Bayesian mixture models (in)consistency for the number of clusters

Louise Alamichel, Daria Bystrova, Julyan Arbel, Guillaume Kon Kam King

Bayesian mixture models (in)consistency for the number of clusters

Louise Alamichel[1], Daria Bystrova[1], Julyan Arbel[1], Guillaume Kon Kam King[2]

[1]Univ. Grenoble Alpes, Inria, France [2]Univ. Paris-Saclay, INRAE, France

Mixture models

$$x \sim \int p(\cdot|\boldsymbol{\theta})G(\mathrm{d}\boldsymbol{\theta}),$$

with G the mixing measure.

$$x \mid \boldsymbol{w}, \boldsymbol{\theta}, K \sim f(\cdot) = \sum_{k=1}^{K} w_k p(\cdot|\theta_k),$$

where K is the number of components, $\boldsymbol{w}$ the weights and p a kernel parametrized by $\boldsymbol{\theta}$.
K may be fixed, infinite or random.

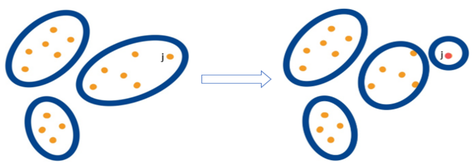
Posterior consistency

The posterior distribution is said to be consistent at θ_0 if $\Pi(U^c \mid X_{1:n}) \underset{n\to\infty}{\longrightarrow} 0$ in P_{θ_0}-probability for all neighborhoods U of θ_0.
G_0 defines a finite mixture with K_0 components; consistency for G with K components:

Quantity of interest	Finite		Infinite	MFM
	$K = K_0$	$K \geq K_0$	$K = \infty$	K random
Density f_0	✓	✓	✓	✓
G_0	✓	✓	✓	✓
K_0	N/A	✗/✓	✗/✓	✓

Partition-based model

- Partitions in k sets:
  $\mathcal{A}_k(n) = \left\{(A_1, \ldots, A_k) : \cup_{i=1}^{k} A_i = 1:n\right\}$
- For $A \in \mathcal{A}_k(n)$, $\mathcal{Z}_A = \{B(A,j),\ j \in 1:n\}$

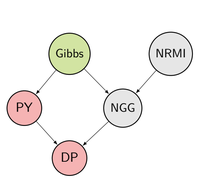


$A \in \mathcal{A}_k(n)$ $\qquad$ $B = B(A,j) \in \mathcal{A}_{k+1}(n)$

- K_n: number of clusters

$$\begin{aligned} p(A,k) &= p(A)\mathbb{I}(A \in \mathcal{A}_k(n)) \\ \theta_i|A,k &\sim \pi, & i &= 1,\ldots,n \\ x_j|A,k,\theta_{1:k} &\sim p(\cdot|\theta_\ell) \text{ if } j \in A_\ell, & j &= 1,\ldots,n \end{aligned}$$

Partition distributions



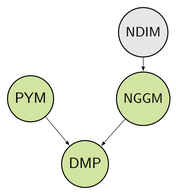
Gibbs-type process with $0 < \sigma < 1$, $A \in \mathcal{A}_k(n)$, $n_j := |A_j|$:

$$p(A) = \frac{V_{n,k}}{k!} \prod_{j=1}^{k} (1-\sigma)_{n_j - 1}.$$

$V_{n,k}$ satisfy the recurrence relation

$$V_{n,k} = (n - \sigma k)V_{n+1,k} + V_{n+1,k+1},\ V_{1,1} = 1.$$

Examples of Gibbs-type processes finite representation:



- Dirichlet multinomial process,
- Pitman–Yor multinomial process,
- NGG multinomial process,

parametrized by $K < \infty$: for $k \leq \min(n, K)$, $A \in \mathcal{A}_k(n)$,

$$p(A) = \Pi_K(n_1, \ldots, n_k).$$

Inconsistency Theorem [1]

Condition 1: Assume for any $k \geq 1$, $\limsup_{n\to\infty} \frac{1}{n} \max_{A \in \mathcal{A}_k(n)} \max_{B \in \mathcal{Z}_A} \frac{p(A)}{p(B)} < \infty$.

Condition 2: Condition on the data distribution which involves a control on the likelihood.

Theorem 1 [1]: Let $X_1, X_2, \ldots \in \mathcal{X}$ be a sequence of r.v., and consider a partition-based model. Then, for any $k \geq 1$ if Conditions 1 and 2 hold, we have

$$\limsup_{n\to\infty} \Pi(K_n = k|X_{1:n}) < 1 \quad \text{with probability 1.}$$

Inconsistency results

Proposition 1: Consider a Gibbs-type process with $0 < \sigma < 1$, then Condition 1 holds for any $k \in \{1, 2, \ldots\}$, so does the inconsistency of Theorem 1.

Proposition 2: Consider any of the following priors:
- Dirichlet multinomial process,
- Pitman–Yor multinomial process,
- Normalized Generalized Gamma multinomial process,

then Condition 1 holds for any $k < \min(n, K)$, and so does the inconsistency of Theorem 1.

Illustration on simulated data:

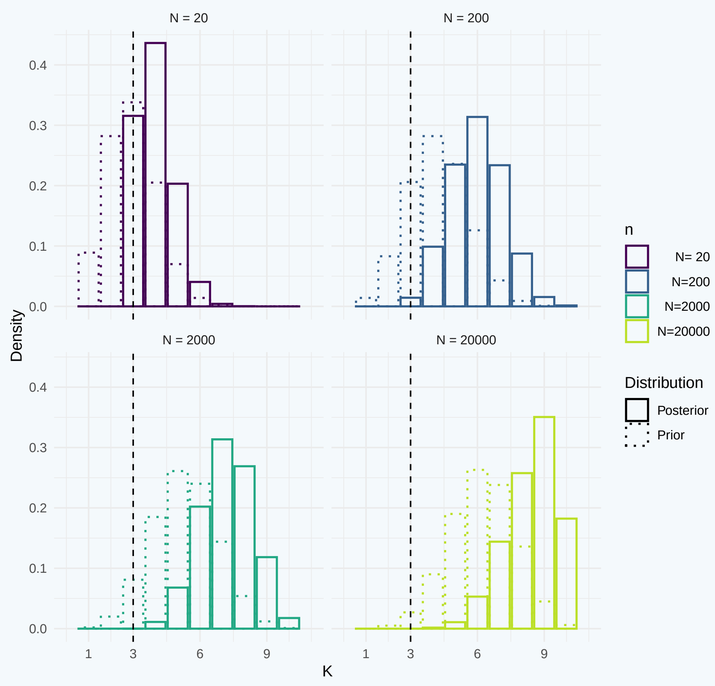
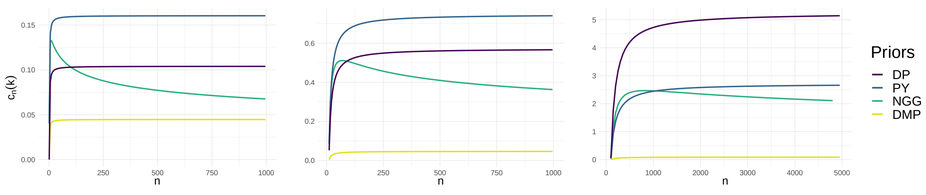


Figure 1: Distribution of K_n for a Dirichlet multinomial process, $\alpha = 0.9$.

Idea of the proof for Gibbs-type processes

- For fixed k, $\forall A \in \mathcal{A}_k(n)$ and $B = B(A,j)$, we want $\frac{1}{n} \max_{A,B} \frac{p(A)}{p(B)} < \infty$, we can show $\frac{1}{n} \frac{p(A)}{p(B)} \leq \frac{V_{n,k}}{V_{n,k+1}}(k+1)$.
- It is sufficient to prove $\left(\frac{V_{n,k}}{V_{n,k+1}}\right)_{n\geq 0}$ is bounded. By an approximation of $V_{n,k}$: $\left(\frac{V_{n,k}}{V_{n,k+1}}\right)_{n\geq 0}$ converges, so is bounded.

Illustration of function $n \mapsto \frac{1}{n} \max_{A \in \mathcal{A}_k(n)} \max_{B \in \mathcal{Z}_A} \frac{p(A)}{p(B)}$, for $k \in \{1, 10, 100\}$:



Consistency results

- *Dirichlet multinomial process*: satisfies assumptions in [2] for some α, so the weights of extra clusters $\to 0$, a form of mixing measure consistency.
- *PY multinomial process*: for $\sigma = 1/2$ and some α satisfies assumptions in [2], the weights of extra clusters $\to 0$.
- *Dirichlet process*: Posterior consistency for K_n by MTM shown in [3].
- *Pitman–Yor process*: from [4] we deduce a contraction rate, hence by using MTM algorithm we have posterior consistency for K_n.
- *Overfitted mixtures*: Posterior consistency for K_n by MTM.

Merge-Truncate-Merge algorithm [3]

Suppose $\omega_n : W_r(G, G_0) = o_p(\omega_n)$ for posterior sample G

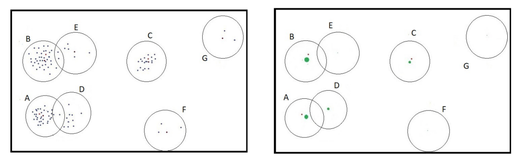


Figure 1: Initial distribution G.

Figure 2: After first stage-"merge".

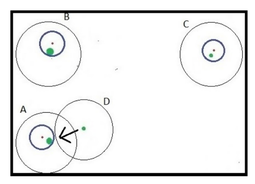


Figure 3: After second stage-"truncation".

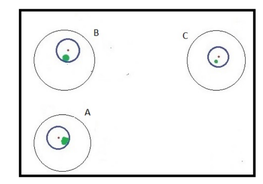


Figure 4: After second stage-"merge".

Figure 2: MTM algorithm illustration (figure from [3]).

$\tilde{k}$ number of support atoms of the output $\tilde{G}$.

$$\Pi\left(\tilde{k} = k_0 \mid X_{1:n}\right) \to 1 \text{ in probability.}$$

Future work

- Do the inconsistency results generalize to NIDM processes ?
- Can consistency be recovered by adding a prior on α, as in [5] ?

Preprint available on arXiv, search for "Alamichel".

References

[1] Jeffrey W Miller and Matthew T Harrison. Inconsistency of Pitman-Yor process mixtures for the number of components. *The Journal of Machine Learning Research*, 15(1):3333–3370, 2014.

[2] Judith Rousseau and Kerrie Mengersen. Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(5):689–710, 2011.

[3] Aritra Guha, Nhat Ho, and XuanLong Nguyen. On posterior contraction of parameters and interpretability in bayesian mixture modeling. *Bernoulli*, 27(4):2159–2188, 2021.

[4] Catia Scricciolo. Adaptive Bayesian density estimation in Lp-metrics with Pitman-Yor or normalized inverse-Gaussian process kernel mixtures. *Bayesian Analysis*, 9(2):475–520, 2014.

[5] Filippo Ascolani, Antonio Lijoi, Giovanni Rebaudo, and Giacomo Zanella. Clustering consistency with Dirichlet process mixtures. *arXiv preprint arXiv:2205.12924*, 2022.