Bayesian mixture models (in)consistency for the number of clusters
Louise Alamichel, Daria Bystrova, Julyan Arbel, Guillaume Kon Kam King

To cite this version:
Louise Alamichel, Daria Bystrova, Julyan Arbel, Guillaume Kon Kam King. Bayesian mixture models (in)consistency for the number of clusters. 13th Bayesian nonparametrics (BNP) conference, Oct 2022, Puerto Varas, Chile. hal-03866441

HAL Id: hal-03866441
https://hal.archives-ouvertes.fr/hal-03866441
Submitted on 22 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
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 \quad \(^2\)Univ. Paris-Saclay, INRAE, France

Mixture models

\[
x \sim \int p(x | \theta) G(d\theta),
\]

with \(G\) the mixing measure.

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

where \(K\) is the number of components, \(w\) the weights and \(p\) a kernel parametrized by \(\theta\).

\(K\) may be fixed, infinite or random.

Posterior consistency

The posterior distribution is said to be consistent at \(\theta_0\) if \(\Pi(U^n | X_{1:n}) \to 0\) in \(P_{\theta_0}\). Probability for all neighborhoods \(U\) of \(\theta_0\).

\(G_0\) defines a finite mixture with \(K_0\) components, consistency for \(G\) with \(K\) components:

- Quantity of interest: \(K = K_0, K \geq K_0, K < \infty\) randomly
- \(\theta_0\)
- \(G_0\)
- \(K_0\)
- MFM

Partition-based model

- Partitions in \(k\) sets:
  \(A_0(n) = \{A_1, \ldots, A_k\} \cup \bigcup_{j=1}^k A_j = 1 : n\}
- For \(A \in A_0(n)\), \(Z_A = \{B(A, j)\}, j \in 1 : n\)
  \(A \in A_0(n)\) \quad \(B = B(A, j) \in A_{i+1}(n)\)

- \(K_n\): number of clusters

\[
p(A, k) = \frac{p(A) \Pi (A \in A_0(n))}{\max_{A \in A_0(n)} \Pi (A \in A_0(n))},
\]

\[
x_j | A, k, \theta_k \sim p(x | \theta_k), j \in A_i, j = 1, \ldots, n
\]

Partition distributions

Gibbs-type process with \(0 < \sigma < 1\), \(A \in A_0(n)\), \(n_j := |A_j|\):

\[
p(A) = \frac{V_{n,k}}{K} \prod_{j=1}^{k} (1 - \sigma)^{n_j - 1}.
\]

\(V_{n,k}\) satisfy the recurrence relation

\[
V_{n,k} = (n - \sigma k) V_{n-1,k} + V_{n-1,k+1}, \quad V_{1,1} = 1.
\]

Examples of Gibbs-type processes finite representation:

- Dirichlet multinomial process,
- Pitman–Yor multinomial process,
- Normalized Generalized Gamma multinomial process, parametrized by \(K < \infty\): for \(k \leq \min(n, K)\), \(A \in A_0(n)\),
  \[
p(A) = \Pi_k (n_1, \ldots, n_k).
\]

Inconsistency Theorem [1]

Condition 1: Assume for any \(k \geq 1\),
  \[
  \lim_{n \to \infty} \frac{\max_{A \in A_0(n)} \max_{i \in A_0(n)} p(A | \theta_i)}{\max_{A \in A_0(n)} \max_{i \in A_0(n)} p(A | \theta_j)} < \infty.
\]

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 i.i.d.,
and consider a partition-based model. Then, for any \(k \geq 1\) if Conditions 1 and 2 hold, we have

\[
\lim_{n \to \infty} \Pi(K_n = k | X_{1:n}) < 1 \text{ with probability } 1.
\]

Partition distributions

Gibbs-type process with \(0 < \sigma < 1\), \(A \in A_0(n)\), \(n_j := |A_j|\):

\[
p(A) = \frac{V_{n,k}}{K} \prod_{j=1}^{k} (1 - \sigma)^{n_j - 1}.
\]

\(V_{n,k}\) satisfy the recurrence relation

\[
V_{n,k} = (n - \sigma k) V_{n-1,k} + V_{n-1,k+1}, \quad V_{1,1} = 1.
\]

Examples of Gibbs-type processes finite representation:

- Dirichlet multinomial process,
- Pitman–Yor multinomial process,
- Normalized Generalized Gamma multinomial process, parametrized by \(K < \infty\): for \(k \leq \min(n, K)\), \(A \in A_0(n)\),
  \[
p(A) = \Pi_k (n_1, \ldots, n_k).
\]

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:

Idea of the proof for Gibbs-type processes

- For fixed \(k, \forall A \in A_0(n)\) and \(B = B(A, j)\), we want \(\max_{A \in A_0(n)} \max_{i \in A_0(n)} p(A | \theta_i) \cdot p(A) < \infty\), we can show \(\frac{\prod p(A | \theta_i)}{\max_{A \in A_0(n)} \max_{i \in A_0(n)} p(A | \theta_j)} \leq \frac{\max_{A \in A_0(n)} \max_{i \in A_0(n)} p(A | \theta_j)}{\max_{A \in A_0(n)} \max_{i \in A_0(n)} p(A | \theta_i)}\) for \(k \geq 1\).

- It is sufficient to prove \(\frac{V_{n,k}}{K} \prod_{j=1}^{k} (1 - \sigma)^{n_j - 1} \leq \frac{V_{n,k+1}}{K} \prod_{j=1}^{k+1} (1 - \sigma)^{n_j - 1}\) is bounded. By an approximation of \(V_{n,k} \cdot \frac{\prod_{j=1}^{k} (1 - \sigma)^{n_j - 1}}{\prod_{j=1}^{k+1} (1 - \sigma)^{n_j - 1}}\) converges, so is bounded.

Illustration of function \(n \mapsto \frac{\max_{A \in A_0(n)} \max_{i \in A_0(n)} p(A | \theta_i)}{\max_{A \in A_0(n)} \max_{i \in A_0(n)} p(A | \theta_j)}\), for \(k \in \{1, 10, 100\}\):

Consistency results

- Dirichlet multinomial process: satisfies assumptions in [2] for some \(\alpha\), so the weights of extra clusters \(\to 0\), a form of mixing measure consistency.
- PY multinomial process: for \(\sigma = 1/2\) and some \(\alpha\) satisfies assumptions in [2], the weights of extra clusters \(\to 0\).
- Dirichlet process: Posterior consistency for \(K_n\) by MTM shown in [3].
- Overfitted mixtures: Posterior consistency for \(K_n\) by MTM.

Merge-Truncate-Merge algorithm [3]

Suppose \(\omega_n = W, G_0 = \emptyset(\omega_n)\) for posterior sample \(G\).

Future work

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

Preprint available on arXiv, search for “Alamichel”.

References