

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

Julyan Arbel, Igor Prünster

▶ **To cite this version:**

Julyan Arbel, Igor Prünster. Truncation error of a superposed gamma process in a decreasing order representation. Argiento, R.; Lanzarone, E.; Antoniano Villalobos, I.; Mattei, A. Bayesian Statistics in Action, 194, , pp.11–19, 2017, Bayesian Statistics in Action. hal-01405580

HAL Id: hal-01405580
https://hal.science/hal-01405580

Submitted on 30 Nov 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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

Julyan Arbel[*]
Inria Grenoble, Université Grenoble Alpes
julyan.arbel@inria.fr

Igor Prünster[†]
Bocconi University, Milan
igor@unibocconi.it

Abstract

Completely random measures (CRM) represent a key ingredient of a wealth of stochastic models, in particular in Bayesian Nonparametrics for defining prior distributions. CRMs can be represented as infinite random series of weighted point masses. A constructive representation due to Ferguson and Klass provides the jumps of the series in decreasing order. This feature is of primary interest when it comes to sampling since it minimizes the truncation error for a fixed truncation level of the series. We quantify the quality of the approximation in two ways. First, we derive a bound in probability for the truncation error. Second, following Arbel and Prünster (2016), we study a moment-matching criterion which consists in evaluating a measure of discrepancy between actual moments of the CRM and moments based on the simulation output. This note focuses on a general class of CRMs, namely the *superposed gamma process*, which suitably transformed have already been successfully implemented in Bayesian Nonparametrics. To this end, we show that the moments of this class of processes can be obtained analytically.

Keywords: Approximate inference, Bayesian Nonparametrics, Completely random measures, Ferguson & Klass algorithm, Normalized random measures, Truncation.

1 Introduction

Completely random measures (CRMs) represent a fundamental building block of countless popular stochastic models and play a prominent role within Bayesian nonparametric modeling (see Lijoi and Prünster, 2010; Jordan, 2010). For instance, the popular Dirichlet process (Ferguson, 1973) can be obtained as normalization or exponentiation of suitable CRMs.

Implementation of CRM-based models usually requires to simulate the CRMs trajectories. As infinite dimensional objects, representable as infinite random series (1), truncation of the series is needed leading to an approximation error. Different strategies for representing the series and choosing the truncation threshold M are available in the literature. A thorough account on CRM *size-biased* representations has been recently provided in Campbell et al. (2016), where the truncation error is examined from the angle of the L^1 distance between marginal data densities under truncated and untruncated models, an approach first considered by Ishwaran and James (2001) for stick-breaking processes.

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

[*]Inria Grenoble - Rhône-Alpes, 655 Avenue de l'Europe, 38330 Montbonnot-Saint-Martin & Laboratoire Jean Kuntzmann, Université Grenoble Alpes. Corresponding author, http://www.julyanarbel.com/

[†]Department of Decision Sciences, BIDSA and IGIER, Bocconi University, via Roentgen 1, 20136 Milan, Italy.

Another important type of representation due to Ferguson and Klass (1972) (see also Walker and Damien, 2000) is arguably one of the most useful ones in that it displays the weights in decreasing order. This yields the very interesting feature that the approximation error is minimized over the whole sample space for a given truncation level. This competitive advantage was exploited in many works, including Argiento et al. (2016); Barrios et al. (2013); De Blasi et al. (2010); Epifani et al. (2003); Griffin and Walker (2011); Nieto-Barajas et al. (2004); Nieto-Barajas and Walker (2004) to cite just a few in Bayesian Nonparametrics. The quality of the approximation, addressed only heuristically in those previous works, is the focus of this note.

In Arbel and Prünster (2016) it shown how moments of the CRMs can be used in order to assess the quality of approximation due to the truncation. Since moments of CRMs are simple to compute, the quality of the approximation can be quantified by evaluating a measure of discrepancy between the actual moments of the CRM at issue and the moments computed based on the sampled realizations of the CRM. The truncation level is selected so that measure of discrepancy does not exceed a given threshold, say 5%. In Arbel and Prünster (2016) the methodology is illustrated on two classes of CRMs, namely the *generalized gamma process* and the *stable-beta process*.

In the proposed talk we review the results of Arbel and Prünster (2016) and, in addition, analyze another broad class called the *superposed gamma process* (see Regazzini et al., 2003; Lijoi et al., 2005). After defining CRMs and the superposed gamma process (Section 2), we display a bound in probability on the truncation error in Section 3.1 and then show the applicability of the moment-matching criterion by deriving analytically the moments of the superposed gamma process in Section 3.2. An illustration of the Ferguson & Klass algorithm and a proof are deferred to the Appendix.

2 Superposed gamma process

Completely random measures (CRM) have been successfully employed in a wide spectrum of modern applications, including survival analysis, random sparse networks, biology, to cite just a few. A CRM $\tilde{\mu}$ on $\mathbb{X}$ is a random measure which spreads out mass independently in the space, which means that the random measure $\tilde{\mu}$, when evaluated on disjoint sets $A_1, \ldots, A_n$, leads to independent random variables $\tilde{\mu}(A_1), \ldots, \tilde{\mu}(A_n)$. Importantly, Kingman (1967) showed that the only way to spread out mass in a *completely random* fashion (without deterministic components) is by randomly scattering point masses in the space. In other words, CRM present the interesting feature that they select (almost surely) discrete measures and hence can be represented as

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

where both the jumps J_i and the locations Z_i are random and are controlled by the so-called Lévy intensity which characterizes the CRM. It is a measure on $\mathbb{R}^+ \times \mathbb{X}$ which can be written as $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v)\alpha(\mathrm{d}x)$ for so-called homogeneous CRM, which are considered here and correspond to the case of jumps independent of the locations. The function ρ controls the intensity of the jumps. The measure α, if the CRM is (almost surely) finite, which is assumed throughout, splits up in $\alpha = aP_0$ where $a > 0$ is called the total mass parameter and the probability distribution P_0 tunes the locations.

Ever-popular CRM include the *generalized gamma process* introduced by Brix (1999) and the *stable-beta process*, or *three-parameter beta process*, defined by Teh and Gorur (2009) as an extension of the beta process (Hjort, 1990). Here we consider another large class of completely random measures called *superposed gamma process*, introduced by Regazzini et al. (2003). It is identified by the jump intensity

$$\rho(\mathrm{d}v) = \frac{1 - \mathrm{e}^{-\eta v}}{1 - \mathrm{e}^{-v}} \frac{\mathrm{e}^{-v}}{v} \mathrm{d}v, \quad \eta > 0. \tag{2}$$

As noted by Lijoi et al. (2005), one usually restricts attention to the case of positive integer η. Under this assumption, the superposed gamma process takes the form of a genuine *superposition* of independent gamma processes with increasing integer-valued scale parameter,

with jump intensity $\rho(\mathrm{d}v) = \frac{1}{v}\left(\mathrm{e}^{-v} + \mathrm{e}^{-2v} + \ldots + \mathrm{e}^{-\eta v}\right)\mathrm{d}v/v$. The specification of integer values for η has also the advantage to lead to analytic computation of the moments. Note that the special case $\eta = 1$ reduces to the gamma process, which gives rise to the Dirichlet process by normalization. Alternatively, the normalization of the superposed gamma process for unspecified η provides the so-called generalized Dirichlet process (Lijoi et al., 2005).

Ferguson and Klass (1972) devise a constructive representation of a CRM which produces the jumps in decreasing order. This corresponds to the (almost surely unique) ordering of the sum elements in (1) where $J_1 > J_2 > \cdots$. Indeed, the jumps are obtained as $\xi_i = N(J_i)$, where $N(v) = \nu([v,\infty),\mathbb{R})$ is a decreasing function, and $\xi_1, \xi_2, \ldots$ are jump times of a standard Poisson process (PP) of unit rate: $\xi_1, \xi_2 - \xi_1, \ldots \overset{\text{i.i.d.}}{\sim} \text{Exp}(1)$. Figure 2 in Appendix illustrates the function $N(\,\cdot\,)$ which takes the following form in the superposed gamma process case

$$N(v) = aE_\eta(v), \text{ where } E_\eta(v) = \sum_{l=1}^{\eta} E_1(kv) \text{ and } E_1(v) = \int_v^\infty u^{-1}\mathrm{e}^{-u}\mathrm{d}u, \tag{3}$$

and where the function E_1 denotes the *exponential integral function*. Since it is impossible to sample an infinite number of jumps, approximate simulation of $\tilde{\mu}$ is in order. This becomes a question of determining the number M of jumps to sample leading to the truncation $\tilde{\mu}_M$ and truncation error T_M as follow

$$\tilde{\mu}_M = \sum_{i=1}^{M} J_i \delta_{Z_i}, \quad T_M = \sum_{i=M+1}^{\infty} J_i. \tag{4}$$

The Ferguson & Klass representation has the key advantage of generating the jumps in decreasing order implicitly minimizing such an approximation error. However, a precise evaluation of T_M, for example in expectation, is a daunting task due to the non independence of the jumps in the Ferguson & Klass representation. The algorithm is summarized in Algorithm 1.

Algorithm 1: Ferguson & Klass algorithm

1: sample $\xi_i \sim \text{PP}$ for $i = 1, \ldots, M$
2: define $J_i = N^{-1}(\xi_i)$ for $i = 1, \ldots, M$
3: sample $Z_i \sim P^*$ for $i = 1, \ldots, M$
4: approximate $\tilde{\mu}$ by $\sum_{i=1}^{M} J_i \delta_{Z_i}$

3 Truncation error of the superposed gamma process

3.1 Bound in probability

We provide an evaluation in probability of the truncation error T_M in (4).

Proposition 1. *Let $(\xi_j)_{j\geq 1}$ be the jump times for a homogeneous Poisson process on $\mathbb{R}^+$ with unit intensity. Then for any $\epsilon \in (0,1)$, the tail sum of the superposed gamma process (4) satisfies*

$$\mathbb{P}\left(T_M \leq t_M^\epsilon\right) \geq 1 - \epsilon, \text{ for } t_M^\epsilon = \frac{C}{(\eta!)^{1/\eta}}\mathrm{e}^{1-\frac{M}{C}}, \text{ where } C = \frac{2\mathrm{e}a\eta}{\epsilon}.$$

The simple proof is deferred to Appendix B. It is interesting to note that the bound t_M^ϵ for the superposed gamma process is equal to its counterpart for the beta process with concentration parameter c set to η, all else things being equal (total mass parameter a and threshold ϵ). See Proposition 1 in Arbel and Prünster (2016). This finding provides a nice connection between both processes otherwise seemingly unrelated.

The bound t_M^ϵ obtained in Proposition 1 is exponentially decreasing with M, which is reminiscent of the results obtained by Brix (1999) and Arbel and Prünster (2016), respectively,

for the generalized gamma process and the stable-beta process with no stable component. As already pointed out by these authors, the bound t^{ϵ}_{M} is very conservative due to a crude lower bound on the quantiles q_j (notation of the proof). The left panel of Figure 1 displays this bound t^{ϵ}_{M}, while the right panel illustrates the truncation level M (in log-scale) required in order to guarantee with 95% probability an upper bound on T_M of $t_{max} \in \{1, 10, 100\}$, for varying values of η. Inspection of the plots demonstrates the rapid increase with η of the number of jumps needed in order to assess a given bound in probability.

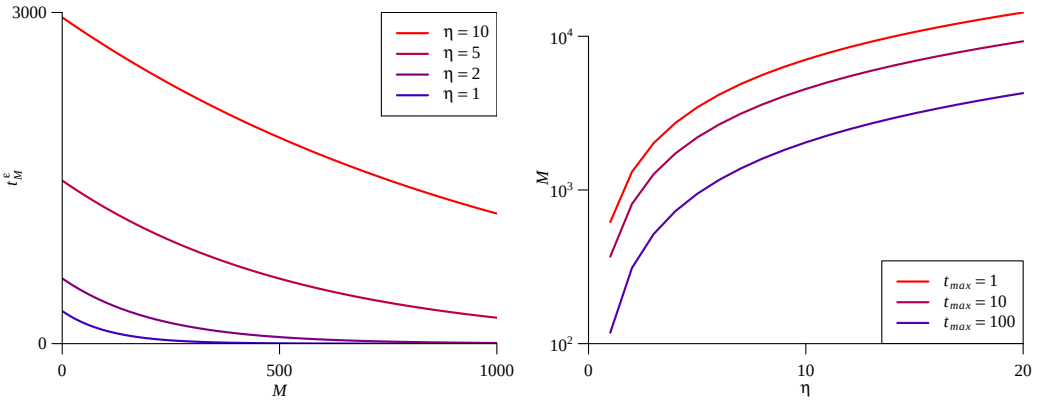


Fig. 1: Left: variation of $M \mapsto t^{\epsilon}_{M}$ for $\eta \in \{1, 2, 5, 10\}$. Right: variation of the threshold function $\eta \mapsto M$ needed to match an error bound of $t_{max} \in \{1, 10, 100\}$ with $\eta \in \{1, \ldots, 20\}$, log scale on y-axis.

3.2 Moment-matching criterion

Given the limited practical usefulness of the bound provided in Proposition 1, we propose to use the alternative route of the moment-matching methodology of Arbel and Prünster (2016) and derive the relevant quantities in order to implement it for the superposed gamma process. Let us consider the n-th (raw) moment of the (random) total mass $\tilde{\mu}(\mathbb{X})$ defined by $m_n = \mathbb{E}\big[\tilde{\mu}^n(\mathbb{X})\big]$. Given the Laplace transform of CRM, the moments (see, e.g., Proposition 1 in Arbel and Prünster, 2016) take on the form

$$m_n = \sum_{(*)} \binom{n}{k_1 \cdots k_n} \prod_{i=1}^{n} \big(\kappa_i / i!\big)^{k_i}, \tag{5}$$

where the sum $(*)$ is over all n-tuples of nonnegative integers $(k_1, \ldots, k_n)$ satisfying the constraint $k_1 + 2k_2 + \cdots + nk_n = n$ and κ_i is the ith cumulant defined by $\kappa_i = a \int_0^{\infty} v^i \rho(\mathrm{d}v)$. Simple algebra leads to the following expression for the cumulants of the superposed gamma process

$$\kappa_i = a(i-1)!\zeta_{\eta}(i) \tag{6}$$

which displays the *incomplete Euler–Riemann zeta function* $\zeta_{\eta}(i) = \sum_{l=1}^{\eta} \frac{1}{l^i}$. Hence the moment-matching methodology introduced by Arbel and Prünster (2016) can be readily applied by making use of (5) and (6).

References

Arbel, J. and Prünster, I. (2016). A moment-matching Ferguson & Klass algorithm. *Stat. Comput., in press*. arxiv:1606.02566.

Argiento, R., Bianchini, I., and Guglielmi, A. (2016). A blocked Gibbs sampler for NGG-mixture models via a priori truncation. *Stat. Comput.*, 26(3):641–661.

Barrios, E., Lijoi, A., Nieto-Barajas, L. E., and Prünster, I. (2013). Modeling with normalized random measure mixture models. *Statist. Sci.*, 28(3):313–334.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Adv. Appl. Prob.*, 31:929–953.

Campbell, T., Huggins, J. H., How, J., and Broderick, T. (2016). Truncated Random Measures. *arXiv preprint arXiv:1603.00861*.

De Blasi, P., Favaro, S., and Muliere, P. (2010). A class of neutral to the right priors induced by superposition of beta processes. *J. Stat. Plan. Inference*, 140(6):1563–1575.

Epifani, I., Lijoi, A., and Prünster, I. (2003). Exponential functionals and means of neutral-to-the-right priors. *Biometrika*, 90(4):791–808.

Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1(2):209–230.

Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increment processes without Gaussian components. *Ann. Math. Stat.*, 43(5):1634–1643.

Griffin, J. E. and Walker, S. G. (2011). Posterior simulation of normalized random measure mixtures. *J. Comput. Graph. Stat.*, 20(1):241–259.

Hjort, N. L. (1990). Nonparametric bayes estimators based on beta processes in models for life history data. *Ann. Statist.*, 18(3):1259–1294.

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

Jordan, M. I. (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*, pages 207–218.

Kingman, J. (1967). Completely random measures. *Pacific J. Math.*, 21(1):59–78.

Lijoi, A., Mena, R. H., and Prünster, I. (2005). Bayesian nonparametric analysis for a generalized Dirichlet process prior. *Stat. Inference Stoch. Process.*, 8(3):283–309.

Lijoi, A. and Prünster, I. (2010). Models beyond the Dirichlet process. In Hjort, N. L., Holmes, C. C., Müller, P., and Walker, S. G., editors, *Bayesian nonparametrics*, pages 80–136. Cambridge University Press, Cambridge.

Nieto-Barajas, L. E., Prünster, I., and Walker, S. G. (2004). Normalized random measures driven by increasing additive processes. *Ann. Statist.*, 32(6):2343–2360.

Nieto-Barajas, L. E. and Walker, S. G. (2004). Bayesian nonparametric survival analysis via Lévy driven Markov processes. *Stat. Sin.*, 14(4):1127–1146.

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

Teh, Y. W. and Gorur, D. (2009). Indian buffet processes with power-law behavior. In *Advances in neural information processing systems*, pages 1838–1846.

Walker, S. G. and Damien, P. (2000). Representations of Lévy processes without Gaussian components. *Biometrika*, 87(2):477–483.

A Ferguson and Klass illustration

From Figure 2 it is apparent that increasing η leads to larger jumps which in turn leads to the need of a higher truncation level in order to match a given precision level. This is not surprising given the CRM at hand can be thought of as a superposition of η gamma CMRs. Such an intuition is made precise in Section 3 in the main text by (i) deriving a bound in probability on T_M and (ii) obtaining the moments of the superposed gamma process, thus showing the applicability of the moment-criterion introduced by Arbel and Prünster (2016).

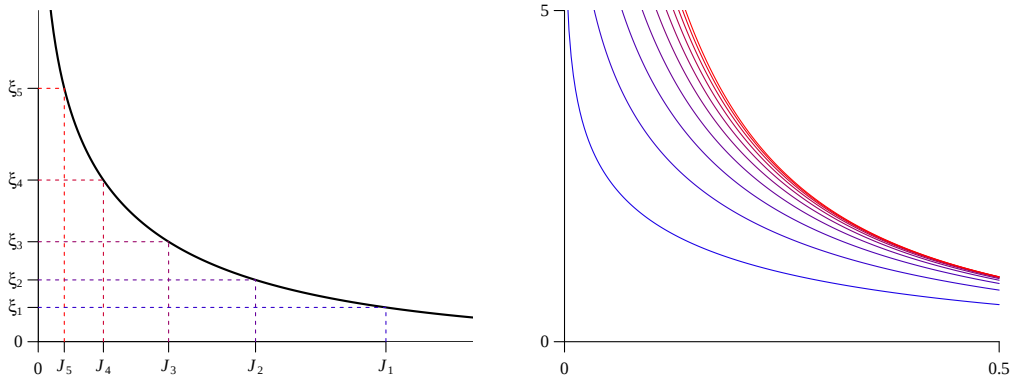


Fig. 2: Left: illustration of Ferguson and Klass representation through the inversion of the jumps times $\xi_1,\ldots,\xi_5$ for a homogeneous Poisson process on $\mathbb{R}_+$ to the jumps $J_1,\ldots,J_5$ of a CRM. Right: Tail of the Lévy measure $N(\cdot)$ of the superposed gamma process with $\eta \in \{1,\ldots,10\}$, $\eta = 1$ for the lowest curve, $\eta = 10$ for the highest curve.

B Proof

Proof of Proposition 1. The proof follows along the same lines as the proof of Theorem A.1. by Brix (1999) for the generalized gamma process and Proposition 4 by Arbel and Prünster (2016) for the stable-beta process. Let q_j denote the $\epsilon 2^{M-j}$ quantile, for $j = M+1, M+2, \ldots$, of a gamma distribution with mean and variance equal to j. Then

$$\mathbb{P}\Big(\sum_{j=M+1}^{\infty} N^{-1}(\xi_j) \leq \sum_{j=M+1}^{\infty} N^{-1}(q_j) \Big) \geq 1-\epsilon.$$

Denote $\tilde{t}_M^{\epsilon} = \sum_{j=M+1}^{\infty} N^{-1}(q_j) = \sum_{j=M+1}^{\infty} E_{\eta}^{-1}(q_j/a)$, and let us upper bound E_{η}^{-1}. By using $E_1(u) \leq 1 - \log(u)$, one gets

$$E_{\eta}(u) = \sum_{l=1}^{\eta} E_1(lu) \leq \eta - \sum_{l=1}^{\eta} \log(lu) = \eta - \log\left(\eta! u^{\eta}\right),$$

which can be inverted to obtain

$$E_{\eta}^{-1}(x) \leq \frac{1}{(\eta!)^{1/\eta}} \mathrm{e}^{1-\frac{x}{\eta}}.$$

Additionally, since the quantiles satisfy $q_j \geq \frac{\epsilon}{2\mathrm{e}} j$, we can conclude that

$$\tilde{t}_M^{\epsilon} \leq \frac{1}{(\eta!)^{1/\eta}} \sum_{j=M+1}^{\infty} \mathrm{e}^{1-\frac{q_j}{a\eta}} \leq \frac{1}{(\eta!)^{1/\eta}} \sum_{j=M+1}^{\infty} \mathrm{e}^{1-\frac{\epsilon j}{2\mathrm{e}a\eta}} \leq \frac{2\mathrm{e}a\eta}{\epsilon(\eta!)^{1/\eta}} \mathrm{e}^{1-\frac{\epsilon M}{2\mathrm{e}a\eta}}.$$

□