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

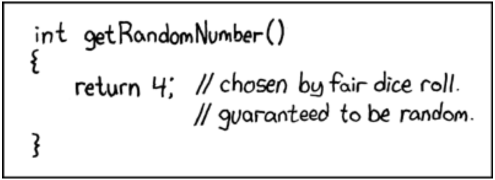
✉ julyan.arbel@inria.fr 🖱 www.julyanarbel.com

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



(xkcd)

Bayesian nonparametric priors

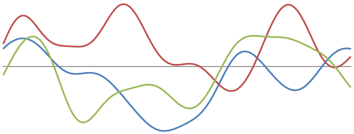
Two main categories of priors depending on parameter spaces

Spaces of functions
random functions

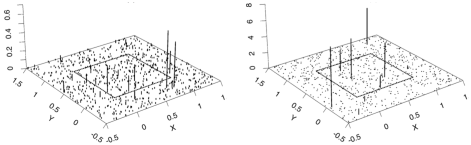
- Stochastic processes
  s.a. Gaussian processes
- Random basis expansions
- Random densities
- Mixtures

Spaces of probability measures
discrete random measures

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

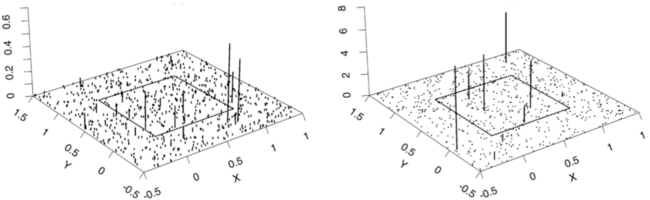


[Wikipedia]



[Brix, 1999]

Completely random measures



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

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

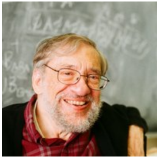
- Independent Increment Processes, Lévy processes
- Popular models with applications in biology, sparse random graphs, survival analysis, machine learning, etc. Pivotal role in BNP (Lijoi and Prünster, 2010, Jordan, 2010)

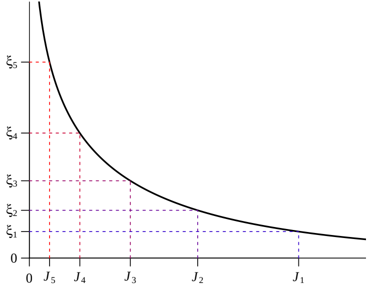
Ferguson and Klass algorithm and goal

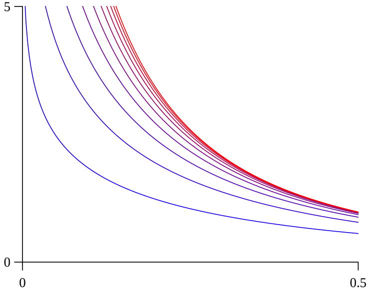
- Jumps in decreasing order in $\tilde{\mu} = \sum_{j=1}^{\infty} J_j \delta_{Z_j}$
- $\longrightarrow$ Minimal error at threshold M $\tilde{\mu}(\mathbb{X}) - \tilde{\mu}_M(\mathbb{X}) = \sum_{j=M+1}^{\infty} J_j$
- BNPdensity R package on CRAN, for F & K mixtures of normalized CRMs

Algorithm 1 Ferguson and Klass algorithm

sample $\xi_j \sim \text{PP}$ for $j = 1, \ldots, M$
define $J_j = N^{-1}(\xi_j)$ for $j = 1, \ldots, M$
sample $Z_j \sim P_0$ for $j = 1, \ldots, M$
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

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

Relative error index

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

Moment-based index

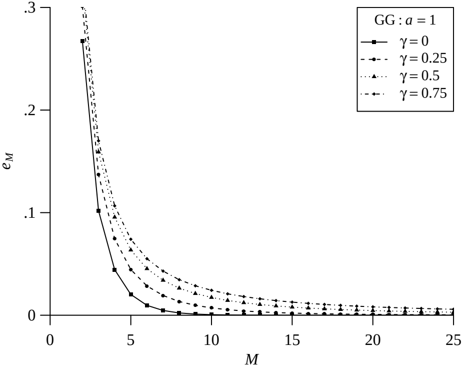
$$\ell_M = \left(\frac{1}{K}\sum_{n=1}^{K}\left(m_n^{1/n} - \hat{m}_n^{1/n}\right)^2\right)^{1/2}$$

Examples of completely random measures

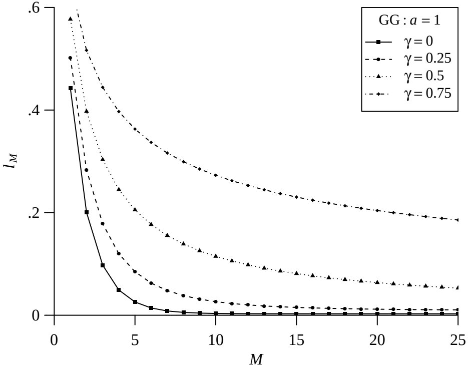
- Generalized gamma process by Brix (1999), $\gamma \in [0,1)$, $\theta \geq 0$
- Superposed gamma process by Regazzini et al. (2003), $\eta \in \mathbb{N}$
- Stable-beta process by Teh and Gorur (2009), $\sigma \in [0,1)$, $c > -\sigma$

Moment matching

Relative error index e_M



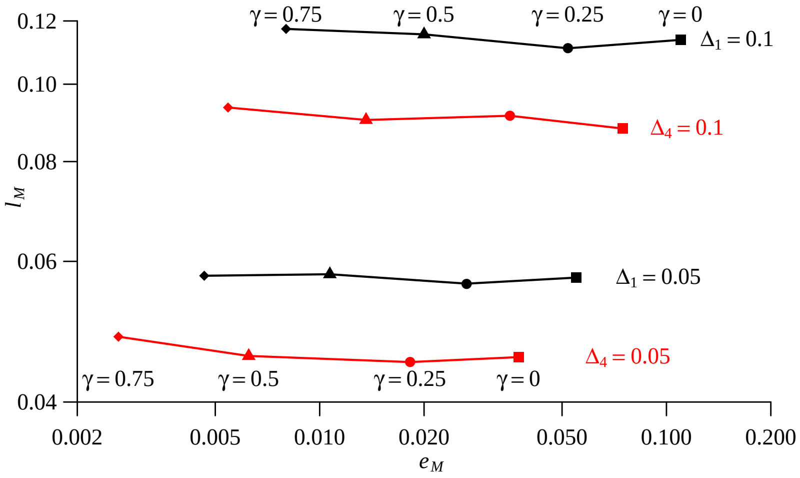
Moment-based index ℓ_M



Evaluation of error on functionals

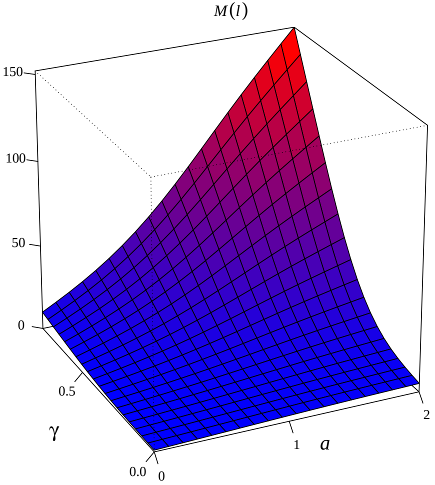
Functional of interest: the total mass, criterion $\Delta_1 = |\tilde{\mu}(\mathbb{X}) - \tilde{\mu}_M(\mathbb{X})|$

Define similarly Δ_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 $\ell = 10\%$

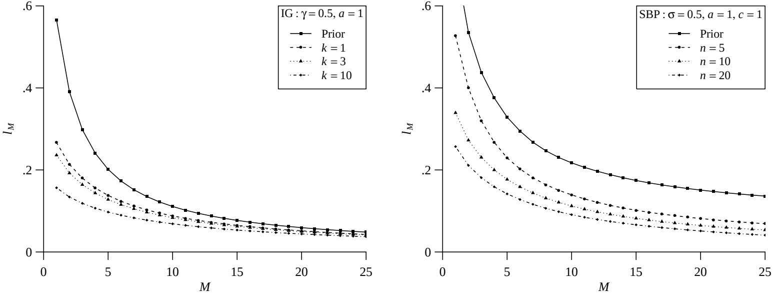


Posterior moment match

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)

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

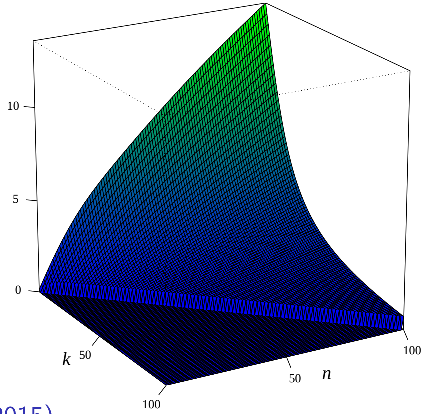


Posterior moment match

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

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

$k \backslash n$	10	30	100
1	3.34	7.30	13.50
n^γ	2.65	4.68	6.05
n	0.89	0.98	0.99



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_∞:

- 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.

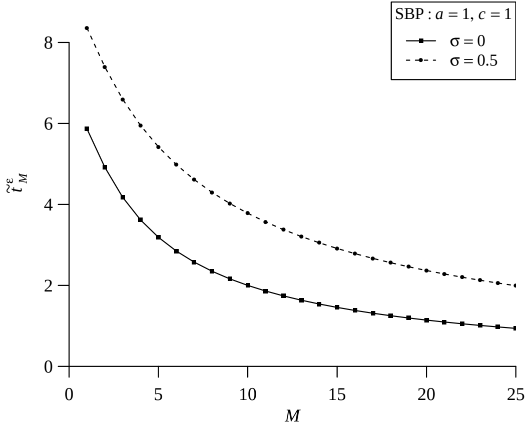
Then for any $\epsilon \in (0,1)$,

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

for

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

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



Density estimation

Mixtures of normalized random measures with independent increments

$$
\begin{aligned}
Y_i|\mu_i,\sigma_i &\overset{\text{ind}}{\sim} k(\cdot|\mu_i,\sigma_i), \quad i=1,\dots,n,\\
(\mu_i,\sigma_i)|\tilde{P} &\overset{\text{iid}}{\sim} \tilde{P}, \quad i=1,\dots,n,\\
\tilde{P} &\sim \text{NRMI},
\end{aligned}
$$

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

γ	$e_M = 0.1$	$e_M = 0.05$	$e_M = 0.01$
0	19.4	15.5	9.2
0.25	31.3	23.7	15.1
0.5	42.4	28.9	18.3
0.75	64.8	41.0	23.2

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:

A. and Prünster (2016). A moment-matching Ferguson and Klass algorithm. **Statistics and Computing**. *arXiv:1606.02566*

References I

Julyan Arbel and Igor Prünster (2016). A moment-matching Ferguson and Klass algorithm. **Statistics and Computing**.

Anders Brix (1999). Generalized gamma measures and shot-noise Cox processes. **Advances in Applied Probability**, pages 929–953.

Trevor Campbell, Jonathan H Huggins, Jonathan How, and Tamara Broderick (2016). Truncated random measures. **arXiv preprint arXiv:1603.00861**.

Thomas S Ferguson and Michael J Klass (1972). A representation of independent increment processes without Gaussian components. **The Annals of Mathematical Statistics**, 43(5):1634–1643.

H. Ishwaran and L.F. James (2001). Gibbs sampling methods for stick-breaking priors. **Journal of the American Statistical Association**, 96(453):161–173.

Michael I Jordan (2010). Hierarchical models, nested models and completely random measures. **Frontiers of Statistical Decision Making and Bayesian Analysis: in Honor of James O. Berger. New York: Springer**.

John Kingman (1967). Completely random measures. **Pacific Journal of Mathematics**, 21(1):59–78.

Antonio Lijoi and Igor Prünster (2010). Models beyond the Dirichlet process. **Bayesian nonparametrics**, 28:80.

Eugenio Regazzini, Antonio Lijoi, and Igor Prünster (2003). Distributional results for means of normalized random measures with independent increments. **Ann. Statist.**, 31(2):560–585.

Yee W. Teh and Dilan Gorur (2009). Indian buffet processes with power-law behavior. In Y. Bengio, D. Schuurmans, J.D. Lafferty, C.K.I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pages 1838–1846. Curran Associates, Inc.