Complexity Bounds for Dirichlet Process Slice Samplers

Beatrice Franzolini

Joint work with:

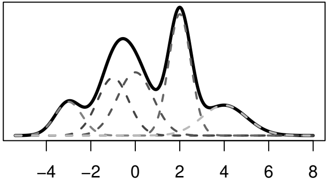
Francesco Gaffi

DP-based models

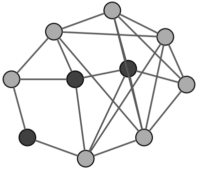
$$Y \mid \eta, \boldsymbol{z}_n \sim \underbrace{\mathcal{L}(\boldsymbol{z}_n, \eta)}_{\text{likelihood depends on } \boldsymbol{z}_n = (z_1,\ldots,z_n)} \qquad z_i \mid G \overset{iid}{\sim} \underbrace{G(\mathrm{d}x) \overset{a.s.}{=} \sum_{k=1}^{\infty} \pi_k \, \delta_{\phi_k}(\mathrm{d}x)}_{\text{discrete random measure}} \qquad \underbrace{G \sim \mathrm{DP}(\alpha, P_0)}_{\text{prior}}$$

$\Longrightarrow$ ρ_n partition of $[n] = \{1, \ldots, n\}$ defined by $i \sim j$ if and only if $z_i = z_j$

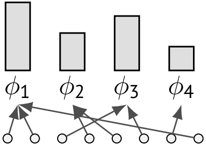
Mixture models



Stochastic block models



Species sampling models



MCMC sampling in DP-based models

$$Y \mid \eta, \boldsymbol{z}_n \sim \underbrace{\mathcal{L}(\boldsymbol{z}_n, \eta)}_{\text{likelihood depends on } \boldsymbol{z}_n=(z_1,\ldots,z_n)}$$

$$z_i \mid G \overset{iid}{\sim} \underbrace{G(\mathrm{d}z) \overset{a.s.}{=} \sum_{k=1}^{\infty} \pi_k \, \delta_{\phi_k}(\mathrm{d}z)}_{\text{discrete random measure}}$$

$$\underbrace{G \sim \mathrm{DP}(\alpha, P_0)}_{\text{prior}}$$

Marginal algorithms

(Neal, 2000)

$$\mathrm{pr}(z_i = z \mid \text{rest}) \propto \begin{cases} n_h \left(\boldsymbol{z}_n^{-(i)}\right) \times \mathcal{L}(\boldsymbol{z}_n, \eta) & \text{if } z \in \boldsymbol{z}_n^{-(i)} \\ \alpha \times \int \mathcal{L}(\boldsymbol{z}_n, \eta) P_0(\mathrm{d}z) & \text{if } z = \text{new} \end{cases}$$

- ✓ Target the posterior
- ✗ Force one-at-a-time updates
- ✗ Bookkeeping

Blocked Gibbs samplers

(Ishwaran and James, 2001)

$$\mathrm{pr}(z_i = z \mid \text{rest}) \propto \pi_{k(z)} \times \mathcal{L}(\boldsymbol{z}_n, \eta)$$

- ✗ Target an approximation of the posterior
- ✓ Allow joint updates
- ✓ Easy/No bookkeeping

Non-vanishing truncation bias of blocked Gibbs samplers

Exponential accuracy of truncated blocked Gibbs samplers (Ishwaran and James, 2002)

Let L be the truncation level, i.e., G is approximated by $\sum_{\ell=1}^{L} \pi_\ell \delta_{\phi_\ell}$ and ρ_n the partition induced by ties in z_i in a Normal mixture model

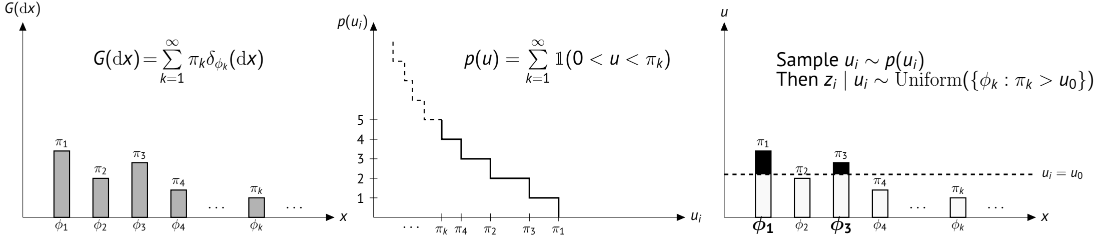
$$\int_{\mathbb{R}^n} \sum_{\rho_n} \Big| \underbrace{p_L(\rho_n \mid \boldsymbol{y}_{1:n})}_{\text{target of the algorithm}} - \underbrace{p_\infty(\rho_n \mid \boldsymbol{y}_{1:n})}_{\text{posterior of the partition}} \Big| \underbrace{m_\infty(\boldsymbol{y}_{1:n})}_{\text{prior predictive}} \, \mathrm{d}\boldsymbol{y}_{1:n} = O(4\, n\, e^{-(L-1)/\alpha}).$$

Extensions in Ishwaran and Zarepour (2002); Campbell et al. (2019); Li and Campbell (2021).

Two main limitations of the result:
✗ It holds **under the generative model**
✗ Trivially: unless $L = n$, $p_L(\rho_n \mid \boldsymbol{y}_{1:n}) = 0$ on all partitions with more than L clusters.

Slice sampling (Neal, 2003; Walker, 2007)

$$z_i \sim G(\mathrm{d}z) \qquad \rightarrow \qquad (u_i, z_i) \sim \mathbb{1}(0 < u < G(\mathrm{d}z))$$



- ✓ Target the posterior
- ✓ Allow joint updates
- ✓ Easy/No bookkeeping

Slice sampling (Neal, 2003; Walker, 2007; Ge et al., 2015)

Posterior slice sampling in DP-based models

1. Sample G, i.e., $(\pi_1, \pi_2, \ldots)$ and $(\phi_1, \phi_2, \ldots)$
2. Sample u_i from $p(u_i \mid z_i = \phi_k) = \mathrm{Uniform}(0, \pi_k) \quad \forall i$
3. Sample z_i from $p(z_i \mid u_i = u_0) = \mathrm{Discrete}\big(\phi_k : \pi_k > u_i,$ with $w_k \propto \mathrm{Lik}(Y; z_i = \phi_k)\big) \quad \forall i$

→ Step 2. requires weights for all currently allocated components.

$$(\underbrace{\pi_1, \ldots, \pi_H}_{H=\text{number of clusters}}, \pi^\star) \sim \mathrm{Dirichlet}(n_1, \ldots, n_H, \alpha) \qquad \text{(Ge et al., 2015)}$$

→ Step 3. requires determination of the slice set $A_i = \{\phi_k : \pi_k > u_i\}$, let $u_{\min} = \min_i u_i$

$$\forall i, \; A_i \subseteq \left\{ \phi_1, \ldots, \phi_K : \sum_{h=K+1}^{\infty} \pi_h < u_{\min} \right\} \qquad \text{(Walker, 2007)}$$

Dynamic truncation level K

$$K = \min \left\{ k \geq H : \sum_{h=k+1}^{\infty} \pi_h < u_{\min} \right\}$$

✓ At each iteration, K components are enough and ***K* is finite**.

✗ Time complexity is driven by at most $K \times n$ likelihood evaluations.

About K

$$K = \min \left\{ k \geq H : \sum_{h=k+1}^{\infty} \pi_h < u_{\min} \right\}$$

✗ If $u_{\min} \approx 0$, **K can be arbitrarily large.**

any finite threshold

$$\mathbb{P}_\pi \left(\exists s \leq T : C_s > C \right) \xrightarrow[T \to \infty]{} 1$$

stationary distribution

computational cost of iteration s
(in some unit of measure)

No deterministic bound on per-iteration time exists.
To assess actual scalability, we need to quantify the "probability of exceeding a threshold" and how this varies with n.

About K (from now K_n)

A related quantity: $$K(x) = \min \left\{ k \geq 1 : \sum_{h=k+1}^{\infty} \underbrace{\pi_h}_{\text{from the prior}} < x \right\} \qquad x \in (0,1)$$

From Muliere and Tardella (1998)

$$K \mid x \sim 1 + \text{Poisson}(-\alpha \log x)$$

But at **each MCMC iteration** (in the improved slice sampler of Ge et al., 2015)

$$K_n = \min \left\{ k \geq H_n : \sum_{h=k+1}^{\infty} \pi_h < u_{\min} \right\}$$

where:

H_n is the number of clusters of the partition configuration

π_h's depend on the partition configuration

$u_{\min}$ depends on the $(\pi_h)_{h \geq 1}$ and the partition configuration

ρ_n, the partition configuration visited by the chain depends on the specific data

Survival probability of $u_{\min}$

Proposition: Merging two clusters increases the survival probability of $u_{\min}$

If $\rho_n^{(r\oplus s)}$ is the partition obtained from ρ_n by merging two (distinct) clusters r and s into a single cluster and leaving the other clusters unchanged, then, for any $x \in (0,1)$

$$\Pr\left(u_{\min} > x \,\middle|\, \rho_n^{(r\oplus s)}\right) \;\geq\; \Pr\left(u_{\min} > x \mid \rho_n\right).$$

Corollary: Singleton partition yields the lowest survival probability of $u_{\min}$

Let ρ_n^{sing} denote the singleton partition of $[n]$, *i.e.*, the partition with n clusters of size 1. Then, for every $x \in (0,1)$ and every partition ρ_n,

$$\Pr\left(u_{\min} > x \mid \rho_n\right) \;\geq\; \Pr\left(u_{\min} > x \,\middle|\, \rho_n^{\mathrm{sing}}\right)$$

Main result

Theorem: High-probability bound on dynamic truncation level (Franzolini and Gaffi, 2026)

Let K_n and H_n be respectively the truncation level and the number of occupied clusters at a given iteration when run for n input data.

Then, for every $\delta \in (0, 1)$ and any $n \geq 2$

$$\Pr\big(K_n - H_n \leq C_\delta \log n \mid \rho_n\big) \;\geq\; 1 - \delta\,, \qquad \forall \rho_n$$

with $C_\delta = B_\alpha^{(1)} + B_\alpha^{(2)} \log(1/\delta)$, and $B_\alpha^{(1)}$, $B_\alpha^{(2)} \asymp \alpha$.

In particular, as $n \to \infty$, $K_n - H_n = O_{\mathbb{P}}(\log n)$ and, in the worst-case scenario of $H_n = O(n)$, $K_n = O_{\mathbb{P}}(n)$.

"In the worst case, the slice sampler requires $O(n)$ active components with high probability."

Proof sketch

$$K_n = \min\left\{k \geq H_n : \sum_{h=k+1}^{\infty} \pi_h < u_{\text{min}}\right\} = \min\left\{k \geq H_n : \pi^\star \times \prod_{i=H_n+1}^{k} (1 - V_i) < u_{\text{min}}\right\}.$$

Proof sketch

$$K_n = \min\left\{k \geq H_n : \sum_{h=k+1}^{\infty} \pi_h < u_{\min}\right\} = \min\left\{k \geq H_n : \pi^\star \times \prod_{i=H_n+1}^{k} (1 - V_i) < u_{\min}\right\}.$$

- Define $K_n^0(x) = \min\left\{k > H_n : \prod_{i=H_n+1}^{k} (1 - V_i) < x\right\}$

Proof sketch

$$K_n = \min\left\{k \geq H_n : \sum_{h=k+1}^{\infty} \pi_h < u_{\min}\right\} = \min\left\{k \geq H_n : \pi^\star \times \prod_{i=H_n+1}^{k} (1 - V_i) < u_{\min}\right\}.$$

- Define $K_n^0(x) = \min\left\{k > H_n : \prod_{i=H_n+1}^{k} (1 - V_i) < x\right\}$
- $\Pr(K_n - H_n > C_\delta \log n \mid \rho_n) \leq \boxed{\Pr\left(K_n^0(x) > H_n + C_\delta \log n \mid \rho_n\right)} + \Pr(K_n > K_n^0(x) \mid \rho_n)$

Proof sketch

$$K_n = \min\left\{k \geq H_n : \sum_{h=k+1}^{\infty} \pi_h < u_{\text{min}}\right\} = \min\left\{k \geq H_n : \pi^\star \times \prod_{i=H_n+1}^{k} (1 - V_i) < u_{\text{min}}\right\}.$$

- Define $K_n^0(x) = \min\left\{k > H_n : \prod_{i=H_n+1}^{k} (1 - V_i) < x\right\}$
- $\Pr(K_n - H_n > C_\delta \log n \mid \rho_n) \leq \boxed{\Pr\left(K_n^0(x) > H_n + C_\delta \log n \mid \rho_n\right)} + \Pr(K_n > K_n^0(x) \mid \rho_n)$
- Show that $\boxed{\Pr\left(K_n^0(x) \geq 1 + H_n + 3\alpha \log(1/x) + \log(2/\delta) \mid \rho_n\right) \leq \frac{\delta}{2}}$ (1)

Proof sketch

$$K_n = \min\left\{k \geq H_n : \sum_{h=k+1}^{\infty} \pi_h < u_{\min}\right\} = \min\left\{k \geq H_n : \pi^\star \times \prod_{i=H_n+1}^{k} (1 - V_i) < u_{\min}\right\}.$$

- Define $K_n^0(x) = \min\left\{k > H_n : \prod_{i=H_n+1}^{k} (1 - V_i) < x\right\}$
- $\Pr(K_n - H_n > C_\delta \log n \mid \rho_n) \leq \boxed{\Pr\left(K_n^0(x) > H_n + C_\delta \log n \mid \rho_n\right)} + \boxed{\Pr(K_n > K_n^0(x) \mid \rho_n)}$
- Show that $\boxed{\Pr\left(K_n^0(x) \geq 1 + H_n + 3\alpha \log(1/x) + \log(2/\delta) \mid \rho_n\right) \leq \frac{\delta}{2}}$ (1)
- Show that $\boxed{\Pr(K_n > K_n^0(x) \mid \rho_n) \leq \Pr(u_{\min} \leq x \mid \rho_n) \leq n\,(\alpha + n - 1)\,x\,[1 + \log(1/x)]}$ (2)

Proof sketch

$$K_n = \min\left\{k \geq H_n : \sum_{h=k+1}^{\infty} \pi_h < u_{\min}\right\} = \min\left\{k \geq H_n : \pi^\star \times \prod_{i=H_n+1}^{k} (1-V_i) < u_{\min}\right\}.$$

- Define $K_n^0(x) = \min\left\{k > H_n : \prod_{i=H_n+1}^{k} (1-V_i) < x\right\}$
- $\Pr\left(K_n - H_n > C_\delta \log n \mid \rho_n\right) \leq \boxed{\Pr\left(K_n^0(x) > H_n + C_\delta \log n \mid \rho_n\right)} + \Pr(K_n > K_n^0(x) \mid \rho_n)$
- Show that $\boxed{\Pr\left(K_n^0(x) \geq 1 + H_n + 3\alpha \log(1/x) + \log(2/\delta) \mid \rho_n\right) \leq \frac{\delta}{2}}$ (1)
- Show that $\Pr(K_n > K_n^0(x) \mid \rho_n) \leq \Pr(u_{\min} \leq x \mid \rho_n) \leq n\,(\alpha + n - 1)\,x\,[1 + \log(1/x)]$ (2)
- Choose $x = x(n, \delta) \in (0, 1)$ such that, for any $n \geq 2$, the lower-bound in the probability in (1) is upper-bounded by $H_n + C_\delta \log n$, and the r.h.s. of (2) is upper-bounded by $\delta/2$.

A few additional results

Corollary: exponential tails of the slice overhead

Let (K_n, H_n) be defined as in the main Theorem.
There exist constants $B_\alpha^{(1)}, B_\alpha^{(2)} > 0$ such that for all $n \geq 2$ and all $t \geq 0$,

$$\Pr\left(\frac{K_n - H_n}{\log n} > B_\alpha^{(1)} + B_\alpha^{(2)}\, t \,\middle|\, \rho_n \right) \leq e^{-t}\,, \qquad \forall \rho_n.$$

Corollary: Almost-sure control of super-logarithmic slice overheads with infinite data coupling

Let (K_n, H_n) be defined as in the main Theorem.
There exists a constant $D_\alpha > 0$ such that

$$\Pr\left(\limsup_{n\to\infty} \left\{ K_n - H_n > D_\alpha \log \frac{1}{\delta_n} \log n \right\} \right) = 0$$

for any summable sequence $(\delta_n)_{n\geq 1} \subset (0, 1/2)$.

A few additional results

Proposition: tightness of the logarithmic slice-overhead bound

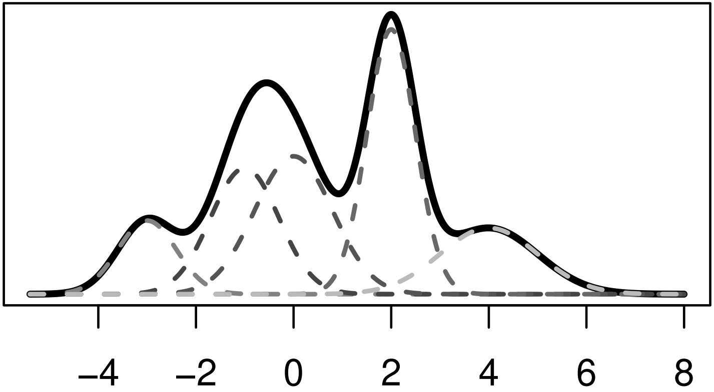
Let (K_n, H_n) be defined as in the main Theorem.
There exist constants $c_\alpha, \eta_\alpha > 0$ such that for all sufficiently large n,

$$\Pr\left(K_n - H_n \geq c_\alpha \log n \,\middle|\, \rho_n^{\mathrm{sing}}\right) \geq \eta_\alpha.$$

Consequently, for any deterministic sequence $r_n = o(\log n)$ and any $M > 0$,

$$\liminf_{n\to\infty} \sup_{\rho_n} \Pr\left(K_n - H_n > M r_n \mid \rho_n\right) > 0.$$

Practical consequences for mixture models



MCMC alternatives comparison

$$\text{for } i \in \{1, \ldots, n\} \qquad Y_i \mid z_i \overset{\text{ind}}{\sim} \mathcal{N}(z_i, \sigma^2), \quad z_i \mid G \overset{\text{iid}}{\sim} G, \quad G \sim \mathrm{DP}(\alpha, P_0),$$

	CRP	BGS−L	BGS−n	Slice
Scalability by posterior cluster growth				
$H_n = O(n)$	$O(n^2)$	$\Theta(n)$	$\Theta(n^2)$	$O_{\mathbb{P}}(n^2)$
$H_n = O(\log n)$	$O(n \log n)$	$\Theta(n)$	$\Theta(n^2)$	$O_{\mathbb{P}}(n \log n)$
$H_n = O(1)$	$O(n)$	$\Theta(n)$	$\Theta(n^2)$	$O_{\mathbb{P}}(n \log n)$
Exact posterior partition target	✓	✗	✗	✓
No hard-threshold bias	✓	✗	✓	✓
No bookkeeping	✗	✓	✓	✓
Joint Updates	✗	✓	✓	✓

Numerical study

Competing algorithms:

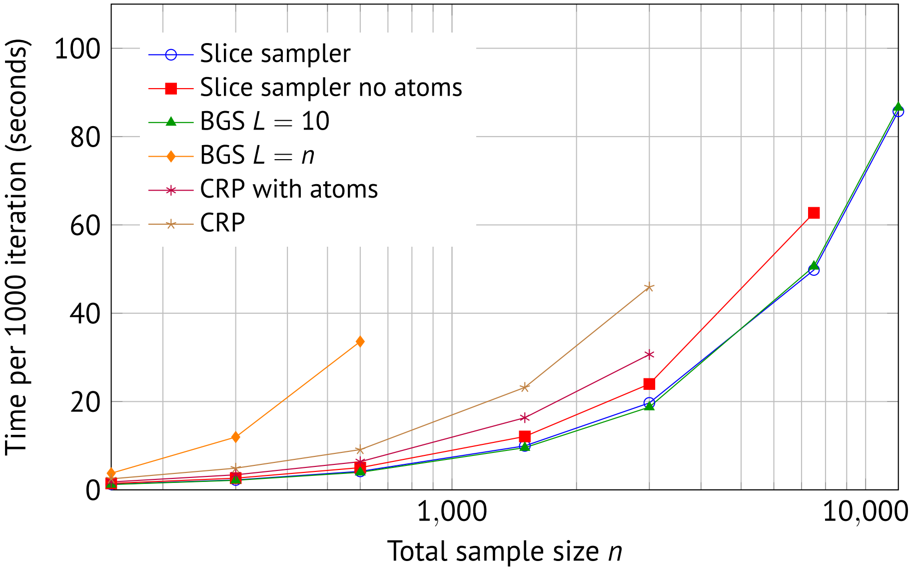
- ⋆ Slice sampler
- ⋆ Slice sampler with no atoms
- ⋆ BGS $L = 10$
- ⋆ BGS $L = n$
- ⋆ Marginal CRP
- ⋆ Marginal CRP with atoms

Simulated data scenario: Three clusters of equal size

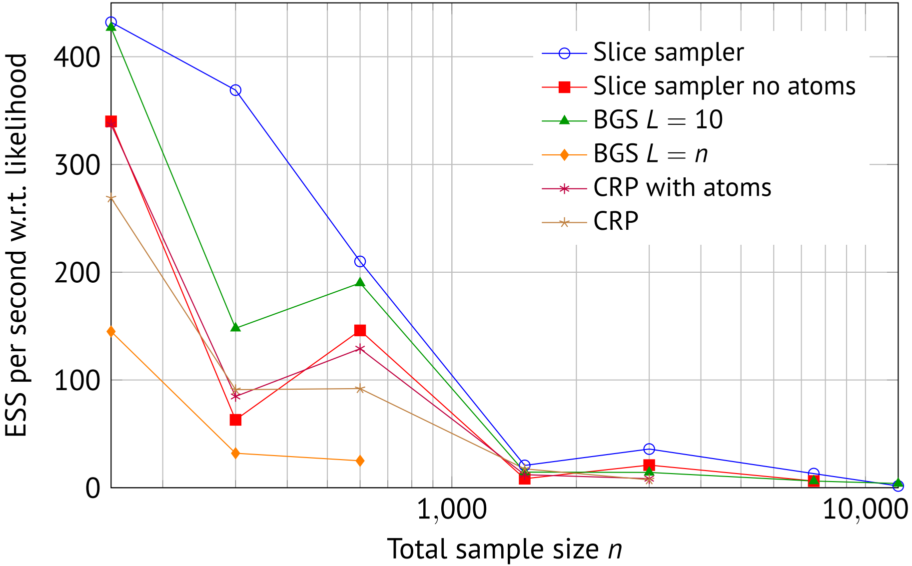
$$y_i \sim \mathcal{N}\left(\phi_{z_i},\, \tau^{-1}\right)$$

with true cluster means $\phi = (-3,\, 0,\, 3)$ and precision $\tau = 1$

Wall-clock time



Mixing



ESS per second w.r.t. likelihood
400
300
200
100
0
1,000
10,000
Total sample size n
Slice sampler
Slice sampler no atoms
BGS L = 10
BGS L = n
CRP with atoms
CRP

Numerical study

Competing algorithms:

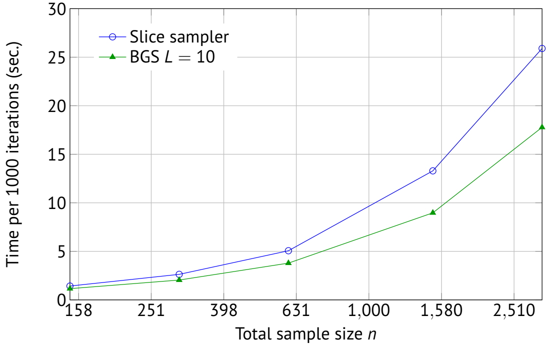
- $\star$ Slice sampler
- $\star$ BGS $L = 10$

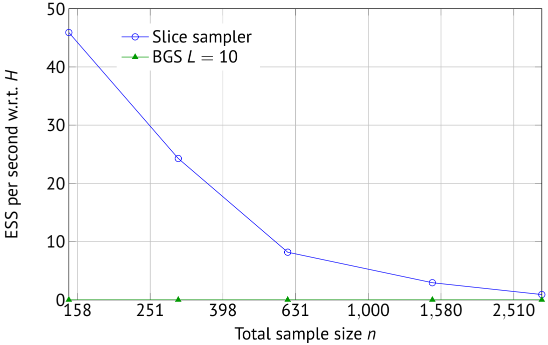
Simulated data scenario: Perturbed Zipf

$$y_i \sim \mathcal{N}\left(\phi_{z_i},\, \tau^{-1}\right)$$

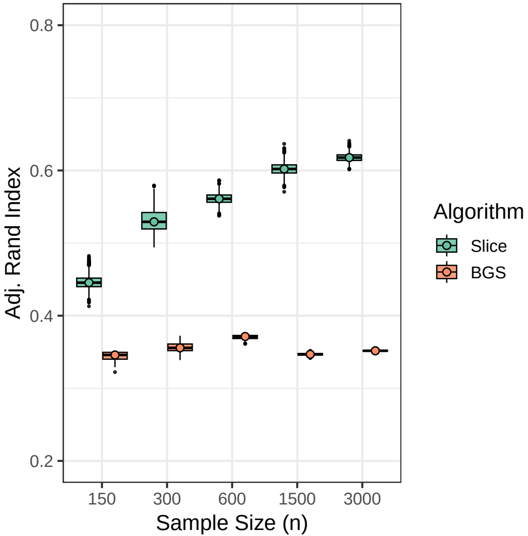
$$\phi_{z_i} = 3 \times z_i \qquad p(z_i) \propto z_i^{-0.5} \quad \text{for } z_i \in \{1, \ldots, 50\}$$

Wall-clock time and mixing





Inferential performance



Conclusions and future directions

Take-home message

- Slice samplers for DP-based models have **unbounded** per-iteration cost:

$$K_n = \min\left\{k \geq H_n : \sum_{h=k+1}^{\infty} \pi_h < u_{\min}\right\}.$$

- Nevertheless, for **any dataset**, the slice overhead is **small with high probability**:

$$K_n - H_n = O_{\mathbb{P}}(\log n), \qquad \text{uniformly over all visited partitions } \rho_n.$$

Future directions

- Extend the bounds to **adaptive DP models**, where the concentration parameter α is learned from the data.
- Study **richer BNP priors**: Pitman–Yor processes (Canale et al., 2022) and general species sampling processes (Mena et al., 2025).
- Develop guarantees for **structured models**: hierarchical DPs, sticky constructions, and dependent or temporally evolving partitions.

References

Campbell, T., Huggins, J. H., How, J. P., and Broderick, T. (2019). Truncated random measures. *Bernoulli*, 25(2):1256–1288.

Canale, A., Corradin, R., and Nipoti, B. (2022). Importance conditional sampling for pitman–yor mixtures. *Statistics and Computing*, 32(3):40.

Franzolini, B. and Gaffi, F. (2026). **Complexity bounds for Dirichlet process slice samplers**. In *International Conference on Machine Learning*, pages 1–22. PMLR (in press).

Ge, H., Chen, Y., Wan, M., and Ghahramani, Z. (2015). Distributed inference for Dirichlet process mixture models. In *International Conference on Machine Learning*, pages 2276–2284. PMLR.

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

Ishwaran, H. and James, L. F. (2002). Approximate Dirichlet process computing in finite normal mixtures: smoothing and prior information. *Journal of Computational and Graphical statistics*, 11(3):508–532.

Ishwaran, H. and Zarepour, M. (2002). Exact and approximate sum representations for the Dirichlet process. *Canadian Journal of Statistics*, 30(2):269–283.

Li, X. and Campbell, T. (2021). Truncated simulation and inference in edge-exchangeable networks. *Electronic Journal of Statistics*, 15(2):5117–5157.

Mena, R. H., Merkatas, C., Nicoleris, T., and Rodríguez, C. E. (2025). Exact finite mixture representations for species sampling processes. *arXiv preprint arXiv:2512.24414*.

Muliere, P. and Tardella, L. (1998). Approximating distributions of random functionals of Ferguson-Dirichlet priors. *Canadian Journal of Statistics*, 26(2):283–297.

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of computational and graphical statistics*, 9(2):249–265.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3):705–767.

Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics—Simulation and Computation*, 36(1):45–54.