Discussion on

“Nested nonparametric processes”

by Federico Camerlenghi

Antonio Canale · canale@stat.unipd.it · @tonycanale_

O’Bayes 2022, University of California Santa Cruz, Sept 9, 2022

Università
degli Studi
di Padova

Brief summary (1)

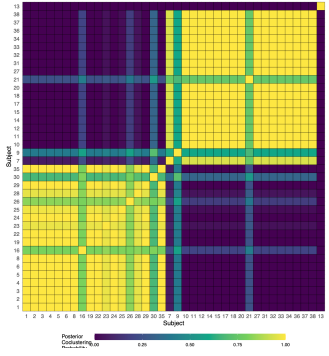
- Federico reviewed the degeneracy property of the nDP presented in Camerlenghi et al (2019, BA), i.e. two random probability measures are either identical or share no common atoms
- To solve the above issue the large class of latent nested processes (LNP) is introduced
- In Denti et al (2022, JASA), instead, the common atom model (CAM) is introduced:

$$y_{i,j} \mid G_j \qquad G_j \mid Q \sim Q, \qquad Q = \sum_{h \geq 1} \pi_h \delta_{G_h^*}, \qquad G_h^* = \sum_{l \geq 1} w_{hl} \delta_{\theta_l^*}.$$

Brief summary (2)

- CAM does not suffer from the degeneracy property and allows a two-layer clustering
  - Distributional clustering: G_j are clustered to the G_h^*
  - Observational clustering $y_{i,j}$ are clustered in the atoms θ_l^*.

- CAM is applied to analyze complex microbiome data
- Data consist of a $n \times J$ abundance table, a matrix formed by n operational taxonomic unit (OTU) measurements (obervations) for each of the J individuals (groups)
- In this case the distributional clustering are grouping the individuals



X

1) Possible CAM generalizations

- In the CAM all the sequences of weights have a DP-like construction, i.e.

$$\pi_h = \nu_h \prod_{\ell<h}(1-\nu_\ell), \qquad \nu_h \sim Beta(1, a)$$

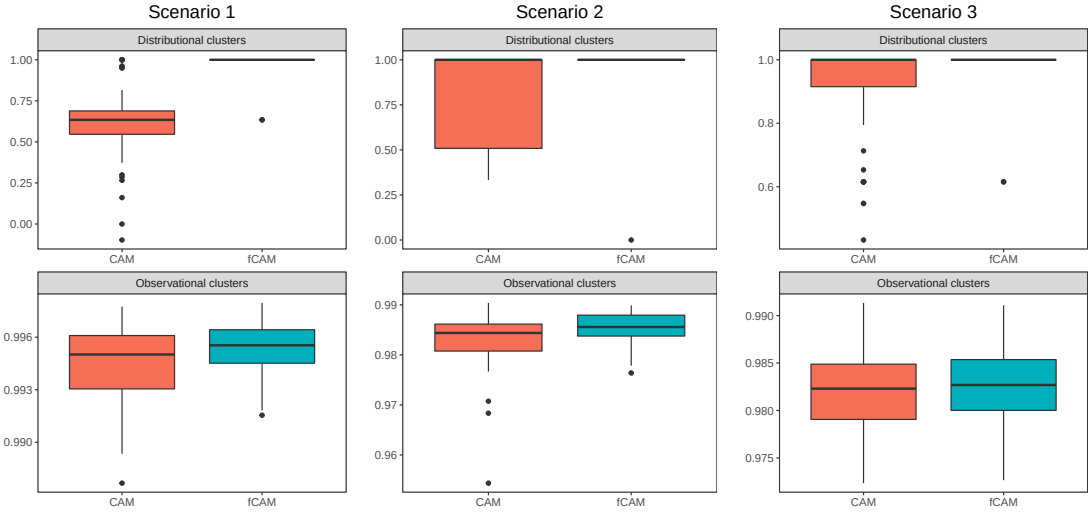
- Natural extensions include general stick-breaking priors, e.g. the Pitman-Yor process

$$\pi_h = \nu_h \prod_{\ell<h}(1-\nu_\ell), \qquad \nu_h \sim Beta(1-\sigma, a+h\sigma)$$

  this would allow more flexible distributional clustering behaviour.

- In D'Angelo et al. (2022, Biometrics) we defined a mixture of finite mixture (MFM) version of the CAM also employing the computational strategies of Frühwirth-Schnatter, et al. (2021, BA)

finite-CAM: clustering performance on simulation



2) Testing group differences

- The CAM is reminiscent of the shared kernel (SK) screening approach by Lock and Dunson (2015, Biometrika) and Canale and Dunson (2017, Stat. Sinica).
- Consider data belonging to two groups (e.g. cases and controls) and assume to measure some outcome $y_{i,1} \sim f_1$ for group 1 and $y_{j,0} \sim f_0$ for group 0 with interest on

$$H_0 : f_0 = f_1 \qquad H_1 : f_0 \neq f_1$$

- Assume a SK mixture model for both cases and controls, e.g.

$$f_h(\cdot) = \sum_{\ell} \pi_{\ell,h} K(\cdot; \theta_\ell)$$

2) Testing group differences

- The CAM is reminiscent of the shared kernel (SK) screening approach by Lock and Dunson (2015, Biometrika) and Canale and Dunson (2017, Stat. Sinica).
- Consider data belonging to two groups (e.g. cases and controls) and assume to measure some outcome $y_{i,1} \sim f_1$ for group 1 and $y_{j,0} \sim f_0$ for group 0 with interest on

$$H_0 : f_0 = f_1 \qquad H_1 : f_0 \neq f_1$$

- Assume a SK mixture model for both cases and controls, e.g.

$$f_h(\cdot) = \sum_{\ell} \pi_{\ell,h} K(\cdot; \theta_\ell)$$

- The hypotheses is equivalent to

$$H_0 : \pi_{\ell,0} = \pi_{\ell,1} \text{ for each } \ell \qquad H_1 : \pi_{\ell,0} \neq \pi_{\ell,1} \text{ for some } \ell$$

2) Testing group differences

- The CAM is reminiscent of the shared kernel (SK) screening approach by Lock and Dunson (2015, Biometrika) and Canale and Dunson (2017, Stat. Sinica).
- Consider data belonging to two groups (e.g. cases and controls) and assume to measure some outcome $y_{i,1} \sim f_1$ for group 1 and $y_{j,0} \sim f_0$ for group 0 with interest on

$$H_0 : f_0 = f_1 \qquad H_1 : f_0 \neq f_1$$

- Assume a SK mixture model for both cases and controls, e.g.

$$f_h(\cdot) = \sum_{\ell} \pi_{\ell,h} K(\cdot; \theta_\ell)$$

- The hypotheses is equivalent to

$$H_0 : \pi_{\ell,0} = \pi_{\ell,1} \text{ for each } \ell \qquad H_1 : \pi_{\ell,0} \neq \pi_{\ell,1} \text{ for some } \ell$$

- Can we consider this a special case of CAM mixture? Can we use CAM mixtures for testing group differences?

3) Really nonparametric?

- In mixture models, assuming a finite or infinite (nonparametric) mixing measure makes a conceptual difference

3) Really nonparametric?

- In mixture models, assuming a finite or infinite (nonparametric) mixing measure makes a conceptual difference
- Assume the simple setting

$$y_i \mid P \sim P, \qquad P = \sum_{h \geq 1} \pi_h \delta_{\theta_h^*}$$

- In both cases, for a finite sample ($i = 1, dots, n$) data cluster into $k_n \leq n$ clusters

3) Really nonparametric?

- In mixture models, assuming a finite or infinite (nonparametric) mixing measure makes a conceptual difference
- Assume the simple setting

$$y_i \mid P \sim P, \qquad P = \sum_{h \geq 1} \pi_h \delta_{\theta_h^*}$$

- In both cases, for a finite sample ($i = 1, dots, n$) data cluster into $k_n \leq n$ clusters
- In finite mixtures y_{n+1} can be assigned in a new cluster but up to a prespecified upper bound.

3) Really nonparametric?

- In mixture models, assuming a finite or infinite (nonparametric) mixing measure makes a conceptual difference
- Assume the simple setting

$$y_i \mid P \sim P, \qquad P = \sum_{h \geq 1} \pi_h \delta_{\theta_h^*}$$

- In both cases, for a finite sample ($i = 1, dots, n$) data cluster into $k_n \leq n$ clusters
- In finite mixtures y_{n+1} can be assigned in a new cluster but up to a prespecified upper bound.
- In CAM, however,

$$Q = \sum_{h \geq 1} \pi_h \delta_{G_h^*}, \qquad G_h^* = \sum_{l \geq 1} w_{hl} \delta_{\theta_l^*}$$

is an infinite sum. Does it really make sense to assume an infinite mixture for the groups?

Questions/comments

1. Generalizations to other type of weights construction. Can we play similar game for classical BNP mixtures? Any special case that makes particularly sense in the CAM settings? Any problem in doing so (e.g. harder computations?)

Questions/comments

1. Generalizations to other type of weights construction. Can we play similar game for classical BNP mixtures? Any special case that makes particularly sense in the CAM settings? Any problem in doing so (e.g. harder computations?)
2. Similarities with the SK approach. Is the SK approach a special case of CAM mixture? Can we use CAM mixtures for testing group differences?

Questions/comments

1. Generalizations to other type of weights construction. Can we play similar game for classical BNP mixtures? Any special case that makes particularly sense in the CAM settings? Any problem in doing so (e.g. harder computations?)
2. Similarities with the SK approach. Is the SK approach a special case of CAM mixture? Can we use CAM mixtures for testing group differences?
3. Do we really need to assume an infinite mixture for $Q = \sum_{h \geq 1} \pi_h \delta_{G_h^*}$?

References

- Camerlenghi, F., Dunson, D, Lijoi A., Prünster, I., Rodrigues, A. (2019). Latent nested nonparametric priors (with discussion). *Bayesian Analysis*.
- Canale, A., and Dunson, D. B. (2016). Multiscale Bernstein polynomials for densities. *Statistica Sinica*
- D'Angelo, L., Canale, A., Yu, Z., Guindani, M. (2022) Bayesian nonparametric analysis for the detection of spikes in noisy calcium imaging data, *Biometrics*
- Denti, F., Camerlenghi, F., Guindani, M., & Mira, A. (2021) A Common Atom Model for the Bayesian Nonparametric Analysis of Nested Data, *Journal of the American Statistical Association*.
- Frühwirth-Schnatter, S., Malsiner-Walli, G., & Grün, B. (2021) Generalized mixtures of finite mixtures and telescoping sampling, *Bayesian Analysis*
- Lock, E. F., and Dunson, D. B. (2015). Shared kernel Bayesian screening. *Biometrika*