On the posterior structure of NRMI

Igor Prünster

University of Turin, Collegio Carlo Alberto and ICER

Joint work with L.F. James and A. Lijoi

Isaac Newton Institute, BNR Programme, 8th August 2007

Outline

CRM and NRMI

- Completely random measures (CRM)
- NRMIs
- Relation to other random probability measures

Outline

Outline

Outline

Completely random measures

DEFINITION (Kingman, 1967).

$\tilde{\mu}$ is a completely random measure (CRM) on $(\mathbb{X}, \mathscr{X})$ if

(i) $\tilde{\mu}(\emptyset) = 0$

(ii) *for any collection of disjoint sets in $\mathscr{X}$, $B_1, B_2, \ldots$, the random variables $\tilde{\mu}(B_1), \tilde{\mu}(B_2), \ldots$ are mutually independent and*

$$\tilde{\mu}(\cup_{j\geq 1} B_j) = \sum_{j\geq 1} \tilde{\mu}(B_j)$$

Completely random measures

DEFINITION (Kingman, 1967).
$\tilde{\mu}$ is a completely random measure (CRM) on $(\mathbb{X}, \mathscr{X})$ if

(i) $\tilde{\mu}(\emptyset) = 0$

(ii) *for any collection of disjoint sets in $\mathscr{X}$, $B_1, B_2, \ldots$, the random variables $\tilde{\mu}(B_1), \tilde{\mu}(B_2), \ldots$ are mutually independent and*

$$\tilde{\mu}(\cup_{j\geq 1} B_j) = \sum_{j\geq 1} \tilde{\mu}(B_j)$$

Let $\mathscr{G}_\nu = \{g : \int_{\mathbb{X}} g(x)\tilde{\mu}(\mathrm{d}x) < \infty\}$. Then, $\tilde{\mu}$ is characterized by its Laplace functional

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}} g(x)\,\tilde{\mu}(\mathrm{d}x)}\right] = \exp\left\{-\int_{\mathbb{R}^+\times\mathbb{X}} [1 - \mathrm{e}^{-v\,g(x)}]\,\nu(\mathrm{d}v, \mathrm{d}x)\right\}$$

for any $g \in \mathscr{G}_\nu$. In the following, denote by $\psi(\,\cdot\,)$ the Laplace exponent $\int_{\mathbb{R}^+\times\mathbb{X}}[1 - \mathrm{e}^{-v\,\cdot}]\,\nu(\mathrm{d}v, \mathrm{d}x)$.

Completely random measures

DEFINITION (Kingman, 1967).

$\tilde{\mu}$ is a completely random measure (CRM) on $(\mathbb{X}, \mathscr{X})$ if

(i) $\tilde{\mu}(\emptyset) = 0$

(ii) *for any collection of disjoint sets in $\mathscr{X}$, $B_1, B_2, \ldots$, the random variables $\tilde{\mu}(B_1), \tilde{\mu}(B_2), \ldots$ are mutually independent and*

$$\tilde{\mu}(\cup_{j\geq 1} B_j) = \sum_{j\geq 1} \tilde{\mu}(B_j)$$

Let $\mathscr{G}_\nu = \{g : \int_{\mathbb{X}} g(x)\tilde{\mu}(\mathrm{d}x) < \infty\}$. Then, $\tilde{\mu}$ is characterized by its Laplace functional

$$\mathbb{E}\left[\mathrm{e}^{-\int_{\mathbb{X}} g(x)\,\tilde{\mu}(\mathrm{d}x)}\right] = \exp\left\{-\int_{\mathbb{R}^+\times\mathbb{X}} [1-\mathrm{e}^{-v\,g(x)}]\;\nu(\mathrm{d}v, \mathrm{d}x)\right\}$$

for any $g \in \mathscr{G}_\nu$. In the following, denote by $\psi(\,\cdot\,)$ the Laplace exponent $\int_{\mathbb{R}^+\times\mathbb{X}} [1-\mathrm{e}^{-v\,\cdot}]\;\nu(\mathrm{d}v, \mathrm{d}x)$.

$\Longrightarrow$ $\tilde{\mu}$ is identified by the intensity ν (which represents the intensity of the underlying Poisson random measure).

Letting α be a non–atomic and σ–finite measure on $\mathbb{X}$.
According to the decomposition of ν we distinguish two classes of CRM:

(a) if $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v)\,\alpha(\mathrm{d}x)$, we say that $\tilde{\mu}$ is homogeneous;

(b) if $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v|x)\,\alpha(\mathrm{d}x)$, we say that $\tilde{\mu}$ is non–homogeneous.

Letting α be a non–atomic and σ–finite measure on $\mathbb{X}$.
According to the decomposition of ν we distinguish two classes of CRM:

(a) if $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v)\,\alpha(\mathrm{d}x)$, we say that $\tilde{\mu}$ is homogeneous;

(b) if $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v|x)\,\alpha(\mathrm{d}x)$, we say that $\tilde{\mu}$ is non–homogeneous.

Necessary assumptions for the normalization to be well defined:

(A) $\tilde{\mu}$ is almost surely finite

$\iff \int_{\mathbb{R}^+\times\mathbb{X}}[1 - \mathrm{e}^{-\lambda v}]\,\nu(\mathrm{d}v, \mathrm{d}x) < \infty$ for every $\lambda > 0$

if $\tilde{\mu}$ is a homogeneous CRM $\iff \alpha$ being a finite measure

Letting α be a non–atomic and σ–finite measure on $\mathbb{X}$.
According to the decomposition of ν we distinguish two classes of CRM:

(a) if $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v)\,\alpha(\mathrm{d}x)$, we say that $\tilde{\mu}$ is homogeneous;
(b) if $\nu(\mathrm{d}v, \mathrm{d}x) = \rho(\mathrm{d}v|x)\,\alpha(\mathrm{d}x)$, we say that $\tilde{\mu}$ is non–homogeneous.

Necessary assumptions for the normalization to be well defined:

(A) $\tilde{\mu}$ is almost surely finite

$\iff \int_{\mathbb{R}^+\times\mathbb{X}}[1 - \mathrm{e}^{-\lambda v}]\,\nu(\mathrm{d}v, \mathrm{d}x) < \infty$ for every $\lambda > 0$

if $\tilde{\mu}$ is a homogeneous CRM $\iff$ α being a finite measure

(B) $\tilde{\mu}$ is almost surely strictly positive

$\iff \nu(\mathbb{R}^+, \mathbb{X}) = \infty \iff$ infinite activity of $\tilde{\mu}$

Normalized random measures with independent increments (NRMIs)

DEFINITION. Let $\tilde{\mu}$ be a CRM on $(\mathbb{X}, \mathscr{X})$ satisfying (A) and (B). Then the random probability measure on $(\mathbb{X}, \mathscr{X})$ given by

$$\tilde{P}(\,\cdot\,) = \frac{\tilde{\mu}(\,\cdot\,)}{\tilde{\mu}(\mathbb{X})}$$

is well defined and termed normalized random measure with independent increments (NRMI).

Normalized random measures with independent increments (NRMIs)

DEFINITION. Let $\tilde{\mu}$ be a CRM on $(\mathbb{X}, \mathscr{X})$ satisfying (A) and (B). Then the random probability measure on $(\mathbb{X}, \mathscr{X})$ given by

$$\tilde{P}(\,\cdot\,) = \frac{\tilde{\mu}(\,\cdot\,)}{\tilde{\mu}(\mathbb{X})}$$

is well defined and termed normalized random measure with independent increments (NRMI).

A NRMI is uniquely characterized by the intensity ν of the corresponding CRM $\tilde{\mu}$: according to the structure of ν we will distinguish homogeneous and non–homogeneous NRMI.

Special cases of NRMI

1. Dirichlet process:
   Let $\tilde{\mu}$ be a gamma CRM with α a finite measure on $\mathbb{X}$ and

$$\nu(\mathrm{d}v, \mathrm{d}x) = \frac{\mathrm{e}^{-v}}{v}\mathrm{d}v \; \alpha(\mathrm{d}x)$$

   $\Longrightarrow$ NRMI is a Dirichlet process with parameter measure α.

Special cases of NRMI

1. Dirichlet process:
   Let $\tilde{\mu}$ be a gamma CRM with α a finite measure on $\mathbb{X}$ and
   $$\nu(\mathrm{d}v, \mathrm{d}x) = \frac{\mathrm{e}^{-v}}{v}\mathrm{d}v\ \alpha(\mathrm{d}x)$$
   $\Longrightarrow$ NRMI is a Dirichlet process with parameter measure α.
2. Normalized generalized gamma (GG) process:
   Let $\tilde{\mu}$ be a GG–CRM (Brix, '99) with α a finite measure on $\mathbb{X}$ and
   $$\nu(\mathrm{d}v, \mathrm{d}x) = \frac{\alpha}{\Gamma(1-\alpha)}\, s^{-1-\alpha}\,\mathrm{e}^{-\tau s}\mathrm{d}s\,\alpha(\mathrm{d}x)$$
   $\Longrightarrow$ NRMI is a normalized GG process with parameter α.

Special cases of NRMI

1. Dirichlet process:
   Let $\tilde{\mu}$ be a gamma CRM with α a finite measure on $\mathbb{X}$ and
   $$\nu(\mathrm{d}v, \mathrm{d}x) = \frac{\mathrm{e}^{-v}}{v} \mathrm{d}v\, \alpha(\mathrm{d}x)$$
   $\Longrightarrow$ NRMI is a Dirichlet process with parameter measure α.
2. Normalized generalized gamma (GG) process:
   Let $\tilde{\mu}$ be a GG–CRM (Brix, '99) with α a finite measure on $\mathbb{X}$ and
   $$\nu(\mathrm{d}v, \mathrm{d}x) = \frac{\alpha}{\Gamma(1-\alpha)}\, s^{-1-\alpha}\, \mathrm{e}^{-\tau s} \mathrm{d}s\, \alpha(\mathrm{d}x)$$
   $\Longrightarrow$ NRMI is a normalized GG process with parameter α.
3. Normalized extended gamma process:
   Let $\tilde{\mu}$ be an extended gamma CRM (Dykstra & Laud, '81) with
   $$\nu(\mathrm{d}v, \mathrm{d}x) = \frac{e^{-b(t)v}}{v}\, \mathrm{d}v\, \mathrm{d}t$$
   with b a strictly positive function and α s.t. $\mu(\mathbb{X}) < \infty$ a.s.
   $\Longrightarrow$ NRMI is a normalized extended gamma process with parameters α and b.

Relation to other random probability measures

Homogeneous NRMI are members of the following families of random probability measures:

(i) Species sampling models (Pitman, ’96) are defined as

$$\tilde{P}(\,\cdot\,) = \sum_{i\geq 1} \tilde{P}_i \, \delta_{X_i}(\,\cdot\,) + \Big(1 - \sum_{i\geq 1} \tilde{P}_i\Big) H(\,\cdot\,)$$

where $0 < \tilde{P}_i < 1$ are random weights such that $\sum_{i\geq 1} \tilde{P}_i \leq 1$, independent of the locations X_i, which are i.i.d. with some non–atomic distribution H.

Relation to other random probability measures

Homogeneous NRMI are members of the following families of random probability measures:

(i) Species sampling models (Pitman, '96) are defined as

$$\tilde{P}(\,\cdot\,) = \sum_{i\geq 1} \tilde{P}_i \, \delta_{X_i}(\,\cdot\,) + \Big(1 - \sum_{i\geq 1} \tilde{P}_i\Big) H(\,\cdot\,)$$

where $0 < \tilde{P}_i < 1$ are random weights such that $\sum_{i\geq 1} \tilde{P}_i \leq 1$, independent of the locations X_i, which are i.i.d. with some non–atomic distribution H.

• Problem: concrete assignment of the random weights $\tilde{P}_i$: Stick-breaking procedure (Ishwaran and James 2001).

Remark: A non–homogeneous NRMI is not a species sampling model: weights and locations are not independent.

Relation to other random probability measures

Homogeneous NRMI are members of the following families of random probability measures:

(i) Species sampling models (Pitman, '96) are defined as

$$\tilde{P}(\cdot) = \sum_{i\geq 1} \tilde{P}_i \, \delta_{X_i}(\cdot) + \Big(1 - \sum_{i\geq 1} \tilde{P}_i\Big) H(\cdot)$$

where $0 < \tilde{P}_i < 1$ are random weights such that $\sum_{i\geq 1} \tilde{P}_i \leq 1$, independent of the locations X_i, which are i.i.d. with some non–atomic distribution H.
• Problem: concrete assignment of the random weights $\tilde{P}_i$: Stick-breaking procedure (Ishwaran and James 2001).
Remark: A non–homogeneous NRMI is not a species sampling model: weights and locations are not independent.

(ii) Poisson–Kingman models (Pitman, '03):
More tractable than general species sampling models, but is still difficult to derive expressions for posterior quantities.

Characterization of the Dirichlet process

$(X_n)_{n\geq 1}$ is a sequence of exchangeable observations with values in $\mathbb{X}$ governed by a NRMI.

Characterization of the Dirichlet process

$(X_n)_{n\geq 1}$ is a sequence of exchangeable observations with values in $\mathbb{X}$ governed by a NRMI.

A sample $X^{(n)} = (X_1, \ldots, X_n)$ will contain:

- $X_1^*, \ldots, X_k^*$ the k distinct observations in $X^{(n)}$
- $n_j > 0$ the number of observations equal to X_j^* ($j = 1, \ldots, k$).

Characterization of the Dirichlet process

$(X_n)_{n\geq 1}$ is a sequence of exchangeable observations with values in $\mathbb{X}$ governed by a NRMI.

A sample $X^{(n)} = (X_1, \ldots, X_n)$ will contain:
- $X_1^*, \ldots, X_k^*$ the k distinct observations in $X^{(n)}$
- $n_j > 0$ the number of observations equal to X_j^* $(j = 1, \ldots, k)$.

Let $\mathscr{P}$ be the set of all NRMIs and let $\tilde{P} \in \mathscr{P}$. The posterior distribution of $\tilde{P}$, given $X^{(n)}$, is still in $\mathscr{P}$ if and only if $\tilde{P}$ is a Dirichlet process.

Characterization of the Dirichlet process

$(X_n)_{n\geq 1}$ is a sequence of exchangeable observations with values in $\mathbb{X}$ governed by a NRMI.

A sample $X^{(n)} = (X_1, \ldots, X_n)$ will contain:
- $X_1^*, \ldots, X_k^*$ the k distinct observations in $X^{(n)}$
- $n_j > 0$ the number of observations equal to X_j^* ($j = 1, \ldots, k$).

Let $\mathscr{P}$ be the set of all NRMIs and let $\tilde{P} \in \mathscr{P}$. The posterior distribution of $\tilde{P}$, given $X^{(n)}$, is still in $\mathscr{P}$ if and only if $\tilde{P}$ is a Dirichlet process.

$\Longrightarrow$ CONJUGACY is a distinctive feature of the Dirichlet process.

Characterization of the Dirichlet process

$(X_n)_{n\geq 1}$ is a sequence of exchangeable observations with values in $\mathbb{X}$ governed by a NRMI.

A sample $X^{(n)} = (X_1, \dots, X_n)$ will contain:
- $X_1^*, \dots, X_k^*$ the k distinct observations in $X^{(n)}$
- $n_j > 0$ the number of observations equal to X_j^* $(j = 1, \dots, k)$.

Let $\mathscr{P}$ be the set of all NRMIs and let $\tilde{\mathrm{P}} \in \mathscr{P}$. The posterior distribution of $\tilde{\mathrm{P}}$, given $X^{(n)}$, is still in $\mathscr{P}$ if and only if $\tilde{\mathrm{P}}$ is a Dirichlet process.

$\Longrightarrow$ CONJUGACY is a distinctive feature of the Dirichlet process.

Nonetheless, conditionally on a latent variable U and the data $X^{(n)}$, the (posterior) NRMI $\tilde{P}|X^{(n)}, U$ is still a NRMI.

The latent variable *U*

U is not an auxiliary variable: it has a precise meaning summarizing the normalization procedure and the distribution of $\tilde{\mu}(\mathbb{X})$.

The latent variable U

U is not an auxiliary variable: it has a precise meaning summarizing the normalization procedure and the distribution of $\tilde{\mu}(\mathbb{X})$.

- $\tau_m(u|x) = \int_{\mathbb{R}^+} s^m \, \mathrm{e}^{-us} \, \rho_x(\mathrm{d}s)$ for any $m \geq 1$ and $x \in \mathbb{X}$

The latent variable U

U is not an auxiliary variable: it has a precise meaning summarizing the normalization procedure and the distribution of $\tilde{\mu}(\mathbb{X})$.

- $\tau_m(u|x) = \int_{\mathbb{R}^+} s^m \,\mathrm{e}^{-us}\, \rho_x(\mathrm{d}s)$ for any $m \geq 1$ and $x \in \mathbb{X}$
- U_0 is a positive random variable with density

$$f_0(u) \propto \mathrm{e}^{-\psi(u)} \int_{\mathbb{X}} \tau(u|x)\, \eta(\mathrm{d}x)$$

The latent variable U

U is not an auxiliary variable: it has a precise meaning summarizing the normalization procedure and the distribution of $\tilde{\mu}(\mathbb{X})$.

- $\tau_m(u|x) = \int_{\mathbb{R}^+} s^m \, \mathrm{e}^{-us} \, \rho_x(\mathrm{d}s)$ for any $m \geq 1$ and $x \in \mathbb{X}$
- U_0 is a positive random variable with density

$$f_0(u) \propto \mathrm{e}^{-\psi(u)} \int_{\mathbb{X}} \tau(u|x) \, \eta(\mathrm{d}x)$$

- U_n is a positive random variable whose density, conditional on the data $X^{(n)}$, is for any $n \geq 1$

$$f(u|X^{(n)}) \propto u^{n-1} \, \mathrm{e}^{-\psi(u)} \prod_{j=1}^{k} \tau_{n_j}(u|X_j^*)$$

The latent variable U

U is not an auxiliary variable: it has a precise meaning summarizing the normalization procedure and the distribution of $\tilde{\mu}(\mathbb{X})$.

- $\tau_m(u|x) = \int_{\mathbb{R}^+} s^m \, \mathrm{e}^{-us} \rho_x(\mathrm{d}s)$ for any $m \geq 1$ and $x \in \mathbb{X}$
- U_0 is a positive random variable with density

$$f_0(u) \propto \mathrm{e}^{-\psi(u)} \int_{\mathbb{X}} \tau(u|x) \, \eta(\mathrm{d}x)$$

- U_n is a positive random variable whose density, conditional on the data $X^{(n)}$, is for any $n \geq 1$

$$f(u|X^{(n)}) \propto u^{n-1} \, \mathrm{e}^{-\psi(u)} \prod_{j=1}^{k} \tau_{n_j}(u|X_j^*)$$

Remark: *The distribution of $(U|X^{(n)})$ is a mixture of gamma distributions with mixing measure the posterior total mass* $(\tilde{\mu}(\mathbb{X})|X^{(n)})$

$$f_{U|X^{(n)}}(u) = \int_{(0,+\infty)} \frac{y^n}{\Gamma(n)} \, u^{n-1} \, \mathrm{e}^{-yu} \, \bar{Q}\left(\mathrm{d}y|X^{(n)}\right)$$

where $\bar{Q}(\,\cdot\,|X^{(n)})$ denotes the posterior distribution of $\tilde{\mu}(\mathbb{X})$.

The posterior distribution of the CRM $\tilde{\mu}$

The posterior distribution of the CRM $\tilde{\mu}$

The posterior distribution of $\tilde{\mu}$, given $X^{(n)}$, is a mixture with respect to $f(u|X^{(n)})$

The posterior distribution of the CRM $\tilde{\mu}$

The posterior distribution of $\tilde{\mu}$, given $X^{(n)}$, is a mixture with respect to $f(u|X^{(n)})$

Given $U_n = u$ and $X^{(n)}$,

$$\tilde{\mu} \stackrel{d}{=} \tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)} \, \delta_{X_i^*}$$

where

The posterior distribution of the CRM $\tilde{\mu}$

The posterior distribution of $\tilde{\mu}$, given $X^{(n)}$, is a mixture with respect to $f(u|X^{(n)})$

Given $U_n = u$ and $X^{(n)}$,

$$\tilde{\mu} \stackrel{d}{=} \tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)} \, \delta_{X_i^*}$$

where

(i) jump $J_i^{(u)}$ at X_i^* has density $f_i(s) \propto s^{n_i} \mathrm{e}^{-us} \rho_{X_i^*}(\mathrm{d}s)$

The posterior distribution of the CRM $\tilde{\mu}$

The posterior distribution of $\tilde{\mu}$, given $X^{(n)}$, is a mixture with respect to $f(u|X^{(n)})$

Given $U_n = u$ and $X^{(n)}$,

$$\tilde{\mu} \stackrel{d}{=} \tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)} \, \delta_{X_i^*}$$

where

(i) jump $J_i^{(u)}$ at X_i^* has density $f_i(s) \propto s^{n_i} \mathrm{e}^{-us} \rho_{X_i^*}(\mathrm{d}s)$

(ii) $\tilde{\mu}_u$ is a CRM with intensity

$$\nu^{(u)}(\mathrm{d}x, \mathrm{d}s) = \mathrm{e}^{-us} \, \rho_x(\mathrm{d}s) \, \eta(\mathrm{d}x)$$

The posterior distribution of the CRM $\tilde{\mu}$

The posterior distribution of $\tilde{\mu}$, given $X^{(n)}$, is a mixture with respect to $f(u|X^{(n)})$

Given $U_n = u$ and $X^{(n)}$,

$$\tilde{\mu} \stackrel{d}{=} \tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)} \, \delta_{X_i^*}$$

where

(i) jump $J_i^{(u)}$ at X_i^* has density $f_i(s) \propto s^{n_i} \mathrm{e}^{-us} \rho_{X_i^*}(\mathrm{d}s)$

(ii) $\tilde{\mu}_u$ is a CRM with intensity

$$\nu^{(u)}(\mathrm{d}x, \mathrm{d}s) = \mathrm{e}^{-us} \, \rho_x(\mathrm{d}s) \, \eta(\mathrm{d}x)$$

(iii) $\tilde{\mu}_u$ and $J_i^{(u)}$ $(i = 1, \ldots, k)$ are independent.

The posterior distribution of the NRMI $\tilde{P}$

It now follows easily that given U_n and $X^{(n)}$, the (posterior) distribution of $\tilde{P}$ is again a NRMI:

The posterior distribution of the NRMI $\tilde{P}$

It now follows easily that given U_n and $X^{(n)}$, the (posterior) distribution of $\tilde{P}$ is again a NRMI:

$$\tilde{P} \mid U_n, X^{(n)}$$

$$= \stackrel{d}{=} \frac{\tilde{\mu}_u + \sum_{i=1}^k J_i^{(u)} \delta_{X_i^*}}{\tilde{\mu}_u(\mathbb{X}) + \sum_{i=1}^k J_i^{(u)}}$$

$$\stackrel{d}{=} w \frac{\tilde{\mu}_u}{\tilde{\mu}_u(\mathbb{X})} + (1-w) \frac{\sum_{i=1}^k J_i^{(u)} \delta_{X_i^*}}{\sum_{r=1}^k J_r^{(u)}}$$

with $w = \tilde{\mu}_u(\mathbb{X})(\tilde{\mu}_u(\mathbb{X}) + \sum_{i=1}^k J_i^{(u)})^{-1}$.

The posterior distribution of the normalized GG–process

Let $\tilde{P}$ be a normalized GG–process. Then the (posterior) distribution of $\tilde{\mu}$ given U_n and $X^{(n)}$, $\tilde{\mu}$ can be represented as

$$\tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)} \, \delta_{X_i^*}$$

where

The posterior distribution of the normalized GG–process

Let $\tilde{P}$ be a normalized GG–process. Then the (posterior) distribution of $\tilde{\mu}$ given U_n and $X^{(n)}$, $\tilde{\mu}$ can be represented as

$$\tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)} \, \delta_{X_i^*}$$

where

(i) $\tilde{\mu}_u$ is a GG CRM with intensity measure $\nu^{(u)}(\mathrm{d}s, \mathrm{d}x) = \frac{\sigma}{\Gamma(1-\sigma)} \, s^{-1-\sigma} \, \mathrm{e}^{-(u+1)s} \mathrm{d}s \, \alpha(\mathrm{d}x)$

(ii) the fixed points of discontinuity coincide with the distinct observations X_i^*,
the jumps $J_i \sim \mathrm{Gamma}(u+1, n_i - \sigma)$, for $i = 1, \ldots, k$;

(iii) $\tilde{\mu}^{(u)}$ and J_i $(i = 1, \ldots, k)$ are independent.

Moreover, the distribution of U, conditional on $X^{(n)}$, is

$$f(u|X^{(n)}) \propto \frac{u^{n-1} \, \mathrm{e}^{-\alpha(\mathbb{X})(1+u)^{\sigma}}}{(u+1)^{n-k\sigma}}.$$

Predictive distributions

The (predictive) distribution of X_{n+1}, given $X^{(n)}$, coincides with

$$P[X_{n+1} \in \mathrm{d}x_{n+1} | X_1, \ldots, X_n] = w^{(n)}\, \alpha(\mathrm{d}x_{n+1}) + \frac{1}{n} \sum_{j=1}^{k} w_j^{(n)}\, \delta_{X_j^*}(\mathrm{d}x_{n+1})$$

Predictive distributions

The (predictive) distribution of X_{n+1}, given $X^{(n)}$, coincides with

$$P[X_{n+1} \in \mathrm{d}x_{n+1} | X_1, \ldots, X_n] = w^{(n)}\, \alpha(\mathrm{d}x_{n+1}) + \frac{1}{n} \sum_{j=1}^{k} w_j^{(n)}\, \delta_{X_j^*}(\mathrm{d}x_{n+1})$$

where

$$w^{(n)} = \frac{1}{n} \int_0^{+\infty} u\, \tau_1(u|x_{n+1})\, f(u|X^{(n)})\, \mathrm{d}u$$

$$w_j^{(n)} = \int_0^{+\infty} u\, \frac{\tau_{n_j+1}(u|X_j^*)}{\tau_{n_j}(u|X_j^*)}\, f(u|X^{(n)})\, \mathrm{d}u$$

Predictive distributions

The (predictive) distribution of X_{n+1}, given $X^{(n)}$, coincides with

$$P[X_{n+1} \in \mathrm{d}x_{n+1} | X_1, \ldots, X_n] = w^{(n)}\, \alpha(\mathrm{d}x_{n+1}) + \frac{1}{n} \sum_{j=1}^{k} w_j^{(n)}\, \delta_{X_j^*}(\mathrm{d}x_{n+1})$$

where

$$w^{(n)} = \frac{1}{n} \int_0^{+\infty} u\, \tau_1(u | x_{n+1})\, f(u | X^{(n)})\, \mathrm{d}u$$

$$w_j^{(n)} = \int_0^{+\infty} u\, \frac{\tau_{n_j+1}(u | X_j^*)}{\tau_{n_j}(u | X_j^*)}\, f(u | X^{(n)})\, \mathrm{d}u$$

For the homogeneous case one obtains the predictive distributions of Pitman (2003).

Sampling from the marginal distribution of the X_is

Note that, conditionally on $U_n = u$, the predictive distribution is

$$\mathbb{P}[X_{n+1} \in \mathrm{d}x_{n+1} | X^{(n)}, U_n = u] \propto \kappa_1(u)\, \tau_1(u|x_{n+1})\, \alpha(\mathrm{d}x_{n+1}) + \sum_{j=1}^{k} \frac{\tau_{n_j+1}(u|X_j^*)}{\tau_{n_j}(u|X_j^*)}\, \delta_{X_j^*}(\mathrm{d}x_{n+1})$$

where $\kappa_1(u) = \int_{\mathbb{X}} \tau_1(u|x)\, \alpha(\mathrm{d}x)$.

Sampling from the marginal distribution of the X_is

Note that, conditionally on $U_n = u$, the predictive distribution is

$$\mathbb{P}[X_{n+1} \in \mathrm{d}x_{n+1} | X^{(n)}, U_n = u] \propto \kappa_1(u)\, \tau_1(u|x_{n+1})\, \alpha(\mathrm{d}x_{n+1}) + \sum_{j=1}^{k} \frac{\tau_{n_j+1}(u|X_j^*)}{\tau_{n_j}(u|X_j^*)}\, \delta_{X_j^*}(\mathrm{d}x_{n+1})$$

where $\kappa_1(u) = \int_{\mathbb{X}} \tau_1(u|x)\, \alpha(\mathrm{d}x)$.
From this one can implement an analogue of the Pólya urn scheme in order to draw a sample $X^{(n)}$ from $\tilde{P}$

Sampling from the marginal distribution of the X_is

Note that, conditionally on $U_n = u$, the predictive distribution is

$$\mathbb{P}[X_{n+1} \in \mathrm{d}x_{n+1} | X^{(n)}, U_n = u] \propto \kappa_1(u)\, \tau_1(u|x_{n+1})\, \alpha(\mathrm{d}x_{n+1}) + \sum_{j=1}^{k} \frac{\tau_{n_j+1}(u|X_j^*)}{\tau_{n_j}(u|X_j^*)}\, \delta_{X_j^*}(\mathrm{d}x_{n+1})$$

where $\kappa_1(u) = \int_{\mathbb{X}} \tau_1(u|x)\, \alpha(\mathrm{d}x)$.
From this one can implement an analogue of the Pólya urn scheme in order to draw a sample $X^{(n)}$ from $\tilde{P}$

- Let

$$m(\mathrm{d}x|u) \propto \tau_1(u|x)\, \alpha(\mathrm{d}x)$$

Sampling from the marginal distribution of the X_is

Note that, conditionally on $U_n = u$, the predictive distribution is

$$\mathbb{P}[X_{n+1} \in \mathrm{d}x_{n+1} | X^{(n)}, U_n = u] \propto \kappa_1(u)\, \tau_1(u|x_{n+1})\, \alpha(\mathrm{d}x_{n+1}) + \sum_{j=1}^{k} \frac{\tau_{n_j+1}(u|X_j^*)}{\tau_{n_j}(u|X_j^*)}\, \delta_{X_j^*}(\mathrm{d}x_{n+1})$$

where $\kappa_1(u) = \int_{\mathbb{X}} \tau_1(u|x)\, \alpha(\mathrm{d}x)$.
From this one can implement an analogue of the Pólya urn scheme in order to draw a sample $X^{(n)}$ from $\tilde{P}$

- Let
$$m(\mathrm{d}x|u) \propto \tau_1(u|x)\, \alpha(\mathrm{d}x)$$
- For any $i \geq 2$ set
$$m(\mathrm{d}x_i | x^{i-1}, u) = \mathbb{P}[X_i \in \mathrm{d}x_i | X^{i-1}, U_{i-1} = u]$$

Generalization of a Pólya urn scheme

1) Sample U_0 from $f_0(u)$

2) Sample X_1 from $m(\mathrm{d}x|U_0)$

Generalization of a Pólya urn scheme

1) Sample U_0 from $f_0(u)$

2) Sample X_1 from $m(\mathrm{d}x|U_0)$

3) At step i

- Sample U_{i-1} from $f(u|X^{(i-1)})$
- Generate ξ_i from $m(\mathrm{d}\xi|U_{i-1})$ and X_i from $m(\mathrm{d}x|X_1,\ldots,X_{i-1},U_{i-1})$

$$X_i = \begin{cases} \xi_i & \text{prob } \propto \kappa_1(U_{i-1}) \\ X^*_{j,i-1} & \text{prob } \propto \tau_{n_{j,i-1}+1}(U_{i-1}|X^*_{j,i-1})/\tau_{n_{j,i-1}+1}(U_{i-1}|X^*_{j,i-1}) \end{cases}$$

where $X^*_{j,i-1}$ is the j–th distinct value among $X_1,\ldots,X_{i-1}$ and

$$n_{j,i-1} = \text{card}\{X_s :\ X_s = X^*_{j,i-1},\ \ s = 1,\ldots,i-1\}$$

Sampling the posterior random measure

Recall that, given $U_n = u$ and $X^{(n)}$, $\tilde{\mu} \stackrel{d}{=} \tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)} \, \delta_{X_i^*}$

Sampling the posterior random measure

Recall that, given $U_n = u$ and $X^{(n)}$, $\tilde{\mu} \stackrel{d}{=} \tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)} \, \delta_{X_i^*}$
Algorithm:

Sampling the posterior random measure

Recall that, given $U_n = u$ and $X^{(n)}$, $\tilde{\mu} \stackrel{d}{=} \tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)} \, \delta_{X_i^*}$
Algorithm:

(1) Sample U_n from $f(u|X^{(i-1)})$

(2) Sample $J_i^{(U_n)}$ from the density

$$f_i(s)\,\mathrm{d}s \propto s^{n_i}\,\mathrm{e}^{-U_n s}\,\rho_{X_i^*}(\mathrm{d}s)$$

Sampling the posterior random measure

Recall that, given $U_n = u$ and $X^{(n)}$, $\tilde{\mu} \overset{d}{=} \tilde{\mu}_u + \sum_{i=1}^k J_i^{(u)} \delta_{X_i^*}$

Algorithm:

(1) Sample U_n from $f(u|X^{(i-1)})$

(2) Sample $J_i^{(U_n)}$ from the density

$$f_i(s)\,\mathrm{d}s \propto s^{n_i}\,\mathrm{e}^{-U_n s}\,\rho_{X_i^*}(\mathrm{d}s)$$

(3) Simulate a realization of the completely random measure $\tilde{\mu}^{(U_n)}$ with intensity measure

$$\nu^{(U_n)}(\mathrm{d}x, \mathrm{d}s) = \mathrm{e}^{-U_n s}\,\rho_x(\mathrm{d}s)\,\eta(\mathrm{d}x)$$

via the Ferguson and Klass algorithm.

The two parameter Poisson–Dirichlet process

The PD(σ, θ) can be represented (Pitman, '96) as species sampling model $\sum_{i=1}^{\infty} \tilde{p}_i \delta_{X_i}$ with stick breaking weights

$$\tilde{p}_i = V_i \prod_{j=1}^{i-1} (1 - V_j) \qquad V_i \overset{\text{ind}}{\sim} \text{beta}(\theta + i\sigma, 1 - \sigma), \; X_i \overset{\text{iid}}{\sim} H$$

The two parameter Poisson–Dirichlet process

The PD(σ, θ) can be represented (Pitman, '96) as species sampling model $\sum_{i=1}^{\infty} \tilde{p}_i \delta_{X_i}$ with stick breaking weights

$$\tilde{p}_i = V_i \prod_{j=1}^{i-1} (1 - V_j) \qquad V_i \stackrel{\text{ind}}{\sim} \text{beta}(\theta + i\sigma, 1 - \sigma), \; X_i \stackrel{\text{iid}}{\sim} H$$

Using this representation, in Pitman ('96), it is shown that

$$\tilde{P} \mid X^{(n)} \stackrel{d}{=} (1 - \sum_{i=1}^{k} p_i^*) \, \tilde{P}^{(k)} + \sum_{j=1}^{k} p_j^* \, \delta_{X_j^*}$$

where $\tilde{P}^{(k)} = \text{PD}(\sigma, \theta + k\sigma)$ and
$(p_1^*, \ldots, p_k^*) \sim \text{Dir}(n_1 - \sigma, \ldots, n_k - \sigma, \theta + k\sigma)$

The two parameter Poisson–Dirichlet process

The PD(σ,θ) can be represented (Pitman, '96) as species sampling model $\sum_{i=1}^{\infty} \tilde{p}_i \delta_{X_i}$ with stick breaking weights

$$\tilde{p}_i = V_i \prod_{j=1}^{i-1}(1 - V_j) \qquad V_i \overset{\text{ind}}{\sim} \text{beta}(\theta + i\sigma, 1-\sigma),\ X_i \overset{\text{iid}}{\sim} H$$

Using this representation, in Pitman ('96), it is shown that

$$\tilde{P} \mid X^{(n)} \overset{d}{=} (1 - \sum_{i=1}^{k} p_i^*)\, \tilde{P}^{(k)} + \sum_{j=1}^{k} p_j^*\, \delta_{X_j^*}$$

where $\tilde{P}^{(k)} = \text{PD}(\sigma, \theta + k\sigma)$ and
$(p_1^*, \ldots, p_k^*) \sim \text{Dir}(n_1 - \sigma, \ldots, n_k - \sigma, \theta + k\sigma)$
The PD(σ,θ) process is also representable as normalized measure

$$\tilde{P}(\,\cdot\,) = \frac{\tilde{\phi}(\,\cdot\,)}{\tilde{\phi}(\mathbb{X})},$$

but $\tilde{\phi}$ does not have independent increments (Pitman and Yor, '97). Indeed, the Laplace functional of $\tilde{\phi}$ is of the form

$$\mathbb{E}[\mathrm{e}^{-\int f(x)\tilde{\phi}(\mathrm{d}x)}] = \frac{1}{\Gamma(\theta)} \int_0^\infty u^{\theta-1}\, \mathrm{e}^{-\int_0^\infty (u+f(x))^\sigma P_0(\mathrm{d}x)} \mathrm{d}u$$

Identify a latent variable U_n such that $U_n|X^{(n)}$ has density

$$f(u|X^{(n)}) = \frac{\alpha}{\Gamma(k+\theta/\alpha)}\, u^{\theta+k\alpha-1}\, \mathrm{e}^{-u^{\alpha}}$$

Then, given U_n and $X^{(n)}$, the (posterior) distribution of $\tilde{\varphi}$ coincides with the distribution of the random measure

$$\tilde{\mu}_u + \sum_{i=1}^{k} J_i^{(u)}\, \delta_{X_i^*}$$

where $\tilde{\mu}_u$ is a GG–CRM with intensity

$$\nu^{(u)}(s) = \frac{\alpha}{\Gamma(1-\alpha)}\, s^{-1-\alpha}\, \mathrm{e}^{-u\,s} \tag{1}$$

The jumps $J_i^{(u)} \sim \mathrm{Gamma}(u, n_i - \alpha)$. Finally, the jumps $J_i^{(u)}$ $(i = 1, \ldots, k)$ and $\tilde{\mu}_u$ are, conditional on U_n, independent.

Hierarchical mixture models

$$
\begin{aligned}
Y_i \,|\, X_i &\overset{\text{ind}}{\sim} f(\,\cdot\,|X_i) \\
X_i | \tilde{P} &\overset{\text{iid}}{\sim} \tilde{P} \\
\tilde{P} &\sim \text{NRMI}
\end{aligned}
$$

Hierarchical mixture models

$$\begin{aligned} Y_i \,|\, X_i &\overset{\text{ind}}{\sim} f(\,\cdot\,|X_i) \\ X_i|\tilde{P} &\overset{\text{iid}}{\sim} \tilde{P} \\ \tilde{P} &\sim \text{NRMI} \end{aligned}$$

Equivalently, $Y^{(n)} = (Y_1, \dots, Y_n)$ are exchangeable draws from the random density

$$\tilde{f}(\,\cdot\,) = \int_{\mathbb{X}} f(\,\cdot\,|x)\,\tilde{P}(\mathrm{d}x).$$

The posterior distribution of the mixture model

The posterior distribution of the mixture model

The posterior density $\tilde{f}$, given the observations $Y^{(n)}$, is

$$\int_{\mathbb{X}} f(\,\cdot\,|x)\,\tilde{P}(\mathrm{d}x|\,Y^{(n)})$$

where $\tilde{P}(\mathrm{d}x|\,Y^{(n)})$ is the (posterior) random probability measure whose distribution is

$$\tilde{P}(\mathrm{d}x|\,Y^{(n)}) \stackrel{d}{=} \int \mathbb{P}(\mathrm{d}p|X^{(n)})\mathbb{P}(\mathrm{d}X^{(n)}|\,Y^{(n)}).$$

with

The posterior distribution of the mixture model

The posterior density $\tilde{f}$, given the observations $Y^{(n)}$, is

$$\int_{\mathbb{X}} f(\,\cdot\,|x)\,\tilde{P}(\mathrm{d}x|Y^{(n)})$$

where $\tilde{P}(\mathrm{d}x|Y^{(n)})$ is the (posterior) random probability measure whose distribution is

$$\tilde{P}(\mathrm{d}x|Y^{(n)}) \stackrel{d}{=} \int \mathbb{P}(\mathrm{d}p|X^{(n)})\mathbb{P}(\mathrm{d}X^{(n)}|Y^{(n)}).$$

with

- $\mathbb{P}(\mathrm{d}p|X^{(n)})$ is the (posterior) distribution of the NRMI $\tilde{P}$, given $X^{(n)}$

The posterior distribution of the mixture model

The posterior density $\tilde{f}$, given the observations $Y^{(n)}$, is

$$\int_{\mathbb{X}} f(\,\cdot\,|x)\, \tilde{P}(\mathrm{d}x|Y^{(n)})$$

where $\tilde{P}(\mathrm{d}x|Y^{(n)})$ is the (posterior) random probability measure whose distribution is

$$\tilde{P}(\mathrm{d}x|Y^{(n)}) \stackrel{d}{=} \int \mathbb{P}(\mathrm{d}p|X^{(n)})\mathbb{P}(\mathrm{d}X^{(n)}|Y^{(n)}).$$

with

- $\mathbb{P}(\mathrm{d}p|X^{(n)})$ is the (posterior) distribution of the NRMI $\tilde{P}$, given $X^{(n)}$
- $\mathbb{P}(\mathrm{d}X^{(n)}|Y^{(n)})$ is determined via Bayes' theorem as

$$\frac{\left\{\prod_{i=1}^n f(Y_i|X_i)\right\}\ m(\mathrm{d}X^{(n)})}{\int \left\{\prod_{i=1}^n f(Y_i|X_i)\right\}\ m(\mathrm{d}X^{(n)})}$$

where $m(\mathrm{d}X^{(n)})$ is the marginal distribution of the latent variables.

The posterior distribution of the mixture model

The posterior density $\tilde{f}$, given the observations $Y^{(n)}$, is

$$\int_{\mathbb{X}} f(\,\cdot\,|x)\; \tilde{P}(\mathrm{d}x|\, Y^{(n)})$$

where $\tilde{P}(\mathrm{d}x|\, Y^{(n)})$ is the (posterior) random probability measure whose distribution is

$$\tilde{P}(\mathrm{d}x|\, Y^{(n)}) \stackrel{d}{=} \int \mathbb{P}(\mathrm{d}p|X^{(n)})\mathbb{P}(\mathrm{d}X^{(n)}|\, Y^{(n)}).$$

with

- $\mathbb{P}(\mathrm{d}p|X^{(n)})$ is the (posterior) distribution of the NRMI $\tilde{P}$, given $X^{(n)}$
- $\mathbb{P}(\mathrm{d}X^{(n)}|\, Y^{(n)})$ is determined via Bayes' theorem as

$$\frac{\left\{\prod_{i=1}^{n} f(Y_i|X_i)\right\}\; m(\mathrm{d}X^{(n)})}{\int \left\{\prod_{i=1}^{n} f(Y_i|X_i)\right\}\; m(\mathrm{d}X^{(n)})}$$

where $m(\mathrm{d}X^{(n)})$ is the marginal distribution of the latent variables.

Remark: In any mixture model, the crucial point is the determination of a tractable expression for $\mathbb{P}(\mathrm{d}p|X^{(n)})$: once available, by following Escobar and West (1995) the derivation of a simulation algorithm is trivial.

Some concluding remarks

- Question 1: is it preferable to specify a GG prior as mixing measure (which includes Dirichlet process as special case) or stick with the Dirichlet process and enrich it with hyperpriors? What about parsimony in model specification?

Some concluding remarks

- Question 1: is it preferable to specify a GG prior as mixing measure (which includes Dirichlet process as special case) or stick with the Dirichlet process and enrich it with hyperpriors? What about parsimony in model specification?
- Question 2: Do we need applied statistical motivations for the introduction of new classes of priors? E.g. beta process (Hjort, '90) introduced for survival analysis, but turned out to be also the de Finetti measure of the Indian Buffet Process. Random probability measures are objects of interest in their own well beyond what we may think: e.g. the distribution of a mean functional of the two parameter PD process is relevant for the study of phylogenetic trees.

Some concluding remarks

- Mixture model is not the only use one can make of discrete nonparametric priors: if the data come from a discrete distribution, then it is reasonable the model the data with a discrete nonparameteric prior (see Ramses' talk). Simpler context and there you get a real feeling of the limitations of the Dirichlet process: prediction is not monotone in the number of observed species.