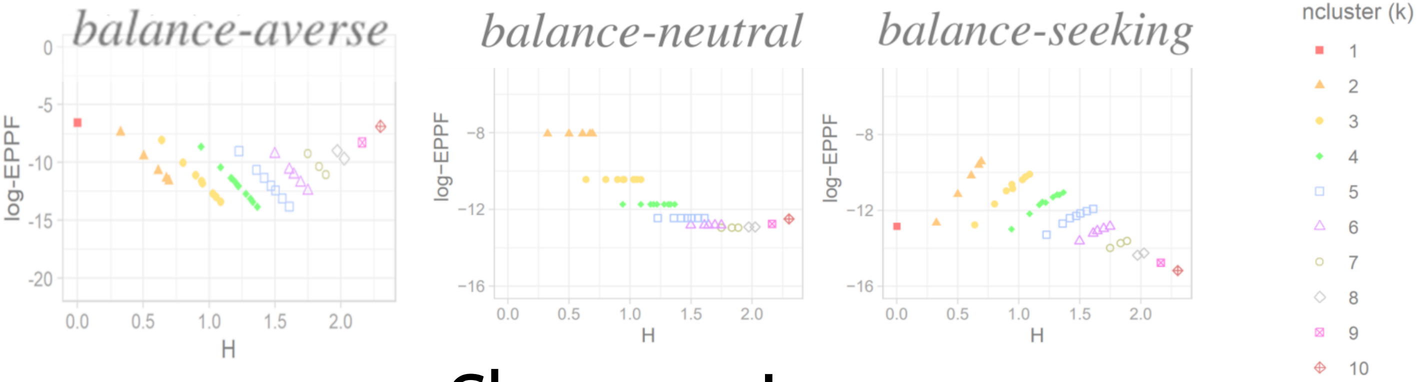
Why the Rich Get Richer?
On the Balancedness of Random Partition Models



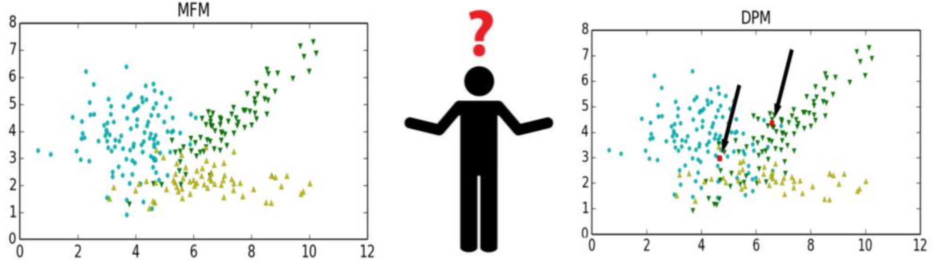
Changwoo Lee
(joint work with Dr. Huiyan Sang)
Department of Statistics, Texas A&M University

ICML 2022

Bayesian Methods for Clustering Problems

Mixture models:	Finite mixture	vs.	Dirichlet process (DP) mixture
Topic models:	Latent Dirichlet allocation (LDA)	vs.	Hierarchical DP
and many others...			

Q. Different models leads to different clustering results. **Why? How to choose?**



Miller and Harrison(2018)

A. We study "***balancedness***" of random partitions to understand those differences.

Random Partition Models

Probability distribution on the space of partition. Ex) Partitons of {1,2,3} :

π	$\{\{1,2,3\}\}$	$\{\{1,2\},\{3\}\}$	$\{\{1,3\},\{2\}\}$	$\{\{1\},\{2,3\}\}$	$\{\{1\},\{2\},\{3\}\}$
$\mathbb{P}(\Pi_3 = \pi)$	p_1	p_2	p_3	p_4	p_5

Act as a prior distribution for the Bayesian clustering methods.

Two important, common assumptions:

- (Finite) **Exchangeability assumption** $p_2 = p_3 = p_4.$
- **Projectivity assumption**

Random Partition Models

Probability distribution on the space of partition. Ex) Partitons of {1,2,3} :

π	$\{\{1,2,3\}\}$	$\{\{1,2\},\{3\}\}$	$\{\{1,3\},\{2\}\}$	$\{\{1\},\{2,3\}\}$	$\{\{1\},\{2\},\{3\}\}$
$\mathbb{P}(\Pi_3=\pi)$	p_1	p_2	p_3	p_4	p_5

Act as a prior distribution for the Bayesian clustering methods.

Two important, common assumptions:

- (Finite) **Exchangeability assumption**
- **Projectivity assumption**

$\downarrow$ (marginalize out 3)

π	$\{\{1,2\}\}$	$\{\{1\},\{2\}\}$
$\mathbb{P}(\Pi_2=\pi)$	p_1+p_2	$p_3+p_4+p_5$

EPPF and Gibbs partition

Under exchangeability, p.m.f can be written as an EPPF $p(n_1, \ldots, n_k)$

π	$\{\{1,2,3\}\}$	$\{\{1,2\},\{3\}\}$	$\{\{1,3\},\{2\}\}$	$\{\{1\},\{2,3\}\}$	$\{\{1\},\{2\},\{3\}\}$
$\mathbb{P}(\Pi_3 = \pi)$	p_1	p_2	p_3	p_4	p_5

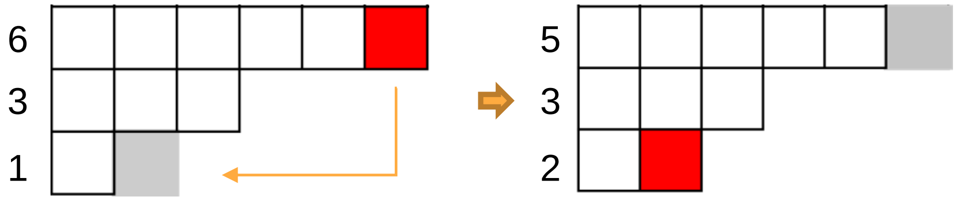
$p(\boldsymbol{n} = (3)) = p_1$ $\quad$ $p(\boldsymbol{n} = (2,1)) = p_2$ $\quad$ $p(\boldsymbol{n} = (1,1,1)) = p_5$

Gibbs partition: $\Pi_n \sim \text{Gibbs}_{[n]}(\boldsymbol{V}, \boldsymbol{W})$ if $p^{(n)}(n_1, \ldots, n_k) = V_{n,k} \prod_{j=1}^{k} W_{n_j}$ (product form)

Examples of Gibbs partition: Dirichlet-multinomial, Chinese restaurant process, …

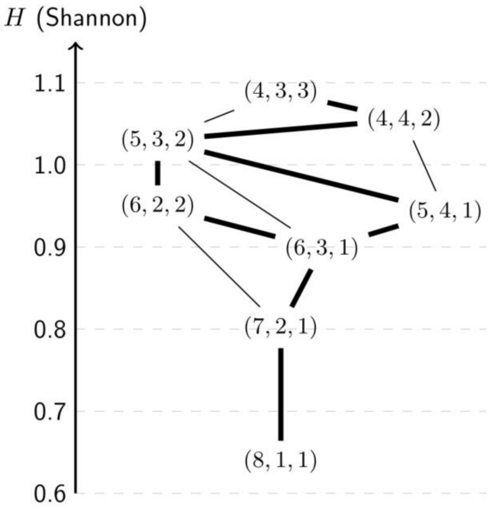
Balancedness of Fixed Partitions

Reverse dominance order $\prec$, based on one-step downshift(s).



(6,3,1) $\prec$ (5,3,2) (more balanced)

$$H(\boldsymbol{n}) := -\sum_{j=1}^{k}(n_j/n)\log(n_j/n)$$



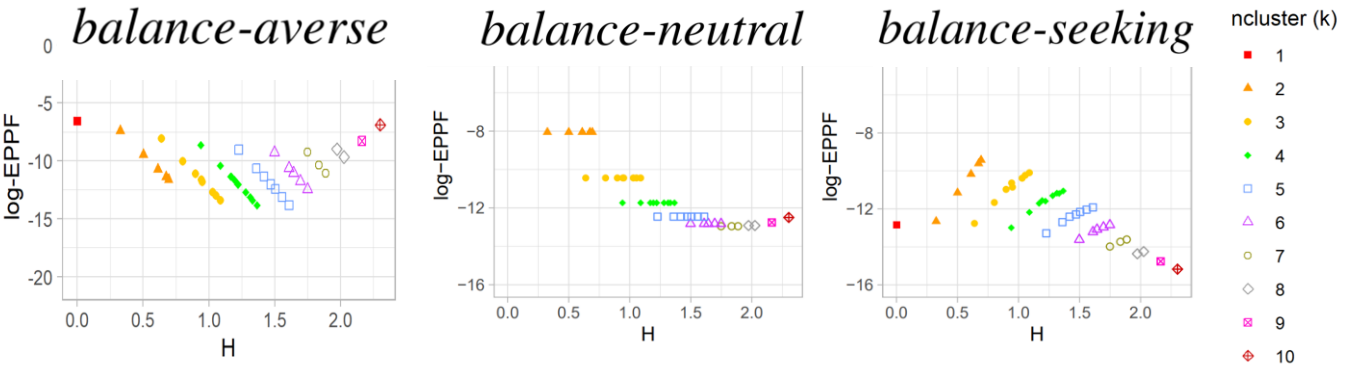
This induces *partial* order relation:

Q. What about the balancedness of RANDOM partition?

Balancedness of Exchageable Random Partitions

Definition 3.1. Let $p^{(n)}$ be an EPPF of a finitely exchangeable random partition Π_n on $[n]$. Then call Π_n and $p^{(n)}$

- *balance-averse* if $\boldsymbol{n} \prec \boldsymbol{n}' \implies p^{(n)}(\boldsymbol{n}) \geq p^{(n)}(\boldsymbol{n}')$
- *balance-seeking* if $\boldsymbol{n} \prec \boldsymbol{n}' \implies p^{(n)}(\boldsymbol{n}) \leq p^{(n)}(\boldsymbol{n}')$



Main Theorem: Balancedness of Gibbs Partitions

Theorem 3.2. *Let* $p^{(n)}$ *be an EPPF of* $\mathsf{Gibbs}_{[n]}(\boldsymbol{V}, \boldsymbol{W})$. *Then for any* $n = 1, 2, \ldots,$ $p^{(n)}$ *is*

- *balance-averse if and only if* $\boldsymbol{W}$ *is log-convex,*
- *balance-seeking if and only if* $\boldsymbol{W}$ *is log-concave,*

Log-convexity of $W = (W_s)_{s=1}^{\infty}$ determines the balancedness of Gibbs partition.

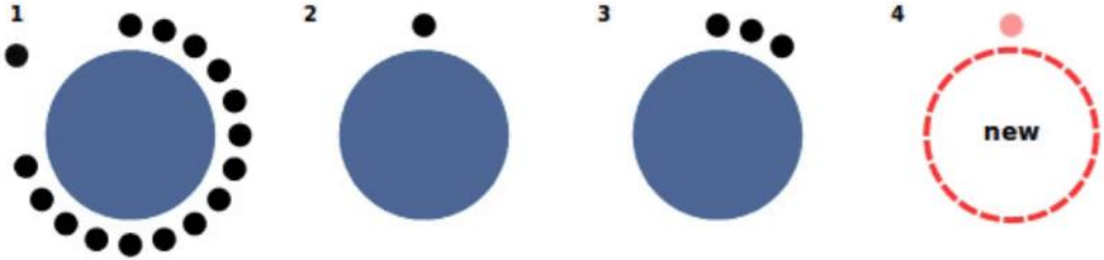
Necessary and sufficient condition, easy-to-check.

“Rich-get-richer” property

“Rich-get-richer” property is shared across many existing random partition models.

Example: Chinese restaurant process (CRP)

https://topicmodels.west.uni-koblenz.de/ckling/tmt/crp.html?parameters=0.5&dp=1



$$\mathbb{P}(z_{i+1} = j \mid z_1, \ldots, z_i) = \begin{cases} \dfrac{n_j}{i+\theta} & \text{for } j\text{th table} \\ \dfrac{\theta}{i+\theta} & \text{new table} \end{cases}$$

Why the rich get richer? *Exchangeability + Projectivity!*

For Gibbs partition (exchangeable with product-form),

Corollary 3.3. *Let* $\Pi_{n+1} \sim \mathsf{Gibbs}_{[n+1]}(\boldsymbol{V}, \boldsymbol{W})$. *Given the cluster memberships of the first* n *datapoints* $\mathbf{z}_{1:n}$ *with* k *clusters, the reallocation rule for the next datapoint is*

$$\mathbb{P}(z_{n+1} = j | \mathbf{z}_{1:n}) \propto \begin{cases} f(n_j) & \text{if } j = 1, \ldots, k \\ g(n,k) & \text{if } j = k+1 \end{cases} \quad (5)$$

Then f *is an increasing (decreasing) function over* $\mathbb{N}$ *if and only if* Π_{n+1} *is balance-averse (seeking) for any* $n \in \mathbb{N}$.

Rich-get-richer ⇔ Balance-averse

Rich-get-poorer ⇔ Balance-seeking

Corollary 3.4. *Let* $\mathrm{p}^{(n)}$ *be an EPPF of infinitely exchangeable Gibbs partition. Then* $\mathrm{p}^{(n)}$ *is always balance-averse; i.e., for two integer partitions* $\boldsymbol{n}, \boldsymbol{n}' \in \mathcal{I}_n^k$ *with same size* $k \leq n$,

$$\boldsymbol{n} \prec \boldsymbol{n}' \implies \mathrm{p}^{(n)}(\boldsymbol{n}) \geq \mathrm{p}^{(n)}(\boldsymbol{n}')$$

with equality holds only if $\sigma = -\infty$.

Product form Exchangeability + Projectivity

⇒ Always balance-averse!

(i.e. Always the rich-get-richer!)

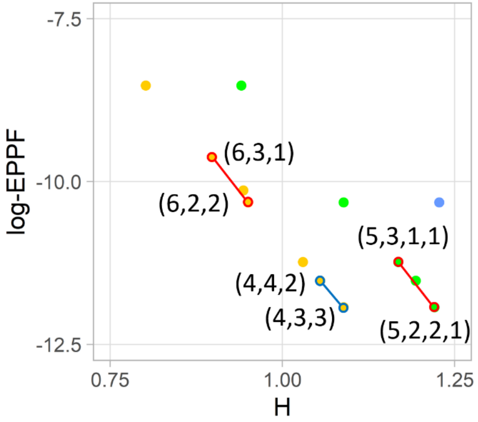
Want an exchangeable, balance-seeking partition? May need to sacrifice projectivity.

Comparing Balancedness between Gibbs Partitions

Definition 3.7 (B-sequence). $(B_s(\boldsymbol{W}))_{s\geq 2}$ is a B-sequence of $\text{Gibbs}_{[n]}(\boldsymbol{V},\boldsymbol{W})$ for $n \in \mathbb{N}$, a sequence of extended real numbers which only depends on $\boldsymbol{W}$, defined as

$$B_s(\boldsymbol{W}) = -s(\log W_{s+1} - 2\log W_s + \log W_{s-1}) \quad (9)$$

with the provision that $B_s(\boldsymbol{W}) = +\infty$ if $W_{s+1} = 0$.



B-sequence is an approximation of the slope of log-EPPF versus Shannon index plot.

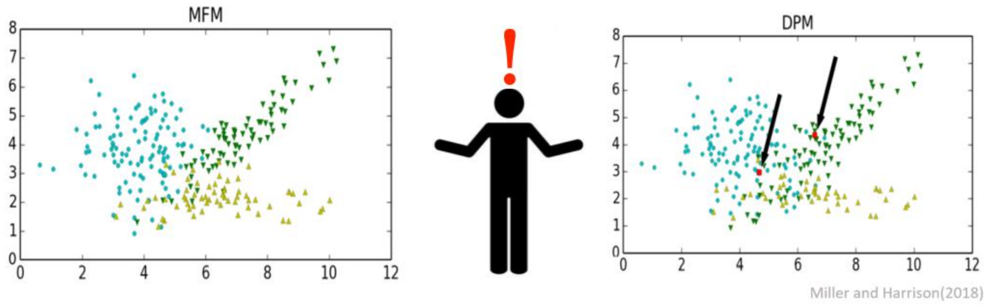
- Negative (Positive) B-sequence: balance-averse (balance-seeking).

$\text{Gibbs}_{[n]}(\boldsymbol{V},\boldsymbol{W})$ is more balanced than $\text{Gibbs}_{[n]}(\boldsymbol{V}',\boldsymbol{W}')$

if and only if $B_s(\boldsymbol{W}) \geq B_s(\boldsymbol{W}')$ *for all* $s \geq 2$.

Comparison: mixture model

DP mixture ($\sigma' = 0$) is more unbalanced than the (mixture of) finite mixture ($\sigma < 0$), by the B-sequence comparison $0 > B_s(\sigma) > B_s(\sigma')$.

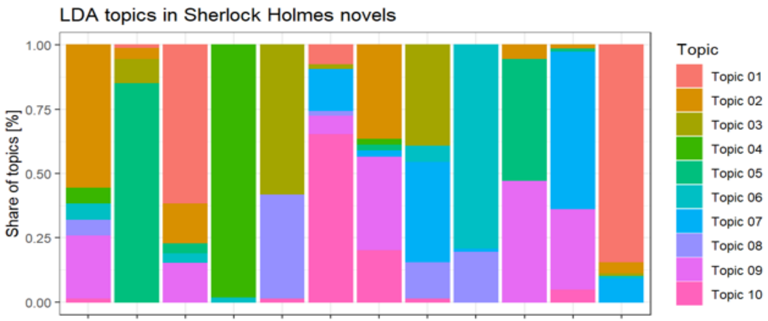


Explains why DP mixture creates tiny extra clusters compared to finite mixture model.

Unbounded number of clusters in DP-based models are often emphasized, however it comes with a cost: the resulting partition becomes more unbalanced.

Insight: topic model

LDA uses Dirichlet-multinomial random partition for per-document topic distribution, which is **balance-averse** (negative B-sequence).



https://content-analysis-with-r.com/6-topic_models.html

Per-document topic distribution is **naturally unbalanced** (e.g. news article).

Demonstrates that LDA is a highly suitable for topic modeling applicaton.

In the paper, we also provide an example of *balance-seeking* random partition and its application to entity resolution tasks.

Thank you!