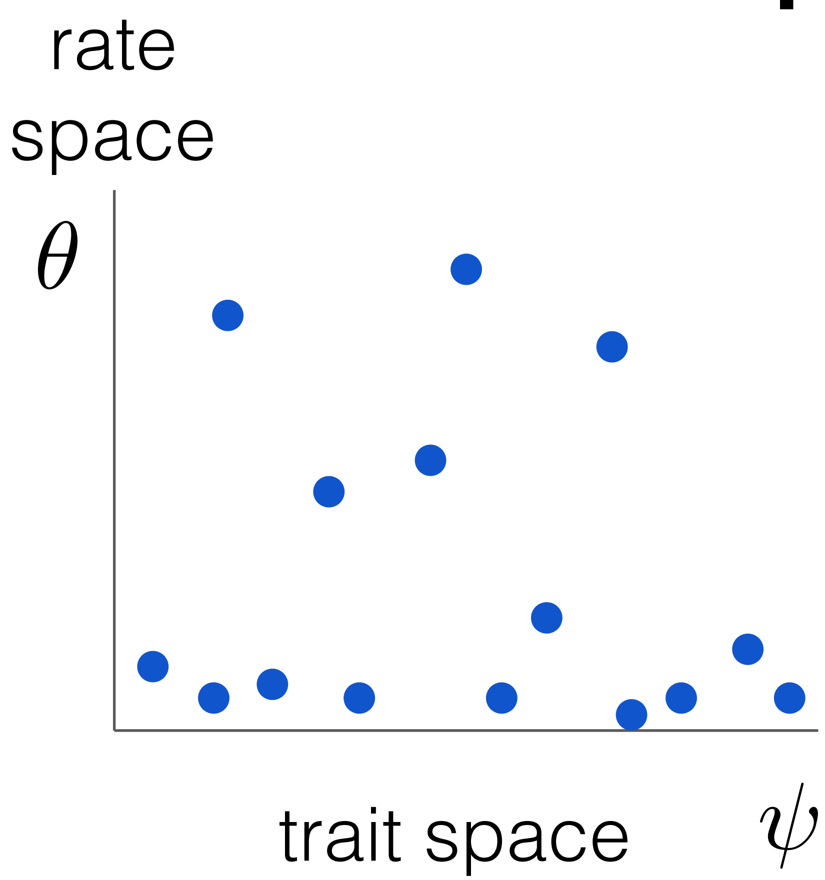


[Kingman 93]

Poisson processes and (N)CRMs

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

Poisson process with intensity measure $\mu(\mathrm{d}\theta \times \mathrm{d}\psi)$
$= \nu(\mathrm{d}\theta) H(\mathrm{d}\psi)$



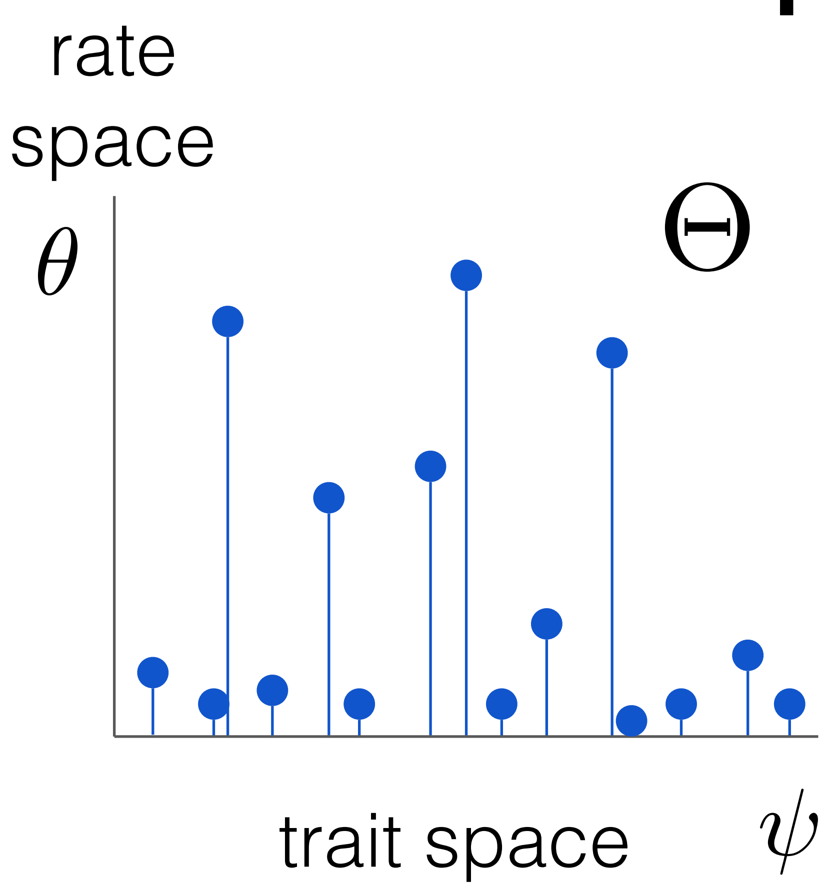
[Kingman 93]

Poisson processes and (N)CRMs

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

Poisson process with intensity measure $\mu(\mathrm{d}\theta \times \mathrm{d}\psi)$

$$= \nu(\mathrm{d}\theta)H(\mathrm{d}\psi)$$



completely random measure (CRM)

(e.g. BP, $\mathbf{\Gamma}$P) $\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$

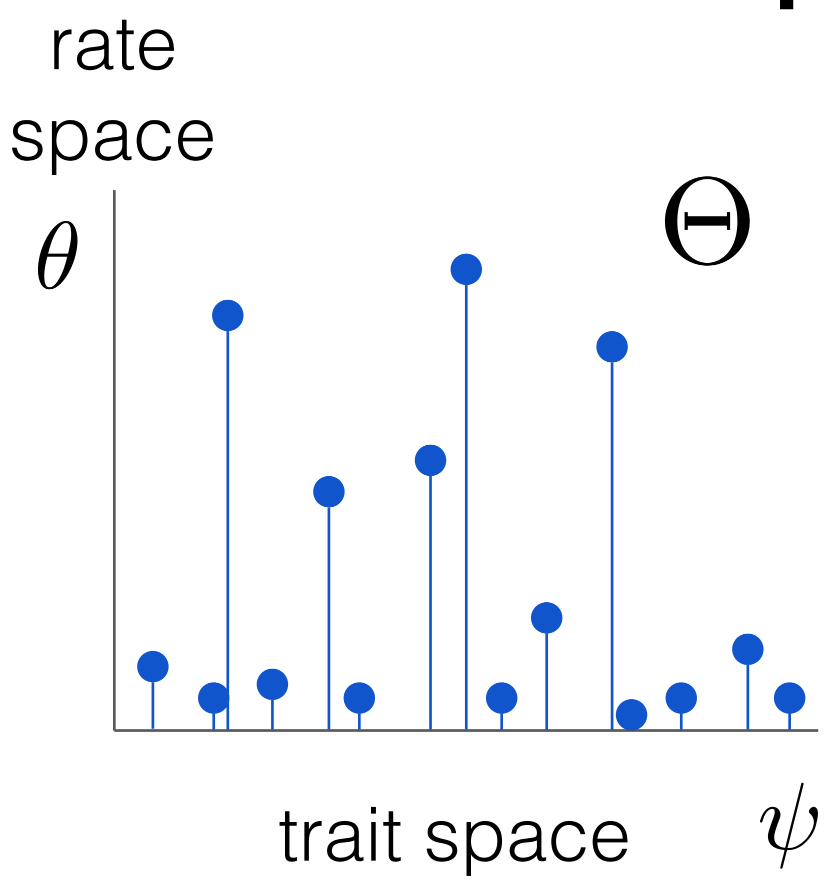
[Kingman 93]

Poisson processes and (N)CRMs

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

Poisson process with intensity measure $\mu(\mathrm{d}\theta \times \mathrm{d}\psi)$

$$= \nu(\mathrm{d}\theta) H(\mathrm{d}\psi)$$



completely random measure (CRM)
(e.g. BP, $\mathbf{\Gamma}$P) $\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$

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

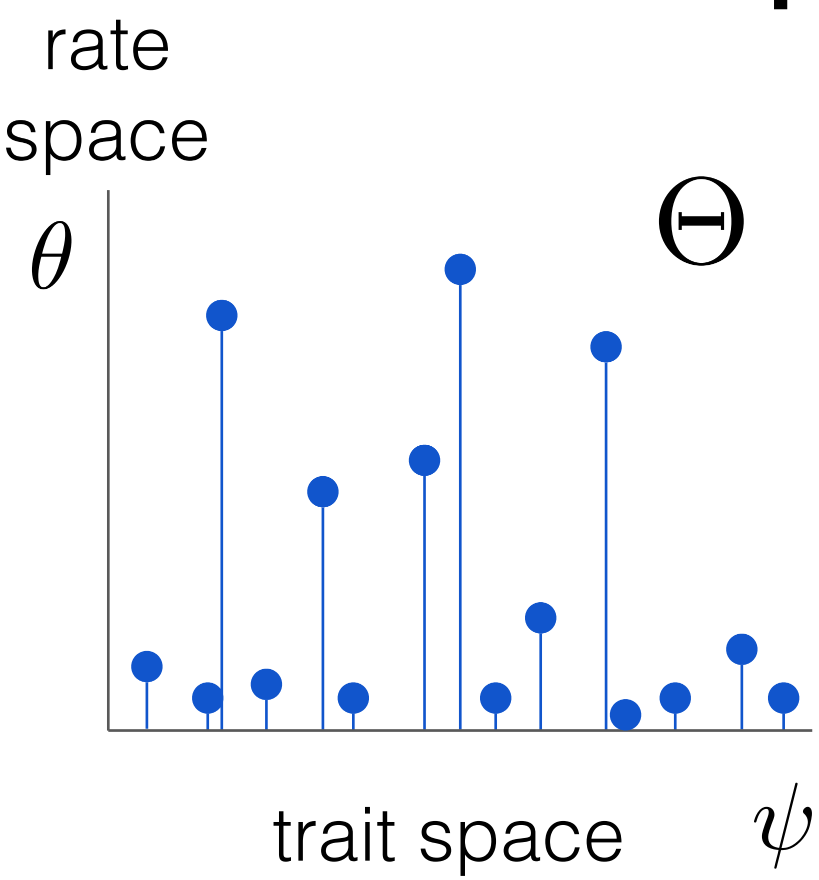
[Kingman 93]

Poisson processes and (N)CRMs

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

Poisson process with intensity measure $\mu(\mathrm{d}\theta \times \mathrm{d}\psi)$

$$= \nu(\mathrm{d}\theta) H(\mathrm{d}\psi)$$



completely random measure (CRM) (e.g. BP, $\mathbf{\Gamma}$P) $\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$

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

Captures a large class of useful priors in BNP

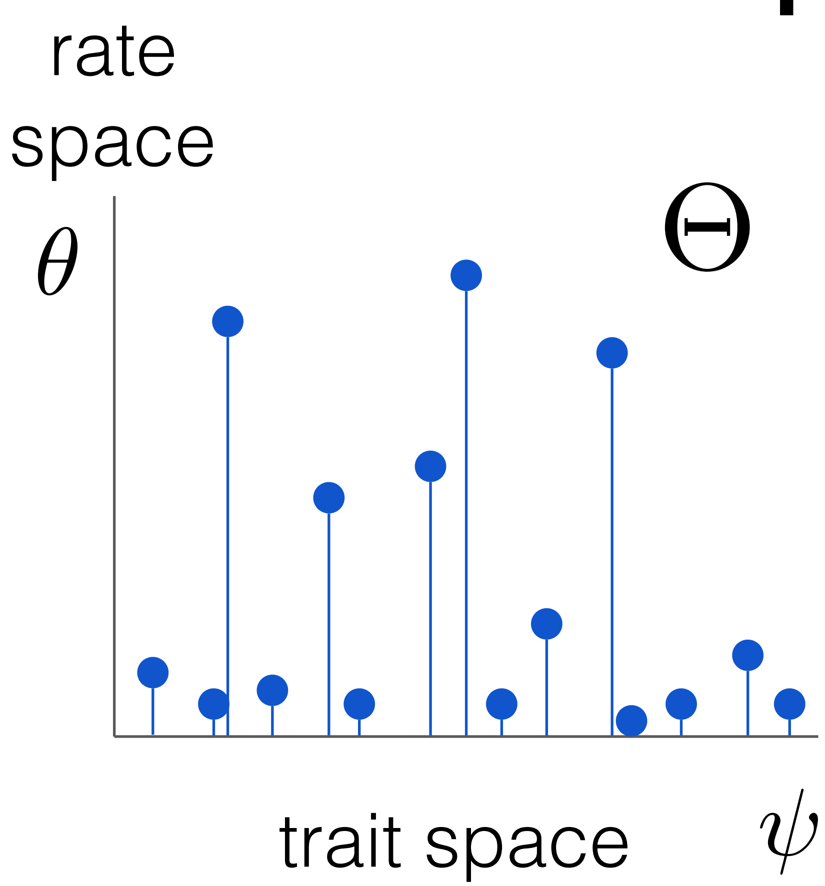
[Kingman 93]

Poisson processes and (N)CRMs

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

Poisson process with intensity measure $\mu(\mathrm{d}\theta \times \mathrm{d}\psi)$

$$= \nu(\mathrm{d}\theta) H(\mathrm{d}\psi)$$



completely random measure (CRM) (e.g. BP, $\boldsymbol{\Gamma}$P) $\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$

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

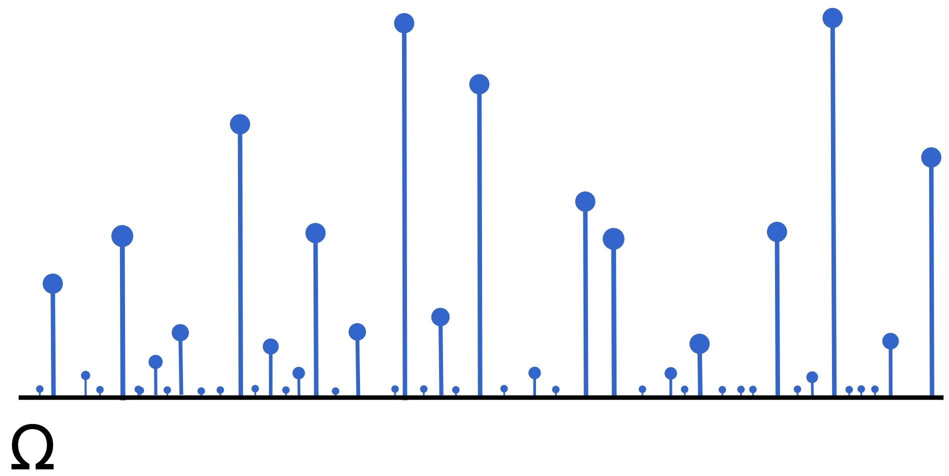
Captures a large class of useful priors in BNP

How do we approximate with finite number of atoms?

[Kingman 93]

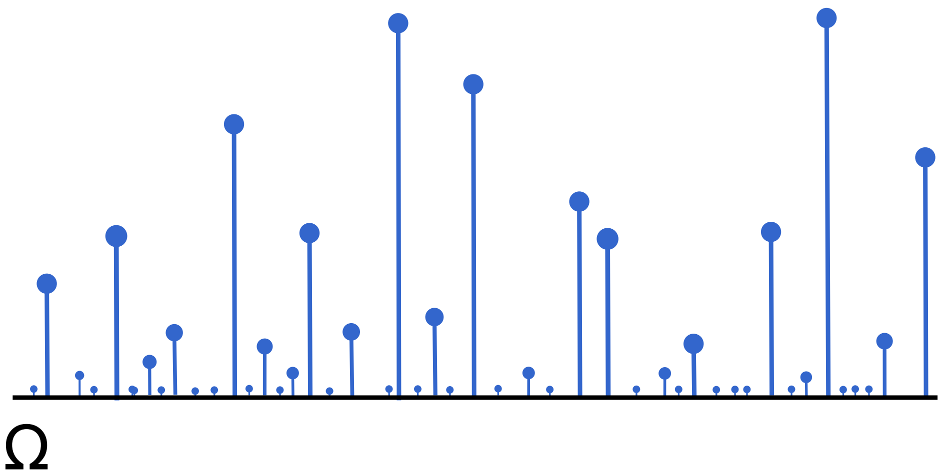
Finite approximation approaches

Finite approximation approaches

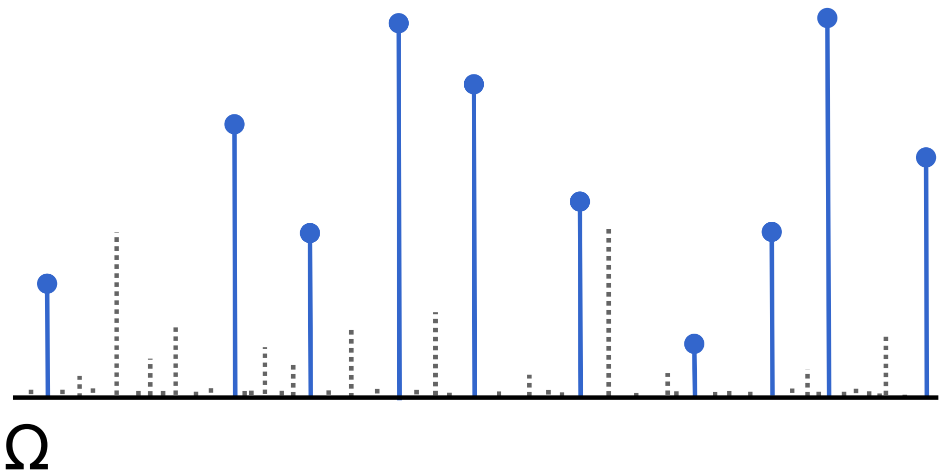


$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

Finite approximation approaches

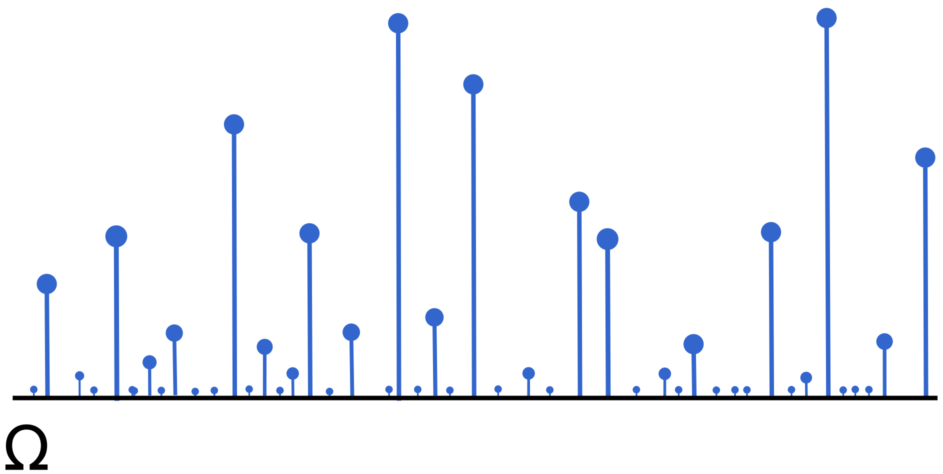


$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

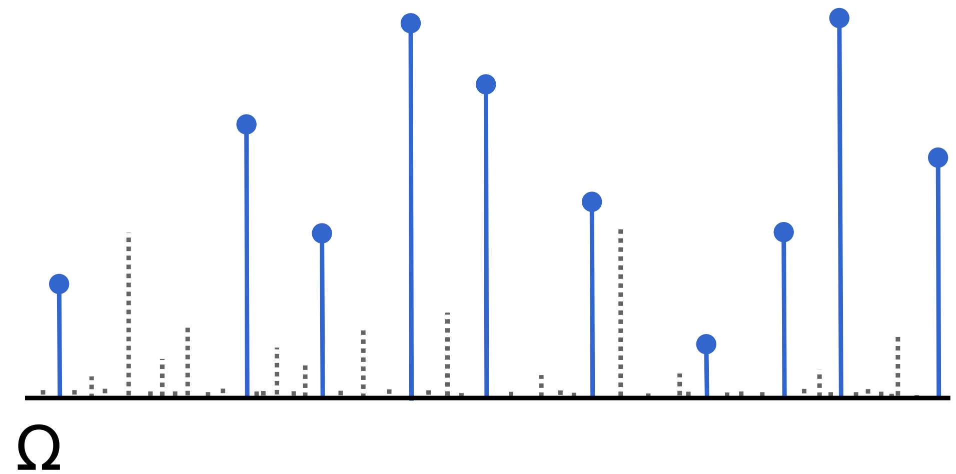


$$\Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

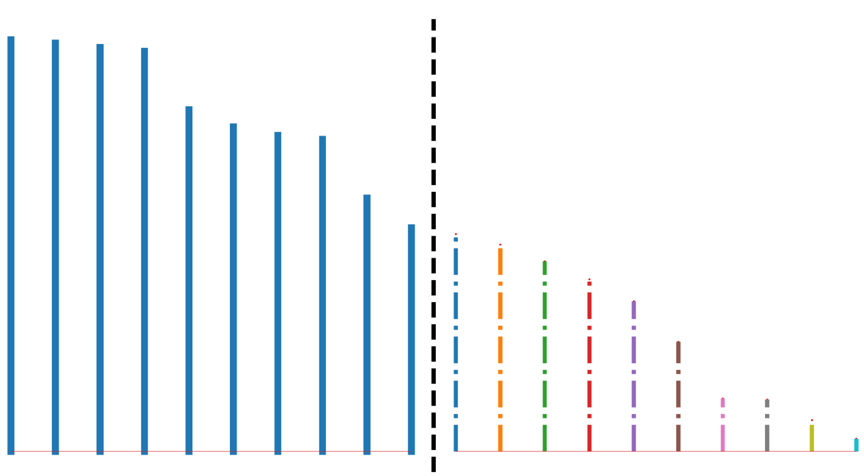
Finite approximation approaches



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



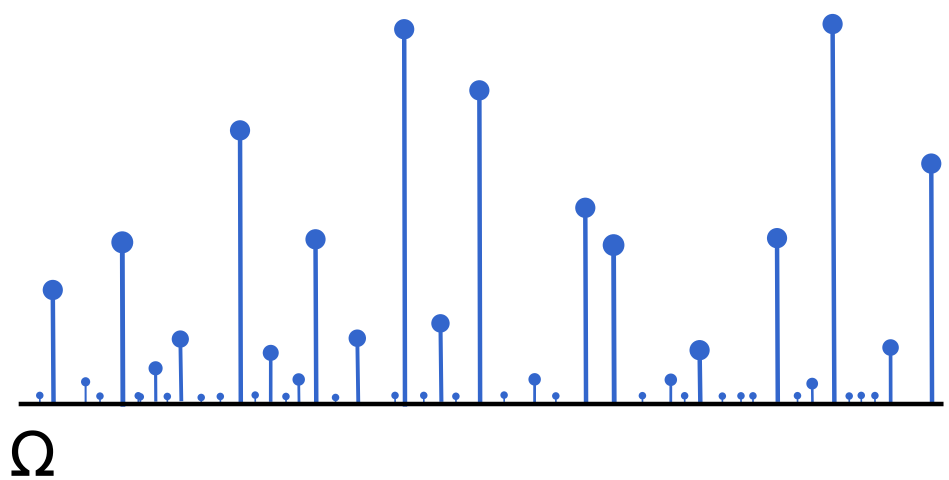
$$\Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$



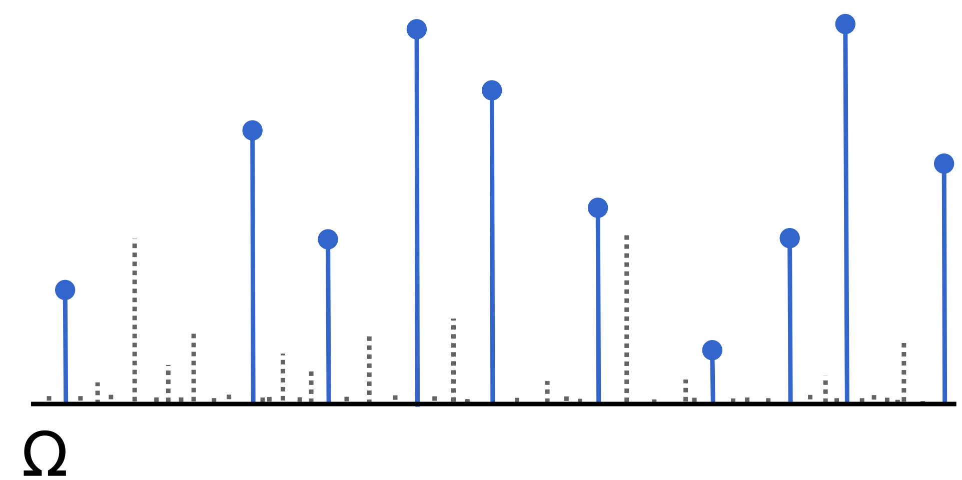
Truncated finite approx.

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

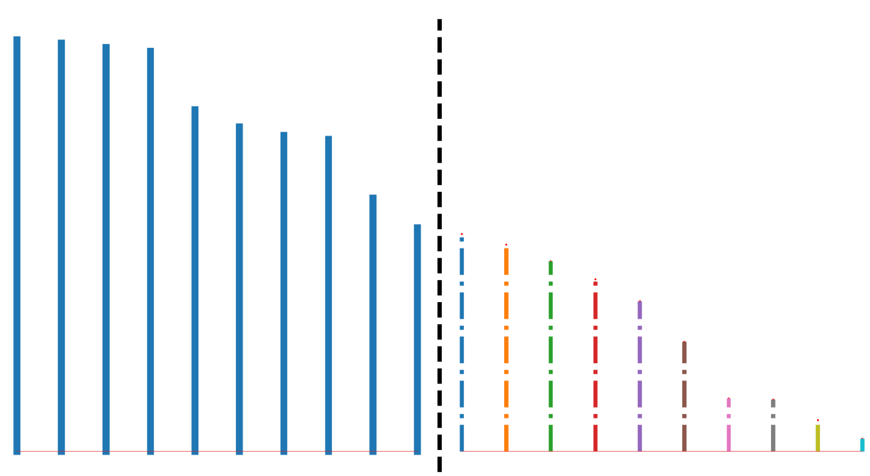
Finite approximation approaches



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

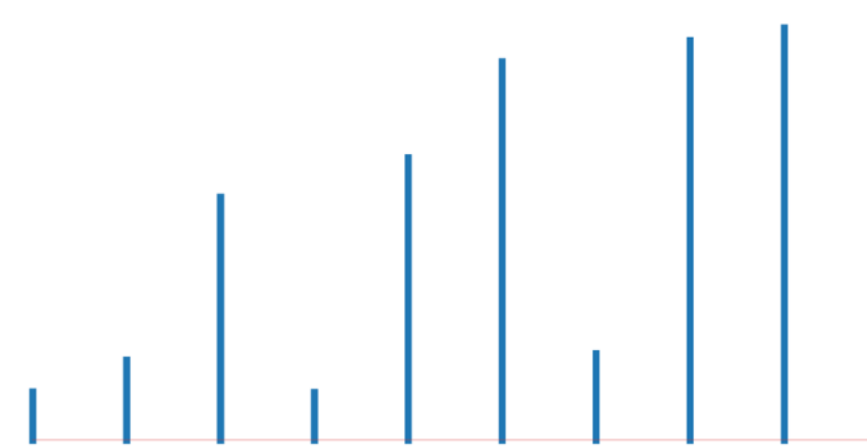


$$\Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$



Truncated finite approx.

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



Non-nested finite approx.

$$\Theta_K = \sum_{k=1}^{K} \theta_{K,k} \delta_{\psi_k}$$

Past work: finite approximations to BNP priors

	Truncated Approximations	Truncation Error Bounds	Non-nested Approximations
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓		
ΓP	✓	✓	✓
(N)CRM	✓		

Past work: finite approximations to BNP priors

	Truncated Approximations	Truncation Error Bounds	Non-nested Approximations
DP	✓ [Sethuraman 94] [Roychowdhury 15]	✓ [Ishwaran 01]	✓ [Ishwaran 02]
BP	✓ [Teh 07] [Paisley 12] [Thibaux 07]	✓ [Doshi-Velez 09] [Paisley 12]	✓ [Paisley 16]
BPP	✓ [Broderick 14]		
ΓP	✓ [Bondesson 82] [Roychowdhury 15]	✓ [Roychowdhury 15]	✓ [Titsias 07]
(N)CRM	✓ [Ferguson 72] [Bondesson 82] [Rosinski 01] [Broderick 14]		

Past work: finite approximations to BNP priors

	Truncated Approximations	Truncation Error Bounds	Non-nested Approximations
DP	✓ [Sethuraman 94] [Roychowdhury 15]	✓ [Ishwaran 01]	✓ [Ishwaran 02]
BP	✓ [Teh 07] [Paisley 12] [Thibaux 07]	✓ [Doshi-Velez 09] [Paisley 12]	✓ [Paisley 16]
BPP	✓ [Broderick 14]		
ΓP	✓ [Bondesson 82] [Roychowdhury 15]	✓ [Roychowdhury 15]	✓ [Titsias 07]
(N)CRM	✓ [Ferguson 72] [Bondesson 82] [Rosinski 01] [Broderick 14]		

Sparse results for a few priors in BNP

Past work: finite approximations to BNP priors

	Truncated Approximations	Truncation Error Bounds	Non-nested Approximations
DP	✓ [Sethuraman 94] [Roychowdhury 15]	✓ [Ishwaran 01]	✓ [Ishwaran 02]
BP	✓ [Teh 07] [Paisley 12] [Thibaux 07]	✓ [Doshi-Velez 09] [Paisley 12]	✓ [Paisley 16]
BPP	✓ [Broderick 14]		
ΓP	✓ [Bondesson 82] [Roychowdhury 15]	✓ [Roychowdhury 15]	✓ [Titsias 07]
(N)CRM	✓ [Ferguson 72] [Bondesson 82] [Rosinski 01] [Broderick 14]		

Sparse results for a few priors in BNP

Incomplete general theory

Outline

- Tractable priors in BNP
- Truncated approximations
  - ➡ **Two forms for sequential representations**
  - Truncation and error analysis
- Non-nested approximations

Ordering of (N)CRM atoms

Ordering of (N)CRM atoms

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

Ordering of (N)CRM atoms

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

Ordering of (N)CRM atoms

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \quad \Rightarrow \quad \Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

2 forms for sequential representations $\nu(\mathrm{d}\theta) H(\mathrm{d}\psi)$

Ordering of (N)CRM atoms

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \quad \Rightarrow \quad \Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

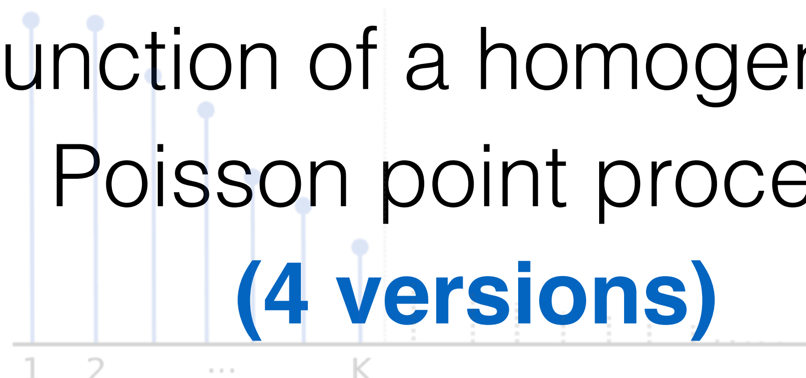
2 forms for sequential representations $\nu(\mathrm{d}\theta) H(\mathrm{d}\psi)$

Series representation

function of a homogenous

Poisson point process

(4 versions)



[Ferguson and Klass 1972, Bondesson 1982, Rosinski 2001]

Ordering of (N)CRM atoms

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \quad \Rightarrow \quad \Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

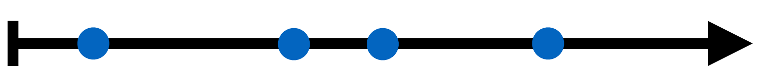
2 forms for sequential representations $\nu(\mathrm{d}\theta)H(\mathrm{d}\psi)$

Series representation

function of a homogenous

Poisson point process

(4 versions)



Γ_k

[Ferguson and Klass 1972, Bondesson 1982, Rosinski 2001]

Ordering of (N)CRM atoms

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \quad \Rightarrow \quad \Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

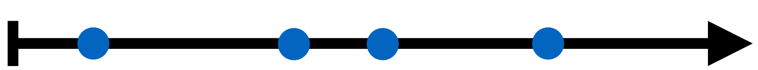
2 forms for sequential representations $\nu(\mathrm{d}\theta)H(\mathrm{d}\psi)$

Series representation

function of a homogenous

Poisson point process

(4 versions)



Γ_k

$V_k \overset{\text{i.i.d.}}{\sim} g$

[Ferguson and Klass 1972, Bondesson 1982, Rosinski 2001]

Ordering of (N)CRM atoms

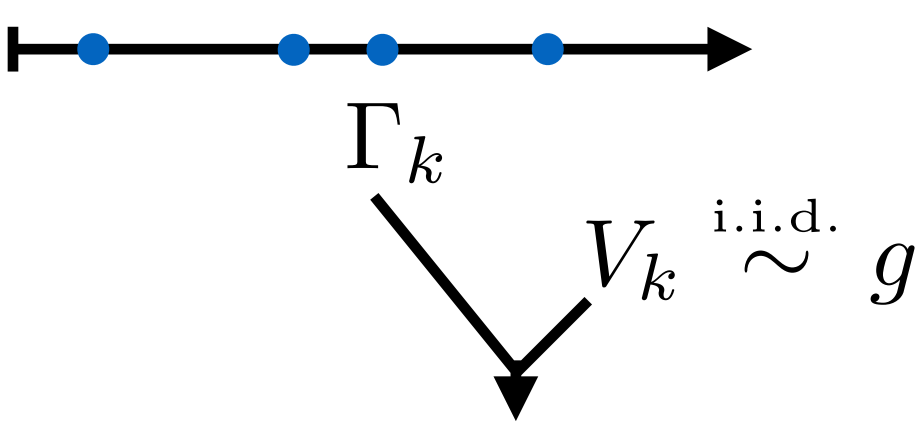
$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \quad \Rightarrow \quad \Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

2 forms for sequential representations $\nu(\mathrm{d}\theta) H(\mathrm{d}\psi)$

Series representation

function of a homogenous Poisson point process

(4 versions)



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

[Ferguson and Klass 1972, Bondesson 1982, Rosinski 2001]

Ordering of (N)CRM atoms

$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \quad \Longrightarrow \quad \Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

2 forms for sequential representations $\nu(\mathrm{d}\theta)H(\mathrm{d}\psi)$

Series representation

function of a homogenous

Poisson point process

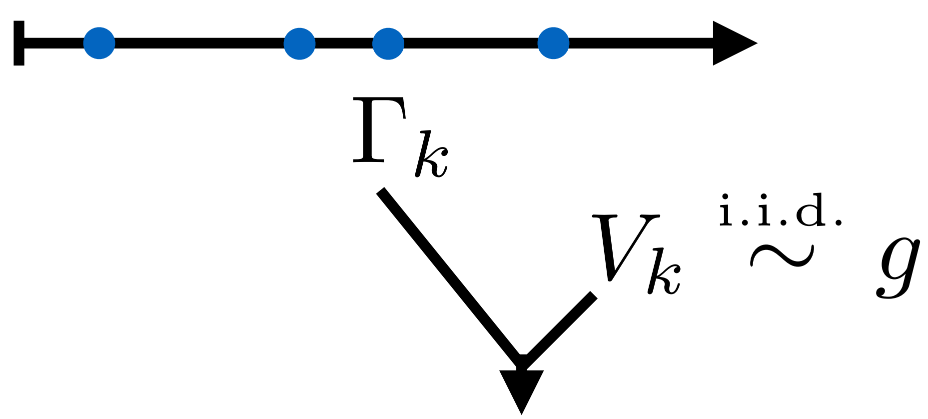
(4 versions)

Superposition representation

infinite sum of CRMs,

each with finite # of atoms

(3 versions)



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

[Ferguson and Klass 1972, Bondesson 1982, Rosinski 2001]

[James 2014, Broderick et al 2017]

Ordering of (N)CRM atoms

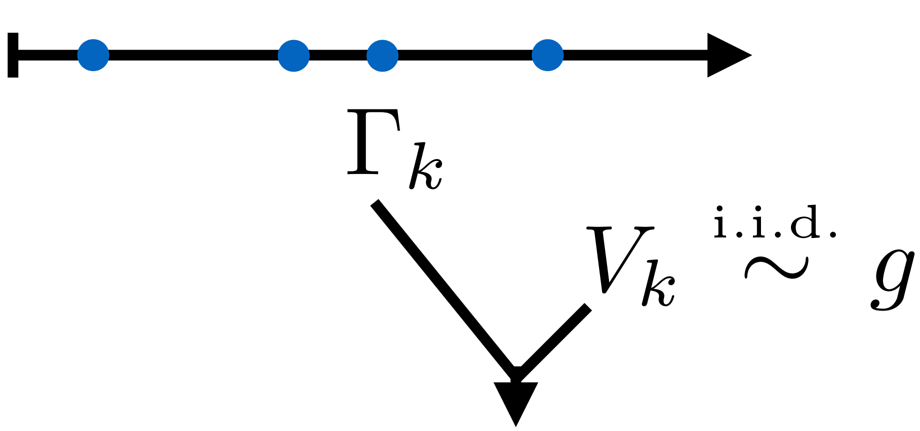
$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \quad \Rightarrow \quad \Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

2 forms for sequential representations $\nu(\mathrm{d}\theta)H(\mathrm{d}\psi)$

Series representation

function of a homogenous Poisson point process

(4 versions)



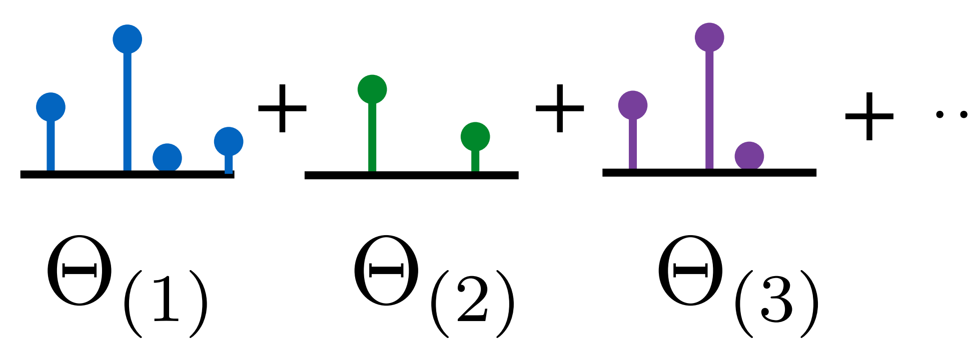
$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

[Ferguson and Klass 1972, Bondesson 1982, Rosinski 2001]

Superposition representation

infinite sum of CRMs, each with finite # of atoms

(3 versions)



[James 2014, Broderick et al 2017]

Ordering of (N)CRM atoms

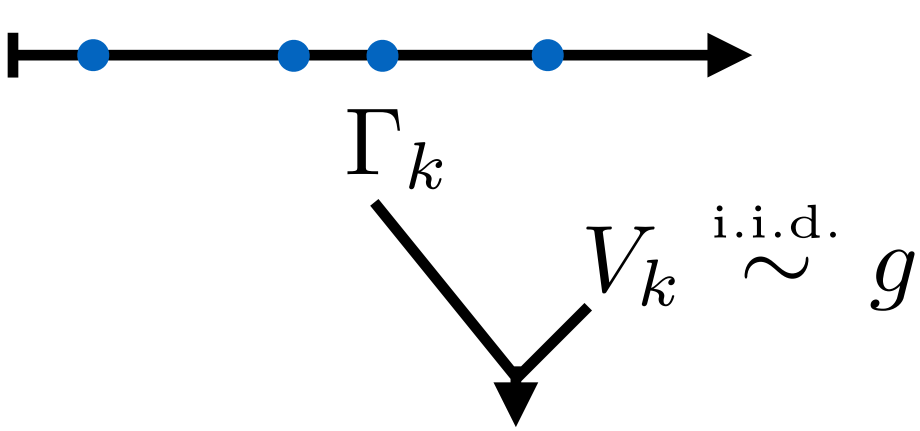
$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \quad \Rightarrow \quad \Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

2 forms for sequential representations $\nu(\mathrm{d}\theta) H(\mathrm{d}\psi)$

Series representation

function of a homogenous Poisson point process

(4 versions)



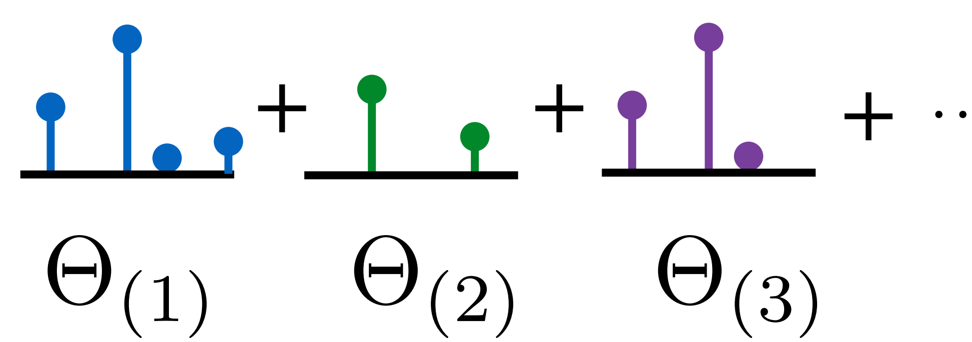
$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}$$

[Ferguson and Klass 1972, Bondesson 1982, Rosinski 2001]

Superposition representation

infinite sum of CRMs, each with finite # of atoms

(3 versions)



$$\Theta = \sum_{k=1}^{\infty} \Theta_{(k)}$$

[James 2014, Broderick et al 2017]

Ordering of (N)CRM atoms

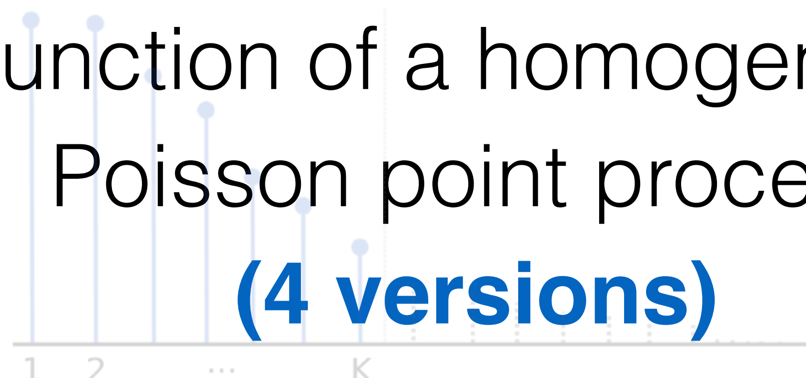
$$\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k} \quad \Rightarrow \quad \Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$

2 forms for sequential representations $\nu(\mathrm{d}\theta)H(\mathrm{d}\psi)$

Series representation

function of a homogenous Poisson point process

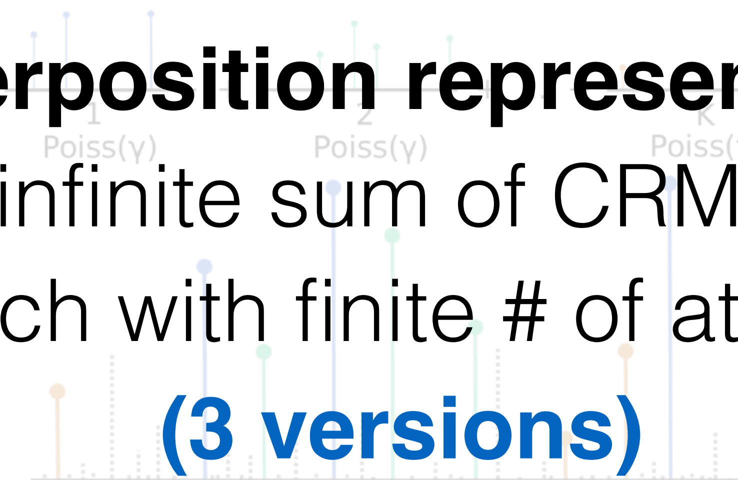
(4 versions)



Superposition representation

infinite sum of 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	✓	✓	✓/✗	✗	✓	✓	✗
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!

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

Outline

✓ Tractable priors in BNP

- Truncated approximations

    ✓ Two forms for sequential representations

    ➡ **Truncation and error analysis**

- Non-nested approximations

Choosing between the seven representations

How close is our finite approximation?

Choosing between the seven representations

How close is our finite approximation?

$$\Pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\Pi_0(\mathrm{d}\Theta)$$

Choosing between the seven representations

How close is our finite approximation?

$$\Pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\Pi_0(\mathrm{d}\Theta)$$

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

Choosing between the seven representations

How close is our finite approximation?

$$\Pi(\mathrm{d}\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\Pi_0(\mathrm{d}\Theta)$$

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

full infinite Θ

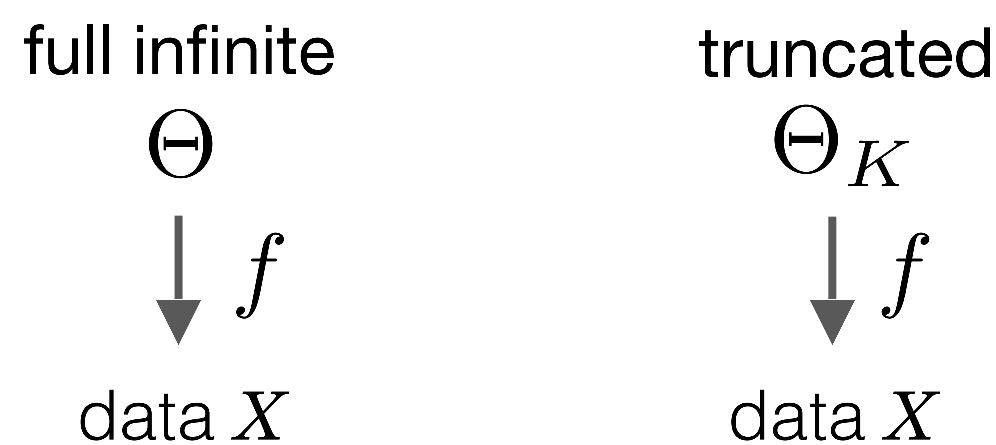
truncated Θ_K

Choosing between the seven representations

How close is our finite approximation?

$$\Pi(\mathrm{d}\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\Pi_0(\mathrm{d}\Theta)$$

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

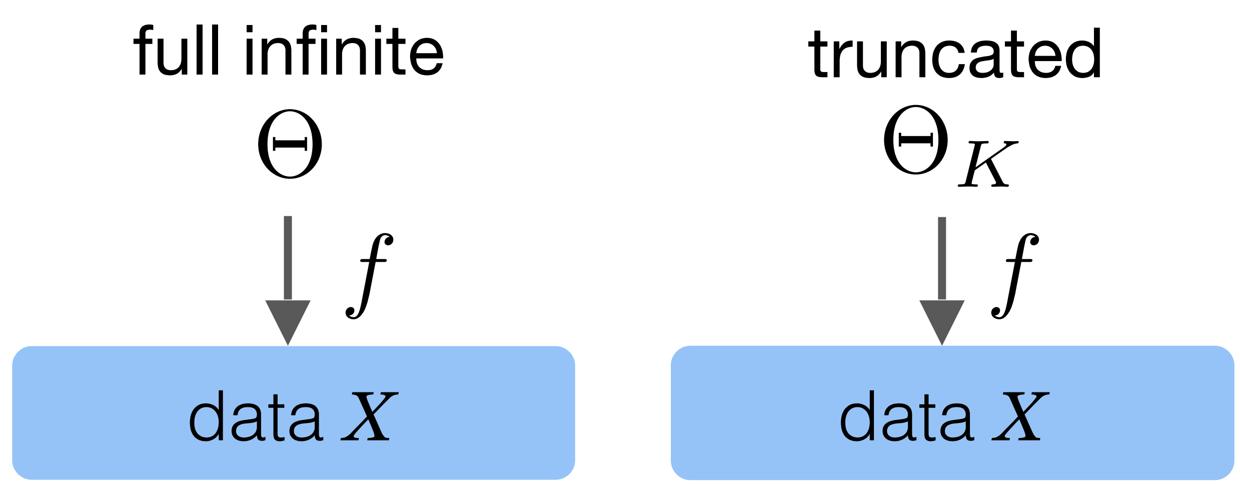


Choosing between the seven representations

How close is our finite approximation?

$$\Pi(\mathrm{d}\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\Pi_0(\mathrm{d}\Theta)$$

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



Compare the distribution of the data under full vs. truncated

Choosing between the seven representations

How close is our finite approximation?

$$\Pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\Pi_0(\mathrm{d}\Theta)$$

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

Choosing between the seven representations

How close is our finite approximation?

$$\Pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\Pi_0(\mathrm{d}\Theta)$$

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

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

Choosing between the seven representations

How close is our finite approximation?

$$\Pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\Pi_0(\mathrm{d}\Theta)$$

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

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?

$$\Pi(\mathrm{d}\Theta \,|\, X) \propto_{\Theta} f(X \,|\, \Theta)\Pi_0(\mathrm{d}\Theta)$$

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

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?

$$\Pi(\mathrm{d}\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\Pi_0(\mathrm{d}\Theta)$$

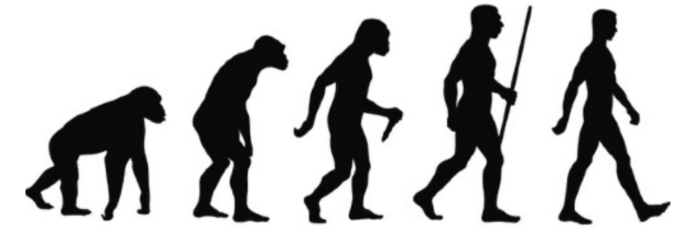
Truncation error: $\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2}\int |p_{N,\infty}(X) - p_{N,K}(X)|\mathrm{d}X$

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

As N gets larger, error increases

As K gets larger, error decreases

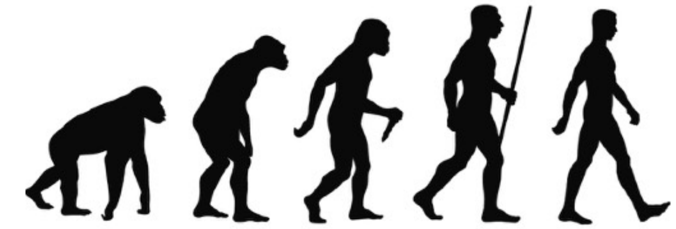
We develop **new upper bounds**

Protobound 

Leads to all the other truncation error bounds in this work

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

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

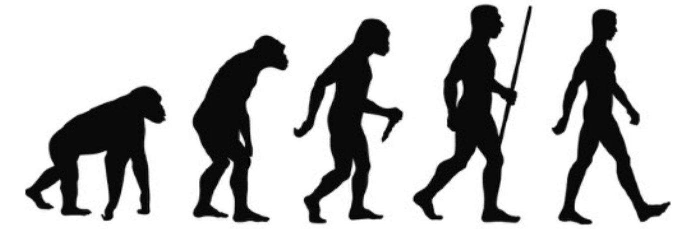
Protobound 

Leads to all the other truncation error bounds in this work

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

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

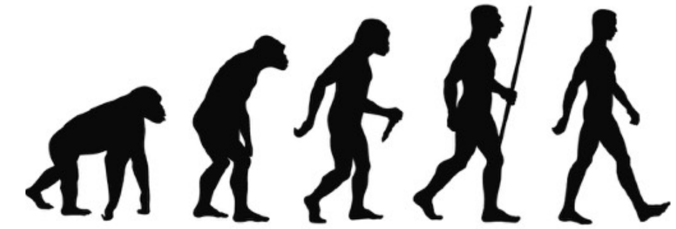
Proposition (HCHB). The protobound is tight

Protobound 

Leads to all the other truncation error bounds in this work

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

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

Protobound 

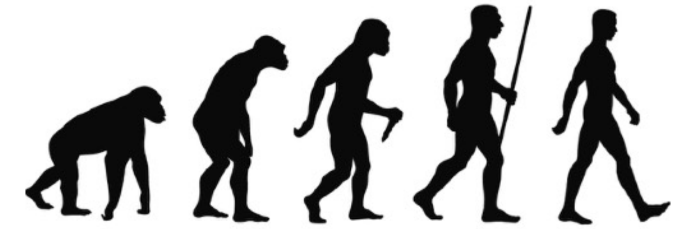
Leads to all the other truncation error bounds in this work

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

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

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

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

Protobound 

Leads to all the other truncation error bounds in this work

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

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

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

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

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

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

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

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

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

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!

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

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!

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

$$N \to \infty, \mathrm{bound} \to 1$$

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!

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

$$N \to \infty, \mathrm{bound} \to 1 \qquad K \to \infty, \mathrm{bound} \to 0$$

Outline

- ✓ Tractable priors in BNP
- ✓ Truncated approximations
    - ✓ Two forms for sequential representations
    - ✓ Truncation and error analysis
- ➡ **Non-nested approximations**

Non-nested CRM approximations

Atom weights are independent

$$\Theta_K = \sum_{k=1}^{K} \theta_{K,k} \delta_{\psi_k}, \quad \theta_{K,k} \overset{\text{ind}}{\sim} \nu_K$$

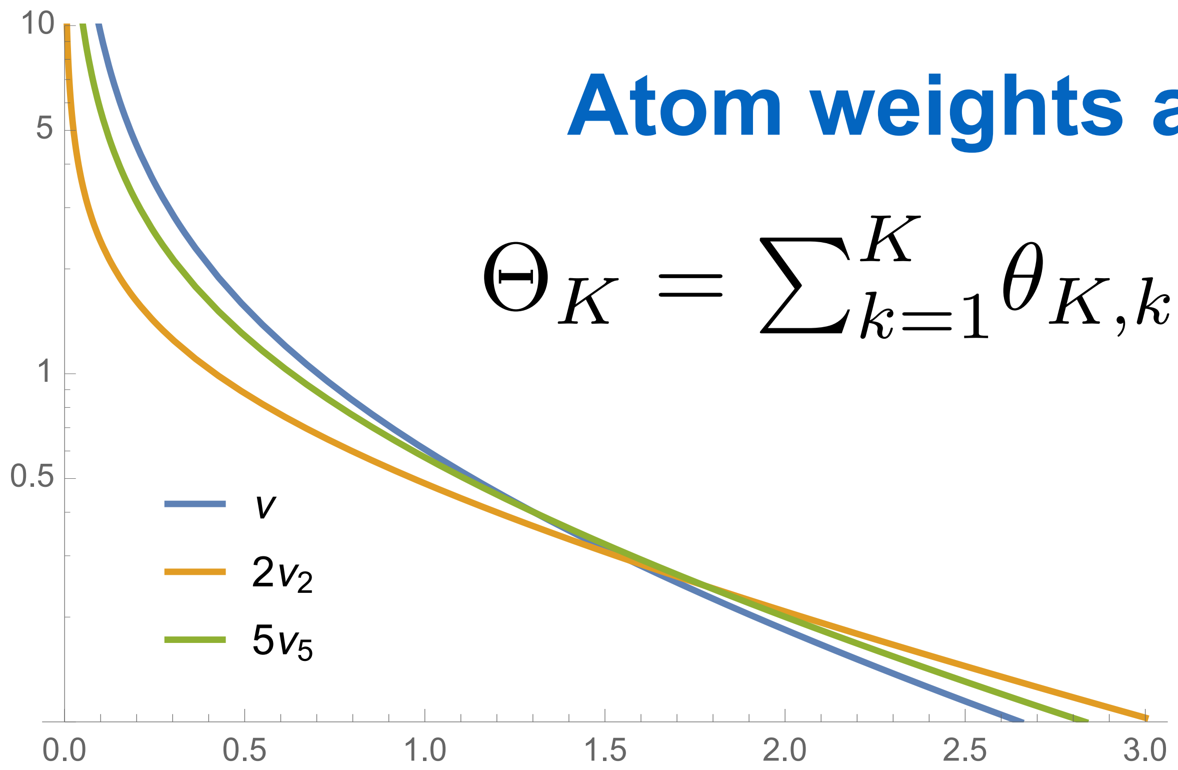
Non-nested CRM approximations

Atom weights are independent

$$\Theta_K = \sum_{k=1}^{K} \theta_{K,k} \delta_{\psi_k}, \quad \theta_{K,k} \overset{\text{ind}}{\sim} \nu_K$$

Intuition: choose ν_K such that
$K\nu_K(\theta) \approx \nu(\theta)$ and $K\nu_K \to \nu$

Non-nested CRM approximations

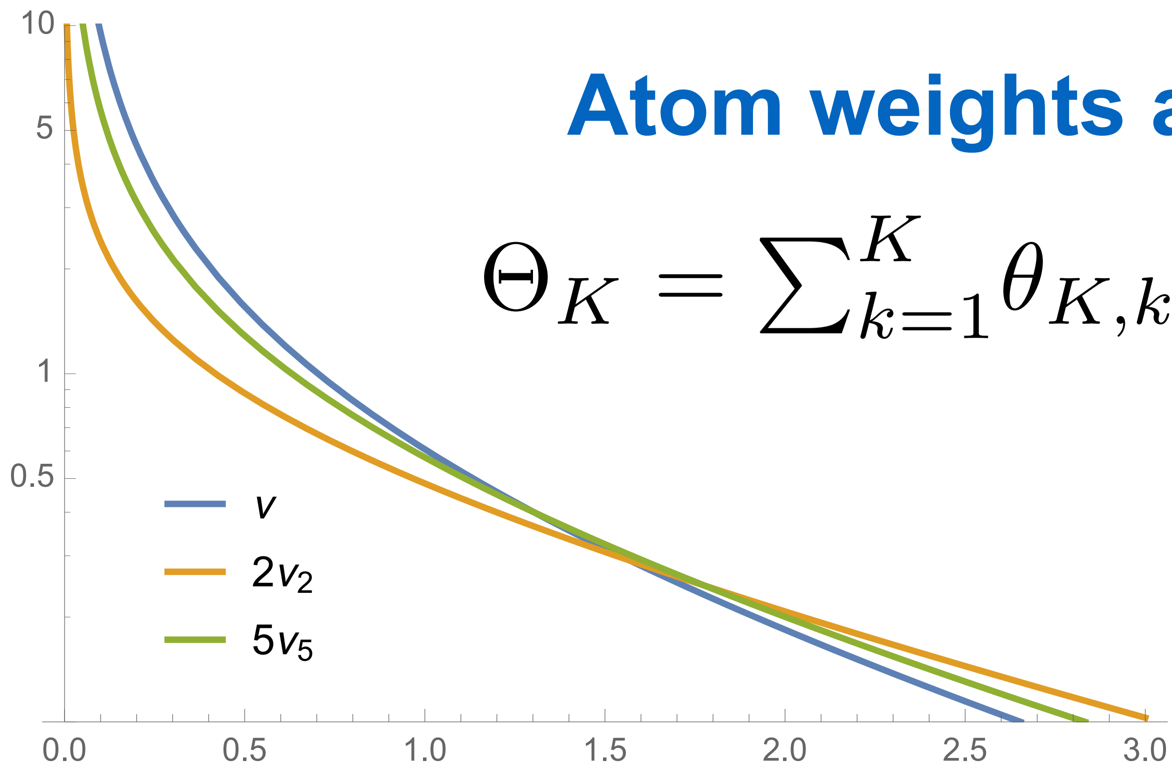


Atom weights are independent

$$\Theta_K = \sum_{k=1}^{K} \theta_{K,k} \delta_{\psi_k}, \quad \theta_{K,k} \overset{\text{ind}}{\sim} \nu_K$$

Intuition: choose ν_K such that $K\nu_K(\theta) \approx \nu(\theta)$ and $K\nu_K \to \nu$

Non-nested CRM approximations



Atom weights are independent

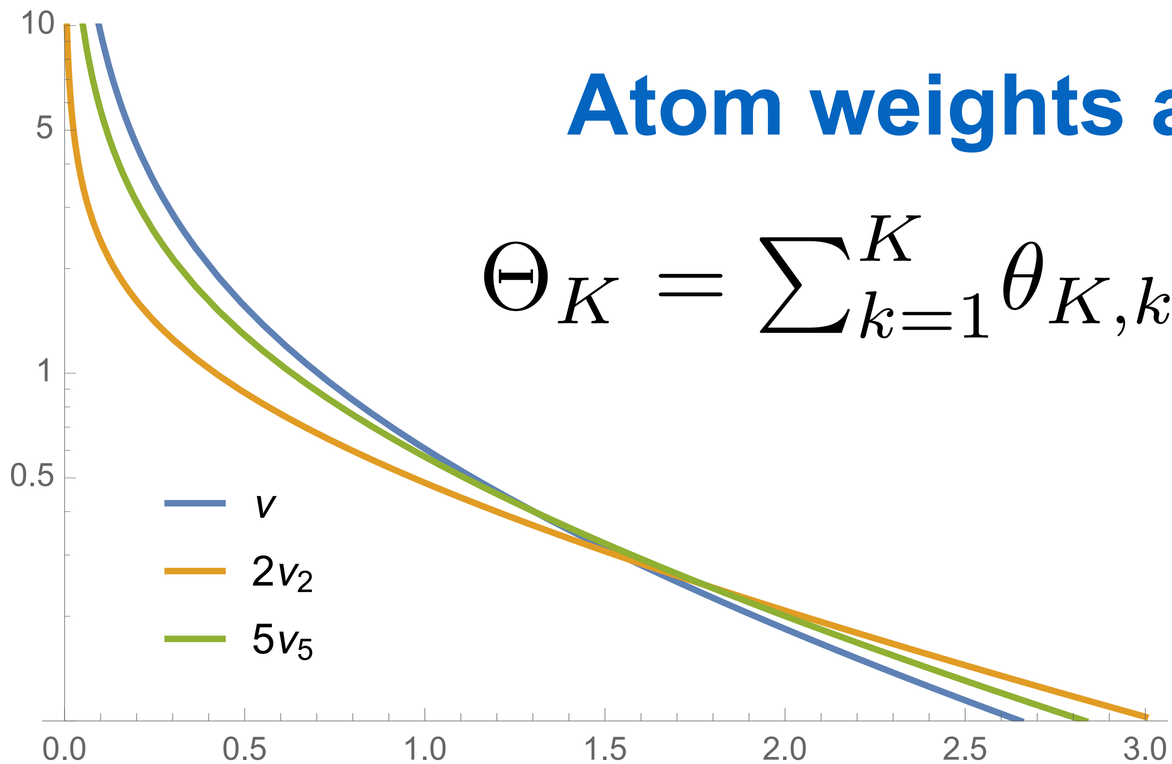
$$\Theta_K = \sum_{k=1}^{K} \theta_{K,k} \delta_{\psi_k}, \quad \theta_{K,k} \overset{\text{ind}}{\sim} \nu_K$$

Intuition: choose ν_K such that $K\nu_K(\theta) \approx \nu(\theta)$ and $K\nu_K \to \nu$

Theorem (H., Masoero, Mackey, Broderick). Assume that

$$\nu(\mathrm{d}\theta; \gamma, d, \eta) = \gamma\theta^{-1-d} g(\theta)^{-d} \frac{h(\theta;\eta)}{Z(1-d,\eta)} \mathrm{d}\theta.$$

Non-nested CRM approximations



Atom weights are independent

$$\Theta_K = \sum_{k=1}^{K} \theta_{K,k} \delta_{\psi_k}, \quad \theta_{K,k} \overset{\text{ind}}{\sim} \nu_K$$

Intuition: choose ν_K such that $K\nu_K(\theta) \approx \nu(\theta)$ and $K\nu_K \to \nu$

Theorem (H., Masoero, Mackey, Broderick). Assume that

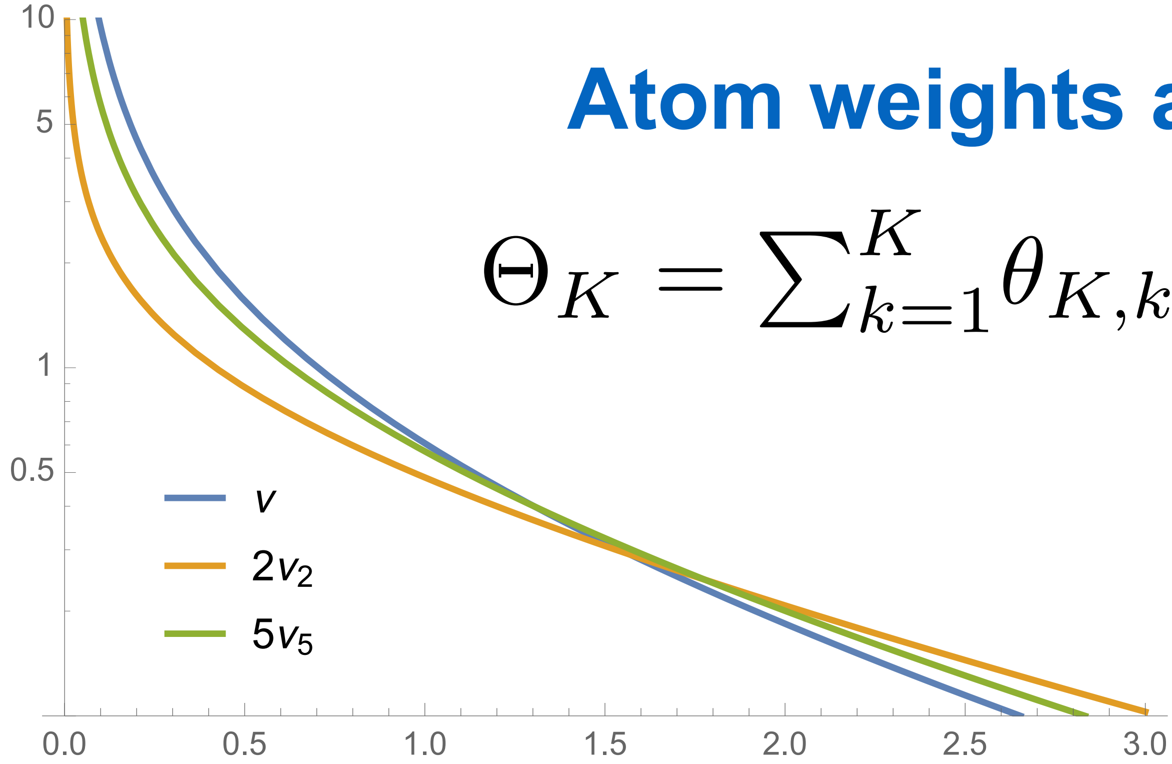
$$\nu(\mathrm{d}\theta; \gamma, d, \eta) = \gamma \theta^{-1-d} g(\theta)^{-d} \frac{h(\theta;\eta)}{Z(1-d,\eta)} \mathrm{d}\theta.$$

Then, under mild regularity conditions, when $d = 0$

$$\nu_K(d\theta) = \theta^{-1+c/K} g(\theta)^{c/K} \frac{h(\theta;\eta)}{Z(c/K,\eta)} d\theta,$$

where $c \triangleq \gamma \frac{h(0;\eta)}{Z(1,\eta)}$.

Non-nested CRM approximations



Atom weights are independent

$$\Theta_K = \sum_{k=1}^{K} \theta_{K,k} \delta_{\psi_k}, \quad \theta_{K,k} \overset{\text{ind}}{\sim} \nu_K$$

Intuition: choose ν_K such that $K\nu_K(\theta) \approx \nu(\theta)$ and $K\nu_K \to \nu$

Theorem (H., Masoero, Mackey, Broderick). Assume that

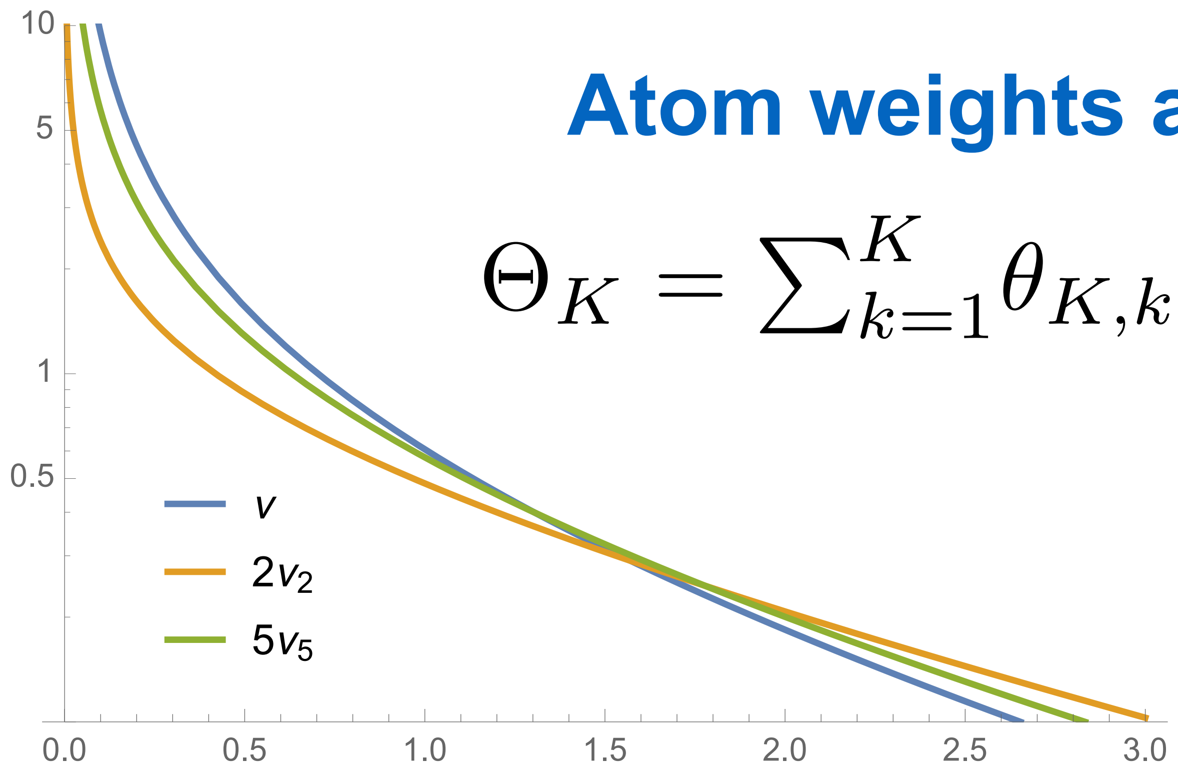
$$\nu(\mathrm{d}\theta; \gamma, d, \eta) = \gamma\theta^{-1-d} g(\theta)^{-d} \frac{h(\theta;\eta)}{Z(1-d,\eta)} \mathrm{d}\theta.$$

Then, under mild regularity conditions, when $d = 0$

$$\nu_K(d\theta) = \theta^{-1+c/K} g(\theta)^{c/K} \frac{h(\theta;\eta)}{Z(c/K,\eta)} d\theta,$$

where $c \triangleq \gamma \frac{h(0;\eta)}{Z(1,\eta)}$.

Non-nested CRM approximations



Atom weights are independent

$$\Theta_K = \sum_{k=1}^{K} \theta_{K,k} \delta_{\psi_k}, \quad \theta_{K,k} \overset{\text{ind}}{\sim} \nu_K$$

Intuition: choose ν_K such that $K\nu_K(\theta) \approx \nu(\theta)$ and $K\nu_K \to \nu$

Theorem (H., Masoero, Mackey, Broderick). Assume that

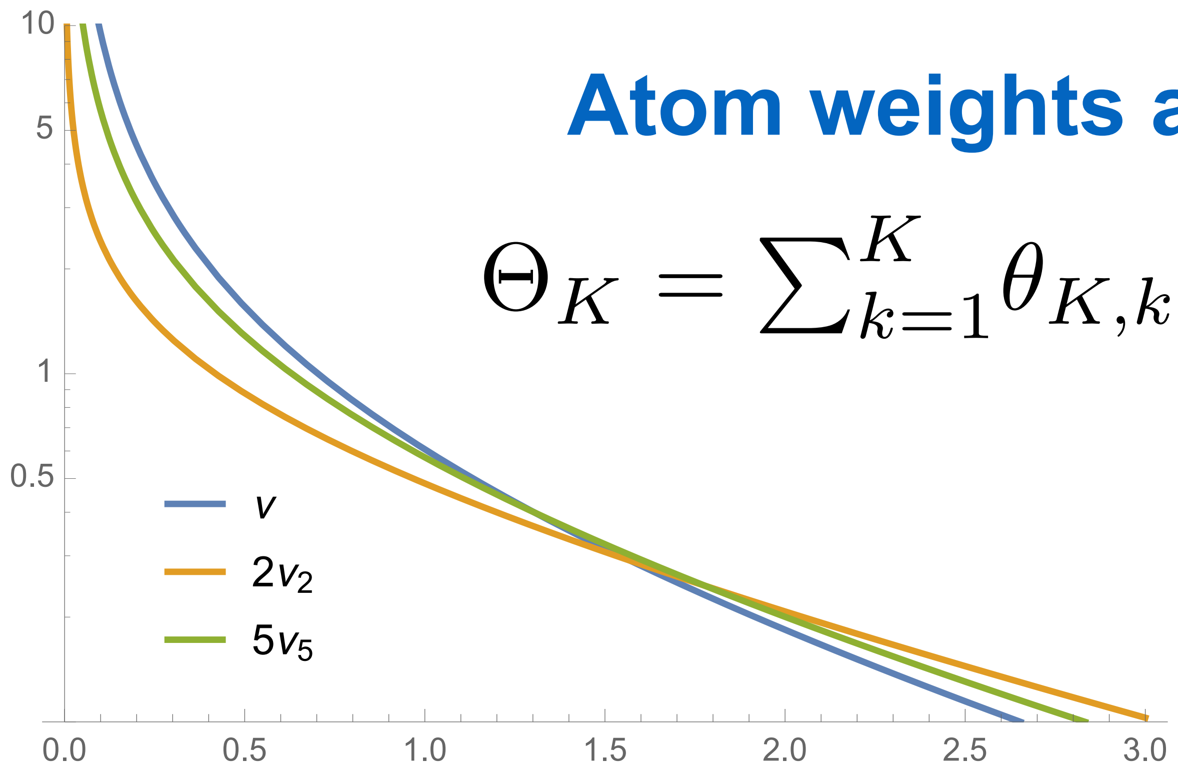
$$\nu(\mathrm{d}\theta; \gamma, d, \eta) = \gamma\theta^{-1-d} g(\theta)^{-d} \frac{h(\theta;\eta)}{Z(1-d,\eta)} \mathrm{d}\theta.$$

Then, under mild regularity conditions, when $d = 0$

$$\nu_K(d\theta) = \theta^{-1+c/K} g(\theta)^{c/K} \frac{h(\theta;\eta)}{Z(c/K,\eta)} d\theta,$$

where $c \triangleq \gamma \frac{h(0;\eta)}{Z(1,\eta)}$.

Non-nested CRM approximations



Atom weights are independent

$$\Theta_K = \sum_{k=1}^{K} \theta_{K,k} \delta_{\psi_k}, \quad \theta_{K,k} \overset{\text{ind}}{\sim} \nu_K$$

Intuition: choose ν_K such that $K\nu_K(\theta) \approx \nu(\theta)$ and $K\nu_K \to \nu$

Theorem (H., Masoero, Mackey, Broderick). Assume that

$$\nu(\mathrm{d}\theta; \gamma, d, \eta) = \gamma \theta^{-1-d} g(\theta)^{-d} \frac{h(\theta;\eta)}{Z(1-d,\eta)} \mathrm{d}\theta.$$

Then, under mild regularity conditions, when $d = 0$

$$\nu_K(d\theta) = \theta^{-1+c/K} g(\theta)^{c/K} \frac{h(\theta;\eta)}{Z(c/K,\eta)} d\theta,$$

where $c \triangleq \gamma \frac{h(0;\eta)}{Z(1,\eta)}$.

Outline

- ✓ Tractable priors in BNP
- ✓ Truncated approximations
  - ✓ Two forms for sequential representations
  - ✓ Truncation and error analysis
- ✓ Non-nested approximations

Conclusion

Previous Work	Truncated Approximations	Truncation Error Bounds	Non-nested Approximations
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓		
ΓP	✓	✓	✓
(N)CRM	✓		

Conclusion

Our Work	Truncated Approximations	Truncation Error Bounds	Non-nested Approximations
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓	✓	✓
ΓP	✓	✓	✓
(N)CRM	✓	✓	✓

Conclusion

Our Work	Truncated Approximations	Truncation Error Bounds	Non-nested Approximations
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓	✓	✓
ΓP	✓	✓	✓
(N)CRM	✓	✓	✓

- Large family of BNP priors that admit efficient inference

Conclusion

Our Work	Truncated Approximations	Truncation Error Bounds	Non-nested Approximations
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓	✓	✓
ΓP	✓	✓	✓
(N)CRM	✓	✓	✓

- Large family of BNP priors that admit efficient inference
- Use of "modern" inference methods (e.g. HMC and VB)

Conclusion

Our Work	Truncated Approximations	Truncation Error Bounds	Non-nested Approximations
DP	✓	✓	✓
BP	✓	✓	✓
BPP	✓	✓	✓
ΓP	✓	✓	✓
(N)CRM	✓	✓	✓

- Large family of BNP priors that admit efficient inference
- Use of “modern” inference methods (e.g. HMC and VB)
- Trade off computational efficiency and statistical accuracy

Conclusion

J. Huggins*, T. Campbell*, J. How, T. Broderick
Truncated random measures
Bernoulli, to appear
Available online: https://arxiv.org/abs/1603.00861

J. Huggins, L. Masoero, L. Mackey, T. Broderick
Generic finite approximations for practical Bayesian nonparametrics
NIPS Workshop on Advances in Approximate Bayesian Inference, 2017
Available online: http://approximateinference.org/2017/accepted/HugginsEtAl2017.pdf

- Large family of BNP priors that admit efficient inference
- Use of “modern” inference methods (e.g. HMC and VB)
- Trade off computational efficiency and statistical accuracy