

# On a class of priors for Bayesian Nonparametrics

*Su una famiglia di distribuzioni iniziali  
per la statistica bayesiana nonparametrica*

Lancelot F. James  
Dep. Inform. Systems Managem.  
Hong Kong University  
of Science and Technology  
lancelot@ust.hk

Antonio Lijoi,<sup>1</sup> Igor Prünster,<sup>2</sup>  
Dip. Econom. Pol. Met. Quantitat.  
Università di Pavia  
lijoi@unipv.it, igor.pruenster@unipv.it

**Riassunto:** Viene presentata una classe di misure di probabilità aleatorie ottenute mediante la normalizzazione di misure aleatorie ad incrementi indipendenti. Le leggi di tali misure di probabilità aleatorie, dette anche *RMI normalizzate*, possono svolgere il ruolo di distribuzioni iniziali per l'inferenza bayesiana nonparametrica. Si dimostra che esse godono di proprietà analoghe a quelle ben note del processo di Dirichlet. In ipotesi di scambiabilità delle osservazioni, è possibile fornire una caratterizzazione della distribuzione finale di una RMI normalizzata in termini di mistura. Sfruttando tale caratterizzazione, si ottiene un'espressione della distribuzione finale della media di una RMI normalizzata. Infine, viene fornita una rappresentazione delle leggi predittive ottenute a partire da una RMI normalizzata. Essa consiste di una combinazione lineare del parametro della RMI e di una distribuzione empirica ponderata.

**Keywords:** Bayesian nonparametric inference, posterior analysis, predictive distributions, random measures with independent increments.

## 1. Introduction.

Since the Dirichlet process has been introduced in the celebrated paper of Ferguson (1973), several attempts have been made in order to generalize it. To this end various approaches have been undertaken: a review of the literature in the area can be found in Walker *et al.* (1999). Although it is well-known that the Dirichlet process can be defined by normalizing the increments of a gamma process, Bayesians have seldom considered “normalization” as a tool for defining random probability measures. Indeed, this idea has been exploited and applied in a variety of contexts not closely related to Bayesian inference such as storage problems, computer science, population genetics, ecology, statistical physics, combinatorics, number theory and excursion theory. See Pitman (2002) for references. Some important theoretical insight on the properties of normalized random measures were first given in Kingman (1975), where a random discrete distribution generated by the so-called stable subordinator is considered. Further developments can be found in Perman *et al.* (1992), where a description of the atoms of random probability measures, obtained by normalizing increasing processes with independent and stationary increments, in terms of a stick-breaking procedure, is provided. From a Bayesian

---

<sup>1</sup>Indirizzo per corrispondenza: Antonio Lijoi, Dipartimento di Economia Politica e Metodi Quantitativi, Università di Pavia, Via San Felice 5, 27100 Pavia

<sup>2</sup>Il lavoro è stato svolto con il supporto dei fondi PRIN 2002 “*Metodi bayesiani nonparametrici e loro applicazioni*”

perspective, the idea of normalization has been taken up again in a paper by Regazzini *et al.* (2003), where a *normalized random measure with independent increments* (normalized RMI) is defined as a random probability measure obtained by normalizing a suitably time-changed increasing process with independent but not necessarily stationary increments. To our purposes, it is useful to consider a slight modification of the definition of normalized RMI in order to cover the case of random probability measures defined on abstract spaces as done in James (2002).

Let  $\mu$  be a random element defined on some probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  and with values in the space of measures on  $(\mathbb{X}, \mathcal{X})$ , where  $\mathbb{X}$  is a Polish space equipped with the corresponding Borel  $\sigma$ -field  $\mathcal{X}$ . According to Kingman (1967),  $\mu$  is a *completely random measure* on  $(\mathbb{X}, \mathcal{X})$  if  $\mu(\emptyset) = 0$  a.s.- $\mathbb{P}$  and, for any collection  $\{A_1, A_2, \dots\}$  of disjoint sets in  $\mathcal{X}$ , the random variables  $\mu(A_1), \mu(A_2), \dots$  are mutually independent with  $\mu(\cup_{j \geq 1} A_j) = \sum_{j \geq 1} \mu(A_j)$  a.s.- $\mathbb{P}$ . Such a completely random measure is characterized by its Laplace transform

$$\mathbb{E} [e^{-\lambda \mu(A)}] = e^{-\int_{A \times \mathbb{R}^+} (1 - e^{-\lambda v}) \nu(dx, dv)} \quad (A \in \mathcal{X}, \lambda \geq 0),$$

where  $\nu$  is a measure on the product  $\sigma$ -field  $\mathcal{X} \otimes \mathcal{B}(\mathbb{R}^+)$  such that  $\int_{\mathbb{R}^+} (v \wedge 1) \nu(dx, dv) < +\infty$  for every  $x \in \mathbb{X}$ . Since we aim at defining random probability measures by means of normalization of completely random measures, we need to require the total mass  $\mu(\mathbb{X})$  being finite and positive, almost surely. This happens if  $\nu(\mathbb{X} \times \mathbb{R}^+) = +\infty$  and the Laplace exponent  $\psi(\lambda) := \int_{\mathbb{X} \times \mathbb{R}^+} (1 - e^{-\lambda v}) \nu(dx, dv)$  is finite for any positive  $\lambda$ . When these conditions hold true, one can define a *normalized RMI* on  $(\mathbb{X}, \mathcal{X})$  as

$$\tilde{P}(\cdot) = \frac{\mu(\cdot)}{\mu(\mathbb{X})}. \quad (1)$$

Note that the Dirichlet process with parameter measure  $\alpha$  is recovered by taking  $\nu(dx, dv) = \alpha(dx) e^{-v} v^{-1} dv$ . In the sequel we assume, without loss of generality, that  $\nu(dx, dv) = \rho(dv|x) \alpha(dx)$  where  $\alpha$  is a positive and finite measure on  $\mathbb{X}$  and  $\rho(\cdot|x)$  is a positive measure on  $\mathbb{R}^+$ , for any  $x$  in  $\mathbb{X}$ .

The results that will be presented in the following sections are contained in James *et al.* (2004).

## 2. Posterior characterizations.

The evaluation of the posterior distribution of a random quantity is the main issue in Bayesian statistics. However, when dealing with random probability measures, the determination of the posterior distribution is a demanding issue. In particular, for normalized RMIs difficulties arise since, apart from the Dirichlet process, they are not conjugate. Here we fill this gap by providing useful characterizations of their posterior distributions. To this end, suppose that the sequence of observations  $(X_n)_{n \geq 1}$  taking values in  $\mathbb{X}$  is *exchangeable*, that is the  $X_n$ 's are, conditionally on  $\tilde{P}$ , independent and identically distributed, their common probability distribution being  $\tilde{P}$ . Moreover, let  $Y_1, \dots, Y_k$  denote the  $k$  distinct observations within the sample  $X^{(n)} := (X_1, \dots, X_n)$ .

If  $\tilde{P}$  is a normalized RMI as in (1), it can be proved that the posterior distribution of  $\mu$ , given  $X^{(n)}$ , is a mixture with respect to the distribution of a latent random variable  $U$ . Conditionally on  $U$  and  $X^{(n)}$ ,  $\mu$  can be represented as

$$\mu_u + \sum_{i=1}^k J_i^{(u)} \delta_{Y_i}$$

where  $\delta_x$  denotes the unit mass at  $x$  in  $\mathbb{X}$ ,  $\mu_u$  is a completely random measure characterized by  $\nu_u(dx, dv) = e^{-uv} \rho(dv|x) \alpha(dx)$ , the  $Y_i$ 's are the fixed points of discontinuity and the  $J_i^{(u)}$ 's are the corresponding jumps. Moreover,  $\mu_u$  and the  $J_i^{(u)}$  are independent and explicit expressions for the distribution of  $U$  and of the jumps are derived.

This result sheds some light over the deep structure of the random measures at issue. It essentially shows that, given some latent variable, *a posteriori*  $\mu$  is still a completely random measure with some fixed points of discontinuity corresponding to the locations of the observations. Since all involved quantities are explicit, the posterior description of  $\mu$  is complete. One may note that this characterization is somehow reminiscent of the (well-known) one for processes with independent and non-negative increments used for defining neutral to the right priors. Besides the analogy, it is worth remarking two substantial differences. The first is due to the non-conjugacy of normalized RMIs: in contrast to the neutral to the right case, here we first have to identify an appropriate latent variable and then, conditionally on it, look for a posterior characterization of  $\mu$ . The second is due to the type of transformation of  $\mu$  employed for defining the random probability measures: normalized RMIs are obtained via normalization whereas neutral to the right priors via an exponential transformation and, this, clearly affects the updating mechanism of the intensity measure  $\nu$  and the distribution of the jumps.

Given the posterior characterization of  $\mu$  we obtain also a mixture representation of the posterior distribution of the normalized RMI. Conditionally on  $U$  and on  $X^{(n)}$ ,  $\tilde{P}$  coincides in distribution with

$$W_u \frac{\mu_u}{\mu_u(\mathbb{R})} + (1 - W_u) \frac{\sum_{i=1}^k J_i^{(u)} \delta_{Y_i}}{\sum_{r=1}^k J_r^{(u)}}$$

where  $W_u = \mu_u(\mathbb{R}) \left[ \mu_u(\mathbb{R}) + \sum_{i=1}^k J_i^{(u)} \right]^{-1}$ . The above mentioned results are useful for dealing with means of normalized RMIs, i.e.

$$\int_{\mathbb{X}} f(x) \tilde{P}(dx)$$

for any real-valued function  $f$  such that  $\int |f| d\tilde{P} < +\infty$  a.s.- $\mathbb{P}$ . An exact expression for the posterior distribution of  $\int f d\tilde{P}$  is obtained thus completing the work of Regazzini *et al.* (2003) where an approximation to this distribution has been given.

### 3. Predictive distributions.

An important goal in inferential procedures is the prediction of future values of a random quantity based on its past outcomes. Once we have characterized the posterior distribution of a normalized RMI, the determination of the corresponding predictive distributions is quite straightforward. The latter have quite intuitive forms, since they consist of a linear combination of the parameter measure of the RMI and of a *weighted* version

of the empirical distribution. Indeed, it is proved that

$$\mathbb{P}(X_{n+1} \in dx_{n+1} | X^{(n)}) = w^{(n)} \alpha(dx_{n+1}) + \frac{1}{n} \sum_{j=1}^k w_j^{(n)} \delta_{Y_j}(dx_{n+1}).$$

Exact forms for the weights  $w^{(n)}$  and  $w_j^{(n)}$  can be obtained and are expressed in terms of the quantity  $\tau_m(u|x) := \int_{\mathbb{R}^+} v^m e^{-uv} \rho(dv|x)$ , where  $m$  is a positive integer. More precisely, one has

$$w^{(n)} = \frac{1}{n} \int_0^{+\infty} u \tau_1(u|x_{n+1}) q(u|X^{(n)}) du, \quad w_j^{(n)} = \int_0^{+\infty} u \frac{\tau_{n_j+1}(u|Y_j)}{\tau_{n_j}(u|Y_j)} q(u|X^{(n)}) du$$

where  $q(\cdot | X^{(n)})$  is a density, with respect to the Lebesgue measure, of the conditional distribution of the latent variable  $U$ , given  $X^{(n)}$ , and it is known up to a proportionality constant. The above prediction rule represents a generalization of the one provided in Pitman (2003).

Finally, it is worth mentioning that the notion of normalized RMI can be extended to what we term *weighted normalized RMI*. With suitable modifications, the previous results providing posterior characterizations and explicit forms for the predictive distributions still apply. Such an extension is important since it allows to recover the two parameter Poisson-Dirichlet family, introduced by Pitman (1995), as a particular case.

## References

- Ferguson T. (1973) A bayesian analysis of some nonparametric problems, *The Annals of Statistics*, 1, 209–230.
- James L. (2002) Poisson process partition calculus with applications to exchangeable models and bayesian nonparametrics, *Mathematics ArXiv*.
- James L., Lijoi A. and Prünster I. (2004) Bayesian nonparametric inference via normalized random measures with independent increments, *Technical Report*.
- Kingman J. (1967) Completely random measures, *Pacific Journal of Mathematics*, 21, 59–78.
- Kingman J. (1975) Random discrete distributions, *Journal of the Royal Statistical Society, Series B*, 37, 1–22.
- Perman M., Pitman J. and Yor M. (1992) Size-biased sampling of poisson point processes and excursions, *Probabability Theory and Related Fields*, 92, 21–39.
- Pitman J. (1995) Exchangeable and partially exchangeable random partitions, *Probabability Theory and Related Fields*, 102, 145–158.
- Pitman J. (2002) *Combinatorial stochastic processes*, Lecture Notes for St. Flour Summer School, Springer.
- Pitman J. (2003) Poisson-kingman partitions, *Science and Statistics: A Festschrift for Terry Speed (I.M.S. Lecture Notes)*, 40, 1–35.
- Regazzini E., Lijoi A. and Prünster I. (2003) Distributional results for means of random measures with independent increments, *The Annals of Statistics*, 31, 560–585.
- Walker S., Damien P., Laud P. and Smith A. (1999) Bayesian nonparametric inference for random distributions and related functions, *Journal of the Royal Statistical Society, Series B*, 61, 485–527.