SPARSE THREE-PARAMETER RESTRICTED INDIAN BUFFET PROCESS FOR UNDERSTANDING INTERNATIONAL TRADE

Melanie F. Pradier[1]
Fernando Perez-Cruz[2]

[1]Harvard University
[2]Swiss Data Science Center
melanie@seas.harvard.edu

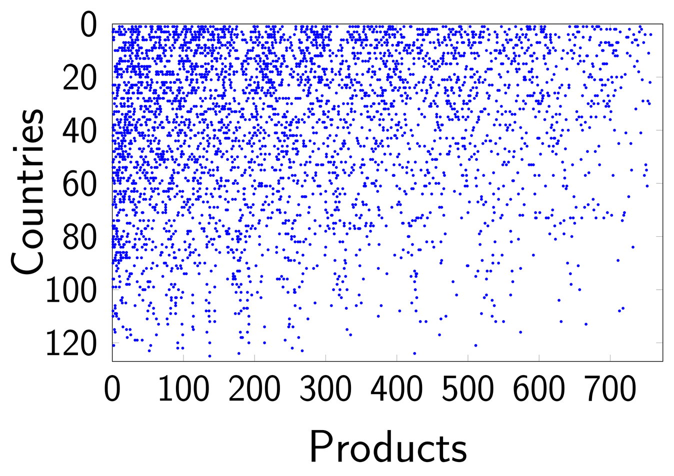
Viktor Stojkoski[3]
Zoran Utkovski[3,4]
Ljupco Kocarev[3,5]

[3]Macedonian Academy of Sciences and Arts
[4]University Goce Delcev in Stip
[5]University of California in San Diego

INTRODUCTION

- **Aim**: Explore high-dimensional count data.
  - a) Increase model interpretability.
  - b) Find structured solutions in latent space.
- **Contribution**: A Bayesian non-parametric Poisson factorization model that gives easy-to-interpret and structured solutions.
- **Key Idea**: Force sparsity in the features and improve prior flexibility to be consistent with reality, by combining the stable-beta process with the restricted Indian Buffet Process.

Motivation: Why some countries are wealthier than others?
Theory of Economic Complexity: Capabilities are "intangible assets which drive the development, wealth and competitiveness of a country" (Cristelli et.al, 2013).



- Triangular structure
- Diversified countries producing exclusive products
- Non-diversified countries producing standard products

THEORETICAL BACKGROUND

Indian-Buffet Process (Ghahramani et.al, 2006)

- Stochastic process defining a probability distribution over equivalent classes of binary matrices. We denote: $\mathbf{Z} \sim \mathrm{IBP}(\alpha)$.
- It corresponds to the limit when $K \to \infty$ of parametric model:

$$\pi_k \sim \mathrm{Beta}(\alpha/K, 1), \quad z_{nk} \sim \mathrm{Bernoulli}(\pi_k) \tag{1}$$

- It can also be constructed based on its underlying De Finetti's representation, i.e., as a mixture of Bernoulli processes directed by a beta process:

$$\mu \sim \mathrm{BP}(1, \alpha, H) \tag{2}$$
$$\mathbf{Z}_{n\cdot} \sim \mathrm{BeP}(\mu) \tag{3}$$

where $\mu = \sum_k \pi_k \delta_{\theta_k}$ is the directing measure, and H is the probability base measure (Thibaux et.al, 2007).

- Disadvantage: Mass parameter α couples both a priori number of ones per row J_n and total number of active features K^+.

$$J_n \sim \mathrm{Poisson}(\alpha) \tag{4}$$
$$K^+ \sim \mathrm{Poisson}\Big(\alpha \sum_{n=1}^{N} \Big(\frac{1}{n}\Big)\Big) \tag{5}$$

Three-parameter IBP (Teh et.al, 2009)

- More flexible distribution for stick weights (power-law behaviors).
- In the De Finetti's representation, it uses a Stable-beta process (SBP).
- Culinary Metaphor:
  - Customer 1 tries $\mathrm{Poisson}(\alpha)$ dishes.
  - Customer n tries:

$$p\left(z_{nk} = 1 | \mathbf{Z}_{\neg n}\right) = \frac{m_k - \sigma}{n + c - 1} \tag{6}$$
$$p\left(J_{new}\right) \sim \mathrm{Poisson}\left(\alpha \frac{\Gamma(1+c)\Gamma(n+c+\sigma-1)}{\Gamma(n+c)\Gamma(c+\sigma)}\right) \tag{7}$$

- Disadvantage: Number of ones per row J_n still Poisson-distributed.

Restricted IBP (Doshi-Velez et.al, 2015)

- Non-exchangeable, with arbitrary marginal prior f over J_n
- In the De Finetti's representation, it uses *restricted* Bernoulli processes:

$$\mathrm{R\text{-}BeP}(\mathbf{Z}_{n\cdot}; \mu, f) = f\left(J_n\right) \cdot \frac{\prod_{k=1}^{\infty} \pi_k^{z_{nk}} (1 - \pi_k^{1-z_{nk}}) \mathbb{1}(\sum_K z_{nk} = J_n)}{\sum_{z' \in \mathcal{Z}} \prod_k \pi_k^{z'_k} (1 - \pi_k)^{(1-z'_k)} \mathbb{1}(\sum_K z'_k = J_n)} \tag{8}$$

- Disadvantage: Stick weights cannot follow power-law behaviors.

SPARSE 3-PARAMETER RESTRICTED IBP (S3R-IBP)

- Combine strengths of three-parameter IBP and restricted IBP:

$$\mu \sim \mathrm{SBP}(1, \alpha, H) \tag{9}$$
$$\mathbf{Z}_{n\cdot} \sim \mathrm{R\text{-}BeP}(\mu, f) \tag{10}$$

We denote this flexible prior as $\mathbf{Z} \sim \mathrm{S3R\text{-}IBP}(\alpha, c, \sigma, f)$.

- Let $\mathbf{X} \in \mathbb{N}^{N \times D}$, N samples, and D dimensions.
- We build a structured infinite latent feature model for count data:

$$x_{nd} \sim \mathrm{Poisson}(\mathbf{Z}_n \mathbf{B}_{\cdot d}), \tag{11}$$
$$B_{kd} \sim \mathrm{Gamma}(\alpha_B, \frac{\mu_B}{\alpha_B}), \tag{12}$$
$$\mathbf{Z} \sim \mathrm{3R\text{-}IBP}(\alpha, c, \sigma, f) \tag{13}$$

where α_B and μ_B are the shape and mean of the prior Gamma distribution.

- Available parameters:
  - mass parameter α
  - stability component $\sigma \in [0, 1)$
  - concentration parameter $c > -\sigma$
  - marginal prior f for J_n
- Features are made sparse by choosing $\alpha_B < 1$.

Inference Scheme

- Model conditionally conjugate: auxiliary variables $x'_{nd,1}, \ldots, x'_{nd,K}$ such that $x_{nd} = \sum_{k=1}^{K} x'_{nd,k}$, and $x'_{nd,k} \sim \mathrm{Poisson}(Z_{nk} B_{kd})$
- For each iteration, do:
  1. Sample each element of matrix $\mathbf{Z}$ using inclusion probabilities (Aires, 1999).
  2. Sample latent measure $\boldsymbol{\pi}$ using Metropolis-Hasting within Gibbs (Doshi-Velez et.al, 2015).
     Let us call $D^K_{J_n}$ the denominator in eq. 8: This value can be computed easily using a dynamic programming approach, since:
     $$D^K_{J_n} = (1 - \pi_K) D^{K-1}_{J_n} + \pi_K D^{K-1}_{J_n - 1}. \tag{14}$$
  3. Sample each element of $\mathbf{B}$ and $\mathbf{X}'$ from their conditional distributions.
  4. Sample hyperparameter α according to (Archambeau, 2015).

RESULTS

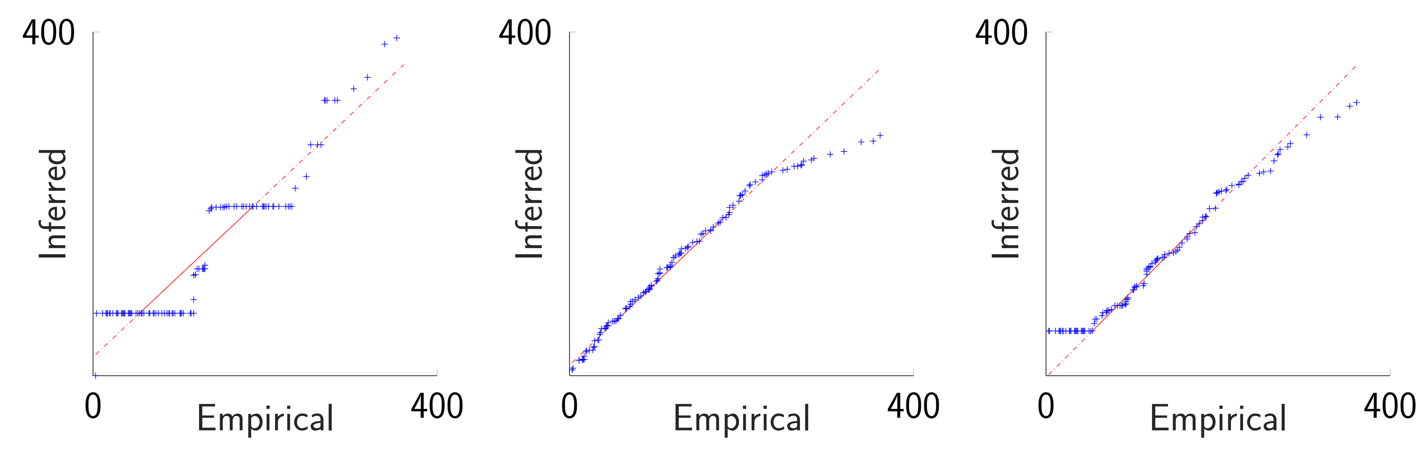
Metric	MF	NMF	IBP	S-IBP	S3R-IBP
Log Perplexity	1.68 ± 0.01	1.61 ± 0.01	$\mathbf{1.59 \pm 0.04}$	3.26 ± 0.17	1.62 ± 0.01
Coherence	-264.60 ± 4.74	-263.27 ± 7.45	-149.36 ± 7.56	-178.44 ± 4.50	$-\mathbf{140.51 \pm 2.73}$

(a) 2010 SITC database ($N = 126$, $D = 744$, 16k non-zero values, 17% sparsity)

Metric	MF	NMF	IBP	S-IBP	S3R-IBP
Log Perplexity	1.48 ± 0.01	$\mathbf{1.47 \pm 0.01}$	1.58 ± 0.01	2.56 ± 0.12	1.57 ± 0.02
Coherence	-264.73 ± 3.11	-264.67 ± 6.22	-148.91 ± 10.57	-168.39 ± 13.16	$-\mathbf{134.51 \pm 4.43}$

(b) 2010 HS database ($N = 123$, $D = 4890$, 77k non-zero values, 13% sparsity)

Table 1: **Quantitative evaluation of accuracy and interpretability.**



(a) IBP (b) S-IBP (c) S3R-IBP

Capturing sparsity structure. S3R-IBP gives the best fit for the distribution of number of non-zero values per row in **X**.

Id	Products with highest weights
F1	misc. animal oils (0.78), bovine entails (0.72), bovine meat (0.68), milk (0.63), equine (0.62), butter (0.58)
F2	synthetic woven, synth. yarn, woven < 85% synth.
F3	parts metalworking, tool parts, polishing stones
F4	Aldehyde–Ketone, glycosides–vaccines, medicaments
F5	synthetic rubber, acrylic polymers, silicones
F6	measuring instruments, math inst., electrical inst.
F7	vehicles parts, cars, iron wire
F8	improved wood, mineral wool, heating equipment
F9	elect. machinery, vehicles stereos, data processing eq.
F10	baked goods, metal containers, misc. edibles
F11	misc. articles of iron, carpentry wood, wood articles
F12	vegetables, fruit–vegetable juices, misc. fruit
F13	misc. pumps, ash–residues, chemical wood pulp
F14	synth. undergarments, feminine outerwear, men's shirts
F15	misc. rotating, electric plant parts, control inst. of gas

Table 2: **Features learned by S3R-IBP.**

IBP
confectionary sugar (0.45)
plastic containers (0.43)
baked goods (0.41)
tissue paper (0.40)
metal containers (0.39)
soaps (0.39)
S-IBP
bovine (0.53)
improved wood (0.51)
misc. vegetable oils (0.50)
butter (0.50)
rape seeds (0.47)
misc. wheat (0.45)

Table 3: **Competitors.**

Id	Weight
F14	0.37
F12	0.32
F10	0.17
F2	0.16
F1	0.14
F9	0.13
F13	0.05
F6	0.04
F5	0.04
F4	0.04
F15	0.04
F7	0.03
F8	0.03
F11	0.02
F3	0.02

(a) M-F0

Id	Weight
F8	0.69
F11	0.68
F15	0.60
F10	0.59
F7	0.52
F6	0.34
F13	0.32
F4	0.31
F3	0.31
F5	0.14
F1	0.05
F9	0.02
F2	0.01
F14	0.00
F12	0.00

(b) M-F1

Table 4: **Meta-features.** A sharp division of the world arises.