Exact finite mixture representations for species sampling processes

Ramsés H. Mena
IIMAS-UNAM, México

Nonparametric Bayesian Inference - Computational Issues
ICERM-Brown University

(joint work with Merkatas, C., Nicoleris, T., and Rodríguez, C.E)

January 14, 2026

Discrete RPMs G key for BNP

Discrete random probability measures G are key for BNP inference, as they serve as building blocks for:

- Flexible **random density** construction

$$f_G(x) = \int f(x \mid \theta)\, G(\mathrm{d}\theta).$$

- **Clustering** inherent to exchangeable seq. driven by G or f_G.
- **Classification** allocated via the mass of f_G on a suitable space.
- **Species sampling problems** and **predictive inference**
- Generalizations of **mixture models**
- Flexible **dependent stochastic structures.**

... etc.

A variety of ways to construct or represent G

- Via infinite dim. distributions with specified fdds of $\{G(A)\}_{A\in\mathcal{X}}$

- Normalized completely random measures $\boldsymbol{\mu}$

$$G(A) = \frac{\boldsymbol{\mu}(A)}{\boldsymbol{\mu}(\mathbb{X})}, \quad A \in \mathcal{X}$$

- Species sampling processes (SSP)

$$G(A) = \sum_{j=1}^{\infty} w_j \, \delta_{\theta_j}(A) + \left(1 - \sum_{j=1}^{\infty} w_j\right) G_0(A), \qquad (1)$$

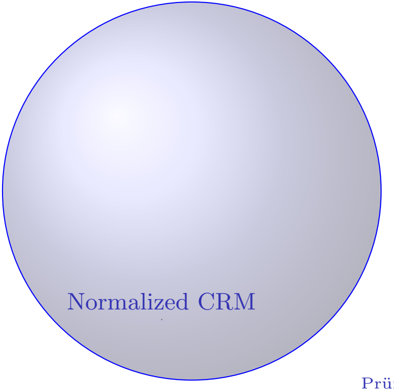
  where $\theta_j \overset{\text{iid}}{\sim} G_0$ and the weights $w_j \geq 0$ satisfy $\sum_j w_j \leq 1$ almost surely (a.s.), and $(\theta_j)_j$ is independent of $(w_j)_j$

- Some others, Pólya trees, NTR, etc.

Kingman (1967,1993), Ferguson (1973), Pitman (1996), Prünster (2002), Nieto-Barajas *et al.*(2004), Lijoi *et al.*(2005,2007), James *et al.* (2006,2009).

SSPs, normalized CRMs and Stick-breaking

Given a CRM $\boldsymbol{\mu}$ satisfying $0 < \boldsymbol{\mu}(\mathbb{X}) < \infty \Rightarrow G(\cdot) = \frac{\boldsymbol{\mu}(\cdot)}{\boldsymbol{\mu}(\mathbb{X})}$

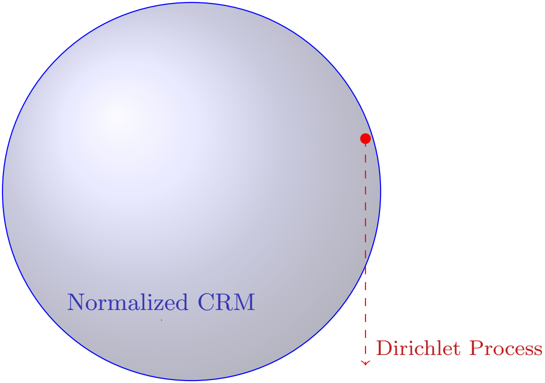


Prünster (2002), Regazzini *et. al.*(2003), James *et. al.* (2006)

SSPs, normalized CRMs and Stick-breaking

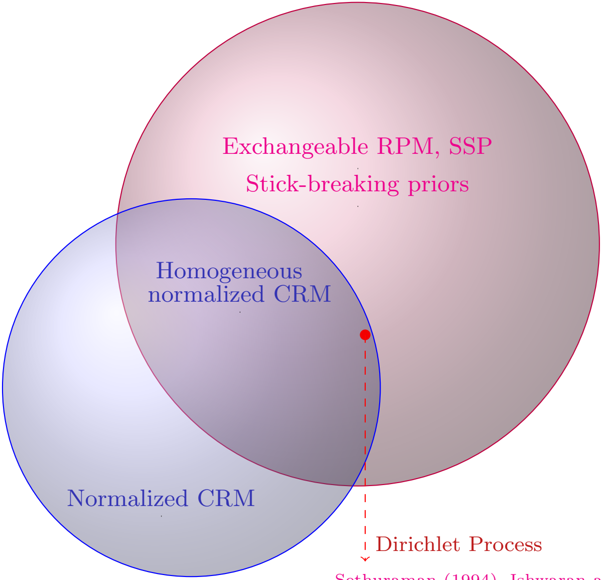
Given a CRM $\boldsymbol{\mu}$ satisfying $0 < \boldsymbol{\mu}(\mathbb{X}) < \infty \Rightarrow G(\cdot) = \frac{\boldsymbol{\mu}(\cdot)}{\boldsymbol{\mu}(\mathbb{X})}$

If $\nu(ds, dx) = s^{-1}e^{-s}\mathrm{d}s\alpha G_0(dx) \quad \Rightarrow \quad G \sim \mathcal{D}(\alpha G_0)$



Ferguson (1973)

SSPs, normalized CRMs and Stick-breaking



$$G = \sum_{i \geq 1} w_i \delta_{\theta_i}$$

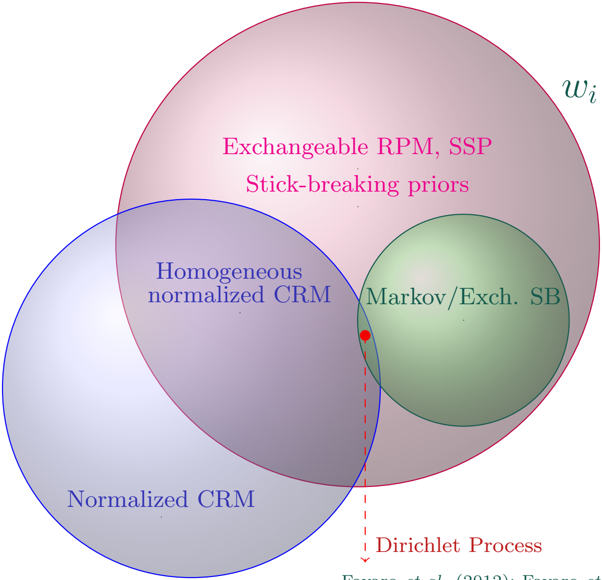
$$\sum_{i \geq 1} w_i = 1$$

$$\theta_i \overset{\text{iid}}{\sim} G_0$$

$$(w_i)_{i \geq 1} \perp (\theta_i)_{i \geq 1}$$

Sethuraman (1994), Ishwaran and James (2001), Kallenberg (2017)

SSPs, normalized CRMs and Stick-breaking



$$w_i = v_i \prod_{j=1}^{i-1}(1 - v_j)$$

Many choices of (v_i)

Favaro *et al.* (2012); Favaro *et al.* (2016); Gil-Leyva *et al.* (2020); Gil-Leyva and M. (2023); Gil-Leyva *et al.*(2026)

Computational bottleneck of SSP

We focus on the rich class of ***proper SSPs***

$$G(A) = \sum_{j=1}^{\infty} w_j \, \delta_{\theta_j}(A) \tag{2}$$

- Infinite series rep. are elegant but computationally awkward.
- Known issues:
  - ▷ fixed truncation ⇒ approximation error.
  - ▷ adaptive/slice schemes ⇒ random truncations, often SSP-specific.
  - ▷ for many SSPs: predictive/EPPF not available in closed form.

Goal

Goal: Recognize a ***generic***, ***distributionally exact*** finite representation for ***any proper SSP*** which allows

- Exact simulation of proper SSP.
- Explicit comparisons with truncation approximations and transparent control of computational cost.
- Prior and posterior machinery (allocations, Gibbs updates) without ad hoc truncations.

Exact finite mixture representation of SSP.

- Taking inspiration from the paper by Kalli *et. al.*(2011).

Theorem 1 (Exact finite representation of proper SSPs)

Fix a strictly decreasing seq. $(\xi_j)_{j\geq 1} \subset (0,1]$ with $\xi_j \downarrow 0$.

$$s_k := \sum_{h=1}^{k} \xi_h^{-1} w_h, \qquad \mathbb{P}(K = k \mid \boldsymbol{w}) := (\xi_k - \xi_{k+1})\, s_k, \qquad \text{for each } k \geq 1$$

Conditionally on $K = k$, define reweighted finite weights

$$\tilde{w}_j^{(k)} := \frac{\xi_j^{-1} w_j}{s_k}, \qquad j = 1, \ldots, k,$$

and the finite random measure

$$G^\star(\cdot \mid K = k, \boldsymbol{\theta}) := \sum_{j=1}^{k} \tilde{w}_j^{(k)}\, \delta_{\theta_j}(\cdot).$$

Then $\boldsymbol{G}^\star \overset{d}{=} \boldsymbol{G}$.

Key idea: Conditionally on $K = k$, say $G_k^\star$, involves only finitely many atoms.

A random choice of (ξ_j)

Corollary 1 (Stick-breaking SSPs)

If $w_j = v_j \prod_{\ell<j}(1-v_\ell)$, choose the *random* decreasing sequence

$$\xi_j := \prod_{\ell<j}(1-v_\ell) \quad \text{(remaining stick after } j-1 \text{ breaks)}.$$

Then $\xi_j^{-1} w_j = v_j$ and $\xi_k - \xi_{k+1} = \xi_k v_k = w_k$, so for $K = k$:

$$\tilde{w}_j^{(k)} = \frac{v_j}{\sum_{h=1}^k v_h}, \qquad \mathbb{P}(K = k \mid \boldsymbol{v}) = w_k \sum_{h=1}^k v_h.$$

Includes DP and Dirichlet stick-breaking variants.

Corollary 2 (Geometric stick-breaking)

If $v_j \equiv v$ and $w_j = v(1-v)^{j-1}$, take $\xi_j = (1-v)^{j-1}$. Then $\tilde{w}_j^{(k)} = 1/k$ &

$$\mathbb{P}(K = k \mid v) = k\, v^2 (1-v)^{k-1}, \qquad k = 1, 2, \ldots$$

TV bounds

▷ For a decreasing sequence (ξ_j) (random or deterministic) define the corresponding ξ–reweighted measure $G^\star_k$.

▷ For a fixed realization $(w_j, \theta_j)_{j\geq 1}$, define the ***tail mass*** $R_k := \sum_{j>k} w_j$ and the renormalized truncation without tail

$$G^{(k)}_{\text{ren}} := \sum_{j=1}^{k} \bar{w}_j \, \delta_{\theta_j}, \qquad \bar{w}_j := \frac{w_j}{1 - R_k}.$$

Coupling-based bounds

$$d_{\text{TV}}(G, G^\star_k) \;\leq\; R_k + D_k, \qquad d_{\text{TV}}\big(G^{(k)}_{\text{ren}}, G^\star_k\big) \;\leq\; D_k,$$

where $D_k = \frac{M_k - 1}{M_k + 1}$, with $M_k = \frac{\xi_1}{\xi_k}$.

- Exponential $\xi_j = e^{-\eta j}$: $\quad D_k = \tanh\big(\eta(k-1)/2\big)$.
- Random $\xi_k = R_{k-1}$: $\quad D_k = (1 - R_{k-1})/(1 + R_{k-1})$.

Averaging over K

Expected-TV

$$\mathbb{E}_{K|\boldsymbol{w},\boldsymbol{\xi}}\big[d_{\mathrm{TV}}(G, G_K^\star)\big] \;\leq\; \mathbb{E}_{K|\boldsymbol{w},\boldsymbol{\xi}}[R_K] \;+\; \mathbb{E}_{K|\boldsymbol{w},\boldsymbol{\xi}}[D_K].$$

For deterministic (ξ_j),

$$\mathbb{P}(K=k) = (\xi_k - \xi_{k+1}) \sum_{h=1}^{k} \frac{\mathbb{E}[w_h]}{\xi_h} \quad \& \quad \mathbb{E}_{K|\boldsymbol{w},\boldsymbol{\xi}}[R_K] = \sum_{1\leq h<j} \left(1 - \frac{\xi_j}{\xi_h}\right) \mathbb{E}[w_h w_j].$$

so

$$\boxed{\mathbb{E}_{K|\boldsymbol{w},\boldsymbol{\xi}}\left[d_{\mathrm{TV}}(G, G_K^\star)\right] \leq \sum_{h<j} \left(1 - \frac{\xi_j}{\xi_h}\right) \mathbb{E}[w_h w_j] + \sum_{k\geq 1} \frac{M_k - 1}{M_k + 1} \mathbb{P}(K=k).}$$

DP_α, Geo, and $\mathrm{PY}(\sigma, \alpha)$ cases and $\xi_k = e^{-\eta k}$

Take $\xi_k = q^k = e^{-\eta k}$ with $q = e^{-\eta} \in (0,1)$. Then

$$D_k = \frac{1 - q^{k-1}}{1 + q^{k-1}} = \tanh\Big(\frac{\eta(k-1)}{2}\Big).$$

Define

$$a_k := \mathbb{E}\Big[\prod_{j=1}^{k}(1 - V_j)\Big], \qquad \mu := \mathbb{E}[V_1], \qquad a_1 = \mathbb{E}[1 - V_1] = 1 - \mu.$$

Then (for $q \neq a_1$) the marginal pmf of K can be written as

$$\boxed{\mathbb{P}(K = k) = \frac{(1-q)\,\mu}{q - a_1}\,\big(q^k - a_k\big), \qquad k \geq 1.}$$

Hence

$$\boxed{\mathbb{E}\big[d_{\mathrm{TV}}(G, G_K^\star)\big] \;\leq\; \mathbb{E}[R_K] \;+\; \sum_{k=1}^{\infty} \tanh\Big(\frac{\eta(k-1)}{2}\Big)\,\frac{(1 - e^{-\eta})\,\mu}{e^{-\eta} - a_1}\,\big(e^{-\eta k} - a_k\big).}$$

DP_α, Geo, and $\mathrm{PY}(\sigma, \alpha)$ particularities

(i) $\mathrm{DP}(\alpha, G_0)$: $V_j \overset{\text{iid}}{\sim} \mathrm{Beta}(1, \alpha)$, $\mu = \frac{1}{\alpha+1}$, $a_1 = \frac{\alpha}{\alpha+1} =: a$, $a_k = a^k$.
When $q \to a$

$$\boxed{\mathbb{P}(K = k) = \frac{k\, a^{k-1}}{(\alpha+1)^2}.}$$

(ii) Geo: $w_k = V(1-V)^{k-1}$ (deterministic or random V).

$$\mu = \mathbb{E}[V], \quad a_1 = \mathbb{E}[1-V] = 1 - \mu, \quad a_k = \mathbb{E}[(1-V)^k].$$

If $q \to 1 - V$, $\boxed{K - 1 \sim \mathrm{NegBin}(2, V)}$.

(iii) $\mathrm{PD}(\sigma, \alpha)$: $V_j \overset{\text{ind}}{\sim} \mathrm{Beta}(1-\sigma, \alpha + j\sigma)$, $\mu = \frac{1-\sigma}{\alpha+1}$, $a_1 = \frac{\alpha+\sigma}{\alpha+1}$,

$$a_k = \frac{\left(\frac{\alpha}{\sigma} + 1\right)_k}{\left(\frac{\alpha+1}{\sigma}\right)_k}.$$

Here $a_k \neq a_1^k$, with singular point $q = a_1$, unless $\sigma = 0$. Otherwise

$$\boxed{\mathbb{P}(K = k) = \frac{(1-q)(1-\sigma)}{(\alpha+1)\left(q - \frac{\alpha+\sigma}{\alpha+1}\right)} \left(q^k - a_k\right), \quad \text{for } q \neq a_1 \quad \& \quad k \geq 1}$$

Calibration of η

- *Default:* choose $q = e^{-\eta}$ close the residual $a_1 = \mathbb{E}[1 - V_1]$:

  $$\eta \approx -\log a_1,$$

  which reduces to $\eta^\star = \log(1 + 1/\alpha)$ for DP and $\eta^\star = -\log(1 - V)$ for geometric weights.

- *Tuning:* choose η to target a desired truncation level (e.g. $\mathbb{E}[K]$ or median(K)):
  - DP and deterministic geometric admit closed forms.
  - PY requires solving $\mathbb{E}_{\sigma,\alpha,\eta}[K] = \kappa$ numerically.

Comparison with a.s. ε-error correction

Take Arbel, De Blasi and Prünster (2019) a.s. error control

$$G_\varepsilon = \sum_{j=1}^{\tau(\varepsilon)} w_j \, \delta_{\theta_j} \; + \; R_{\tau(\varepsilon)} \, \delta_{\theta_0}, \qquad \theta_0 \sim G_0. \tag{3}$$

where $\tau(\varepsilon) := \min\{n \geq 1 : R_n < \varepsilon\}$, which satisfies

$$d_{\mathrm{TV}}(G, G_\varepsilon) \leq R_{\tau(\varepsilon)} < \varepsilon \qquad \text{a.s.}$$

where G is the SSP corresponding to PY process.

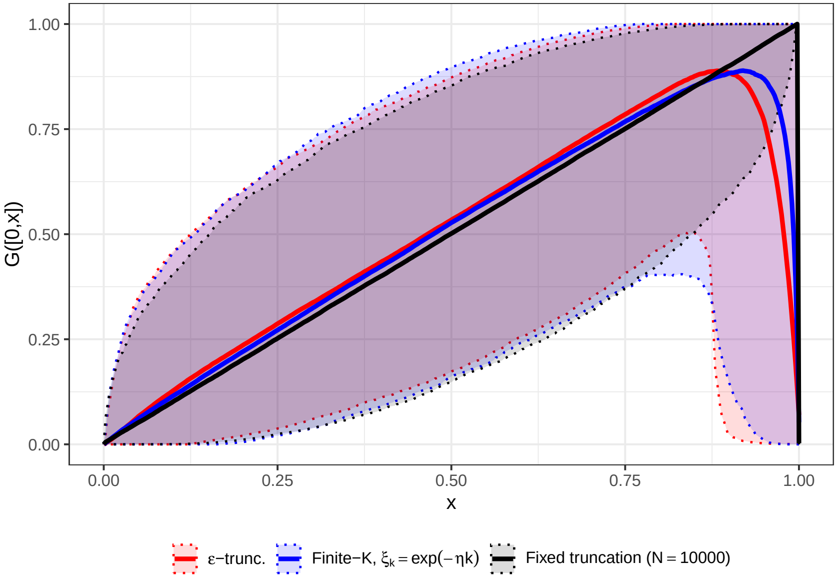
Comparison to an ε–truncation G_ε (same coupling)

If $d_{\mathrm{TV}}(G, G_\varepsilon) < \varepsilon$ a.s., then

$$d_{\mathrm{TV}}(G_K^\star, G_\varepsilon) \;\leq\; \varepsilon + R_K + D_K \quad \text{a.s.},$$

$$\Rightarrow \quad \mathbb{E}[d_{\mathrm{TV}}(G_K^\star, G_\varepsilon)] \leq \varepsilon + \mathbb{E}[R_K] + \mathbb{E}[D_K].$$

DP simulation under four scenarios, $G_0 = \mathsf{U}(0,1)$.



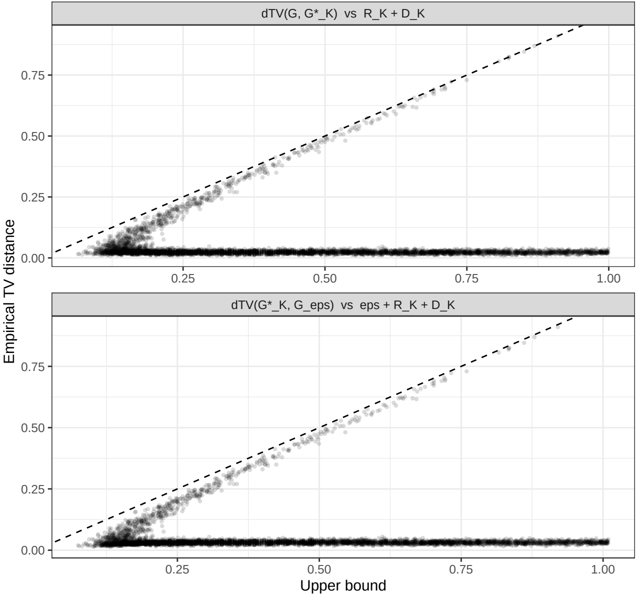


Figure: DP $\alpha = 6$, $G_0 = \text{Unif}(0, 1)$, $\varepsilon = 0.01$, $\eta = 0.01$.

Asymptotics of K as function of $\{\xi_k\}$

DP, $\xi_k = e^{-\eta k}$ $(G \sim \mathrm{DP}(\alpha, G_0))$

Let $q = e^{-\eta}$ and $a = \alpha/(\alpha+1)$. For $q \neq a$,

$$\mathbb{P}(K = k) = \frac{1-q}{(\alpha+1)(q-a)}(q^k - a^k), \qquad \eta^\star = \log\left(1 + \frac{1}{\alpha}\right) \ (q = a).$$

Scaling:

$$\eta \downarrow 0: \ \mathbb{E}[K] \sim \frac{1}{\eta}, \qquad \eta = \eta^\star: \ \mathbb{E}[K] = 2\alpha + 1, \qquad \eta \uparrow \infty: \ \mathbb{E}[K] \to \alpha + 1.$$

Geometric weights, natural $\xi_k = R_{k-1}$

If $w_k = v(1-v)^{k-1}$, then $K - 1 \sim \mathrm{NegBin}(2, v)$ and $\tilde{w}_j \equiv 1/k$ for $j \leq k$.

Pitman–Yor, exponential $\xi_k = q^k$ $(G \sim \mathrm{PY}(\sigma, \alpha, G_0))$

$\mathbb{E}[w_k] \sim C_{\sigma,\alpha} k^{-1/\sigma} \ \Rightarrow \ \mathbb{P}(K = k) \sim C_{\sigma,\alpha} k^{-1/\sigma}$, so $\mathbb{E}[K] < \infty \iff \sigma < 1/2$.

SSP mixture model and latent finite augmentation

- Mixture model with SSP mixing measure:

$$x_i \mid G \overset{\text{iid}}{\sim} f_G, \qquad f_G(x) = \sum_{j \geq 1} w_j f(x \mid \theta_j).$$

- Latent variables (per observation): $(z_i, k_i) \in \mathbb{N} \times \mathbb{N}$.

▷ (Allocation/component-label, Truncation/available components for z_i)

Hierarchical model

$$\boldsymbol{w} \sim p(\boldsymbol{w}), \qquad \theta_j \overset{\text{iid}}{\sim} G_0,\ j \geq 1,$$

$$\mathbb{P}(k_i = k \mid \boldsymbol{w}) = (\xi_k - \xi_{k+1})\, s_k, \quad s_k = \sum_{h=1}^{k} w_h/\xi_h,$$

$$\mathbb{P}(z_i = j \mid k_i, \boldsymbol{w}) \propto \xi_j^{-1} w_j, \qquad j = 1, \ldots, k_i,$$

$$x_i \mid z_i, \boldsymbol{\theta} \sim f(\cdot \mid \theta_{z_i}).$$

▷ Conditional on (k_i) the likelihood ***only involves finitely many components***, yet the model matches the original SSP mixture.

Mean-variance mixtures, 4 components.

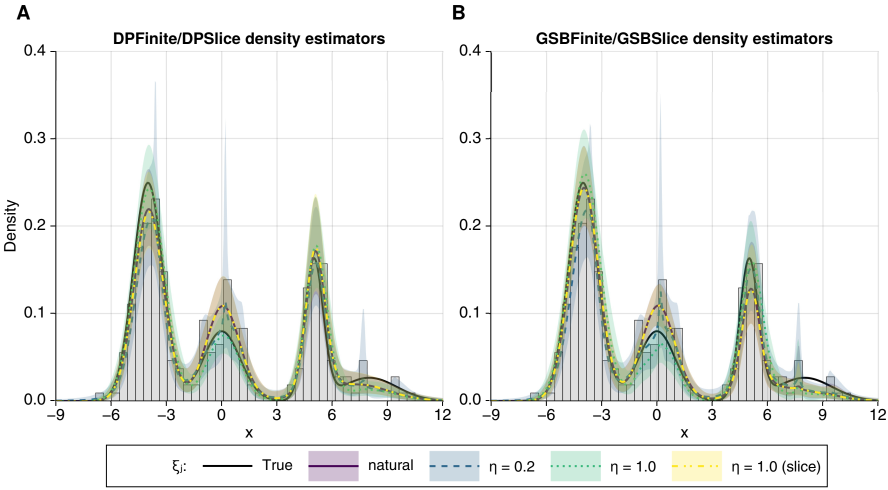


Figure: Panel A: DPFinite and DPSlice models, Panel B: GSBFinite and GSBSlice models.

Galaxy data.

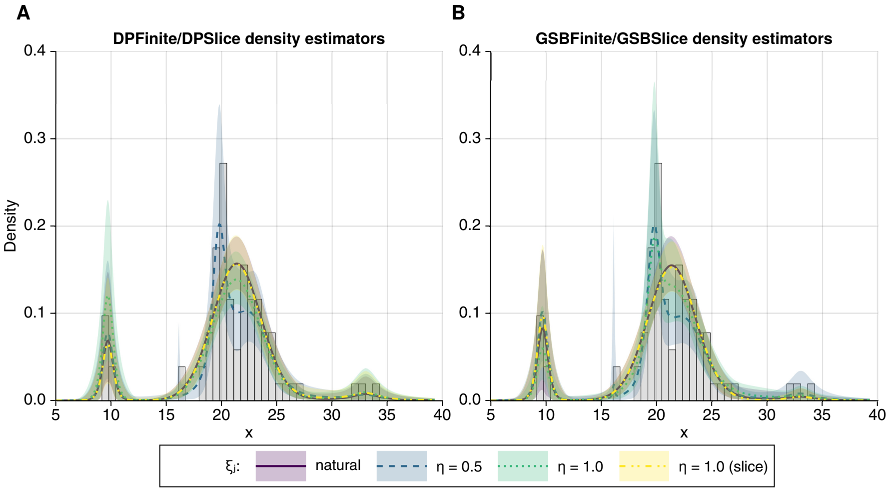


Figure: Panel A: DPFinite and DPSlice models, Panel B: GSBFinite and GSBSlice models.

Conclusions

- Any proper SSP admits an ***exact*** finite-mixture representation via a latent truncation variable and atom reweighting.

- Some payoffs:
  - ▷ **Exact simulation** of arbitrary SSP priors (no truncation),
  - ▷ **Posterior computation** for SSP mixture models via standard finite-mixture machinery (allocations + Gibbs).

- Total variation bounds help compare finite constructions and understand the role of (ξ_j).

- It opens the question of principled choices of sequences ξ_j for targeted mixing/tail exploration.

Note

$$K \text{ (representation)} \neq c_n \text{ (occupancy)} \neq m \text{ (finite model dimension)}.$$

References. . .

Arbel, J., De Blasi, P. and Prünster, I. (2019). Stochastic Approximations to the Pitman–Yor Process. *Bayesian Analysis* **14**(4), 1201–1219.

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

Ishwaran, H. and James, L.F. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Stat. Assoc.*, **96**, 161–173.

James, L.F., Lijoi A., and Prünster (2006). Conjugacy as a distinctive feature of the dirichlet process. *Scand. J. Stat.* **33**, 105–120.

James, L.F., Lijoi A., and Prünster (2009). Posterior analysis for normalized random measures with independent increments. *Scand. J. Stat.* **36**, 76–97.

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

Kingman, J.F.C. (1993). *Poisson processes.* Oxford University Press.

Lijoi, A., Mena, R.H. and Prünster, I. (2005). Bayesian nonparametric estimation of the probability of discovering new species. *J. Amer. Stat. Assoc.*, **100**, 1278–1291.

Lijoi, A., Mena, R.H. and Prünster, I. (2007). Controlling the reinforcement in Bayesian nonparametric mixture models. *J. R. Statist. Soc. B*, **69**, 715–740.

Kalli, M., Griffin, J.E. and Walker, S.G. (2011). Slice sampling mixture models. *Statistics and Computing* **21**(1), 93–105.

Pitman, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In *Statistics, Probability and Game Theory. Papers in honor of David Blackwell* (Eds. Ferguson, T.S., *et al.*). Lecture Notes, Monograph Series, **30**, 245–267. Institute of Mathematical Statistics, Hayward.

Prünster, I. (2002). Random probability measures derived from increasing additive processes and their application to Bayesian statistics. PhD thesis, Università degli Studi di Pavia, Pavia, Italy.

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

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

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statist. Sinica* **4**, 639–650.