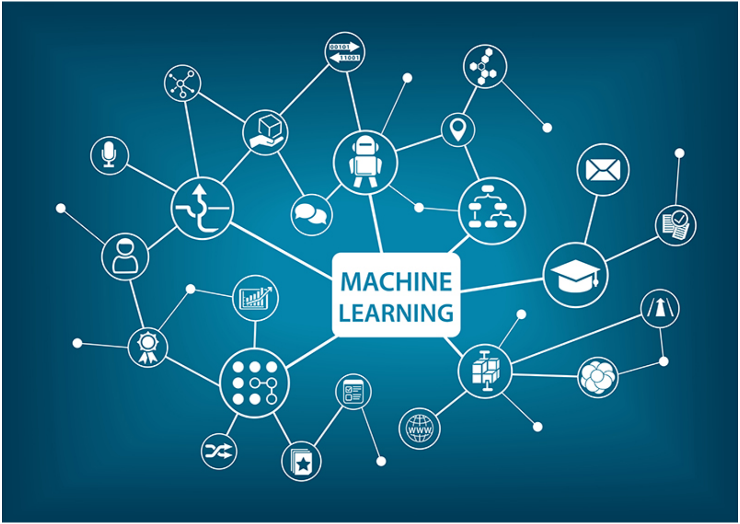
Bayesian Nonparametric Models:
An Application to International Trade

Melanie F. Pradier

Wednesday 13$^{\text{th}}$ September, 2017

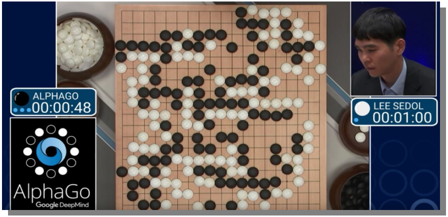
Motivation

Motivation

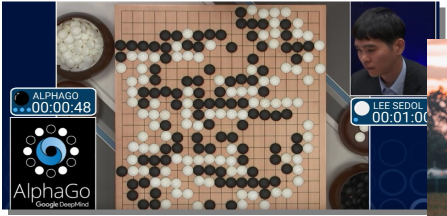


Motivation

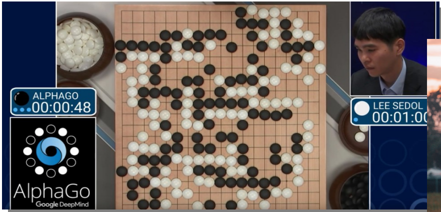
Motivation

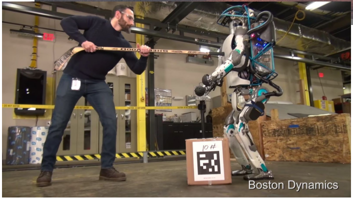


Motivation



Motivation


ALPHAGO
00:00:48
AlphaGo
Google DeepMind
LEE SEDOL


Boston Dynamics

Motivation


ALPHAGO
00:00:48
AlphaGo
Google DeepMind
LEE SEDOL
THINK
$0
KEN
$1,200
WATSON
$0
BRAD
Boston Dynamics

Motivation

Data Exploitation Age

Motivation

Data Exploitation Age

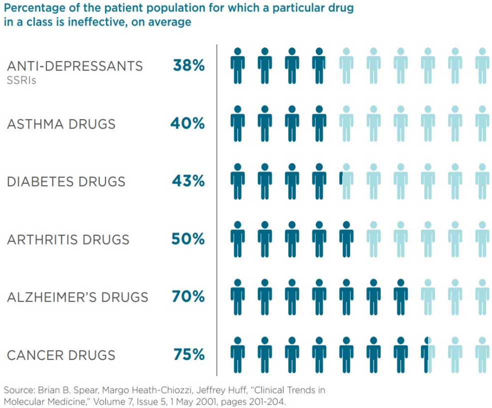
··· but are we making the outmost out of data?

Motivation

An example: personalized medicine

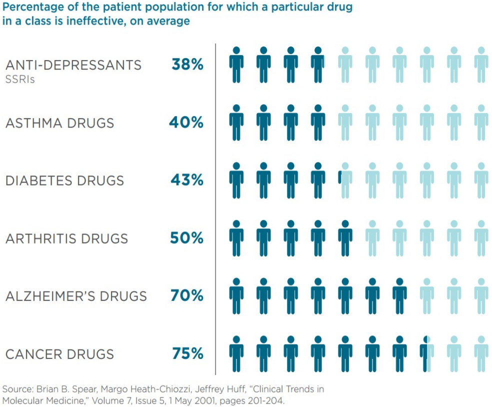
Motivation

An example: personalized medicine



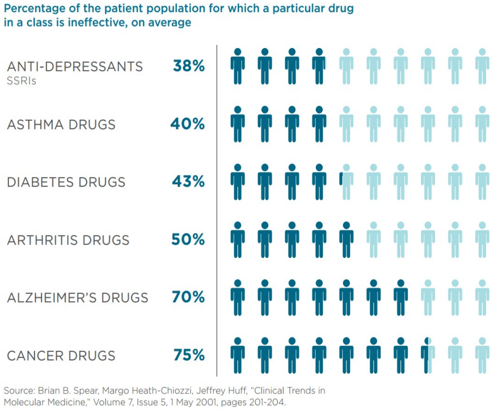
Motivation

An example: personalized medicine



Motivation

An example: personalized medicine

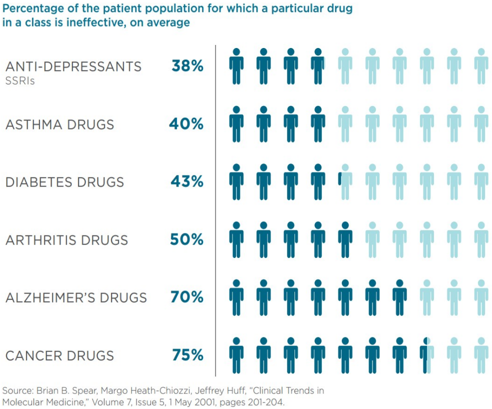


Challenges

- Complexity

Motivation

An example: personalized medicine

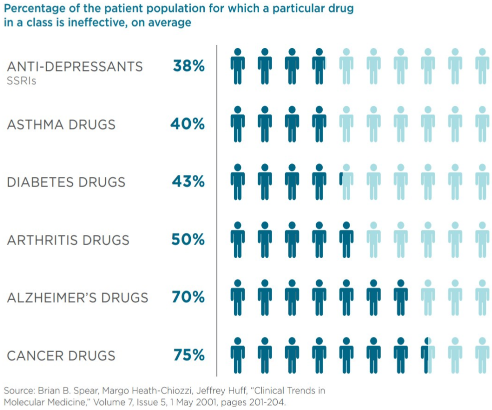


Challenges

- Complexity
- Missing data

Motivation

An example: personalized medicine

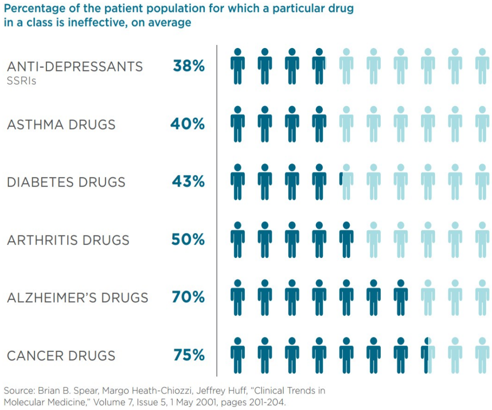


Challenges

- Complexity
- Missing data
- *Small data within big data*
- ...

Motivation

An example: personalized medicine

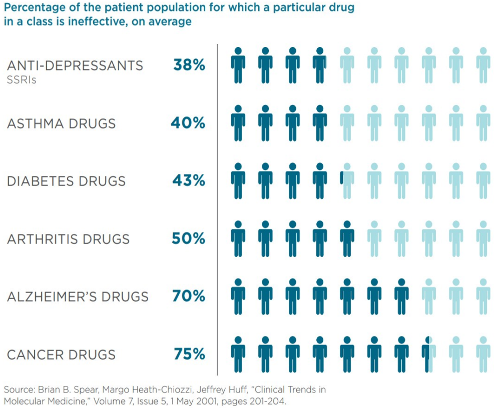


Challenges

- Complexity
- Missing data
- *Small data within big data*
- ...
- Research focus
  $\rightarrow$ data exploration

Motivation

An example: personalized medicine



Challenges

- Complexity
- Missing data
- *Small data within big data*
- ...
- Research focus
  → data exploration
- Interpretability

Motivation

Focus: data exploration

Interpretability

[F. Doshi-Velez, B. Kim, *Towards A Rigorous Science of Interpretable Machine Learning*]

- Understandable for humans (Doshi-Velez, 2017)
- "Right to explanation" (EU General Data Protection Regulation, 2018)

Motivation

Focus: data exploration

Interpretability

[F. Doshi-Velez, B. Kim, *Towards A Rigorous Science of Interpretable Machine Learning*]

- Understandable for humans (Doshi-Velez, 2017)
- "Right to explanation" (EU General Data Protection Regulation, 2018)

Main goal

- Knowledge discovery
- Hypothesis generation

Motivation

Focus: data exploration

Interpretability

[F. Doshi-Velez, B. Kim, *Towards A Rigorous Science of Interpretable Machine Learning*]

- Understandable for humans (Doshi-Velez, 2017)
- "Right to explanation" (EU General Data Protection Regulation, 2018)

Main goal

- Knowledge discovery
- Hypothesis generation

Our Approach

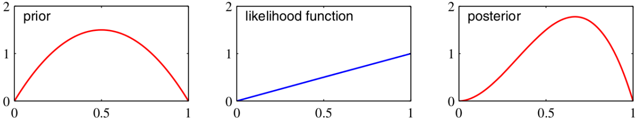
Bayesian nonparametrics

Why Bayesian nonparametrics?

- **Bayesian**: combine prior knowledge with data evidence

Why Bayesian nonparametrics?

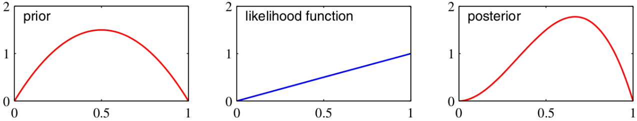
- **Bayesian**: combine prior knowledge with data evidence



[Bishop, 2006]

Why Bayesian nonparametrics?

- **Bayesian**: combine prior knowledge with data evidence

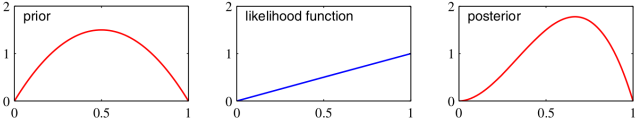


[Bishop, 2006]

- **Nonparametric**
  - actually... really large parametric model

Why Bayesian nonparametrics?

- **Bayesian**: combine prior knowledge with data evidence

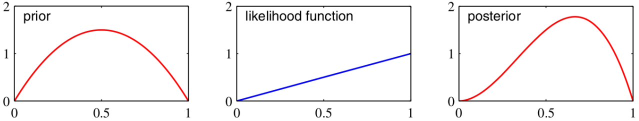


[Bishop, 2006]

- **Nonparametric**
  - actually... really large parametric model
  - number of latent variables grows with data

Why Bayesian nonparametrics?

- **Bayesian**: combine prior knowledge with data evidence



[Bishop, 2006]

In this talk...

- **Nonparametric**
  - actually... really large parametric model
  - number of latent variables grows with data

Outline

1. Introduction
2. Bayesian nonparametrics
3. BNP models for international trade
4. Conclusion

Bayesian nonparametrics (BNPs)

- Bayesian framework for **model selection**
- Nonparametric: number of parameters grows with the amount of data:
  - Prior over **infinite-dimensional** parameter space
  - Only a **finite subset** of parameters is used for any finite dataset

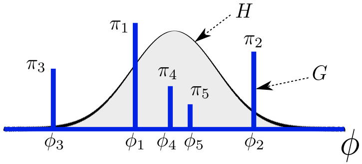
Bayesian nonparametrics (BNPs)

- Bayesian framework for **model selection**
- Nonparametric: number of parameters grows with the amount of data:
  - Prior over **infinite-dimensional** parameter space
  - Only a **finite subset** of parameters is used for any finite dataset

- Rely on stochastic processes:
  - Dirichlet process
  - Beta process
  - Gaussian process
  - ...

Dirichlet process (DP)

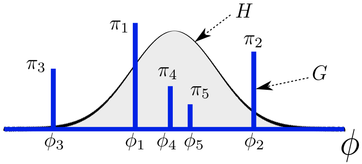
$$G \sim \mathrm{DP}(\alpha, H)$$



$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

Dirichlet process (DP)

$$G \sim \mathrm{DP}(\alpha, H)$$

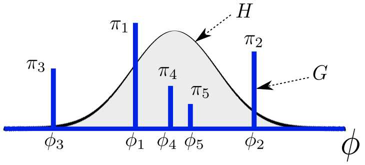


$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

- often used in mixture models

Dirichlet process (DP)

$$G \sim \mathrm{DP}(\alpha, H)$$



$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

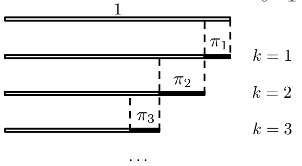
- often used in mixture models

Stick-breaking representation

(Ishwaran et.al, 2001)

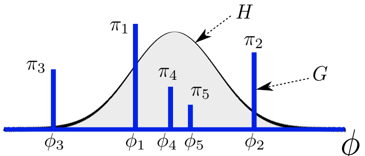
For $k = 1, \cdots, \infty$

$$v_k \sim \mathrm{Beta}(\alpha, 1),\ \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$



Dirichlet process (DP)

$$G \sim \mathrm{DP}(\alpha, H)$$



$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$$

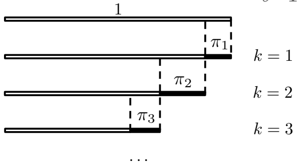
- often used in mixture models

Stick-breaking representation

(Ishwaran et.al, 2001)

For $k = 1, \cdots, \infty$

$$v_k \sim \mathrm{Beta}(\alpha, 1),\ \pi_k = v_k \prod_{\ell=1}^{k-1} (1 - v_\ell)$$



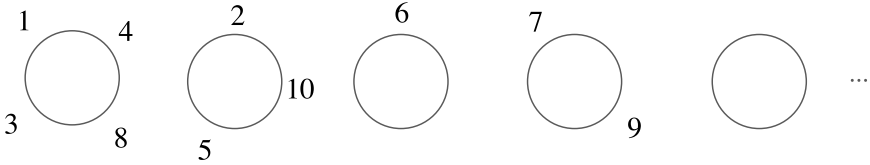
For $k = 1, \cdots, \infty$

$$\phi_k \sim H$$

Chinese restaurant process (CRP)

$$\boldsymbol{c} \sim \mathrm{CRP}(\alpha)$$

where $\boldsymbol{c} \equiv$ infinite sequence of natural numbers.

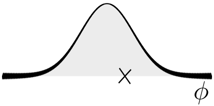


(Pitman et.al, 2002)

$$p\left(c_i = m | \boldsymbol{c}^{\neg i}, \alpha\right) \begin{cases} |m|^{\neg i}, & m \in \boldsymbol{c}^{\neg i} \\ \alpha, & m \notin \boldsymbol{c}^{\neg i} \end{cases}$$

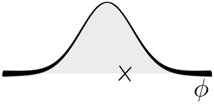
Indian Buffet Process (Ghahramani et.al, 2006)

Indian Buffet Process (Ghahramani et.al, 2006)

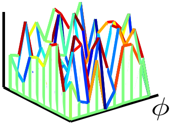


$$x \sim \mathcal{N}(\mu_0, \sigma_0^2) \longrightarrow x = 0.356$$

Indian Buffet Process (Ghahramani et.al, 2006)

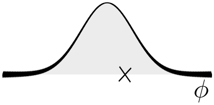


$$\xrightarrow[x \sim \mathcal{N}(\mu_0, \sigma_0^2)]{} \qquad x = 0.356$$

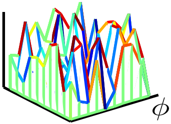


$$\xrightarrow[Z \sim \mathrm{IBP}(\alpha)]{} \qquad Z = \begin{bmatrix} 1 & 0 & 1 & \\ 0 & 1 & 0 & \vdots \\ 1 & 1 & 0 & \\ 1 & 0 & 1 & \end{bmatrix}$$

Indian Buffet Process (Ghahramani et.al, 2006)



$$\xrightarrow[x \sim \mathcal{N}(\mu_0, \sigma_0^2)]{} \quad x = 0.356$$



$$\xrightarrow[Z \sim \mathrm{IBP}(\alpha)]{} \quad Z = \begin{bmatrix} 1 & 0 & 1 & \\ 0 & 1 & 0 & \vdots \\ 1 & 1 & 0 & \\ 1 & 0 & 1 & \end{bmatrix}$$

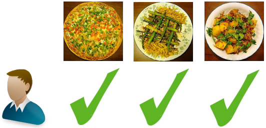
- IBP: distribution over binary matrices $Z_{N \times K}$
- Model chooses number of hidden features, $K \to \infty$

Indian Buffet Process (IBP)

(Slide from F. J.R. Ruiz)

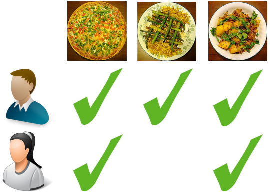
Indian Buffet Process (IBP)

(Slide from F. J.R. Ruiz)



Indian Buffet Process (IBP)

(Slide from F. J.R. Ruiz)



Indian Buffet Process (IBP)

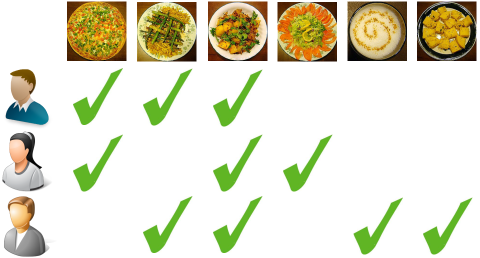
(Slide from F. J.R. Ruiz)

Indian Buffet Process (IBP)

(Slide from F. J.R. Ruiz)

Indian Buffet Process (IBP)

(Slide from F. J.R. Ruiz)



Indian Buffet Process (IBP)

(Slide from F. J.R. Ruiz)

Indian Buffet Process (IBP)

(Slide from F. J.R. Ruiz)

1	1	1	0	0	0
1	0	1	1	0	0
0	1	1	0	1	1

...

⋮

Indian Buffet Process (IBP)

(Slide from F. J.R. Ruiz)

. . .

1	1	0	1	0	1
1	0	1	0	0	1
0	0	1	0	1	1

⋮

Indian buffet process (IBP)

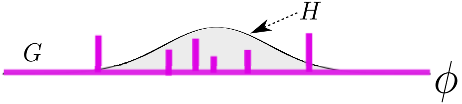
- Prior over binary matrices with infinite number of columns
- Rows $\equiv$ observations; columns $\equiv$ features
- $\mathbf{Z} \sim \mathrm{IBP}(\alpha)$
- α: concentration parameter
- Each element z_{nk} indicates whether the k-th feature contributes to observation n

Indian buffet process (IBP)

An alternative construction

hierarchy of a Beta process (BP) with multiple Bernoulli processes (BeP)

$\Rightarrow$ infinite latent feature model



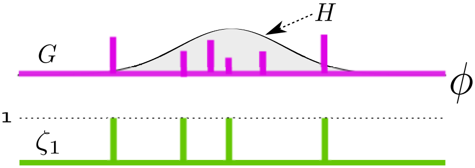
$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \sim \mathrm{BP}(c, \alpha, H)$$

Indian buffet process (IBP)

An alternative construction

hierarchy of a Beta process (BP) with multiple Bernoulli processes (BeP)

$\Rightarrow$ infinite latent feature model



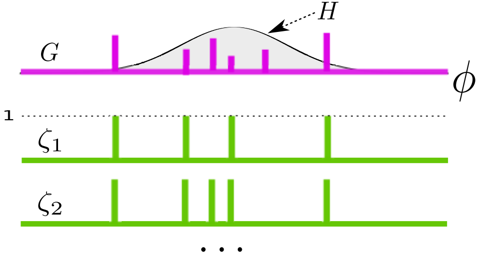
$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \sim \mathrm{BP}(c, \alpha, H)$$

Indian buffet process (IBP)

An alternative construction

hierarchy of a Beta process (BP) with multiple Bernoulli processes (BeP)

$\Rightarrow$ infinite latent feature model



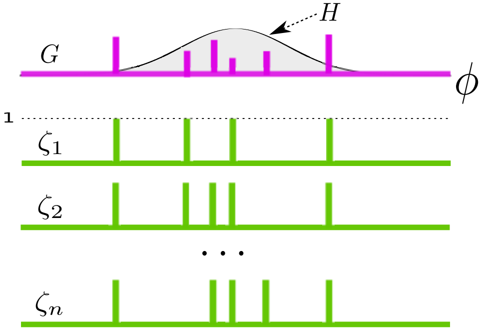
$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \sim \mathrm{BP}(c, \alpha, H)$$

Indian buffet process (IBP)

An alternative construction

hierarchy of a Beta process (BP) with multiple Bernoulli processes (BeP)

$\Rightarrow$ infinite latent feature model



$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \sim \mathrm{BP}(c, \alpha, H)$$

For $n = 1, \cdots, \infty$

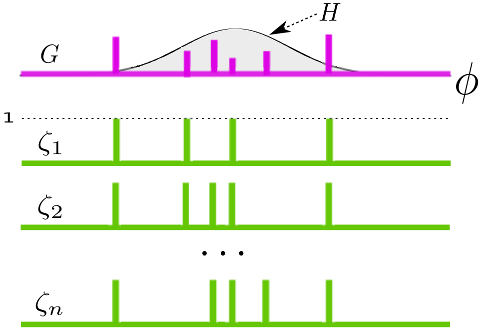
$$\zeta_n = \sum_{k=1}^{\infty} z_{nk} \delta_{\phi_k} \sim \mathrm{BeP}(G)$$

Indian buffet process (IBP)

An alternative construction

hierarchy of a Beta process (BP) with multiple Bernoulli processes (BeP)

$\Rightarrow$ infinite latent feature model



$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k} \sim \mathrm{BP}(c, \alpha, H)$$

For $n = 1, \cdots, \infty$

$$\zeta_n = \sum_{k=1}^{\infty} z_{nk} \delta_{\phi_k} \sim \mathrm{BeP}(G)$$

$$\mathbf{Z} \sim \mathrm{IBP}(\alpha)$$

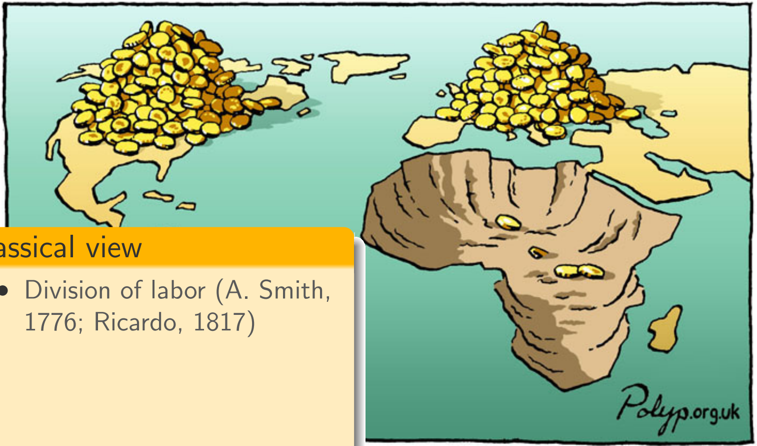
Outline

Motivation: wealth of nations

What makes some countries wealthier than others?

Motivation: wealth of nations

What makes some countries wealthier than others?

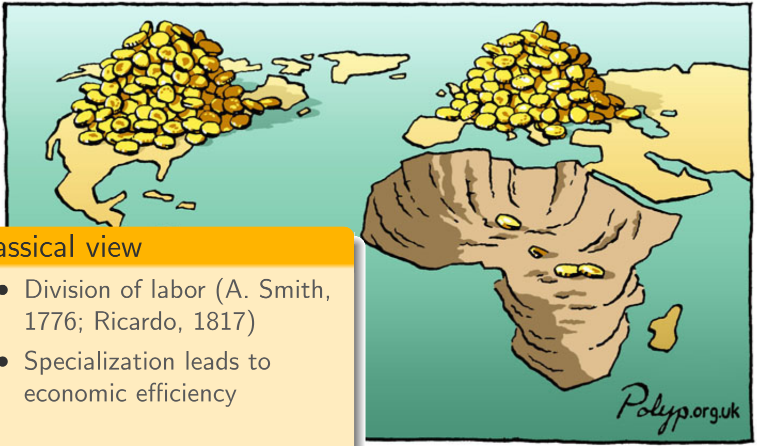


Classical view

- Division of labor (A. Smith, 1776; Ricardo, 1817)

Motivation: wealth of nations

What makes some countries wealthier than others?

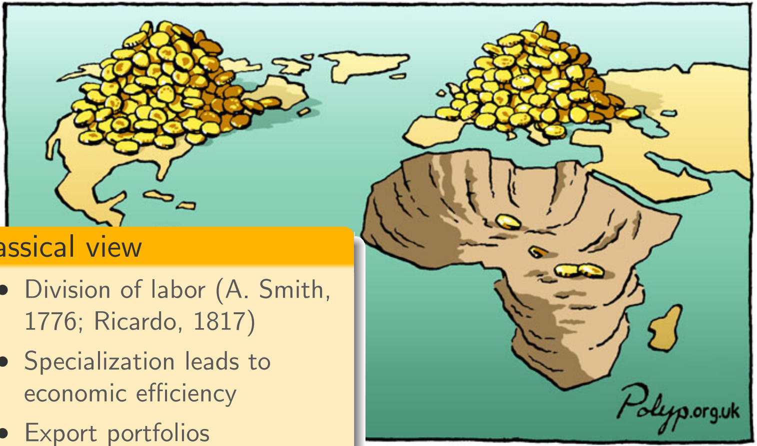


Classical view

- Division of labor (A. Smith, 1776; Ricardo, 1817)
- Specialization leads to economic efficiency

Motivation: wealth of nations

What makes some countries wealthier than others?

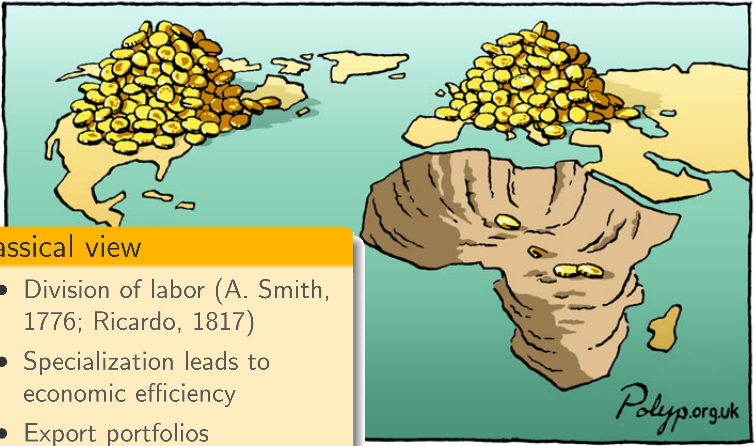


Classical view

- Division of labor (A. Smith, 1776; Ricardo, 1817)
- Specialization leads to economic efficiency
- Export portfolios

Motivation: wealth of nations

What makes some countries wealthier than others?



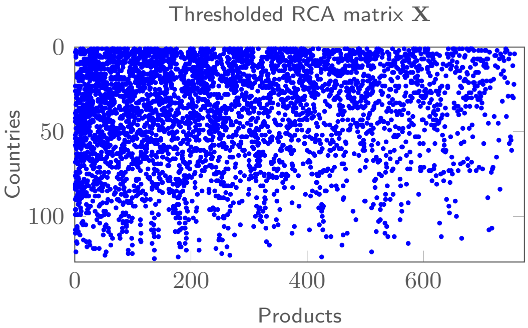
Classical view

- Division of labor (A. Smith, 1776; Ricardo, 1817)
- Specialization leads to economic efficiency
- Export portfolios

  $\rightarrow$ block-structure

Motivation: wealth of nations

The reality:



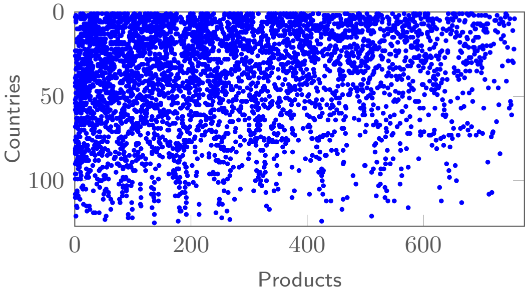
$$\mathrm{RCA}_{nd} = \frac{E_{nd} / \sum_p E_{nd}}{\sum_n E_{nd} / \sum_{n,d} E_{nd}}$$

$$x_{nd} = \begin{cases} 1, & \text{if } \mathrm{RCA}_{nd} \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

Motivation: wealth of nations

The reality:

Thresholded RCA matrix $\mathbf{X}$



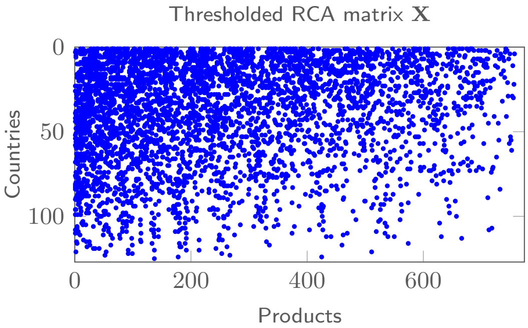
Properties:

$$\mathrm{RCA}_{nd} = \frac{E_{nd} / \sum_p E_{nd}}{\sum_n E_{nd} / \sum_{n,d} E_{nd}}$$

$$x_{nd} = \begin{cases} 1, & \text{if } \mathrm{RCA}_{nd} \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

Motivation: wealth of nations

The reality:



Properties:

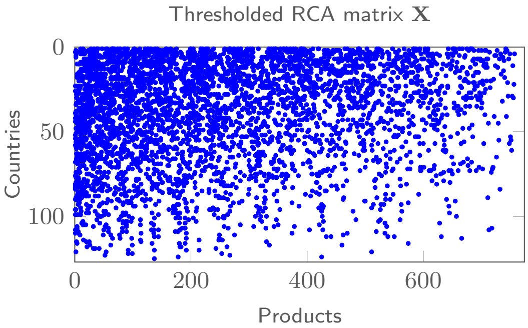
1. Triangularity

$$\mathrm{RCA}_{nd} = \frac{E_{nd}/\sum_p E_{nd}}{\sum_n E_{nd}/\sum_{n,d} E_{nd}}$$

$$x_{nd} = \begin{cases} 1, & \text{if } \mathrm{RCA}_{nd} \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

Motivation: wealth of nations

The reality:



Properties:

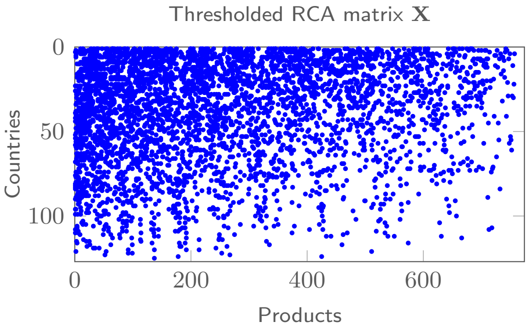
1. Triangularity
2. $D \gg N$

$$\mathrm{RCA}_{nd} = \frac{E_{nd} / \sum_p E_{nd}}{\sum_n E_{nd} / \sum_{n,d} E_{nd}}$$

$$x_{nd} = \begin{cases} 1, & \text{if } \mathrm{RCA}_{nd} \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

Motivation: wealth of nations

The reality:



$$\mathrm{RCA}_{nd} = \frac{E_{nd} / \sum_p E_{nd}}{\sum_n E_{nd} / \sum_{n,d} E_{nd}}$$

$$x_{nd} = \begin{cases} 1, & \text{if } \mathrm{RCA}_{nd} \geq 1 \\ 0, & \text{otherwise} \end{cases}$$

Properties:

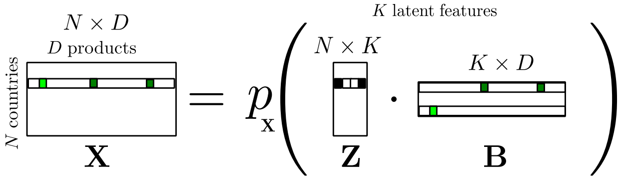
1. Triangularity
2. $D \gg N$

Our Approach

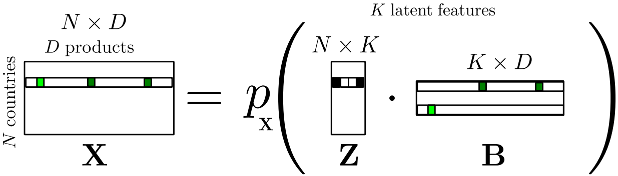
Develop an infinite Poisson factor analysis model $\cdots$

- flexible prior
- feature sparsity

Bernoulli process Poisson factor analysis (BeP-PFA)



Bernoulli process Poisson factor analysis (BeP-PFA)

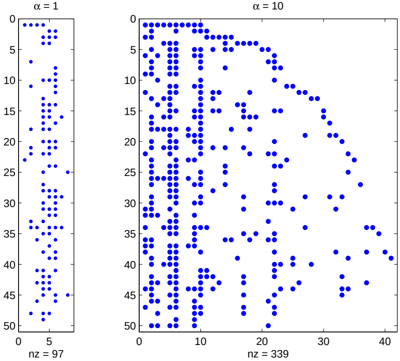


Generative Model

$$
\begin{aligned}
x_{nd} &\sim \text{Poisson}\left(\mathbf{Z}_{n\bullet}\mathbf{B}_{\bullet d}\right) \\
B_{kd} &\sim \text{Gamma}\left(\alpha_B, \frac{\mu_B}{\alpha_B}\right) \\
\mathbf{Z} &\sim \text{IBP}(\alpha)
\end{aligned}
$$

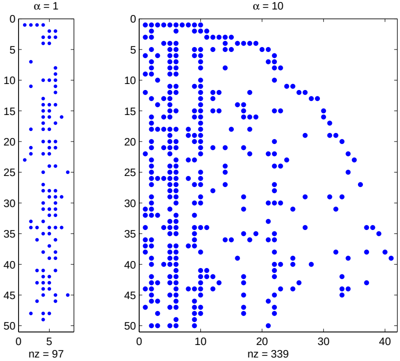
Limitations of the IBP

- Mass parameter α couples both J_n and K^+



Limitations of the IBP

- Mass parameter α couples both J_n and K^+



Beyond the standard IBP

Three-parameter IBP (Teh et.al, 2007)

- More flexible distribution for feature weights

$$\mathbf{Z}_{n\bullet} \sim \mathrm{BeP}(\mu) \qquad (3.1)$$

$$\mu \sim \mathrm{SBP}(1, \alpha, H, c, \sigma) \qquad (3.2)$$

$$p\left(J_{new}\right) \quad \sim \quad \mathrm{Poisson}\left(\boldsymbol{\alpha}\frac{\Gamma(1+\mathbf{c})\Gamma(n+\mathbf{c}+\boldsymbol{\sigma}-1)}{\Gamma(n+\mathbf{c})\Gamma(\mathbf{c}+\boldsymbol{\sigma})}\right)$$

Beyond the standard IBP

Three-parameter IBP (Teh et.al, 2007)

- More flexible distribution for feature weights

$$\mathbf{Z}_{n\bullet} \sim \mathrm{BeP}(\mu) \quad (3.1)$$

$$\mu \sim \mathrm{SBP}(1, \alpha, H, c, \sigma) \quad (3.2)$$

$$p\left(J_{new}\right) \sim \text{Poisson}\left(\boldsymbol{\alpha}\frac{\Gamma(1+\mathbf{c})\Gamma(n+\mathbf{c}+\boldsymbol{\sigma}-1)}{\Gamma(n+\mathbf{c})\Gamma(\mathbf{c}+\boldsymbol{\sigma})}\right)$$

Beyond the standard IBP

Three-parameter IBP (Teh et.al, 2007)

- More flexible distribution for feature weights

$$\mathbf{Z}_{n\bullet} \sim \mathrm{BeP}(\mu) \quad (3.1)$$

$$\mu \sim \mathrm{SBP}(1, \alpha, H, c, \sigma) \quad (3.2)$$

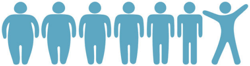
$$p\left(J_{new}\right) \sim \text{Poisson}\left(\boldsymbol{\alpha}\frac{\Gamma(1+\mathbf{c})\Gamma(n+\mathbf{c}+\boldsymbol{\sigma}-1)}{\Gamma(n+\mathbf{c})\Gamma(\mathbf{c}+\boldsymbol{\sigma})}\right)$$

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

- Arbitrary prior f over J_n

$$\mathbf{Z}_{n\bullet} \sim \text{R-BeP}(\mu, f) \quad (3.3)$$

$$\mu \sim \mathrm{BP}(1, \alpha, H) \quad (3.4)$$



Beyond the standard IBP

Three-parameter IBP (Teh et.al, 2007)

- More flexible distribution for feature weights

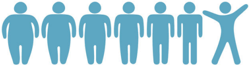
$$\mathbf{Z}_{n\bullet} \sim \mathrm{BeP}(\mu) \quad (3.1)$$
$$\mu \sim \mathrm{SBP}(1, \alpha, H, c, \sigma) \quad (3.2)$$

$$p\,(J_{new}) \sim \mathrm{Poisson}\left(\boldsymbol{\alpha}\frac{\Gamma(1+\mathbf{c})\Gamma(n+\mathbf{c}+\boldsymbol{\sigma}-1)}{\Gamma(n+\mathbf{c})\Gamma(\mathbf{c}+\boldsymbol{\sigma})}\right)$$

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

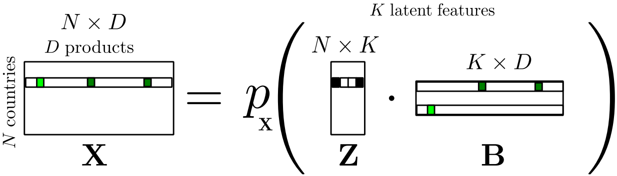
- Arbitrary prior f over J_n

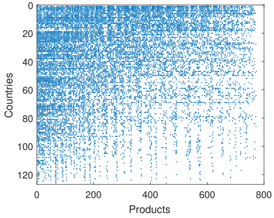
$$\mathbf{Z}_{n\bullet} \sim \text{R-BeP}(\mu, f) \quad (3.3)$$
$$\mu \sim \mathrm{BP}(1, \alpha, H) \quad (3.4)$$



- Combination of both
- Flexible prior

Our Approach





Generative Model

$$x_{nd} \sim \mathrm{Poisson}(\mathbf{Z}_{n\bullet}\mathbf{B}_{\bullet d}) \quad (3.5)$$

$$B_{kd} \sim \mathrm{Gamma}\left(\alpha_B, \frac{\mu_B}{\alpha_B}\right) \quad (3.6)$$

$$\mathbf{Z}_{n\bullet} \sim \mathrm{3R\text{-}IBP}(\alpha, c, \sigma, f) \quad (3.7)$$

Results in static scenario

Quantitative analysis: accuracy Vs interpretability

Metric	PMF	NNMF	BeP-PFA	SBeP-PFA	3RBeP-PFA
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	$\mathbf{-140.51 \pm 2.73}$

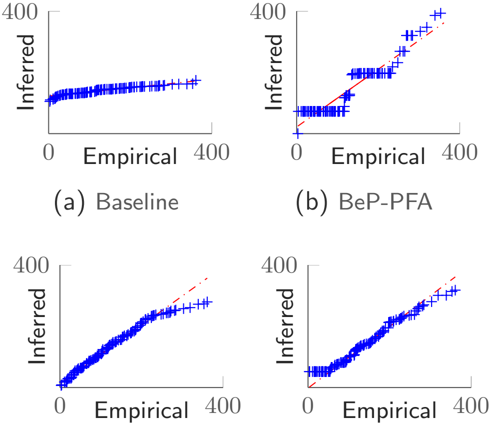
(a) 2010 SITC database ($N = 126$, $D = 744$)

Metric	PMF	NNMF	BeP-PFA	SBeP-PFA	3RBeP-PFA
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 ($N = 123$, $D = 4890$)

Results in static scenario

Capturing input sparsity structure



(a) Baseline (b) BeP-PFA

(c) sBeP-PFA (d) 3RBeP-PFA

Results

Interpretability

F0: Bias	F1: Agriculture	F2: Clothing I	F3: Farming	F4: Clothing II
Non-Coniferous Worked Wood	Vegetables	Synthetic Knitted Undergarments	Misc. Animal Oils	Synthetic Woven Fabrics
Bran and Other Cereals Residues	Fruit or Vegetable Juices	Misc. Feminine Outerwear	Bovine and Equine Entrails	Non-retail Synthetic Yarn
Misc. Non-Iron Waste	Misc. Fruit	Misc. Knitted Outerwear	Bovine meat	Woven Fabric < 85% Discontinuous Synthetic Fibres

F5: Electronics I	F6: Processed Materials	F7: Electronics II	F8: Materials I	F9: Machinery I
Misc. Electrical Machinery	Baked Goods	Measuring Controlling Instruments	Misc. Articles of Iron	Misc. Rotating Electric Plant Parts
Vehicles Stereos	Metal Containers	Mathematical Calculation Instruments	Carpentry Wood	Control Instruments of Gas or Liquid
Misc. Data Processing Equipment	Misc. Edibles	Misc. Electrical Instruments	Misc. Manufactured Wood Articles	Valves

F10: Materials II	F11: Automobile	F12: Chemicals I	F13: Chemicals II	F14: Machinery II	F15: Miscellaneous
Improved Wood	Vehicles Parts - Accessories	Synthetic Rubber	Aldehyde, Ketone	Parts of Metalworking Machine Tools	Misc. Pumps
Mineral Wool	Cars	Acrylic Polymers	Glycosides, Vaccines	Interchangeable Tool Parts	Ash and Residues
Central Heating Equipment	Iron Wire	Silicones	Medicaments	Polishing Stones	Chemical Wood Pulp of sulphite

Results

Interpretability

Top Products (decay 30%)	B_{kd}
Bovine	0.49
Miscellaneous Refrigeration Equipment	0.43
Radioactive Chemicals	0.41
Blocks of Iron and Steel	0.41
Rape Seeds	0.40
Animal meat, misc	0.39
Refined Sugars	0.38
Miscellaneous Tire Parts	0.38
Leather Accessories	0.38
Liquor	0.38
Bovine meat	0.38
Embroidery	0.37
Unmilled Barley	0.37
Dried Vegetables	0.36
Textile Fabrics Clothing Accessories	0.36
Horse Meat	0.35
Iron Bars and Rods	0.35
Analog Navigation Devices	0.35

(c) SVD

Top Products (decay 30%)	B_{kd}
Miscellaneous Animal Oils	0.78
Bovine and Equine Entrails	0.72
Bovine meat	0.68
Preserved Milk	0.63
Equine	0.62
Butter	0.58
Misc. Animal Origin Materials	0.57
Glues	0.56

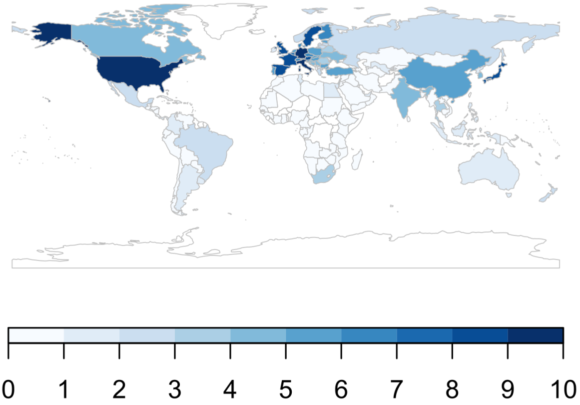
(d) S3R-IBP

Deep S3R-IBP: using a 2nd layer

1. "Simple" and "advanced" capabilities
2. Countries divided in two big groups: "quiescence" trap.

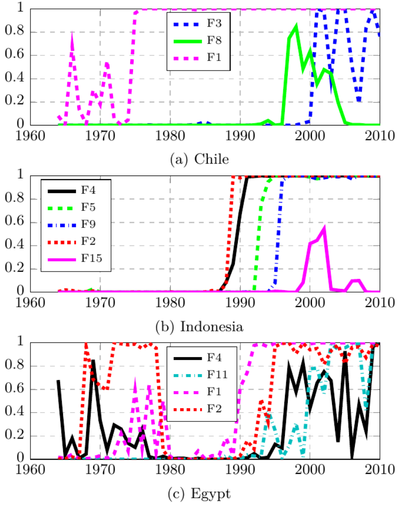
Deep S3R-IBP: using a 2nd layer

1. "Simple" and "advanced" capabilities
2. Countries divided in two big groups: "quiescence" trap.



Temporal Dynamics

	Capabilities
F0	Bias
F1	Agriculture
F2	Clothing I
F3	Farming
F4	Clothing II
F5	Electronics I
F6	Processed Materials
F7	Electronics II
F8	Materials I
F9	Machinery I
F10	Materials II
F11	Automobile
F12	Chemicals I
F13	Chemicals II
F14	Machinery II
F15	Miscellaneous



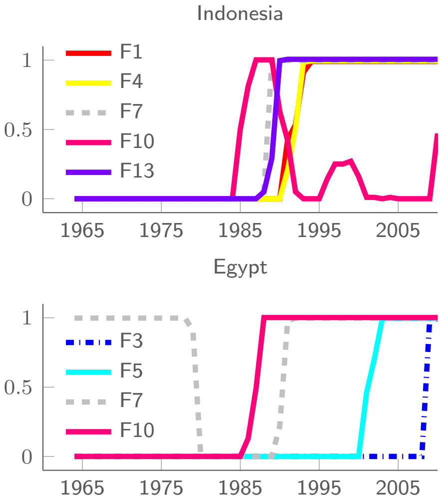
Model extension: Dynamic PFA

Model extension

$$x_{nd}^{(t)} \sim \text{Poisson}\left(\mathbf{Z}_{n\bullet}^{(t)}\mathbf{B}_{\bullet d}\right)$$

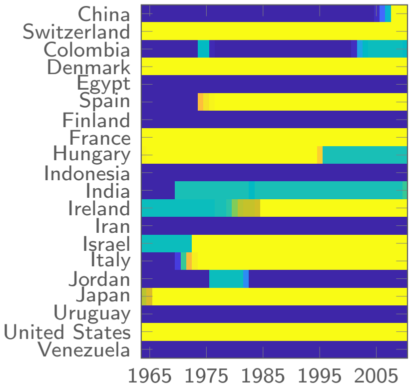
$$B_{kd} \sim \text{Gamma}\left(\alpha_B, \frac{\mu_B}{\alpha_B}\right)$$

$$\mathbf{Z}_{n\bullet}^{(\bullet)} \sim \text{mIBP}(\alpha, \gamma, \delta)$$



Model extension: dynamic PFA

Id	Top-3 products with highest weights
F0	(bias) crude petroleum, crustaceans, cereals
F1	light fixtures, locksmith hardw., misc. ceramic ornaments
F2	inorganic esters, chemical products, nitrogen compound
F3	iron sheets, iron wire, thin iron sheets
F4	misc. elect. machinery, typewriters, misc. office equipment
F5	soaps, confectionary sugar, baked goods
F6	bovine – equine entrails, bovine meat, misc. prepared meats
F7	knit clothing accessories, linens, leather accessor.
F8	glazes, textiles fabrics for machinery, mineral wool
F9	misc. vegetables, grapes – raisins, misc. fruit
F10	inorganic bases, nitrogenous fertilizers, lubricating petrol. oils
F11	imitation jewellery, embroidery, synth. precious stones
F12	coffee, non-coniferous worked wood, cane sugar
F13	copper ores, chemical wood pulp, misc. non-ferrous ores
F14	pepper, vegetable planting materials, natural rubber
F15	raw cotton, cotton linters, green groundnuts



Conclusion

1. BNP model for data exploration in high-dim count data.
2. **interpretable** and **structured** solutions.
3. Analysis of productive structure of world economies.
4. **Time-varying** feature activation.

Conclusion

1. BNP model for data exploration in high-dim count data.
2. **interpretable** and **structured** solutions.
3. Analysis of productive structure of world economies.
4. **Time-varying** feature activation.

Future works

- Improve inference in **dynamic** scenario.

Conclusion

1. BNP model for data exploration in high-dim count data.
2. **interpretable** and **structured** solutions.
3. Analysis of productive structure of world economies.
4. **Time-varying** feature activation.

Future works

- Improve inference in **dynamic** scenario.

Thank you for listening! Any question?

melanie@tsc.uc3m.es

Sources and References

- C. Bishop: *Pattern Recognition and Machine Learning,* 2006.
- K. P. Murphy: *Machine Learning: a Probabilistic Perspective*, 2012.
- D. J.C. MacKay: *Information Theory, Inference, and Learning Algorithms*, 2003.
- S. J. Gershman, D.M. Blei: A tutorial on Bayesian nonparametric models, 2012.
- Y.W. Teh: Slides for *Probabilistic and Bayesian Machine Learning*, UC3M, 2010.
- M. N. Schmidt & M. Morup: *Advanced Topics in Machine Learning*, MLSS, DTU, 2013.
- D. B. Dunson: *Nonparametric Bayes Applications to Biostatistics*, 2010.

Appendix: About inference

- Markov Chain Monte Carlo approach.
- Conditional conjugacy using auxiliary variables.

$$x_{nd} = \sum^{K} x'_{nd,k} \quad \text{where} \quad x'_{nd,k} \sim \text{Poisson}(\mathbf{Z}_{n\bullet}\mathbf{B}_{\bullet d})$$

- Truncated approximation of feature weights
- In 3RBeP-PFA, dynamic programming to compute likelihood (Doshi-Velez et.al, 2015)
- In dBeP-PFA, forward-filtering backward-sampling procedure (Gael et.al, 2009)

Appendix: Results

Interpretability

Countries in latent space

Appendix: Results

Interpretability

Countries in latent space

- France = Belgium + ?
- Germany - ? = Austria
- Malaysia (Electronics) + ? → Phillipines
- Phillipines + ? → Indonesia, Vietnam

Appendix: Results

Interpretability

Countries in Capability Space

- France = Belgium + Industrial Machinery
- Germany - Chemical = Austria
- Malaysia (Electronics) + Clothing → Phillipines
- Phillipines + Basic Processing → Indonesia, Vietnam

Appendix: modeling in dynamic scenario

Dynamic PFA

- T timestamps (years)
- markov IBP to account for temporal dynamics (Gael et.al, 2009)
- Generative model:

$$
\begin{aligned}
x_{nd}^{(t)} &\sim \text{Poisson}\left(\mathbf{Z}_{n\bullet}^{(t)}\,\mathbf{B}_{\bullet d}\right)\\
B_{kd} &\sim \text{Gamma}(\alpha_B, \frac{\mu_B}{\alpha_B})\\
a_k &\sim \text{Beta}(\frac{\alpha}{K}, 1),\\
b_k &\sim \text{Beta}(\gamma, \delta),\\
z_{nk}^{(t)}|a_k, b_k &\sim \text{Bernoulli}\left(a_k^{1-z_{nk}^{(t-1)}}\, b_k^{z_{nk}^{(t-1)}}\right)
\end{aligned}
$$

The transition matrix Q_k for feature k is given by:

$$
Q_k = \begin{pmatrix} 1-a_k & a_k \\ 1-b_k & b_k \end{pmatrix}
$$

Appendix: inference in dynamic scenario

Inference

- MCMC approach, e.g., Gibbs sampler + slice sampler for the IBP
- K Poisson-distributed auxiliary random variables, i.e., $x_{nd}^{(t)} = \sum_{k=1}^{K} r_{nd,k}^{(t)}$
- Forward Filtering Backward Sampling (FFBS) to approximate $p(\mathbf{Z}|\mathbf{X}, \mathbf{B})$

$$p(\mathbf{X}_{n\bullet}^{(1:t)}, z_{nk}^{(t)}|-) = p(\mathbf{X}_{n\bullet}^{(t)}|z_{nk}^{(t)}, -) \sum_{z_{nk}^{(t-1)}} p(\mathbf{X}_{n\bullet}^{(1:t-1)}, z_{nk}^{(t-1)}|-) p(z_{nk}^{(t)}|z_{nk}^{(t-1)})$$

  - Forward step: compute $p(z_{nk}^{(t)}|\mathbf{X}_{n\bullet}^{(1:t)}, \mathbf{Z}_{n,\neg k}^{(t)}, \mathbf{B})$
  - Backward step: sample from $p(z_{nk}^{(t)}|z_{nk}^{(t+1)}, \mathbf{X}_{n\bullet}^{(1:t)}, \mathbf{Z}_{n,\neg k}^{(t)}, \mathbf{B})$