Modeling the Productive Structure of Economies: A nonparametric Bayesian Approach

Melanie F. Pradier, Zoran Utkovski
Collaborators: Viktor Stojkoski and Fernando Perez-Cruz

September 23, 2016

Index

Index

The Wealth of Nations

Question: What makes a country wealthy?

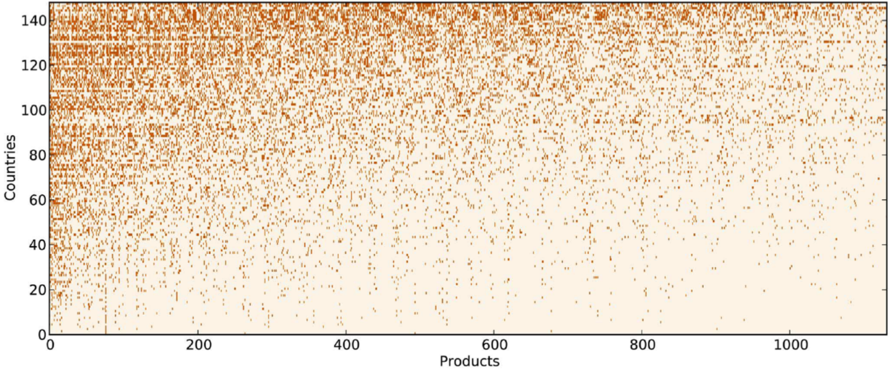
Which elements drive competitiveness of countries?

Classical view

- Division of labor [A. Smith, 1976], Ricardian Paradigm [Ricardo, 1817]
- Specialization leads to economic efficiency
- Wealthy countries producing few products with high degree of specialization

The "classical approach" predicts a block-diagonal structure of the country-product trade matrix

The Reality
[Cristelli, et.al. 2013]

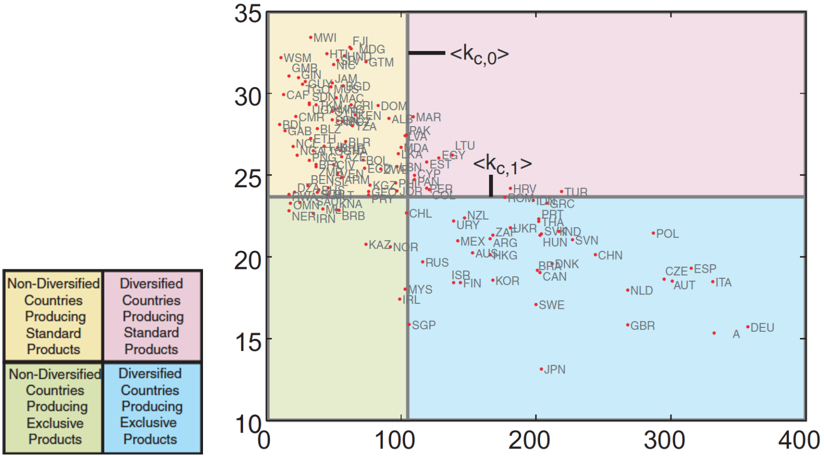


Matrix Triangularity

- Diversification: Number of ones per row
- Ubiquity: Number of ones per column

Diversification vs Ubiquity

[Hidalgo, et.al. 2009]



Index

Theory of Economic Complexity

- There is a robust and stable relationship between a country's productive structure and its economic growth.
- Economic complexity - introduced in [Hidalgo et. al, 2007], [Hidalgo & Hausmann, 2009] to reflect the amount of knowledge that is embedded in the *productive structure* of an economy.

Beyond GDP!

Non-monetary and non-income-based measures which uncover countries' hidden potential for development and growth.

The Country-Product Matrix

- Relationship between countries and the products they export is represented as a bipartite graph $\mathcal{G} = (\mathcal{C}, \mathcal{P}, \mathcal{E})$
- An edge (i, j) between a country $i \in \mathcal{C}$ and a product $p \in \mathcal{P}$ is present in $\mathcal{E}$ if the country has a revealed comparative advantage (RCA) [Balassa, 1964] in the export of that product.

$$R_{ij} = \frac{E_{ij} / \sum_j E_{ij}}{\sum_i E_{ij} / \sum_{i,j} E_{ij}}, \tag{1}$$

- E_{ij} is the export of product j by country i, $i \in \mathcal{C}$, $j \in \mathcal{P}$.
- $R_{ij} > 1$ if country i's share of product j is larger than the product's share of the entire world market
- The country-product matrix

$$M_{ij} = \begin{cases} 1, & \text{if } R_{ij} \geq 1 \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

The Method of Reflections (MR)

Economic Complexity

- A product is "complex" if it is exported by a "complex" country
- Similarity with the Pagerank algorithm

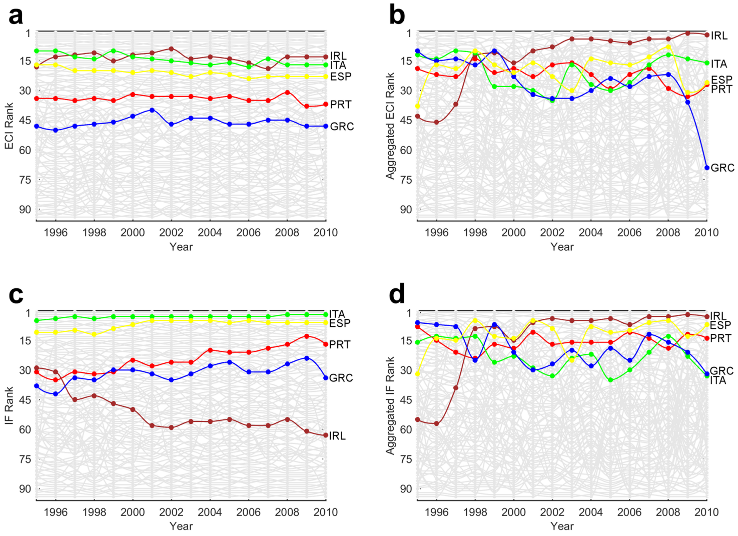
The Method of Reflections (MR) [Hidalgo, Hausmann 2009] - an iterative linear procedure that produces complexity indices of countries and products.

$$\begin{cases} c_{i,n} = \frac{1}{d_i} \sum_j M_{ij} p_{j,n-1} \\ p_{j,n} = \frac{1}{u_j} \sum_i M_{ij} c_{i,n-1} \end{cases}, \quad (3)$$

Initial conditions

- $c_{i,0} = d_i$ is country i's diversity - number of products for which the country has $RCA > 1$
- $p_{j,0} = u_j$ is product j's ubiquity - number of countries which have RCA in that product

Country Rankings: Examples



The Fitness-Complexity Method (FCM)

FCM [Cristelli et. al, 2013]

- Country fitness = (weighted) sum of the complexities of the exported products
- Product complexity ≠ average fitness of the countries producing it.

A strong nonlinear relationship between the complexity of an exported product and the competitiveness of its producers.

$$\begin{cases} \widetilde{c}_{i,n} = \sum_j M_{ij} p_{j,n-1} \\ \widetilde{p}_{j,n} = \frac{1}{\sum_i M_{ij} \frac{1}{c_{i,n-1}}} \end{cases} \longrightarrow \begin{cases} c_{i,n} = \frac{\widetilde{c}_{i,n}}{\langle \widetilde{c}_{i,n} \rangle_i} \\ p_{j,n} = \frac{\widetilde{p}_{j,n}}{\langle \widetilde{p}_{j,n} \rangle_j} \end{cases}, \quad (4)$$

- $\widetilde{c}_{i,n}$ - intermediate fitness (country complexity)
- $\widetilde{p}_{j,n}$ - intermediate product complexity
- Initial conditions: $\widetilde{c}_{i,0} = 1$, $\widetilde{p}_{j,0} = 1$
- Normalization in each step

Modified Fitness-Complexity Method (M-FCM)

Convergence issue of FCM

Country-product matrices obtained from real trade data often exhibit an "unfavorable" structure resulting in some country fitness and product complexity scores converging to zero

M-FCM [Stojkoski et. al] - A modification of FCM

$$\begin{cases} \widetilde{c}_{i,n} = \sum_j M_{ij} p_{j,n-1} \\ \widetilde{p}_{j,n} = \frac{1}{\sum_i M_{ij}(N_c - c_{i,n-1})} \end{cases} \longrightarrow \begin{cases} c_{i,n} = \frac{\widetilde{c}_{i,n}}{\langle \widetilde{c}_{i,n} \rangle_i} \\ p_{j,n} = \frac{\widetilde{p}_{j,n}}{\langle \widetilde{p}_{j,n} \rangle_j} \end{cases} . \quad (5)$$

- The term $\frac{1}{c_{i,n-1}}$ in FCM is substituted with $(N_c - c_{i,n-1})$ (N_c is the number of countries)
- The complexity of a product is still (mostly) determined by the complexity of the least competitive exporting countries.

Index

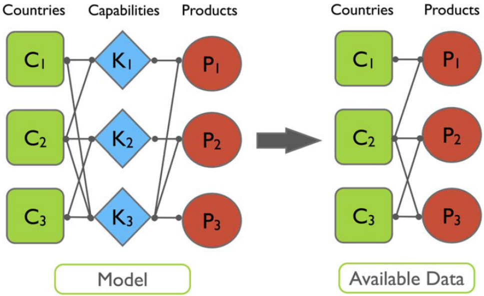
Theory of Hidden Capabilities

Hidden Capabilities

The theory of economic complexity implicitly resides on the premise of "hidden capabilities" behind the productive structure of an economy

Measuring the "intangibles"

Capabilities are "intangible assets which drive the development, the wealth and the competitiveness of a country" [Cristelli et. al, 2013]



Capability-based Interpretation of EC

The Binomial Model [Hidalgo& Hausmann, 2011]

Country i has RCA in product j iff it is endowed with **all** capabilities required to produce the product.

- Z - a country-capability matrix
- B - a product-capability matrix

$$M_{ij} = Z_{ik} \odot B_{jk}, \tag{6}$$

where

$$Z_{ik} \odot B_{jk} = \begin{cases} 1, & \text{if } \sum_k Z_{ik} B_{jk} = \sum_k Z_{ik} \\ 0, & \text{otherwise} \end{cases}. \tag{7}$$

- Operator $\odot$ resembles a (binary) Leontief production function.

Problem Interpretation

This is essentially a stochastic matrix-factorization problem

- Probabilistic interpretation of the country-product matrix M
- Capabilities as hidden variables that relate countries and products

Ideally, we need a model that

- explains the data
- incorporates sparsity
- is consistent with "well accepted" findings in economy
- provides interpretation of the extracted features (capabilities)

Poisson factorization based on the Restricted-Indian Buffet Process

$$P(M_{ij} = 1 | Z_{i\cdot}, B_{\cdot j}) = \text{Poisson}\Big(Z_{i\cdot} B_{\cdot j}\Big)$$

Index

Index

Indian Buffet Process

Definition

Stochastic process defining prob. distribution over sparse binary matrices with finite number of rows and infinite number of columns.

$$Z \sim \text{IBP}(\alpha)$$

Culinary Metaphor

Iteratively, for each customer i:

- For $k = 1, \ldots, K^+$:

$$p(z_{ik} = 1 | Z_{\neg ik}) \propto \frac{m_{-i,k}}{i} \quad (8)$$

- Sample new dishes:

$$J_{new} \sim \text{Poisson}(\frac{\alpha}{i}) \quad (9)$$

Indian Buffet Process

(Slides 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

(Slides from F.J.R.Ruiz)

. . .

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

⋮

Some Measure Theory (I)

Random Measure μ

- Distribution over measures in measurable space $(\Theta, \mathcal{A})$.
- Stochastic process indexed by sigma algebra $\mathcal{A}$, i.e., collection of r.v. $\mu(A) \in [0, \infty]$ for each $A \in \mathcal{A}$.

Some Measure Theory (I)

Random Measure μ

- Distribution over measures in measurable space $(\Theta, \mathcal{A})$.
- Stochastic process indexed by sigma algebra $\mathcal{A}$, i.e., collection of r.v. $\mu(A) \in [0, \infty]$ for each $A \in \mathcal{A}$.

Completely Random Measure

- Random measure such that, $\forall A_1, A_2, \ldots, A_n \subset \mathcal{A}$ disjoint sets, $\mu(A_1), \mu(A_2), \ldots, \mu(A_n)$ are independent.
- Beta Process, Gamma Process, Bernoulli Process, etc, ...

$$\mu = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k}$$

Some Measure Theory (II)

De Finetti's Theorem

Any infinitely exchangeable sequence can be written as a mixture of i.i.d samples

$$p(X_1, X_2, \ldots, X_n) = \int \prod_{i=1}^{n} \mathcal{Q}_\mu(X_i) P(d\mu) \quad (10)$$

Some Measure Theory (II)

De Finetti's Theorem

Any infinitely exchangeable sequence can be written as a mixture of i.i.d samples

$$p(X_1, X_2, \ldots, X_n) = \int \prod_{i=1}^{n} \mathcal{Q}_\mu(X_i) P(d\mu) \qquad (10)$$

De Finetti's Mixing distribution for IBP

$$\mu = \sum_k \pi_k \delta_{\theta_k} \sim \mathrm{BP}(\alpha, H) \qquad (11)$$

$$\zeta_n = \sum_k z_{nk} \delta_{\theta_k} \sim \mathrm{BeP}(\mu) \qquad (12)$$

$$\mu \sim \mathrm{BP}(\alpha, H) \qquad (13)$$

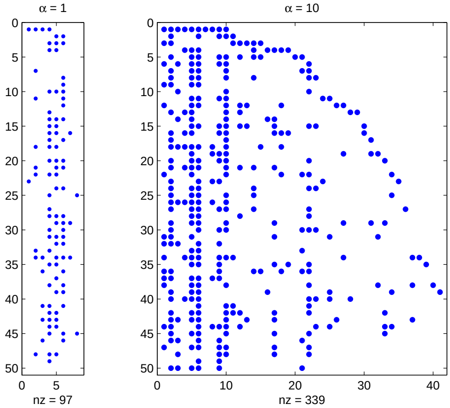
$$Z_n \sim \mathrm{BeP}(\mu) \qquad (14)$$

$$\Updownarrow$$

$$Z \sim \mathrm{IBP}(\alpha) \qquad (15)$$

Assumptions underlying the IBP

- Number of ones per row $r_n \sim \text{Poisson}(\alpha)$.
- Number of non-empty columns $K^+ \sim \text{Poisson}(\alpha \sum_{j=1}^{N} \frac{1}{j})$.



Index

Restricted Indian Buffet Process

[Doshi-Velez et. al, 2015]

- IBP with arbitrary distribution f over $r_n = \sum_i z_{ni}$.

$$Z \sim \text{R-IBP}(\alpha, f) \tag{16}$$

$$\Updownarrow$$

$$\mu \sim \text{BP}(\alpha, H) \tag{17}$$

$$Z_n \sim \text{R-BeP}(\mu, f) \tag{18}$$

Restricted Bernoulli Process, case $f = \delta_J$

$$\text{R-BeP}(Z_n; \mu, f = \delta_J) \propto \begin{cases} \text{BeP}(Z_n; \mu) \text{ if } r_n = J \\ 0 \qquad \text{otherwise} \end{cases} \tag{19}$$

Restricted Indian Buffet Process

[Doshi-Velez et. al, 2015]

- IBP with arbitrary distribution f over $r_n = \sum_i z_{ni}$.

$$Z \sim \text{R-IBP}(\alpha, f) \tag{16}$$

$$\Updownarrow$$

$$\mu \sim \text{BP}(\alpha, H) \tag{17}$$

$$Z_n \sim \text{R-BeP}(\mu, f) \tag{18}$$

Restricted Bernoulli Process, general f

$$\text{R-BeP}(Z_n; \mu, f) = f\left(\sum_k Z_{nk}\right) \text{R-BeP}(Z_n; \mu, f = \delta_{\sum_k Z_{nk}}) \tag{19}$$

Index

Sparse Poisson Factorization Model

Generative Model

$$Z \sim \text{R-IBP}(\alpha, f) \tag{20}$$

$$B_{\cdot j} \sim \text{Gamma}(\alpha_B, \frac{\mu_B}{\alpha_B}) \tag{21}$$

$$M_{ij} | Z, B \sim \text{Poisson}(Z_{i\cdot} B_{\cdot j}) \tag{22}$$

- M: country-product matrix.
- Z: country-capability matrix.
- B: capability-product matrix.
- Double sparsity (by choosing shape parameter $\alpha_B < 1$).

Index

Preliminary Results

Simulation Settings

- Trade data from SITC database, year 2010.
- $N = 126$ countries, $D = 744$ products.
- We choose $\alpha_B = 0.01$, $f =$ Neg. Binomial$(r = 2, p = 0.05)$.
- Burn-out: 45.000 iterations, estimates using 5.000 last iterations.

- IBP finds 11 capabilities on average, whereas R-IBP finds 15 capabilities.
- Concerning prediction accuracy:

	SVD	IBP	R-IBP
MSE	0.1087	0.1142	0.1154
MAE	0.2123	0.2171	0.2095

Inferred Capabilities using SVD

Interpretability

"Bovine"	0.49	"Sulphur"	0.40
"Miscellaneous Refrigeration Equipment"	0.43	"Fuel Wood and Charcoal"	0.34
"Radioactive Chemicals"	0.41	"Miscellaneous Unmilled Cereals"	0.33
"Blocks of Iron and Steel"	0.41	"Household Refrigeration"	0.33
"Rape Seeds"	0.40	"Decorative Wood"	0.33
"Animal meat, misc"	0.39	"Frozen Fish Fillets"	0.32
"Refined Sugars"	0.38	"Rail Freight Transport"	0.32
"Miscellaneous Tire Parts"	0.38	"Wool Undergarments"	0.31
"Leather Accessories"	0.38	"Cheese"	0.31
"Liquor"	0.38	"Ships and Boats"	0.31
"Bovine meat"	0.38	"Miscellaneous Animal Origin Materials"	0.31
"Embroidery"	0.37	"Worked Tin and Alloys"	0.30
"Unmilled Barley"	0.37	"Transmission Belts"	0.29
"Dried Vegetables"	0.36	"Copper"	0.29
"Textile Fabrics Clothing Accessories"	0.36	"Men's Jackets"	0.29
"Horse Meat"	0.35	"Electrical Transformers"	0.29
"Iron Bars and Rods"	0.35	"Polishing Stones"	0.28
"Analog Navigation Devices"	0.35	"Tea"	0.28
"Cocoa Butter"	0.34	"Linens"	0.28
"Miscellaneous Live Animals"	0.34	"Miscellaneous Parts of Steam Power Units"	0.28

Inferred Capabilities using R-IBP (I)

Interpretability

"Cotton Undergarments"	0.93
"Knit Clothing Accessories"	0.91
"Lingerie"	0.84
"Miscellaneous Feminine Outerwear"	0.81
"Men's Shirts"	0.81
"Men's Pants"	0.80
"Knitted Outerwear"	0.78
"Blouses"	0.67
"Miscellaneous Men's Outerwear"	0.64
"Women's Suits"	0.64
"Synthetic Knitted Undergarments"	0.62
"Women's Knitted Outerwear"	0.61
"Men's Coats"	0.60
"Miscellaneous Knitted Outerwear"	0.60
"Men's Jackets"	0.60
"Dresses"	0.60
"Skirts"	0.58
"Men's Suits"	0.56
"Women's Underwear"	0.54
"Headgear"	0.53

"Miscellaneous Electrical Machinery"	1.23
"Miscellaneous Electronic Circuit Parts"	0.71
"Worked Tin and Alloys"	0.69
"Miscellaneous Data Processing Equipment"	0.67
"Miscellaneous Articles of Precious Metals"	0.59
"Vehicles Stereos"	0.58
"Electronic Microcircuits"	0.57
"Plastic or Rubber Clothing"	0.55
"Musical Instrument Parts"	0.54
"Electrical Resistors"	0.53
"Telecom Parts and Accessories"	0.49
"Steam Water Boilers"	0.49
"Miscellaneous Telecom Equipment"	0.47
"Bicycles"	0.44
"Miscellaneous Office Equipment"	0.43
"Prepared Fruit"	0.42
"Cameras"	0.41
"Diodes, Transistors and Photocells"	0.41
"TV Tubes and Cathode Rays"	0.40
"Optical Lenses"	0.35

Inferred Capabilities using R-IBP (II)

Interpretability

	m_k	Capability	Repr. Countries
F1	19	Machinery: rotary	-
F2	27	Industrial parts	-
F3	18	Farmaceutics	-
F4	35	Agriculture/Farming	Paraguay
F5	18	Electronics	Malaysia
F6	26	Car industry	-
F7	23	Chemical treatments (e.g. pesticides)	Peru
F8	42	Basic processing (food, material)	Kenya
F9	24	Synthetic fibers	–
F10	9	Minery (nickel, coal...)	Kazakhstan
F11	24	Machinery, general industry	-
F12	10	Chemical (polymerization, silicons...)	-
F13	29	Minery (iron, copper...)	-
F14	32	Miscellaneous	-
F15	45	Clothing	Morocco, Bangladesh, ...

Inferred Capabilities using R-IBP (III)

Interpretability

Competitive Advantages of each country

- Norway: Minery (nickel, coal) + Rotary machinery
- Russia: Minery (nickel, coal) + Minery (iron, copper)

Inferred Capabilities using R-IBP (III)

Interpretability

Competitive Advantages of each country

- Norway: Minery (nickel, coal) + Rotary machinery
- Russia: Minery (nickel, coal) + Minery (iron, copper)
- Switzerland: Machinery + Car Industry + Chemicals + Farmaceutics

Inferred Capabilities using R-IBP (III)

Interpretability

Competitive Advantages of each country

- Norway: Minery (nickel, coal) + Rotary machinery
- Russia: Minery (nickel, coal) + Minery (iron, copper)
- Switzerland: Machinery + Car Industry + Chemicals + Farmaceutics

Countries in Capability Space

Inferred Capabilities using R-IBP (III)

Interpretability

Competitive Advantages of each country

- Norway: Minery (nickel, coal) + Rotary machinery
- Russia: Minery (nickel, coal) + Minery (iron, copper)
- Switzerland: Machinery + Car Industry + Chemicals + Farmaceutics

Countries in Capability Space

- France = Belgium + ?
- Germany - ? = Austria
- Malaysia (Electronics) + ? $\rightarrow$ Phillipines
- Phillipines + ? $\rightarrow$ Indonesia, Vietnam
- Turkey $\rightarrow$ Italy?
- Italy $\rightarrow$ Spain?

Inferred Capabilities using R-IBP (III)

Interpretability

Competitive Advantages of each country

- Norway: Minery (nickel, coal) + Rotary machinery
- Russia: Minery (nickel, coal) + Minery (iron, copper)
- Switzerland: Machinery + Car Industry + Chemicals + Farmaceutics

Countries in Capability Space

- France = Belgium + Industrial Machinery
- Germany - Chemical = Austria
- Malaysia (Electronics) + Clothing → Phillipines
- Phillipines + Basic Processing → Indonesia, Vietnam
- Turkey → Italy? (Machinery + Chemical)
- Italy → Spain? (Agriculture/Farming)

Deep IBP: 2nd layer of IBP

1. Countries divided in two big groups: “quiescence” trap.

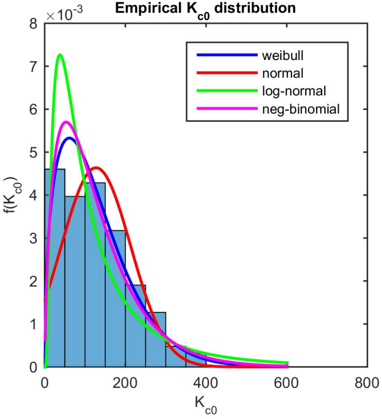
Deep IBP: 2nd layer of IBP

1. Countries divided in two big groups: "quiescence" trap.

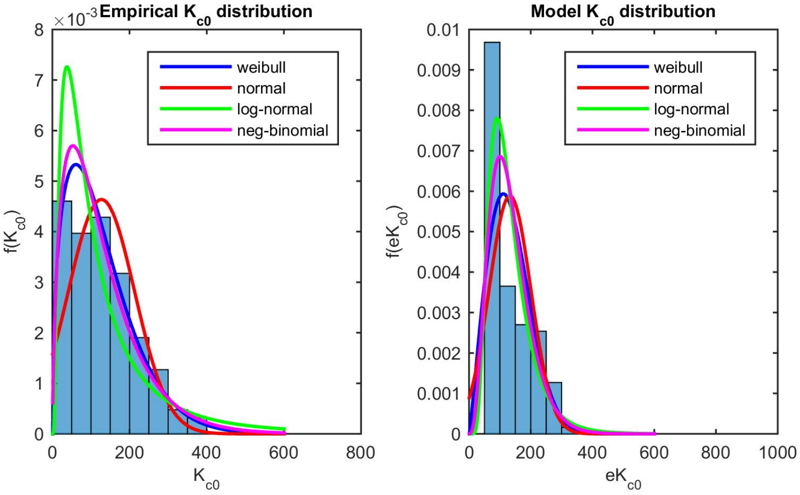
2. Capabilities can be clustered in 3 sets:

Basic	Mixed	Advanced
Clothing	Basic processing	Car industry
Synthetic Fibers	Chemical treatments	Minery (iron, copper...)
Minery(nickel, coal)	Agriculture/farming	Farmaceutics
Electronics		Industrial parts
Chemical (Silicons...)		Machinery: general
		Machinery: specialized
		Miscellaneous

Distribution of countries diversification: IBP



Distribution of countries diversification: R-IBP



Index

Conclusion

So far...

- Application of the R-IBP.
- Bayesian non-parametric latent feature model for sparse count data:
  - High interpretability.
  - Modeling of structured sparsity.

Future works

- Dynamic evolution of capabilities
  - Varying per country activation over time.
  - Smooth variation of capabilities along history.

Restricted Indian Buffet Process

[Doshi-Velez et. al, 2015]

Restricted Bernoulli Process, case $f = \delta_J$

$$\text{R-BeP}(Z_n; \mu, f = \delta_J) = \frac{\prod_{k=1}^{\infty} \pi_k^{z_{nk}} (1 - \pi_k^{1-z_{nk}}) \mathbb{1}(\sum_K z_{nk} = J)}{\sum_{z' \in \mathcal{Z}} \prod_k \pi_k^{z'_k} (1 - \pi_k)^{(1-z'_k)} \mathbb{1}(\sum_K z'_k = J)} \tag{23}$$

Appendix

A few words about inference

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

$$x_{nd} = \sum^{K} x'_{nd,k} \quad \text{where} \quad x'_{nd,k} \sim \text{Poisson}(Z_{n\cdot} B_{\cdot d})$$

- Exact inference for IBP using slice sampler [Teh, 2007]. Truncate approximation for R-IBP.
- Dynamic programming to compute R-IBP likelihood [Doshi-Velez, 2015].

Appendix

Other capability examples using R-IBP

"Glycosides and Vaccines"	1.08	"Miscellaneous Wheat"	0.65
"Polyamides"	0.85	"Preserved Milk"	0.65
"Aldehyde, Ketone and Quinone-Function Compounds"	0.84	"Maize"	0.62
"Analog Navigation Devices"	0.83	"Improved Wood"	0.62
"Other Nitrogen Function Compounds"	0.73	"Simply Shaped Wood"	0.59
"Centrifuges"	0.70	"Carpentry Wood"	0.59
"Printing Ink"	0.68	"Miscellaneous Animal Oils"	0.58
"Cyclic Alcohols"	0.64	"Glues"	0.57
"Antibiotics"	0.64	"Cheese"	0.57
"Miscellaneous Printing Machines"	0.62	"Pulpwood"	0.55
"Sound Recording Media"	0.61	"Miscellaneous Vegetable Oils"	0.55
"Scented Mixtures"	0.60	"Unmilled Barley"	0.54
"Centrifugal Pumps"	0.60	"Soil Preparation Machinery"	0.52
"X-Ray Equipment"	0.60	"Electric Current"	0.51
"Organo-Sulphur Compounds"	0.56	"Rape Seeds"	0.51
"Hormones"	0.56	"Unmilled Rye"	0.50
"Heterocyclic Compounds"	0.54	"Bovine meat"	0.49
"Cellulose Derivates"	0.53	"Bovine"	0.47
"Orthopedic Devices"	0.52	"Coniferous Wood"	0.47
"Analog Instruments for Physical Analysis"	0.51	"Raw Sheep Skin with Wool"	0.46

Appendix

Ranking of countries