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)
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

*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

**Bayesian nonparametrics:** achieves growing model size via infinite parameters

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

movie text medicine robotics genetics
finance astronomy traffic agriculture pathology

[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

Bayesian nonparametrics: achieves growing model size via infinite parameters

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

movie  text  medicine  robotics  genetics

finance  astronomy  traffic  agriculture  pathology

hard work!

[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

**Bayesian nonparametrics:**
achieves growing model size via infinite parameters

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

- movie
- text
- medicine
- robotics
- genetics
- finance
- astronomy
- traffic
- agriculture
- pathology

**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**

All BNP priors

Previously studied 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

All BNP priors
Previously studied priors
with finite approx (past work)
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

<table>
<thead>
<tr>
<th></th>
<th>Finite Approximation</th>
<th>Approximation Error Bounds</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BPP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΓP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(N)CRM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Past work: finite approximations to BNP priors

<table>
<thead>
<tr>
<th></th>
<th>Finite Approximation</th>
<th>Approximation Error Bounds</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>✓ [Sethuraman 94]</td>
<td>✓ [Ishwaran 01]</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>[Roychowdhury 15]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>✓ [Teh 07]</td>
<td>✓ [Doshi-Velez 09]</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>[Paisley 12]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Thibaux 07]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPP</td>
<td>✓ [Broderick 14]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΓP</td>
<td>✓ [Bondesson 82]</td>
<td>✓ [Roychowdhury 15]</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>[Roychowdhury 15]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N)CRM</td>
<td>✓ [Broderick 14]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Past work: finite approximations to BNP priors

<table>
<thead>
<tr>
<th></th>
<th>Finite Approximation</th>
<th>Approximation Error Bounds</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Sethuraman 94]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Roychowdhury 15]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Ishwaran 01]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>[Teh 07]</td>
<td>[Ishwaran 15]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Paisley 12]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Doshi-Velez 09]</td>
<td>[Paisley 15]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPP</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Broderick 14]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Bondesson 82]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Roychowdhury 15]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Roychowdhury 15]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N)CRM</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[Broderick 14]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sparse results for a few priors in BNP
Past work: finite approximations to BNP priors

<table>
<thead>
<tr>
<th></th>
<th>Finite Approximation</th>
<th>Approximation Error Bounds</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BPP</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΓP</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>(N)CRM</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

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

\[ \sum_{k=1}^{K} \theta_k \delta \psi_k \]

Truncation and error analysis

\[ \sum_{k=1}^{\infty} \theta_k \delta \psi_k \]
Truncation Roadmap

Tractable models in BNP

Two forms for sequential representations

\[ \sum_{k=1}^{K} \theta_k \delta \psi_k \]

Truncation and error analysis

\[ \sum_{k=1}^{\infty} \theta_k \delta \psi_k \]
The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th></th>
<th>sports</th>
<th>politics</th>
<th>food</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>343</td>
<td></td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>Doc 2</td>
<td></td>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 3</td>
<td>854</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 4</td>
<td>342</td>
<td>584</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.5</td>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>
## The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th>Doc</th>
<th>Words</th>
<th>Sports</th>
<th>Politics</th>
<th>Food</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>532</td>
<td>343</td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 2</td>
<td>210</td>
<td>210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 3</td>
<td>854</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 4</td>
<td>926</td>
<td>342</td>
<td>584</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0.7</th>
<th>0.5</th>
<th>0.2</th>
</tr>
</thead>
</table>

**Topic Space**

**Frequency Space**
### The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th>Topic</th>
<th>Frequency</th>
<th>Topic</th>
<th>Frequency</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>sports</td>
<td>343</td>
<td>politics</td>
<td>189</td>
<td>food</td>
</tr>
<tr>
<td>Doc 1 (532 words)</td>
<td></td>
<td>Doc 2 (210 words)</td>
<td></td>
<td>Doc 3 (854 words)</td>
</tr>
<tr>
<td>854</td>
<td>342</td>
<td>584</td>
<td>0.7</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Frequency space and topic space are shown in the diagram.
The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th></th>
<th>sports</th>
<th>politics</th>
<th>food</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1 (532 words)</td>
<td>343</td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 2 (210 words)</td>
<td>210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 3 (854 words)</td>
<td>854</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 4 (926 words)</td>
<td>342</td>
<td>584</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Keywords:
- sports
- politics
- food
- ...

Frequency space

Topic space

Values:
- 0.7
- 0.5
- 0.2
The Standard Model in BNP (By Example)

```
| Doc 1 (532 words) | 343 | 189 |
| Doc 2 (210 words) | 210 |
| Doc 3 (854 words) | 854 |
| Doc 4 (926 words) | 342 | 584 |
```

**Frequency space**

- Sports: 0.7
- Politics: 0.5
- Food: 0.2

**Topic space**

- Sports
- Politics
- Food

---

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

<table>
<thead>
<tr>
<th></th>
<th>sports</th>
<th>politics</th>
<th>food</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1 (532 words)</td>
<td>343</td>
<td></td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>Doc 2 (210 words)</td>
<td></td>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 3 (854 words)</td>
<td>854</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 4 (926 words)</td>
<td>342</td>
<td>584</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Frequency space

-头发
-政治
-体育
-食物
...

0.7

0.7 0.5 0.2
The Standard Model in BNP (By Example)

- Doc 1 (532 words): 343 sports, 189 political, 210 food
- Doc 2 (210 words): 210 sports
- Doc 3 (854 words): 854 sports, 342 political, 584 food
- Doc 4 (926 words):...

Frequency space:
- 0.7

Topic space:
- sports

The graph shows the frequency of topics across different documents.
The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th></th>
<th>sports</th>
<th>politics</th>
<th>food</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1 (532 words)</td>
<td>343</td>
<td></td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>Doc 2 (210 words)</td>
<td></td>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 3 (854 words)</td>
<td>854</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 4 (926 words)</td>
<td>342</td>
<td>584</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Frequency space**
  - Sports: 0.7
  - Politics: 0.5
  - Food: 0.2

- **Topic space**
The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th></th>
<th>sports</th>
<th>politics</th>
<th>food</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>343</td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 2</td>
<td>210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 3</td>
<td>854</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 4</td>
<td>342</td>
<td>584</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The numbers represent the frequency of topics in each document.
The Standard Model in BNP (By Example)

θ is a random discrete measure on the topics

<table>
<thead>
<tr>
<th></th>
<th>sports</th>
<th>politics</th>
<th>food</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>343</td>
<td></td>
<td>189</td>
</tr>
<tr>
<td>(532 words)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 2</td>
<td></td>
<td>210</td>
<td></td>
</tr>
<tr>
<td>(210 words)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 3</td>
<td>854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(854 words)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 4</td>
<td>342</td>
<td>584</td>
<td></td>
</tr>
<tr>
<td>(926 words)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

θ is a random discrete measure on the topics
The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th>Document</th>
<th>Words</th>
<th>Sports</th>
<th>Politics</th>
<th>Food</th>
</tr>
</thead>
<tbody>
<tr>
<td>Doc 1</td>
<td>532</td>
<td>343</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>Doc 2</td>
<td>210</td>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Doc 3</td>
<td>854</td>
<td>342</td>
<td>584</td>
<td></td>
</tr>
<tr>
<td>Doc 4</td>
<td>926</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\Theta$ is a random discrete measure on the topics.
The Standard Model in BNP (By Example)

\[ \psi_1 \quad \psi_2 \quad \psi_3 \quad \ldots \]

\[ \begin{array}{ccc}
\text{Obs 1} & 343 & 189 \\
\text{Obs 2} & 210 & \quad \\
\text{Obs 3} & 854 & \quad \\
\text{Obs 4} & 342 & 584 \\
\end{array} \]

\[ \theta_1 \quad \theta_2 \quad \theta_3 \]

\[ \Theta \]

\( \Theta \) is a random discrete measure on the topics traits.
The Standard Model in BNP (By Example)

\[ \psi_1 \quad \psi_2 \quad \psi_3 \quad \ldots \]

\[ \begin{array}{ccc}
   \text{Obs 1} & 343 & 189 \\
   \text{Obs 2} & 210 & \\
   \text{Obs 3} & 854 & \\
   \text{Obs 4} & 342 & 584 \\
   \vdots & \theta_1 & \theta_2 & \theta_3 \\
\end{array} \]

\[ \Theta \]

\( \Theta \) is a random discrete measure on the topics traits.

"rates"
Poisson processes and (N)CRMs

How do we generate infinitely many trait/rate points $(\psi, \theta)$?
Poisson processes and (N)CRMs

How do we generate infinitely many trait/rate points \((\psi, \theta)\)?

**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 \((\psi, \theta)\)?

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

\[\Theta\]

**completely random measure** (CRM) (e.g. BP, GP)

[Kingman 93]
Poisson processes and (N)CRMs

How do we generate infinitely many trait/rate points \((\psi, \theta)\)?

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

- **completely random measure** (CRM) (e.g. BP, \(\Gamma P\))
- Normalize rates: **normalized CRM** (NCRM) (e.g. DP)

[Kingman 93]
Poisson processes and (N)CRMs

How do we generate infinitely many trait/rate points \((\psi, \theta)\)?

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

- **Completely random measure** (CRM) (e.g. BP, \(\Gamma P\))
- 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 \((\psi, \theta)\)?

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

- **completely random measure (CRM)** (e.g. BP, GP)
  - 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

\[ \sum_{k=1}^{\infty} \theta_k \delta \psi_k \]

Truncation and error analysis

\[ \sum_{k=1}^{K} \theta_k \delta \psi_k \]
Truncation Roadmap

Tractable models in BNP

two forms for sequential representations

\[ \sum_{k=1}^{\infty} \theta_k \delta \psi_k \]

Truncation and error analysis

\[ \sum_{k=1}^{K} \theta_k \delta \psi_k \]
Sequential representation & truncation

We pick a finite subset of atoms $(\psi, \theta)$ by:
Sequential representation & truncation

We pick a finite subset of atoms \((\psi, \theta)\) by:

1) ordering the atoms \((\text{sequential representation})\)
Sequential representation & truncation

We pick a finite subset of atoms \((\psi, \theta)\) by:

1) ordering the atoms (sequential representation)
Sequential representation & truncation

We pick a finite subset of atoms \((\psi, \theta)\) by:

1) ordering the atoms (sequential representation)

![Graph showing the rate space and trait space with points labeled 1 and 2]
Sequential representation & truncation

We pick a finite subset of atoms \((\psi, \theta)\) by:

1) ordering the atoms \(\text{(sequential representation)}\)
Sequential representation & truncation

We pick a finite subset of atoms $(\psi, \theta)$ by:

1) ordering the atoms *(sequential representation)*
Sequential representation & truncation

We pick a finite subset of atoms $(\psi, \theta)$ by:

1) ordering the atoms \textbf{(sequential representation)}

\[\begin{array}{cccc}
1 & 2 & 3 & \Theta \\
K & & & \\
\end{array}\]
Sequential representation & truncation

We pick a finite subset of atoms \((\psi, \theta)\) by:

1) ordering the atoms \(\textbf{(sequential representation)}\)

\[
\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}
\]
Sequential representation & truncation

We pick a finite subset of atoms \((\psi, \theta)\) by:

1) ordering the atoms \textbf{(sequential representation)}

2) removing any atoms beyond the K-th \textbf{(truncation)}

\[
\Theta = \sum_{k=1}^{\infty} \theta_k \delta \psi_k
\]
Sequential representation & truncation

We pick a finite subset of atoms $(\psi, \theta)$ 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 \((\psi, \theta)\) 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 \((\psi, \theta)\) by:

1) ordering the atoms \(\textbf{(sequential representation)}\)
2) removing any atoms beyond the K-th \(\textbf{(truncation)}\)

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

\[ \begin{array}{cccc}
1 & 4 & 3 & \Theta \\
\end{array} \]

rate space

trait space
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)
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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Bound Decay</td>
<td>✓</td>
<td>✓</td>
<td>✓/✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>(exp)</td>
<td>(exp)</td>
<td></td>
<td></td>
<td>(exp)</td>
<td>(exp)</td>
<td></td>
</tr>
<tr>
<td>Ease of Analysis</td>
<td>✗</td>
<td>✗✗</td>
<td>✗</td>
<td>✗</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generality</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Known # Atoms</td>
<td>✓</td>
<td>✓</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
<td>✗</td>
</tr>
</tbody>
</table>
Given Gamma process: \( \nu(\theta) = \gamma \lambda \theta^{-1} e^{-\lambda \theta} \)
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 \( f(\theta) := -c^{-1} \frac{d}{d\theta} [\theta \nu(\theta)] \)
Sequential representation example

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

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

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

Exponential(\(\lambda\)) density!
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 $f(\theta) := -c^{-1} \frac{d}{d\theta} [\theta \nu(\theta)] = \lambda e^{-\lambda \theta}$

Step 3: plug in! $\Theta = \sum_{k=1}^{\infty} V_k e^{-\Gamma_k} \delta_{\psi_k}$, $V_k \overset{iid}{\sim} f$, $\Gamma \sim \text{Poisson}(P(c))$
Truncation Roadmap

Tractable models in BNP

two forms for sequential representations

\[ \sum_{k=1}^{\infty} \theta_k \delta \psi_k \]

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
Choosing between the seven representations

How close is our finite approximation?
Choosing between the seven representations

How close is our finite approximation?

**Truncation error:** \[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(y) - p_{N,K}(y)| \, dy \]
Choosing between the seven representations

How close is our finite approximation?

**Truncation error:** \[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(y) - p_{N,K}(y)| \, dy \]

full infinite $\Theta$  
truncated $\Theta_K$
Choosing between the seven representations

*How close is our finite approximation?*

**Truncation error:** \[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(y) - p_{N,K}(y)| \, dy \]

- full infinite \( \Theta \)
- generated data
- truncated \( \Theta_K \)
- generated data
Choosing between the seven representations

How close is our finite approximation?

**Truncation error:** \[ \| p_{N,\infty} - p_{N,K} \|_1 = \frac{1}{2} \int | p_{N,\infty}(y) - p_{N,K}(y) | \, dy \]

- full infinite \( \Theta \)
- truncated \( \Theta_K \)

Compare the distribution of the data under full vs. truncated
Choosing between the seven representations

How close is our finite approximation?

**Truncation error:** \[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(y) - p_{N,K}(y)| \, dy \]

Depends on **number of observations** \( N \) and **truncation level** \( K \)
Choosing between the seven representations

How close is our finite approximation?

**Truncation error:** \[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(y) - p_{N,K}(y)| \, dy \]

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:** \[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(y) - p_{N,K}(y)| \, dy \]

Depends on **number of observations** \(N\) and **truncation level** \(K\)

As \(N\) gets larger, error increases

As \(K\) gets larger, error decreases

\[ \varepsilon \]
Choosing between the seven representations

*How close is our finite approximation?*

**Truncation error:** \[ \| p_{N,\infty} - p_{N,K} \|_1 = \frac{1}{2} \int |p_{N,\infty}(y) - p_{N,K}(y)| \, dy \]

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

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

The truncation error

i.e. \( P(\text{whoops!}) \)
Protobound

Leads to all the other truncation error bounds in this work

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

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

The truncation error

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

\[
\| p_{N,\infty} - p_{N,K} \|_1 \\
\leq 1 - e^{-\int_0^\infty \mathbb{E}[\pi(\tau(V,u+G_K))^N] \, du}
\]
Lemma (H., Campbell, How, Broderick).
\[ \| p_{N,\infty} - p_{N,K} \|_1 \leq \mathbb{P} \text{ (any datum selects a removed trait)} \]

The truncation error

Theorem (HCHB). The series rep error is bounded by
\[ \| p_{N,\infty} - p_{N,K} \|_1 \leq 1 - e^{-\int_0^\infty \mathbb{E}[\bar{\pi}(\tau(V,u+G_K))^N] \, du} \]

Theorem (HCHB). The superposition rep error is bounded by
\[ \| p_{N,\infty} - p_{N,K} \|_1 \leq 1 - e^{-\int_0^\infty \bar{\pi}(\theta)^N \nu_K^+(d\theta)} \]
Error bound example

**Given** Gamma-Poisson process: \( \nu(\theta) = \gamma \lambda \theta^{-1} e^{-\lambda \theta} \quad \pi(\theta) = e^{-\theta} \)
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  
\[ G_K \sim \text{Gamma}(K, c) : \]

\[ \int_0^\infty (1 - \mathbb{E}[\pi(\theta e^{-G_K})]) \nu(d\theta) \]
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 \( G_K \sim \text{Gamma}(K, c) \):

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

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

Integration by parts

\[
\leq \gamma \mathbb{E} [e^{-G_K}] \]

\[ \log(1 + x) \leq x \]
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 \( G_K \sim \text{Gamma}(K, c) \):

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

Integration by parts \quad \log(1 + x) \leq x \quad \text{Gamma expectation}
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 \( G_K \sim \text{Gamma}(K, c) \):

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

Integration by parts  
\( \log(1 + x) \leq x \)  
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

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

Truncation and error analysis

\[ \sum_{k=1}^{K} \theta_k \delta_{\psi_k} \]
Tractable models in BNP

two forms for sequential representations

Truncation and error analysis

Truncation Roadmap

\[ \sum_{k=1}^{\infty} \theta_k \delta \psi_k \]
## Previous Work

<table>
<thead>
<tr>
<th></th>
<th>Finite Approximation</th>
<th>Approximation Error Bounds</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BPP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GP</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>(N)CRM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Our Work

<table>
<thead>
<tr>
<th></th>
<th>Finite Approximation</th>
<th>Approximation Error Bounds</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DP</strong></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td><strong>BP</strong></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td><strong>BPP</strong></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ΓP</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>(N)CRM</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
## Our Work

<table>
<thead>
<tr>
<th></th>
<th>Finite Approximation</th>
<th>Approximation Error Bounds</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DP</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>BP</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>BPP</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>ΓP</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>(N)CRM</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
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