Conversation With an LLM

The research project you're developing sits at a novel and complex intersection of **Multiple Systems Estimation (MSE)** and **causal inference**, particularly under **intervention-driven changes** to data-generating processes. The good news is that while this intersection is largely unexplored, several studies do address components of your conceptual framework. Here's a curated synthesis of the most relevant literature, structured around your research themes:

---

1. MSE Under Intervention and SUTVA Violations

- **Binette & Steorts (2021)** directly examine MSE's reliability for estimating modern slavery prevalence. They simulate violations of assumptions (e.g., independence between lists) and find substantial bias under changing overlap, a core concern in your project (Binette & Steorts, 2021).
- **Lum, Price, & Banks (2013)** discuss MSE in human rights research and introduce simulations to evaluate the impact of assumption violations. This is directly relevant for your simulation design and highlights the cascading effect of SUTVA-like violations on estimates (Lum et al., 2013).
- **Sharifi Far et al. (2020)** explore how list omission and combination affect estimates in MSE for modern slavery. Their simulation studies on Romania and the UK provide a blueprint for testing robustness against structural data changes (Sharifi Far et al., 2020).

---

2. Causal Inference and Diff-in-Diff Under SUTVA Violation

- **Boesche (2020)** critiques quasi-experiments in policy evaluation due to the implausibility of SUTVA in real-world settings. He proposes a weaker alternative assumption (SMUTVA), which could be valuable for formalizing your theoretical framework (Boesche, 2020).
- **Kazunari (2017)** reviews assumptions in Difference-in-Differences (DID), focusing on SUTVA violations and proposing bias-correction methods in observational settings. The methodology is applicable to your staggered intervention context (Kazunari, 2017).
- **Ma & Tresp (2020)** examine causal inference under interference using network models. While not MSE-focused, they rigorously tackle the consequences of treatment spillovers—a form of SUTVA violation—in estimating individual treatment effects (Ma & Tresp, 2020).

---

3. MSE Modeling and List Structure Dependencies

- **Cruyff et al. (2020)** address how model selection under sparsity and inter-list dependence (e.g., police and immigration data) affects the validity of MSE results. Their practical strategies for managing list structure are immediately relevant to your agency coordination scenarios (Cruyff et al., 2020).
- **Manrique-Vallier et al. (2013)** discuss advanced MSE techniques for human rights and war casualty estimation, emphasizing how subtle violations (e.g., regional misrepresentation, list dependency) distort population estimates (Manrique-Vallier et al., 2013).

---

4. Simulation and Counterfactual Testing

- Several of the above papers incorporate simulations, but **Binette & Steorts (2021)** and **Sharifi Far et al. (2020)** are most explicit about building simulation frameworks that could be adapted for your strawman and compound scenarios.

---

5. Additional MSE-Centric Papers on Human Trafficking and Policy Use

- **Banks & Mokel (2023)** reflect on advances in MSE for human trafficking and emphasize the importance of moving beyond naive assumptions when applying MSE in policy contexts (Banks & Mokel, 2023).

- **Zhang & Larsen (2021)** discuss using MSE for policy advocacy and service design in the context of human trafficking, highlighting its value but also its limitations when data structures evolve over time (Zhang & Larsen, 2021).

Conclusion

Your project is highly original in bridging a methodological gap between **MSE robustness under dynamic interventions** and **causal inference frameworks**. The literature above confirms that individual strands (e.g., MSE reliability, SUTVA violations, quasi-experiments) are well-covered, but their intersection remains largely unexplored—positioning your work for real contribution.

Possible DGP / Models for Simulation Study

You:

I am interested in probability distributions that are over binary vectors – vectors of length K with either zero one as values. Here is the setting I have in mind: There is a 'hidden population', H with size specified, $|H|$. There are K lists, each of which has a marginal probability of including individual $i\in H$, given by $p_k, k\in \{1, \ldots, K\}$. But I also want to model dependence across lists. I want a general, flexible way of modeling this dependence.

ChatGPT:

This is a rich and important modeling problem—effectively you're describing a *dependent Bernoulli vector model* for list inclusion patterns over a hidden population. This is central in *multiple systems estimation (MSE)* or *capture-recapture* settings, where modeling *dependencies across capture lists* is crucial for valid inference on the total population size and substructures.

🧠 Core Problem Setup

You observe or postulate:

- A hidden population H of known or fixed size $N = |H|$

- K *capture lists* (or sources), each individual $i \in H$ has a binary inclusion pattern:

$$\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iK}) \in \{0,1\}^K$$

- Marginal inclusion probabilities:

  $$\mathbb{P}(Z_{ik} = 1) = p_k$$

- But you *also want to model dependence* among the entries of $\mathbf{Z}_i$

🎯 Modeling Goal

Define a **flexible probability distribution over binary vectors of length K** that:

- Allows **marginal control** over p_k
- Models **dependence structure** among components $Z_{i1}, \dots, Z_{iK}$

📚 Canonical and Flexible Approaches

1. Log-linear Models (Frequentist & Bayesian)

Widely used in classical capture-recapture literature.

Let:

$$\mathbb{P}(\mathbf{Z}_i = \mathbf{z}) \propto \exp\left(\sum_{s \subseteq \{1,\dots,K\}} \theta_s \prod_{k \in s} z_k \right)$$

- $\theta_{\{k\}}$ controls marginal probability
- $\theta_{\{k,\ell\}}$ controls pairwise interaction
- Higher-order terms for triple-wise or more interactions

Bayesian form: Put priors on the θ_s, e.g. Gaussian priors for shrinkage (see Madigan and York, 1997).

Limitation: Parameter count grows exponentially in K; not ideal if $K \gtrsim 5$.

---

2. Multivariate Probit or Logistic Models

Model latent continuous variables $\mathbf{X}_i \sim \mathcal{N}(\mu, \Sigma)$ and define:

$$Z_{ik} = \mathbb{I}(X_{ik} > 0)$$

This induces **correlated binary vectors** with flexible dependence (cf. Emrich & Piedmonte, 1991).

- Fit via MCMC using data augmentation (Albert & Chib 1993).
- **Priors** on μ_k, Σ (e.g., inverse-Wishart or LKJ for correlation matrix)

---

3. Gaussian Copula Models for Binary Data

Specify marginal probabilities p_k, then use a Gaussian copula:

$$Z_{ik} = \mathbb{I}(X_{ik} \leq \Phi^{-1}(p_k)), \quad \mathbf{X}_i \sim \mathcal{N}(0, \Sigma)$$

Here, dependence is through Σ, and margins are preserved.

- Copula techniques handle large K
- See Hoff (2007) *Extending the rank likelihood for semiparametric copula estimation*

---

4. Latent Class Models (Finite Mixtures)

Let:

$$\mathbb{P}(\mathbf{Z}_i = \mathbf{z}) = \sum_{c=1}^C \pi_c \prod_{k=1}^K \text{Bern}(z_k \mid \theta_{kc})$$

Where each latent class c has its own vector of Bernoulli parameters.

- Can be extended to **Dirichlet Process mixtures** for non-parametric flexibility
- Rich dependence captured through latent heterogeneity
- Reference: Dunson & Xing (2009) *Nonparametric Bayes modeling of multivariate categorical data*

---

5. Restricted Boltzmann Machines (RBMs) or Energy-based Models

Use a latent binary variable model with energy function:

$$\mathbb{P}(\mathbf{Z}_i) \propto \sum_{\mathbf{h}} \exp\left(\mathbf{Z}_i^\top W \mathbf{h} + b^\top \mathbf{Z}_i + c^\top \mathbf{h} \right)$$

Can approximate complex dependence structures efficiently; modern versions are used in deep generative modeling. But harder to interpret and estimate in a Bayesian setting without variational methods.

🔧 Recommendation

For **Bayesian modeling of dependence across lists in hidden populations**, I recommend:

➤ A latent Gaussian copula or finite mixture of product Bernoullis:

These allow control over marginal probabilities, flexible dependence, and can be adapted into hierarchical Bayesian models over the hidden population.

📝 Actionable Next Step:

Would you prefer a **conjugate, computationally easy approach** (like log-linear or latent class), or are you open to **non-conjugate but flexible models** like Gaussian copula or DP mixtures? I can sketch a generative model + prior structure tailored to your preference.

Let me know your preferred modeling axis: interpretability, computational ease, or nonparametric flexibility?

You:

Please provide complete references of all sources cited. Can we explore possible exchangeable models that might not suffer from the 'combinatorics explosion' of considering all subsets $s \subseteq \{1,\ldots, K\}$?

ChatGPT:

Excellent direction. Exchangeable models provide a way to avoid the exponential complexity of full log-linear models, especially appealing when we want to model dependencies over binary vectors with a growing number of dimensions K.

First, here are the **complete references** for all sources cited in the previous message:

Exchangeable Models to Avoid Combinatorial Explosion

When we assume **exchangeability** of the binary vectors $\mathbf{Z}_i \in \{0,1\}^K$, we can drastically reduce model complexity by relying on **projective or partially exchangeable structures**.

Here are several promising Bayesian approaches:

---

1. de Finetti-Type Mixture Models

For **fully exchangeable binary vectors** (i.e., across indices k), de Finetti's theorem implies the joint distribution is a mixture of i.i.d. Bernoulli sequences:

$$\mathbf{Z}_i \sim \text{i.i.d. Bernoulli}(\theta), \quad \theta \sim \pi(\theta)$$

- This can be extended: each individual i could have a vector $\mathbf{Z}_i \sim \text{Ber}(\theta_i)$, with $\theta_i \sim G$, and $G \sim \text{DP}(\alpha, G_0)$
- Nonparametric, exchangeable across i, and parsimonious across k

But this is too restrictive for marginal control at each k or interactions.

---

2. Exchangeable Binary Arrays with Finite Sufficient Statistics

Inspired by work on binary exchangeable sequences:

$$\mathbb{P}(\mathbf{Z}_i = \mathbf{z}) = f\left(\sum_k z_k \right)$$

Only depends on the **number of ones**, i.e., the Hamming weight.

Generalization: Use **weighted sums** or **summary statistics** such as:

$$\mathbb{P}(\mathbf{Z}_i = \mathbf{z}) = f\left(\sum_k \alpha_k z_k + \sum_{k<\ell} \beta_{k\ell} z_k z_\ell \right)$$

Still avoids full 2^K parameter space.

See: Diaconis & Freedman (1980).
Finite exchangeable sequences.
Annals of Probability, **8**(4), 745–764.
https://projecteuclid.org/euclid.aop/1176994719

---

3. Pólya-Gamma Augmented Multivariate Probit

Allows an **exchangeable prior on correlation** among latent dimensions:

- Let $\mathbf{X}_i \sim \mathcal{N}(0, \rho \mathbf{1}\mathbf{1}^\top + (1 - \rho) \mathbf{I})$
- $$Z_{ik} = \mathbb{I}(X_{ik} > 0)$$

This imposes **exchangeability across the K dimensions** with a single parameter ρ.

A similar trick is used in:

Ghosh, J., & Dunson, D. B. (2009).
Default prior distributions and efficient posterior computation in Bayesian factor analysis.
Journal of Computational and Graphical Statistics, **18**(2), 306–320.
https://doi.org/10.1198/jcgs.2009.07145

---

4. Indian Buffet Process (IBP) for Binary Matrix Factorization

An exchangeable nonparametric prior over binary matrices $Z \in \{0,1\}^{N \times K}$:

- Each row $\mathbf{Z}_i$ encodes features (lists) possessed by individual i
- Defines a distribution over potentially infinite latent binary vectors
- Can be interpreted as assigning individuals to latent “factors” with inclusion patterns

See:

Griffiths, T. L., & Ghahramani, Z. (2011).
The Indian Buffet Process: An Introduction and Review.
JMLR, **12**, 1185–1224.
https://jmlr.org/papers/v12/griffiths11a.html

✅ Suggested Next Step

Would you like to **build a Bayesian exchangeable model** for $\mathbf{Z}_i \in \{0,1\}^K$ based on:

1. A **Pólya-Gamma augmented probit model** with exchangeable correlation, or

2. A **nonparametric latent feature model** using the Indian Buffet Process?

Each balances **complexity, flexibility, and exchangeability** differently. I can sketch the generative model, priors, and possible MCMC steps. Which direction would you prefer to explore first?

You:

Thank you. I think there are some errors in some of your notation. Also, please tell me more about latent class models that would be useful for this task.

ChatGPT:

You're absolutely right to flag notation precision—thank you. Let's briefly correct the notational slips first, then I'll dive into **latent class models** tailored for your problem.

🔍 Notational Clarifications

You are modeling binary vectors $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iK}) \in \{0,1\}^K$, where each coordinate corresponds to individual i's inclusion in list k. Your goal is to model the joint distribution:

$$\mathbb{P}(\mathbf{Z}_i = \mathbf{z}) \text{ over } \{0,1\}^K$$

In a model with dependence, the notation:

$$\mathbb{P}(\mathbf{Z}_i = \mathbf{z}) \propto \exp\left(\sum_{s \subseteq \{1,\dots,K\}} \theta_s \prod_{k \in s} z_k \right)$$

is correct as a **log-linear model** over binary vectors. However, for larger K, summing over all subsets s is computationally prohibitive due to 2^K parameters.

🎯 Latent Class Models (Finite Mixtures)

Latent class models are **dimension-reducing** approaches that preserve **marginal control** and **introduce dependence** through shared unobserved structure.

💡 Basic Model

Let each individual belong to one of C latent classes:

- Latent class membership: $C_i \sim \text{Categorical}(\pi_1, \dots, \pi_C)$
- Conditional on class c, the list indicators are **conditionally independent**:

  $$Z_{ik} \mid C_i = c \sim \text{Bernoulli}(\theta_{kc}), \quad \text{independently for } k=1,\dots,K$$

Then the marginal distribution of $\mathbf{Z}_i$ is:

$$\mathbb{P}(\mathbf{Z}_i = \mathbf{z}) = \sum_{c=1}^C \pi_c \prod_{k=1}^K \theta_{kc}^{z_k} (1 - \theta_{kc})^{1 - z_k}$$

🔧 Bayesian Specification

- $$\boldsymbol{\pi} \sim \text{Dirichlet}(\alpha_1, \dots, \alpha_C)$$
- $$\theta_{kc} \sim \text{Beta}(a,b)$$

Posterior inference via Gibbs sampling or variational Bayes is well-studied.

🧠 Interpretation

- Each class encodes a **canonical capture pattern** with different inclusion probabilities
- Dependencies among $Z_{i1}, \dots, Z_{iK}$ arise because marginalizing over C_i induces correlation

---

🔄 Extensions

1. Dirichlet Process Mixture (Infinite Latent Classes)

Use a Dirichlet Process prior on the latent classes:

$$G \sim \text{DP}(\alpha, G_0), \quad \theta^{(i)} \sim G, \quad \mathbf{Z}_i \mid \theta^{(i)} \sim \prod_k \text{Bern}(\theta^{(i)}_k)$$

This allows *data-driven discovery* of the number of latent classes.

Reference:

> Dunson, D. B., & Xing, C. (2009). *Nonparametric Bayes modeling of multivariate categorical data.*
> *Journal of the American Statistical Association*, **104**(487), 1042–1051.
> https://doi.org/10.1198/jasa.2009.0028

---

2. Dependent Latent Class Models

If the marginal probabilities p_k must be matched (or made interpretable), one can **reparameterize** θ_{kc} to ensure this:

$$\sum_c \pi_c \theta_{kc} = p_k$$

This requires a **constrained prior** or **post-hoc adjustment**.

---

3. Sparse or Structured Mixture Priors

To reduce overfitting or improve interpretability:

- Use **sparse Dirichlet priors** on π to favor fewer classes
- **Hierarchical priors** on $\theta_{kc} \sim \text{Beta}(\mu_k \phi, (1 - \mu_k)\phi)$, with $\mu_k \sim \text{Beta}(a_0, b_0)$

This leads to sharing of information across classes.

✅ Advantages for Your Setting

- Captures *complex dependencies* with relatively few parameters
- Easily adapted to **varying K** and interpretable via latent class summaries
- Works well with *partially observed* $\mathbf{Z}_i$ patterns (e.g., some lists missing)

🔬 Suggested Next Step

Would you like to **construct a Dirichlet Process latent class model**, or would a **finite C-class model with hierarchical priors on θ_{kc}** be a better starting point? I can write out the full generative model and outline inference steps accordingly.

You:

But how would you determine the number of latent classes, C? This seems arbitrary to me. Isn't theree a more principled and canonical way to model the kind of 'list overlap' we have been discussing?

ChatGPT:

You're absolutely correct—specifying the number of latent classes C a priori is **arbitrary** and often **statistically unsatisfying**, especially when we aim for a flexible yet *principled* approach to capturing list overlap patterns (i.e., dependencies among binary indicators Z_{ik}).

📌 Canonical Bayesian Nonparametric Alternative: Dirichlet Process Mixture Models

Generative Model:

$$\begin{aligned} \mathbf{Z}_i \mid \theta_i &\sim \prod_{k=1}^K \text{Bernoulli}(\theta_{ik}) \\ \theta_i &\sim G \\ G &\sim \text{DP}(\alpha, G_0) \\ \theta_{ik} &\sim G_0 = \text{Beta}(a, b) \end{aligned}$$

Here:

- You do **not** fix the number of components.
- The DP prior clusters individuals with **similar list-inclusion propensities**.
- The model induces dependence across list indicators **marginally** via the mixture over θ_i's.

This is **fully nonparametric**, exchangeable across i, and computationally feasible via standard **Chinese Restaurant Process** (CRP) Gibbs sampling or **stick-breaking constructions** (Sethuraman, 1994).

📌 Canonical Model for *List Overlap Structure*: Indian Buffet Process (IBP)

A different but highly canonical prior over **binary matrices**:

$$Z \in \{0,1\}^{N \times K}, \quad Z_{ik} = \text{Indicator of list inclusion}$$

The **IBP** (Griffiths & Ghahramani, 2005/2011) defines a distribution over binary matrices with an unbounded number of columns (lists or latent features):

Generative Process:

For each individual i:

- For existing features $k = 1, \ldots, K^+$, include feature k with probability m_k / i
- Draw $\text{Poisson}(\alpha / i)$ new features

The resulting Z matrix is **exchangeable over rows** (individuals) and allows **latent structure with sharing**.

Adapted to your context, the IBP could model *latent causes of list inclusion*, where each latent factor influences **multiple list inclusions** jointly.

IBP Extensions:

- **Linear Gaussian or logistic likelihoods**
- **Dependent IBP** (e.g., correlated features across lists)

🔄 Comparison

Model	Fixed C?	Exchangeable?	Captures Dependence?	Computational Cost
Finite Latent Class	✅	✅ (across i)	✅ (via C_i)	Moderate
DP Mixture of Bernoullis	❌	✅	✅	Moderate
IBP	❌	✅	✅ (via shared features)	High

🧪 Suggestion

To avoid the arbitrariness of selecting C, I recommend:

1. **Dirichlet Process Mixture of Product Bernoullis** for direct modeling of inclusion probabilities.
2. **Indian Buffet Process** if you want a latent *binary feature* model that allows overlap, structure, and partial exchangeability.

Would you like a **concrete implementation plan** (generative model + priors + inference strategy) for either of these? I can prepare one or both depending on your research goals (e.g., full population modeling, partial observability, etc.).

You:

But how would dependence ('correlation') across lists be modeled?

ChatGPT:

Excellent question—this gets at the heart of the matter: **how does a latent class or nonparametric model induce statistical dependence among binary list-inclusion indicators** $Z_{i1}, \ldots, Z_{iK}$? That is, **how do we get correlation across lists** (columns of the Z matrix), despite conditionally independent modeling?

Let's walk through this carefully.

🧠 Key Idea: Dependence via Latent Structure

In both latent class models and their nonparametric extensions, the joint dependence structure among Z_{ik} is **induced by marginalizing over latent variables**.

General Form:

Suppose:

$$\begin{aligned} \mathbf{Z}_i \mid \theta_i &\sim \prod_{k=1}^K \text{Bernoulli}(\theta_{ik}) \\ \theta_i &\sim G \end{aligned}$$

Then:

- **Conditional on θ_i**: Z_{ik} are independent.
- **Marginally**: Z_{ik} and $Z_{i\ell}$ are *dependent*, due to shared θ_i.

This is a form of **explaining away**—typical in Bayesian hierarchical models.

🔄 Concrete Case 1: Dirichlet Process Mixture of Product Bernoullis

Each individual has:

$$\mathbf{Z}_i \mid \theta^{(c_i)} \sim \prod_k \text{Bernoulli}(\theta^{(c_i)}_k), \quad \theta^{(c)}$$

$$\sim \text{Beta}(a,b)^K$$

where:

- $c_i \sim \text{CRP}(\alpha)$: individuals are clustered into classes
- **Within a class**, list inclusions are independent
- **Across classes**, different patterns occur

➤ Marginal Covariance

Let's define:

$$\mathbb{E}[Z_{ik} Z_{i\ell}] = \sum_{c} \pi_c \theta_{kc} \theta_{\ell c}, \quad \mathbb{E}[Z_{ik}] = \sum_c \pi_c \theta_{kc}$$

Then:

$$\text{Cov}(Z_{ik}, Z_{i\ell}) = \sum_c \pi_c \theta_{kc} \theta_{\ell c} - \left(\sum_c \pi_c \theta_{kc} \right)\left(\sum_c \pi_c \theta_{\ell c} \right)$$

This is **non-zero** whenever θ_{kc} and $\theta_{\ell c}$ co-vary across latent classes c. Thus, even though within each class the components are independent, **marginal dependence arises**.

Concrete Case 2: Indian Buffet Process (IBP)

Let:

$$Z_{ik} = \mathbb{I}(\sum_{j=1}^\infty f_{kj} h_{ij} > \epsilon)$$

Where:

- $h_{ij} \in \{0,1\}$: latent binary features (individual i's usage of feature j)
- $f_{kj} \in \mathbb{R}$: loading of feature j on list k

This induces:

- **Shared latent causes** for multiple lists
- Lists are dependent to the extent they are influenced by the *same latent features*

See:

> Knowles, D. A., & Ghahramani, Z. (2011). *Nonparametric Bayesian sparse factor models with application to gene expression modeling*.
> *Annals of Applied Statistics*, **5**(2A), 1534–1552.
> https://doi.org/10.1214/10-AOAS397

Reference:

> Griffiths, T. L., & Ghahramani, Z. (2011). *The Indian Buffet Process: An Introduction and Review.*

Journal of Machine Learning Research, **12**, 1185–1224.
https://jmlr.org/papers/v12/griffiths11a.html

🔗 Analogy to Random Effects Models

In both cases, you can view the latent structure as **random effects** shared across lists. For instance:

- In DP mixtures, each latent class acts like a random intercept
- In IBP, latent binary features are shared *random effect groupings*

This induces **positive or negative dependence** among list inclusion probabilities.

🧮 Summary of How Dependence is Induced

Model Type	Latent Structure	How Dependence is Induced
Latent Class (Fixed C)	Discrete class label C_i	Marginalizing over classes correlates lists
DP Mixture	Infinite latent clusters	Shared θ_i across lists induces correlation
IBP	Binary latent features	Lists influenced by overlapping features

---

📚 Complete References

1. **Madigan, D., & York, J. (1997)**.
   Bayesian methods for estimation of the size of a closed population.
   Biometrika, **84**(1), 19–31.
   https://doi.org/10.1093/biomet/84.1.19

2. **Emrich, L. J., & Piedmonte, M. R. (1991)**.
   A method for generating high-dimensional multivariate binary variates.
   The American Statistician, **45**(4), 302–304.
   https://doi.org/10.1080/00031305.1991.10475864

3. **Albert, J. H., & Chib, S. (1993)**.
   Bayesian analysis of binary and polychotomous response data.
   Journal of the American Statistical Association, **88**(422), 669–679.
   https://doi.org/10.2307/2290350

4. **Hoff, P. D. (2007)**.
   Extending the rank likelihood for semiparametric copula estimation.
   Annals of Applied Statistics, **1**(1), 265–283.
   https://projecteuclid.org/euclid.aoas/1183143727

5. **Dunson, D. B., & Xing, C. (2009)**.
   Nonparametric Bayes modeling of multivariate categorical data.
   Journal of the American Statistical Association, **104**(487), 1042–1051.
   https://doi.org/10.1198/jasa.2009.0028

This marginal dependence through latent variables is well understood in:

- **Dunson & Xing (2009)**: Nonparametric Bayes modeling of multivariate categorical data
- **Fox & Dunson (2011)**: *Bayesian nonparametric modeling of multivariate categorical data*
  Journal of the American Statistical Association, **106**(495), 1046–1060.

See:

Knowles, D. A., & Ghahramani, Z. (2011). *Nonparametric Bayesian sparse factor models with application to gene expression modeling*.
Annals of Applied Statistics, **5**(2A), 1534–1552.
https://doi.org/10.1214/10-AOAS397

Reference:

Griffiths, T. L., & Ghahramani, Z. (2011). *The Indian Buffet Process: An Introduction and Review.*
Journal of Machine Learning Research, **12**, 1185–1224.
https://jmlr.org/papers/v12/griffiths11a.html