Truncation error of a superposed gamma process in a decreasing order representation

✉ julyan.arbel@inria.fr  🌐 www.julyanarbel.com

Inria, Mistis, Grenoble, France

Joint work with Igor Prünster (Bocconi University, Milan, Italy)

Advances in Approximate Bayesian Inference NIPS Workshop, December 9, 2016
Table of Contents

Completely random measures

Ferguson and Klass moment-based algorithm

Simulation study

Theoretical results

Application to density estimation

```c
int getRandomNumber()
{
    return 4;  // chosen by fair dice roll.
    // guaranteed to be random.
}
```

(xkcd)
Bayesian nonparametric priors

Two main categories of priors depending on parameter spaces

Spaces of functions

*random functions*

- Stochastic processes
  s.a. Gaussian processes
- Random basis expansions
- Random densities
- Mixtures

Spaces of probability measures

*discrete random measures*

- Dirichlet process ⊂ Pitman–Yor ⊂ Gibbs-type ⊂ Species sampling processes
- Completely random measures

[Wikipedia] [Brix, 1999]
Completely random measures

Completely random measures $\tilde{\mu} = \sum_{i \geq 1} J_i \delta_{Z_i}$

where the jumps $(J_i)_{i \geq 1}$ and the jump points $(Z_i)_{i \geq 1}$ are independent

Definition (Kingman, 1967)

Random measure $\tilde{\mu}$ s.t. $\forall A_1, \ldots, A_d$ disjoint sets
$\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_d)$ are mutually independent

- Independent Increment Processes, Lévy processes
- Popular models with applications in biology, sparse random graphs, survival analysis, machine learning, etc. Pivotal role in BNP (Lijoi and Prünster, 2010, Jordan, 2010)
Ferguson and Klass algorithm and goal

- Jumps in decreasing order in $\tilde{\mu} = \sum_{j=1}^{\infty} J_j \delta_{Z_j}$
- $\xrightarrow{\longrightarrow}$ Minimal error at threshold $M$ $\tilde{\mu}(X) - \tilde{\mu}_M(X) = \sum_{j=M+1}^{\infty} J_j$
- BNPdensity R package on CRAN, for F & K mixtures of normalized CRM

**Algorithm 1 Ferguson and Klass algorithm**

1: sample $\xi_j \sim \text{PP}$ for $j = 1, \ldots, M$
2: define $J_j = N^{-1}(\xi_j)$ for $j = 1, \ldots, M$
3: sample $Z_j \sim P_0$ for $j = 1, \ldots, M$
4: approximate $\tilde{\mu}$ by $\tilde{\mu}_M = \sum_{j=1}^{M} J_j \delta_{Z_j}$
Moment matching

Assessing the error of truncation at threshold \( M \)

\[
T_M = \tilde{\mu}(X) - \tilde{\mu}_M(X) = \sum_{j=M+1}^{\infty} J_j
\]

Relative error index

\[
e_M = \mathbb{E}_{FK} \left[ \frac{J_M}{\sum_{j=1}^{M} J_j} \right]
\]

Moment-based index

\[
\ell_M = \left( \frac{1}{K} \sum_{n=1}^{K} (m_n^{1/n} - \hat{m}_n^{1/n})^2 \right)^{1/2}
\]

Examples of completely random measures

- Generalized gamma process by Brix (1999), \( \gamma \in [0, 1), \theta \geq 0 \)
- Superposed gamma process by Regazzini et al. (2003), \( \eta \in \mathbb{N} \)
- Stable-beta process by Teh and Gorur (2009), \( \sigma \in [0, 1), c > -\sigma \)
Moment matching

Relative error index $e_M$

Moment-based index $\ell_M$

- $GG : a = 1$
- $\gamma = 0$
- $\gamma = 0.25$
- $\gamma = 0.5$
- $\gamma = 0.75$
Evaluation of error on functionals

Functional of interest: the total mass, criterion $\Delta_1 = |\tilde{\mu}(X) - \tilde{\mu}_M(X)|$

Define similarly $\Delta_k$ for higher order moments of the total mass
Moment matching

Reverse moment index $M(\ell) = M \leftrightarrow \ell_M = \ell$
Number of jumps $M$ needed to achieve a given precision, here of $\ell = 10\%$
Posterior moment match

Theorem (James et al., 2009, Teh and Gorur, 2009)

In mixture models with normalized generalized gamma (left) and the Indian buffet process based on the stable beta process (right) the posterior distribution of $\tilde{\mu}$ is essentially (conditional on some latent variables)

$$\tilde{\mu}^* + \sum_{j=1}^{k} J_j^* \delta_{Y_j^*}$$

IG : $\gamma = 0.5$, $a = 1$

Prior:
- $k = 1$
- $k = 3$
- $k = 10$

SBP : $\sigma = 0.5$, $a = 1$, $c = 1$

Prior:
- $n = 5$
- $n = 10$
- $n = 20$
Posterior moment match

Posterior of inverse-Gaussian process:
\[ \tilde{\mu}^* + \sum_{j=1}^{k} J_j^* \delta_{Y_j^*} \]

\[ \mathbb{E} \left( \sum_{j=1}^{k} J_j^* \right) / \mathbb{E}(\tilde{\mu}^*(X)) \]

<table>
<thead>
<tr>
<th>(k) (n)</th>
<th>10</th>
<th>30</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.34</td>
<td>7.30</td>
<td>13.50</td>
</tr>
<tr>
<td>(n^\gamma)</td>
<td>2.65</td>
<td>4.68</td>
<td>6.05</td>
</tr>
<tr>
<td>(n)</td>
<td>0.89</td>
<td>0.98</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Theorem (Arbel, De Blasi, Prünster, 2015)

Denote by \(P_0\) the true data distribution. In the NRMI model with prior guess \(P^*\), the posterior of \(\tilde{P}\) converges weakly to \(P_\infty\):

- if \(P_0\) is discrete, then \(P_\infty = P_0\)
- if \(P_0\) is diffuse, then \(P_\infty = \sigma P^* + (1 - \sigma)P_0\)
Bounding $T_M$ in probability

Proposition (Arbel and Prünster, 2016, Brix, 1999)

Let $T_M$ be the truncation error for the Generalized Gamma or the Stable Beta Process.

Then for any $\epsilon \in (0, 1)$,

$$\mathbb{P}(T_M \leq t_M^\epsilon) \geq 1 - \epsilon$$

for

$$t_M^\epsilon \sim \begin{cases} 
  e^{-CM} & \text{if } \sigma = 0, \\
  \frac{1}{M^{1/\sigma-1}} & \text{if } \sigma \neq 0,
\end{cases}$$

with ugly explicit constants depending on $\epsilon, \gamma, \theta, \sigma$ and $c$. 

\[ \text{SBP: } a = 1, c = 1 \]

\[ \text{σ = 0 } \]

\[ \text{σ = 0.5 } \]
Density estimation

Mixtures of normalized random measures with independent increments

\[ Y_i | \mu_i, \sigma_i \overset{\text{ind}}{\sim} k(\cdot | \mu_i, \sigma_i), \quad i = 1, \ldots, n, \]

\[ (\mu_i, \sigma_i) | \tilde{P} \overset{\text{iid}}{\sim} \tilde{P}, \quad i = 1, \ldots, n, \]

\[ \tilde{P} \sim \text{NRMI}, \]

Galaxy dataset. Kolmogorov–Smirnov distance \( d_{KS}(\hat{F}_{\ell M}, \hat{F}_{eM}) \) between estimated cdfs \( \hat{F}_{\ell M} \) and \( \hat{F}_{eM} \) under, respectively, the moment-match (with \( \ell_M = 0.01 \)) and the relative error (with \( e_M = 0.1, 0.05, 0.01 \)) criteria.

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>( e_M = 0.1 )</th>
<th>( e_M = 0.05 )</th>
<th>( e_M = 0.01 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19.4</td>
<td>15.5</td>
<td>9.2</td>
</tr>
<tr>
<td>0.25</td>
<td>31.3</td>
<td>23.7</td>
<td>15.1</td>
</tr>
<tr>
<td>0.5</td>
<td>42.4</td>
<td>28.9</td>
<td>18.3</td>
</tr>
<tr>
<td>0.75</td>
<td>64.8</td>
<td>41.0</td>
<td>23.2</td>
</tr>
</tbody>
</table>
Discussion

• Methodology based on moments for assessing quality of approximation in Ferguson and Klass algorithm, a conditional algorithm
• Should be preferred to relative error
• All-purpose criterion: validates the samples of a CRM rather than a transformation of it
• Going to be included in a new release of BNPdensity R package
• Future work: compare $L^1$ type bounds (Ishwaran and James, 2001) in the Ferguson & Klass context and in size biased settings (see the review by Campbell et al., 2016)

For more details and for extensive numerical illustrations:


