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

Bayesian nonparametrics: achieves growing model size via infinite parameters

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.

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]
Bayesian nonparametrics

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

Bayesian nonparametrics: achieves growing model size via infinite parameters

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

\[ \Pi(d\Theta \mid X) \propto \Theta f(X \mid \Theta) \Pi_0(d\Theta) \]

[movies] [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]
Bayesian nonparametrics

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

Bayesian nonparametrics: achieves growing model size via infinite parameters.

\[ \Pi(\text{d}\Theta | X) \propto f(X | \Theta)\Pi_0(\text{d}\Theta) \]

E.g. keep learning new topics from a stream of documents in various fields such as:
- 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]
Bayesian nonparametrics

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

Bayesian nonparametrics: achieves growing model size via infinite parameters

\[ \Pi(d\Theta \mid X) \propto f(X \mid \Theta)\Pi_0(d\Theta) \]

[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]
Bayesian nonparametrics

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

Bayesian nonparametrics: achieves growing model size via infinite parameters.

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

Parameter: \( \Pi(d\Theta | X) \propto f(X | \Theta)\Pi_0(d\Theta) \)

Likelihood: \( f(X | \Theta) \)

Data: \( X \)

Bayesian nonparametrics

Need models that can extract new, useful information from unbounded 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

\[ \Pi(d\Theta \mid X) \propto f(X \mid \Theta)\Pi_0(d\Theta) \]

[Go2014] [The 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

Bayesian nonparametrics: achieves growing model size via infinite parameters

\[
\Pi(d\Theta \mid X) \propto \theta f(X \mid \Theta) \Pi_0(d\Theta)
\]

[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]
Inference in BNP models

\[ \pi(d\Theta \mid X) \propto f(X \mid \Theta) \pi_0(d\Theta) \]
Inference in BNP models

\[ \pi(d\Theta \mid X) \propto_\Theta f(X \mid \Theta)\pi_0(d\Theta) \]

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

\[ \pi(d\Theta \mid X) \propto_{\Theta} f(X \mid \Theta) \pi_0(d\Theta) \]

- Option #1: Integrate out the parameter \( \Theta \) (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

\[ \pi(d\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\pi_0(d\Theta) \]

- Option #1: Integrate out the parameter \( \Theta \) (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

\[ \pi(d\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\pi_0(d\Theta) \]

• Option #1: Integrate out the parameter \( \Theta \) (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

\[ \pi(d\Theta \mid X) \propto_\Theta f(X \mid \Theta)\pi_0(d\Theta) \]

- Option #1: Integrate out the parameter \( \Theta \) (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

\[ \pi(d\Theta \mid X) \propto f(X \mid \Theta) \pi_0(d\Theta) \]

- Option #1: Integrate out the parameter \( \Theta \) (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

\[ \pi(d\Theta \mid X) \propto f(X \mid \Theta)\pi_0(d\Theta) \]

• Option #1: Integrate out the parameter \( \Theta \) (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

\[ \pi(d\Theta \mid X) \propto f(X \mid \Theta)\pi_0(d\Theta) \]

- **Option #1:** Integrate out the parameter \( \Theta \) (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)

<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>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th>Topic</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></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>
</tbody>
</table>

frequency space

<table>
<thead>
<tr>
<th>Topic</th>
<th>Doc 1</th>
<th>Doc 2</th>
<th>Doc 3</th>
<th>Doc 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>sports</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>politics</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>food</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Standard Model in BNP (By Example)

- **Doc 1**: 532 words
  - Sports: 343
  - Politics: 189

- **Doc 2**: 210 words
  - Sports: 210

- **Doc 3**: 854 words
  - Sports: 854
  - Politics: 342
  - Food: 584
  - Others...

- **Doc 4**: 926 words
  - Sports: 0.7
  - Politics: 0.5
  - Food: 0.2
  - Others...

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

<table>
<thead>
<tr>
<th>Topic</th>
<th>Sports</th>
<th>Politics</th>
<th>Food</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Doc 1</strong> (532 words)</td>
<td>343</td>
<td>189</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Doc 2</strong> (210 words)</td>
<td>210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Doc 3</strong> (854 words)</td>
<td>854</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Doc 4</strong> (926 words)</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)

- **Doc 1** (532 words)
  - sports: 343
  - politics: 189
- **Doc 2** (210 words)
  - sports: 210
- **Doc 3** (854 words)
  - sports: 854
- **Doc 4** (926 words)
  - sports: 342
- **frequency space**
  - topic space
  - 0.7
  - 0.5
  - 0.2
The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th>Doc 1 (532 words)</th>
<th>Doc 2 (210 words)</th>
<th>Doc 3 (854 words)</th>
<th>Doc 4 (926 words)</th>
</tr>
</thead>
<tbody>
<tr>
<td>sports 343</td>
<td>politics 210</td>
<td>food 189</td>
<td>...</td>
</tr>
<tr>
<td>sports 854</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sports 342</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td>0.2</td>
<td></td>
</tr>
</tbody>
</table>

frequency space

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 (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>

![Graph showing topics and their frequencies]

- **Frequency Space**: The distribution of topics across documents.
- **Sports**: A topic with significant frequency in some documents.
- **Politics**, **Food**: Other topics with varying frequencies.

Scores 0.7, 0.5, 0.2 indicate the relative frequency of each topic across 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>0.7</td>
</tr>
<tr>
<td>Doc 2 (210 words)</td>
<td></td>
<td>210</td>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>Doc 3 (854 words)</td>
<td>854</td>
<td></td>
<td></td>
<td>0.2</td>
</tr>
<tr>
<td>Doc 4 (926 words)</td>
<td>342</td>
<td>584</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table shows the frequency of words in different documents, with the frequency of the word 'sports' in the sports column and the frequency of the word 'sports' in the sports topic space. The diagram on the right represents the topic space with 'sports' and other topics.
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>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>

\[ \Theta \text{ is a random discrete measure on the topics} \]
The Standard Model in BNP (By Example)

<table>
<thead>
<tr>
<th>Topic</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>

$\Theta$ is a random discrete measure on the topics.

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

<table>
<thead>
<tr>
<th>Obs</th>
<th>ψ₁</th>
<th>ψ₂</th>
<th>ψ₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs 1</td>
<td>343</td>
<td>189</td>
<td></td>
</tr>
<tr>
<td>Obs 2</td>
<td>210</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs 3</td>
<td>854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obs 4</td>
<td>342</td>
<td>584</td>
<td></td>
</tr>
</tbody>
</table>

θ is a random discrete measure on the topics traits

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

<table>
<thead>
<tr>
<th>Obs 1</th>
<th>Obs 2</th>
<th>Obs 3</th>
<th>Obs 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>343</td>
<td>210</td>
<td>854</td>
<td>342</td>
</tr>
</tbody>
</table>

\( \psi_1 \) \( \psi_2 \) \( \psi_3 \) "traits"

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

rate space

trait space

\( \Theta \) 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 process** with intensity measure \(\mu(d\theta \times d\psi)\)

[Kingman 93]
Poisson processes and (N)CRMs

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

**Poisson process** with intensity measure

\[
\mu(d\theta \times d\psi) = \nu(d\theta) H(d\psi)
\]

[Kingman 93]
Poisson processes and (N)CRMs

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

**Poisson process** with intensity measure 
\[
\mu(d\theta \times d\psi) = \nu(d\theta)H(d\psi)
\]

**completely random measure** (CRM) 
(e.g. BP, GP) 
\[
\Theta = \sum_{k=1}^{\infty} \theta_k \delta_{\psi_k}
\]

[Kingman 93]
Poisson processes and (N)CRMs

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

**Poisson process** with intensity measure $\mu(d\theta \times d\psi)$

\[
\mu(d\theta \times d\psi) = \nu(d\theta)H(d\psi)
\]

**completely random measure** (CRM)

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

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

*Kingman 93*
Poisson processes and (N)CRMs

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

**Poisson process** with intensity measure \(\mu(d\theta \times d\psi)\)

\[
\mu(d\theta \times d\psi) = \nu(d\theta) H(d\psi)
\]

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

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

**Poisson process** with intensity measure \(\mu(d\theta \times d\psi)\)

\[
\mu(d\theta \times d\psi) = \nu(d\theta)H(d\psi)
\]

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

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

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.
Past work: finite approximations to BNP priors

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

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

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

$$\Theta_K = \sum_{k=1}^{K} \theta_k \delta_{\psi_k}$$
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 \( \nu(d\theta)H(d\psi) \)
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 \( \nu(d\theta)H(d\psi) \)

Series representation
function of a homogenous
Poisson point process

(4 versions)

Ordering of (N)CRM atoms

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

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

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

Series representation
function of a homogenous Poisson point process

(4 versions)

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 \( \nu(d\theta)H(d\psi) \)

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

\[ V_k \overset{\text{i.i.d.}}{\sim} g \]

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 \( \nu(d\theta)H(d\psi) \)

**Series representation**
function of a homogenous
Poisson point process

(4 versions)

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

Ordering of (N)CRM atoms

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

2 forms for sequential representations \( \nu(d\theta) H(d\psi) \)

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

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

[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 \( \nu(d\theta)H(d\psi) \)

Series representation

function of a homogenous Poisson point process

(4 versions)

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

Superposition representation

infinite sum of CRMs, each with finite # of atoms

(3 versions)

\[ \Theta^{(1)} + \Theta^{(2)} + \Theta^{(3)} + \cdots \]


[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 \( \nu(d\theta) H(d\psi) \)

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

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

**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 \( \nu(d\theta) H(d\psi) \)

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

<table>
<thead>
<tr>
<th>Error Bound Decay</th>
<th>Series Reps</th>
<th>Superposition Reps</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ease of Analysis</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Generality</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Known # Atoms</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
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 \( c := \lim_{\theta \to 0} \theta \nu(\theta) \)
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! Exponential(\(\lambda\)) density!

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

\[ \Pi(d\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\Pi_0(d\Theta) \]
Choosing between the seven representations

*How close is our finite approximation?*

\[
\Pi(d\Theta \mid X) \propto f(X \mid \Theta) \Pi_0(d\Theta)
\]

**Truncation error:**

\[
\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(X) - p_{N,K}(X)|dX
\]
Choosing between the seven representations

How close is our finite approximation?

\[
\Pi(d\Theta \mid X) \propto f(X \mid \Theta)\Pi_0(d\Theta)
\]

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

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

*How close is our finite approximation?*

\[ \Pi(d\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\Pi_0(d\Theta) \]

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

Diagram:

- Full infinite\( \Theta \)
  - \( \downarrow f \)
  - Data \( X \)

- Truncated\( \Theta_K \)
  - \( \downarrow f \)
  - Data \( X \)
Choosing between the seven representations

*How close is our finite approximation?*

\[
\Pi(d\Theta \mid X) \propto f(X \mid \Theta) \Pi_0(d\Theta)
\]

**Truncation error:**

\[
\|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(X) - p_{N,K}(X)|dX
\]

Compare the distribution of the data under full vs. truncated

![Diagram showing full and truncated data distributions](image-url)
Choosing between the seven representations

How close is our finite approximation?

\[
\Pi(d\Theta \mid X) \propto_{\Theta} f(X \mid \Theta) \Pi_0(d\Theta)
\]

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

How close is our finite approximation?

\[ \Pi(d\Theta \mid X) \propto_{\Theta} f(X \mid \Theta)\Pi_0(d\Theta) \]

**Truncation error:**

\[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(X) - p_{N,K}(X)|dX \]

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

How close is our finite approximation?

\[ \Pi(d\Theta \mid X) \propto \Theta f(X \mid \Theta) \Pi_0(d\Theta) \]

**Truncation error:**

\[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(X) - p_{N,K}(X)|dX \]

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?*

\[ \Pi(d\Theta \mid X) \propto_{\Theta} f(X \mid \Theta) \Pi_0(d\Theta) \]

**Truncation error:**

\[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(X) - p_{N,K}(X)|dX \]

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?

\[ \Pi(d\Theta \mid X) \propto f(X \mid \Theta)\Pi_0(d\Theta) \]

**Truncation error:**

\[ \|p_{N,\infty} - p_{N,K}\|_1 = \frac{1}{2} \int |p_{N,\infty}(X) - p_{N,K}(X)|dX \]

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).

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

Leads to all the other truncation error bounds in this work

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

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

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

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

Leads to all the other truncation error bounds in this work

**Lemma (H., Campbell, How, Broderick).**
\[
\|p_{N,\infty} - p_{N,K}\|_1 \leq \mathbb{P} \text{ (any datum selects a removed trait)}
\]

**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 \( G_K \sim \text{Gamma}(K, c) \):

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

Integration by parts
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 \( G_K \sim \text{Gamma}(K, c) \):

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

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

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

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

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

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

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

Integration by parts

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

Gamma expectation

\[
= \gamma \left( \frac{\gamma \lambda}{1 + \gamma \lambda} \right)^K
\]

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

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

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

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!

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

$$N \to \infty, \text{ bound } \to 1$$
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 \( G_K \sim \text{Gamma}(K, c) \):

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

Integration by parts

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

log(1 + x) \leq x

\[
= \gamma \left( \frac{\gamma \lambda}{1 + \gamma \lambda} \right)^K \]

Gamma expectation

**Step 2:** plug in!

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

\( N \to \infty, \text{bound} \to 1 \quad K \to \infty, \text{bound} \to 0 \)
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 \( \nu_K \) 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 \( \nu_K \) 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 \( \nu_K \) such that

\( K \nu_K(\theta) \approx \nu(\theta) \) and \( K \nu_K \to \nu \)

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

\[ \nu(d\theta; \gamma, d, \eta) = \gamma \theta^{-1-d} g(\theta) d \frac{h(\theta; \eta)}{Z(1-d, \eta)} \, d\theta. \]
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 \( \nu_K \) such that \( K
\nu_K(\theta) \approx \nu(\theta) \) and \( K \nu_K \to \nu \)

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

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

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 \( \nu_K \) such that \( K\nu_K(\theta) \approx \nu(\theta) \) and \( K\nu_K \to \nu \)

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

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

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 $\nu_K$ such that $K \nu_K(\theta) \approx \nu(\theta)$ and $K \nu_K \to \nu$

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

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

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 \( \nu_K \) such that \( K \nu_K(\theta) \approx \nu(\theta) \) and \( K \nu_K \to \nu \)

---

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

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

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

<table>
<thead>
<tr>
<th>Previous Work</th>
<th>Truncated Approximations</th>
<th>Truncation Error Bounds</th>
<th>Non-nested Approximations</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>
## Conclusion

<table>
<thead>
<tr>
<th>Our Work</th>
<th>Truncated Approximations</th>
<th>Truncation Error Bounds</th>
<th>Non-nested Approximations</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>
## Conclusion

### Our Work

<table>
<thead>
<tr>
<th>Method</th>
<th>Truncated Approximations</th>
<th>Truncation Error Bounds</th>
<th>Non-nested Approximations</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>

- Large family of BNP priors that admit efficient inference
## Conclusion

Our Work

<table>
<thead>
<tr>
<th></th>
<th>Truncated Approximations</th>
<th>Truncation Error Bounds</th>
<th>Non-nested Approximations</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>

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

<table>
<thead>
<tr>
<th></th>
<th>Truncated Approximations</th>
<th>Truncation Error Bounds</th>
<th>Non-nested Approximations</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**

- 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](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

- 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