Bayesian dimensionality reduction

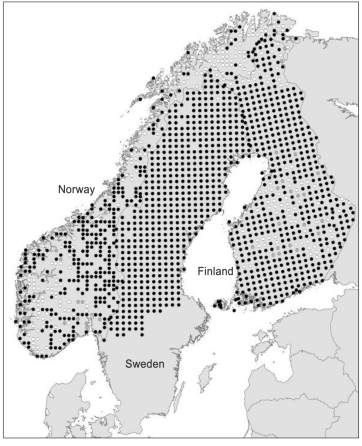
Antonio Canale · canale@stat.unipd.it · @tonycanale_
International Workshop on Statistical Modelling
Trieste, July 22, 2022

Università degli Studi di Padova

Big data: blessing or curse?

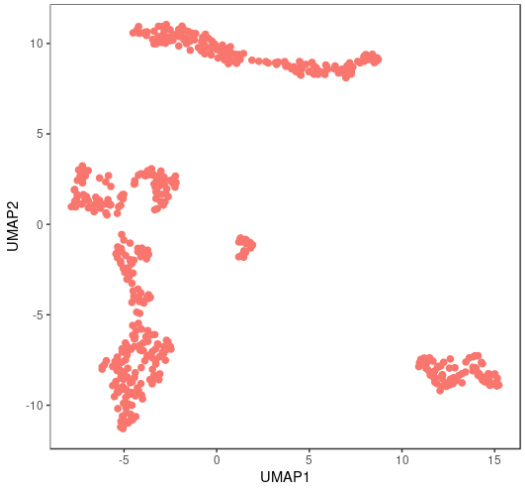
Reccomendation systems

Bird species occurrencies



Source: Lindström et al. (2015)

Genomics



Source: Chandra et al. (2020)

Curse of dimensionality

- A notorious issue in these context is the so called *curse of dimensionality*

Curse of dimensionality

- A notorious issue in these context is the so called *curse of dimensionality*
- A common solution consists in assuming that the *signal* contained in the data is actually concentrated in a low-dimensional subspace

Curse of dimensionality

- A notorious issue in these context is the so called *curse of dimensionality*
- A common solution consists in assuming that the *signal* contained in the data is actually concentrated in a low-dimensional subspace
- In this talk I will focus on *Bayesian dimensionality reduction* techniques.

Curse of dimensionality

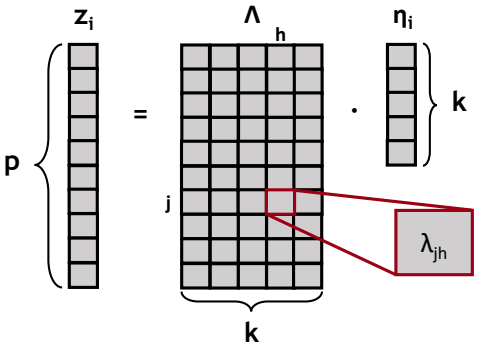
- A notorious issue in these context is the so called *curse of dimensionality*
- A common solution consists in assuming that the *signal* contained in the data is actually concentrated in a low-dimensional subspace
- In this talk I will focus on *Bayesian dimensionality reduction* techniques.
- Side benefits include ease ease of interpretation of the lower dimensional variables as *latent traits* driving the phenomena under study.

Outline

Factor Models (FM)

$$z_i = \Lambda \eta_i + \epsilon_i \qquad \epsilon_i \sim f_\epsilon.$$

- z_i: i-th p-variate random variable;
- Λ: $p \times k$ factor loadings matrix;
- η_i: i-th vector of k latent factors.



FM - Just dimensionality reduction?

- FM can be seen merely as dimensionality reduction tool

FM - Just dimensionality reduction?

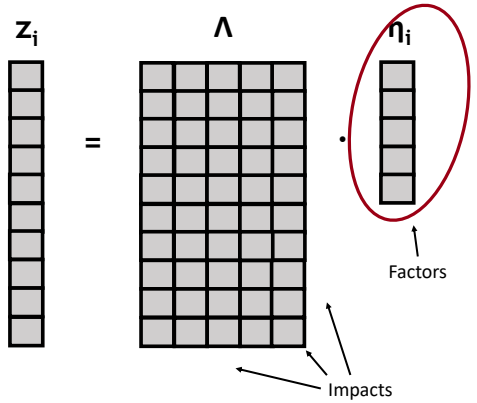
- FM can be seen merely as dimensionality reduction tool
- FM rooted back in psychometrics where the latent factors represent some interpretable latent trait (Spearman, 1904).

FM - Just dimensionality reduction?

- FM can be seen merely as dimensionality reduction tool
- FM rooted back in psychometrics where the latent factors represent some interpretable latent trait (Spearman, 1904).
- Widely adopted and generalized: Gaussian copula FM (Murray et al., 2013), probabilistic matrix factorizations (Mnih & Salakhutdinov, 2008); functional data (Montagna et al., 2012) and (Kowal and Canale, 2022)

Interpretability

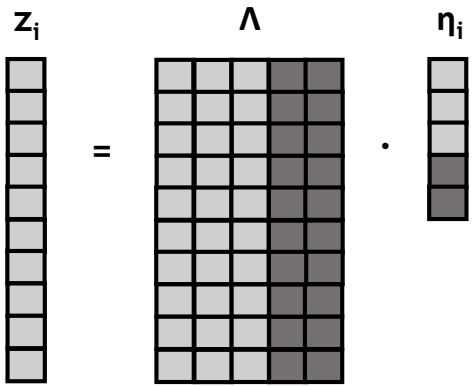
Interpretation of factor models is assigning a meaning to the latent factors and then to their impact on the observed data.



Interpretability

Interpretation of loadings matrix and factors is strongly favoured by

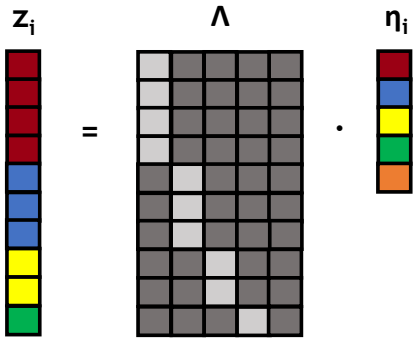
- a limited number *k* of factors;



Interpretability

Interpretation of loadings matrix and factors is strongly favoured when

- each factor has an impact only on a small group of components of z_i.

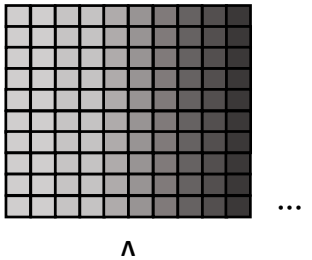


Infinite factor models (IFM)

Bayesian (nonparametric) approach introduced by Bhattacharya and Dunson (2011).

Infinitely many factors, with the **impact** of these factors **decreasing** with the factor index.

Accomplished with **increasing shrinkage priors**, that allow to approximate the IFM through a finite number of factors.

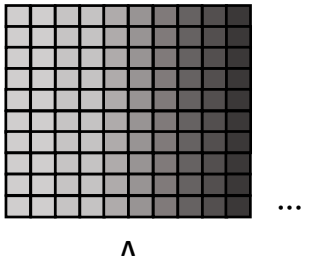


Infinite factor models (IFM)

Bayesian (nonparametric) approach introduced by Bhattacharya and Dunson (2011).

Infinitely many factors, with the **impact** of these factors **decreasing** with the factor index.

Accomplished with **increasing shrinkage priors**, that allow to approximate the IFM through a finite number of factors.



No structure **constraints** are imposed on the number of factors or on the sparsity pattern!

Inference and sparsity in IFM

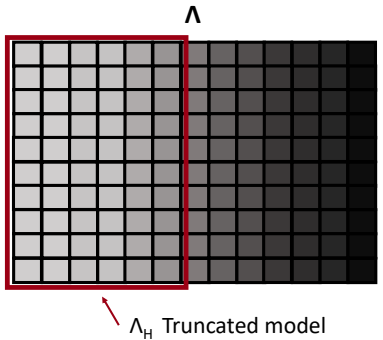
Posterior inference is conducted via **Monte Carlo Markov Chains**.

Inference and sparsity in IFM

Posterior inference is conducted via **Monte Carlo Markov Chains**.

Truncating out the negligible **columns** of Λ, those really **close to zero** $\Rightarrow$ small number of latent factors.

Priors on loadings elements with sufficiently mass concentration **around zero** $\Rightarrow$ Sparse pattern on Λ.



Current IFM

- Multiplicative gamma process (**MGP**) - Bhattacharya & Dunson, 2011.
- Cumulative shrinkage process (**CUSP**) - Legramanti et al., 2020.
- In general

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh})$$

Current IFM

- Multiplicative gamma process (**MGP**) - Bhattacharya & Dunson, 2011.
- Cumulative shrinkage process (**CUSP**) - Legramanti et al., 2020.
- In general

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh})$$

Problems:

1. lack of careful consideration of the within component sparsity structure
2. no accommodation for grouped variables and other non-exchangeable structure.

Generalized Infinite Factorization (GIF)

$$\lambda_{jh} \mid \theta_{jh} \sim \mathsf{N}(0, \theta_{jh})$$

Generalized Infinite Factorization (GIF)

$$\lambda_{jh} \mid \theta_{jh} \sim \mathsf{N}(0, \theta_{jh})$$

$$\theta_{jh} = \tau_0$$

- $\tau_0 \sim f_{\tau_0}$: global scale;

Generalized Infinite Factorization (GIF)

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh})$$

$$\theta_{jh} = \tau_0 \gamma_h$$

- $\tau_0 \sim f_{\tau_0}$: global scale;
- $\gamma_h \sim f_{\gamma_h}$: column scale;

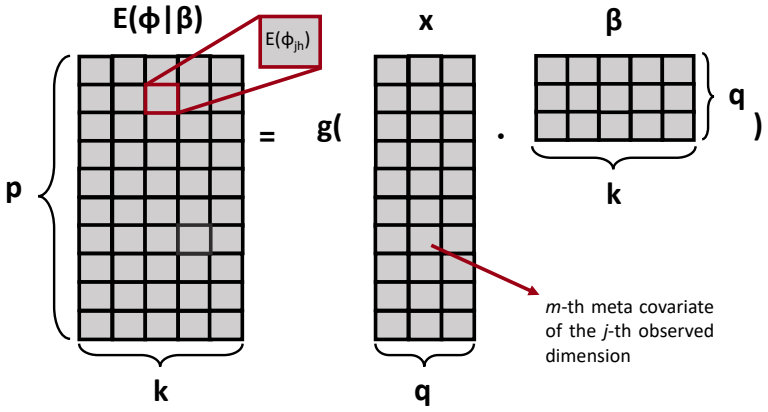
Generalized Infinite Factorization (GIF)

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh})$$

$$\theta_{jh} = \tau_0 \, \gamma_h \, \phi_{jh}$$

- $\tau_0 \sim f_{\tau_0}$: global scale;
- $\gamma_h \sim f_{\gamma_h}$: column scale;
- $\phi_{jh} \sim f_{\phi_j}$: local scale. That depends on meta covariates: $E(\phi_{jh}) = g(x_j^\top \beta_h)$

Exogenous information about the sparsity structure



$$E(\phi_{jh} \mid \beta_h) = g(x_j^\top \beta_h), \qquad \beta_h = (\beta_{1h}, \ldots, \beta_{qh})^\top, \qquad \beta_{mh} \sim f_\beta$$

Bird species occurrence example (1)

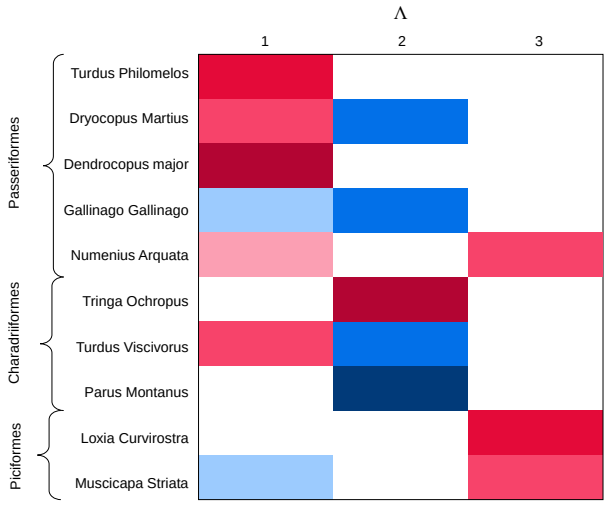
- **y**: occurrence of **p species** in **n** different **environments**;
- η: **k** latent **factors**;
- Λ: **impact of the latent factors** on the species occurrence;
- **x**: q **species characteristics** (taxonomy, size, migratory strategy...), providing similarities between different species.

Bird species occurrence example (1)

- **y**: occurrence of **p species** in **n** different **environments**;
- η: **k** latent **factors**;
- Λ: **impact of the latent factors** on the species occurrence;
- **x**: q **species characteristics** (taxonomy, size, migratory strategy...), providing similarities between different species.

Considering x indicating the phylogenetic order of each species.
If the h-th factor does not impact the occurrence of the species j ($\lambda_{jh} = 0$), it could not even impact the other species s belonging to the same order of j ($\lambda_{sh} = 0$).

Bird species occurrence example (2)



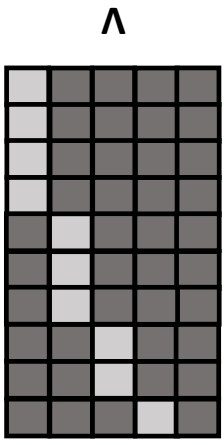
Theoretical prior properties

We define desirable properties for the GIF class including

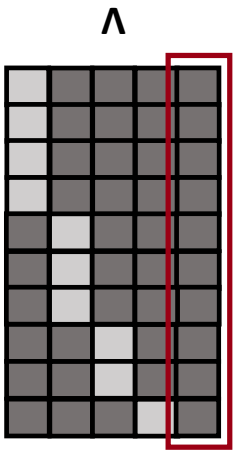
- Increasing shrinkage ($\text{var}(\lambda_{jh}) < \text{var}(\lambda_{j(h-1)})$ for any h)
- Robustness to large signals (not overshrinking)
- Asymptotic increasing sparsity (for $p \to \infty$ the sparsity rate increases)

We provide conditions for the properties to hold.

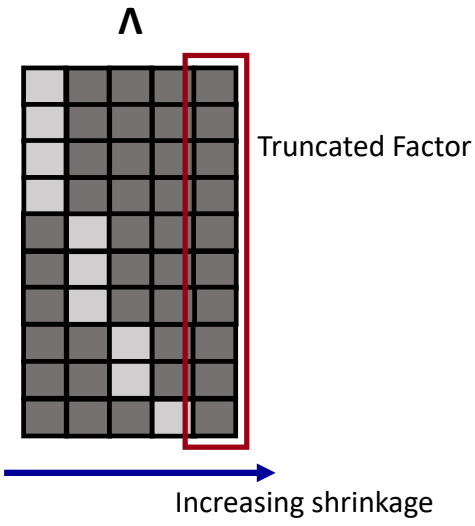
Practical consequences: sparsity and interpretability



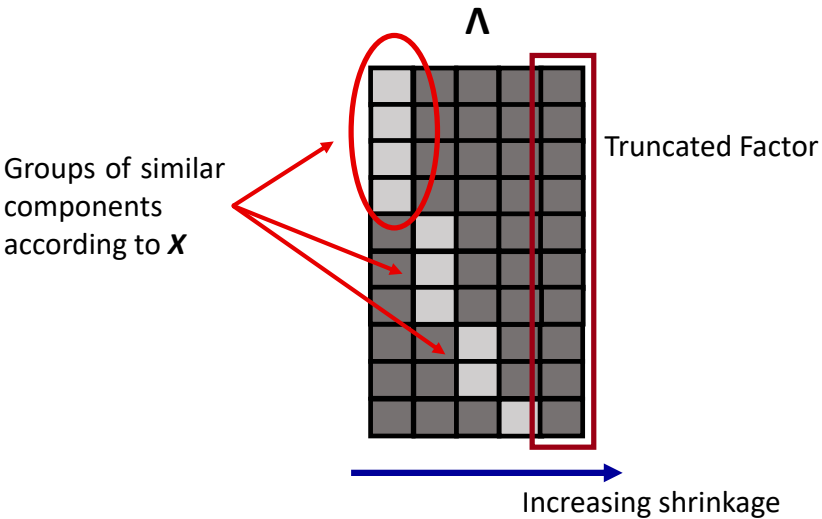
Coming back to the origin: sparsity and interpretation



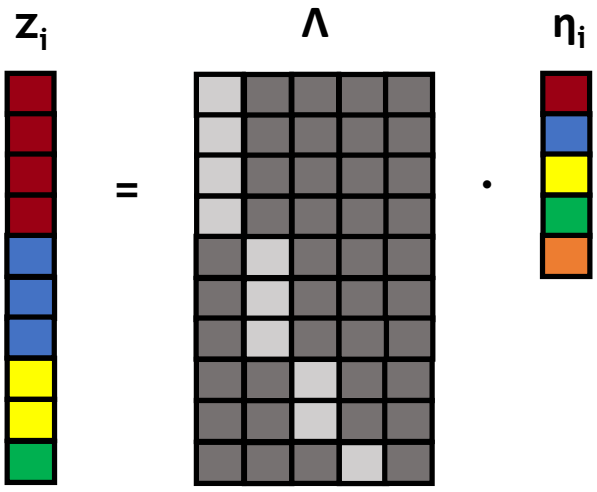
Practical consequences: sparsity and interpretability



Practical consequences: sparsity and interpretability



Practical consequences: sparsity and interpretability



Structured Increasing Shrinkage prior

$$\lambda_{jh} \mid \theta_{jh} \sim N(0, \theta_{jh}) \qquad \theta_{jh} = \tau_0 \, \gamma_h \, \phi_{jh}$$

Central GIF equations

$$\tau_0 = 1, \qquad \gamma_h = \vartheta_h \rho_h, \qquad \vartheta_h^{-1} \sim \text{Ga}(a_\theta, b_\theta),$$

Power law tail column scale

$$\rho_h = \text{Ber}\,(1 - \pi_h)\,, \qquad \pi_h = \textstyle\sum_{l=1}^{h} w_l, \qquad w_l = v_l \prod_{m=1}^{l-1}(1 - v_m), \qquad v_m \sim \text{Be}(1, \alpha),$$

Increasing shrinkage via cumulative stick-breaking process (Legramanti et al. 2020)

$$\phi_{jh} \mid \beta_h \sim \text{Ber}\{\text{logit}(X_j^\top \beta_h)\} \, \log(p)/p \qquad \beta_h \sim N_q(0, \sigma_\beta^2 I_q),$$

Meta covariates inclusion that impacts the sparsity pattern

Empirical assessment

- We compare the performance of our proposal with current approaches (Bhattacharya & Dunson, 2011, Legramanti et al., 2020)
- Scenario (a) increasing shrinkage FM (no local sparsity; Scenario (b) locally sparse FM (no increasing shrinkage); Scenario (c) is a) + b); Scenario (d) is b) + c) + metacovariate-dependence in sparsity
- Performance measures: LPML, posterior mean of k (estimated number of columns of Λ), **MSE of** Ω

Results

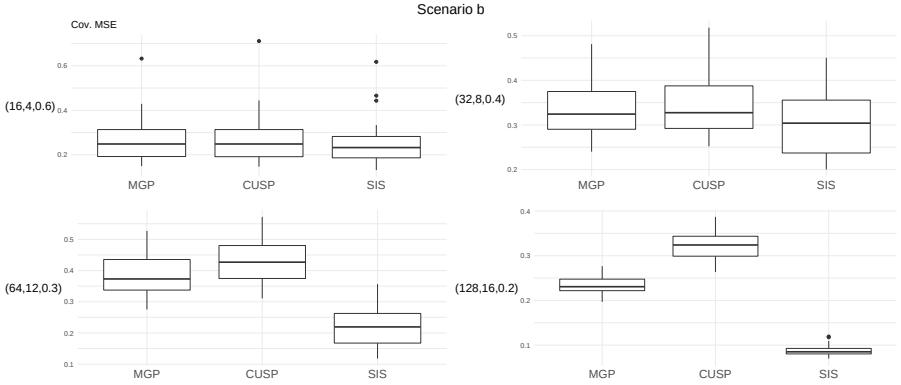


Figure: MSE for $\Omega = \Lambda\Lambda^T + \Sigma$ for different combination of (p, k, s)

The co-occurrence model

- $\boldsymbol{y}$: $n \times p$ binary matrix of occurrence of p **species** in n different **environments**.
- $\boldsymbol{w}$: $n \times c$ **covariate matrix** including habitat type and the 'spring temperature'.
- $\boldsymbol{x}$: $p \times q$ **meta covariate matrix** including **species traits**: the species log body mass, the species migratory strategy and species superfamily.

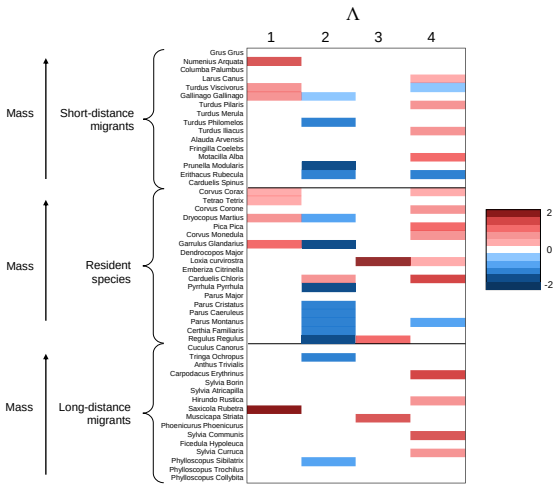
The co-occurrence model

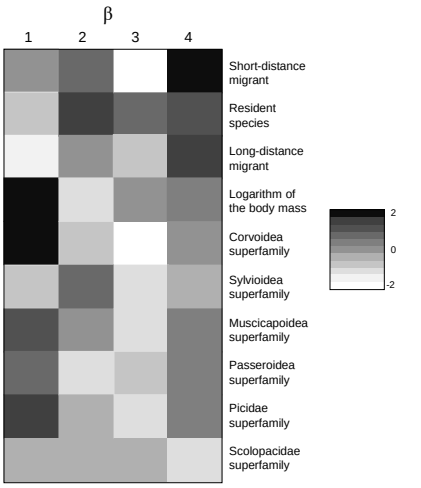
- $\boldsymbol{y}$: $n \times p$ binary matrix of occurrence of p **species** in n different **environments**.
- $\boldsymbol{w}$: $n \times c$ **covariate matrix** including habitat type and the 'spring temperature'.
- $\boldsymbol{x}$: $p \times q$ **meta covariate matrix** including **species traits**: the species log body mass, the species migratory strategy and species superfamily.

$$y_{ij} = \mathbb{1}(z_{ij} > 0), \quad z_{ij} = w_i^T \mu_j + \epsilon_{ij}, \quad \epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^T \sim N_p(0, \Lambda\Lambda^T + I_p),$$

- $\boldsymbol{z}$: $n \times p$ underlying continuous matrix.
- Λ: loadings matrix with structured increasing shrinkage prior such that the species traits x can impact the covariance structure across species.

Loadings

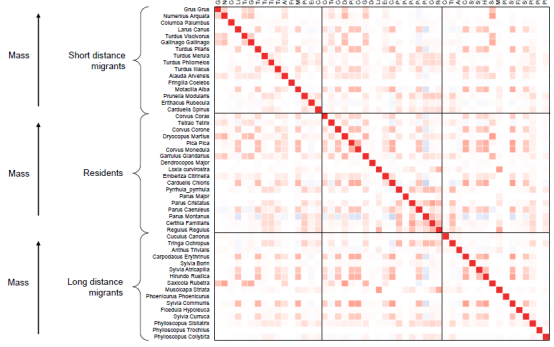

Λ
1
2
3
4
Mass
Short-distance migrants
Mass
Resident species
Mass
Long-distance migrants
Grus Grus
Numenius Arquata
Columba Palumbus
Larus Canus
Turdus Viscivorus
Gallinago Gallinago
Turdus Pilaris
Turdus Merula
Turdus Philomelos
Turdus Iliacus
Alauda Arvensis
Fringilla Coelebs
Motacilla Alba
Prunella Modularis
Erithacus Rubecula
Cardueli Spinus
Corvus Corax
Tetrao Tetrix
Corvus Corone
Dryocopus Martius
Pica Pica
Corvus Monedula
Garrulus Glandarius
Dendrocopos Major
Loxia curvirostra
Emberiza Citrinella
Cardueli Chloris
Pyrrhula Pyrrhula
Parus Major
Parus Cristatus
Parus Caeruleus
Parus Montanus
Certhia Familiaris
Regulus Regulus
Cuculus Canorus
Tringa Ochropus
Anthus Trivialis
Carpodacus Erythrinus
Sylvia Borin
Sylvia Atricapilla
Hirundo Rustica
Saxicola Rubetra
Muscicapa Striata
Phoenicurus Phoenicurus
Sylvia Communis
Ficedula Hypoleuca
Sylvia Curruca
Phylloscopus Sibilatrix
Phylloscopus Trochilus
Phylloscopus Collybita
2
0
-2

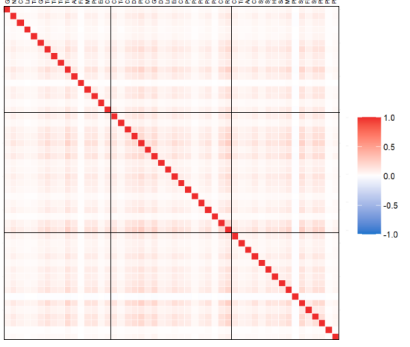

β
1
2
3
4
Short-distance migrant
Resident species
Long-distance migrant
Logarithm of the body mass
Corvoidea superfamily
Sylvioidea superfamily
Muscicapoidea superfamily
Passeroidea superfamily
Picidae superfamily
Scolopacidae superfamily
2
0
-2

Posterior mean of correlation matrices

Structured Increasing Shrinkage prior

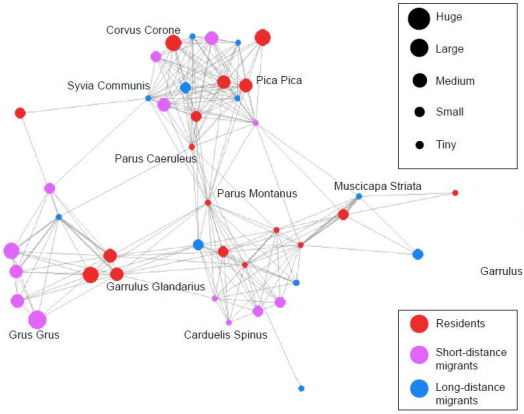
Multiplicative Gamma Process



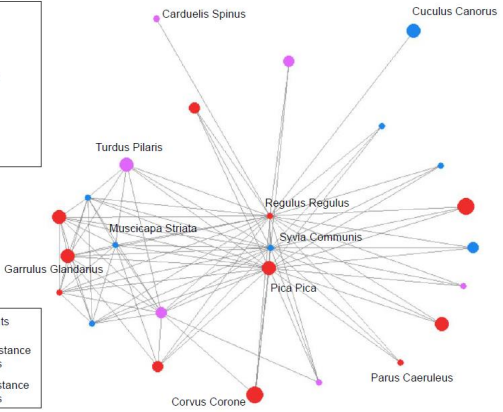


Graph representation of covariance structure

Structured Increasing Shrinkage prior



Multiplicative Gamma Process



Generalizations (1)

- We can play similar games for any high-dimensional matrix

$$Z = H\Lambda$$

  and model it as a factorization of "simpler objects"
- Assuming similar structure for H and Λ we can exploit both metacovariates and covariates helping shrinkage **(see Lorenzo Schiavon's talk of Tuesday)**

Generalizations (1)

- We can play similar games for any high-dimensional matrix

$$Z = H\Lambda$$

  and model it as a factorization of "simpler objects"
- Assuming similar structure for H and Λ we can exploit both metacovariates and covariates helping shrinkage **(see Lorenzo Schiavon's talk of Tuesday)**
- Similarly in multivariate regression

$$y = \mu + \beta X + \varepsilon \tag{1}$$

  where y is a r-dimensional response and x is a p-dimensional vector of covariates we can factorize

$$\beta = \Gamma\eta$$

Generalizations (2)

- Considering the multivariate regression model

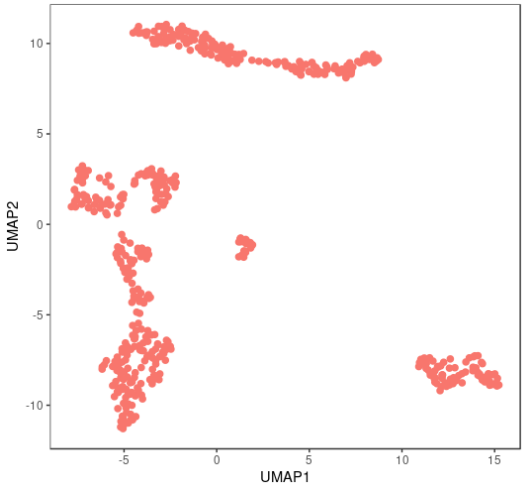
$$y = \mu + \beta X + \varepsilon \tag{2}$$

  where y is a r-dimensional response and x is a p-dimensional vector of covariates.
- Assuming $\beta = \Gamma\eta$ with Γ of dimension $u < \max\{r, p\}$ we obtain a *dimensionality reduction*
- In Envelop models (Cook et al., 2010) a set of particular assumptions on Γ and on the dependence of y from X lead to interesting results beyond dimensionality reduction.

Bayesian Envelope Models

- The rationale behind envelope models is that not all linear combinations of y are influenced by X
- This intuition is formalized assuming that there exist two matrices Γ and Γ_0 such that $O = [\Gamma, \Gamma_0]$ is orthogonal and
  1. $\Gamma_0^T y | X \sim \Gamma_0^T y$
  2. $\Gamma^T y \perp \Gamma_0^T y | X$
- See **Andrea Mascaretti's talk** on Tuesday.

Single-cell RNA-Seq data



Source: Chandra et al. (2020)

Clustering strategies in high p

- In high dimensions, a common pragmatic approach is to apply first stage dimension reduction (e.g., sparse PCA)
- Once we obtain a reduced dimensional set of features, we can cluster using classical methods such as k-means
- Such methods lack of formalism and it is not clear how to characterize uncertainty/ do inference

Model-based clustering

- Model-based clustering typically relies on mixtures:

$$y_i \sim f, \quad f(y) = \sum_{h=1}^{k} \pi_h \mathcal{K}(y; \theta_h).$$

- $\mathcal{K}(y; \theta)$ is a kernel density
- θ_h are different component-specific parameters
- Parametric mixture fix k (there are k potential clusters)
- Bayesian nonparametrics allows $k = \infty$

Model-based clustering in high p

- In high-dimensional applications, $\mathcal{K}(y; \theta_h)$ will be a high-dimensional multivariate density
- Multivariate location-scale mixture let

$$\mathcal{K}(y; \theta_h) = N(y; \mu_h, \Sigma_h)$$

- But Σ_h is $p \times p$

Model-based clustering in high p

- In high-dimensional applications, $\mathcal{K}(y; \theta_h)$ will be a high-dimensional multivariate density
- Multivariate location-scale mixture let

$$\mathcal{K}(y; \theta_h) = N(y; \mu_h, \Sigma_h)$$

- But Σ_h is $p \times p \rightarrow$ we need some sort of **dimensionality reduction**

Dimensionality reduction

- Parsimonious representation of Σ_h (Bouveyron and Brunet-Saumard, 2014)
- Factor analysis models

$$\Sigma_h = \Lambda_h \Lambda_h^T + S_h,$$

with S_h diagonal and Λ_h $(p \times d)$-dimensional $\rightarrow$ Mixture of factor analyzers by (Ghahramani et al. 1996)

Practical experience in high-dimensional clustering

- In implementing Bayes model-based clustering in high-dimensions, we typically use MCMC samplers
- As p increases, common to obtain poor performance for a variety of model specifications & data
- Solutions to avoid MCMC issues (Celeux et al., 2018, Fruehwirth-Schnatter, 2006)
- We often see either **way too many clusters or way too few**

Practical experience in high-dimensional clustering

- In implementing Bayes model-based clustering in high-dimensions, we typically use MCMC samplers
- As p increases, common to obtain poor performance for a variety of model specifications & data
- Solutions to avoid MCMC issues (Celeux et al., 2018, Fruehwirth-Schnatter, 2006)
- We often see either **way too many clusters or way too few**

Poor MCMC mixing or an intrinsic property of the posterior in high dimensions?

Limiting behavior of model-based clustering

- We studied the limiting clustering posterior as $p \to \infty$ with n fixed

Limiting behavior of model-based clustering

- We studied the limiting clustering posterior as $p \to \infty$ with n fixed
- Consider a broad class of mixtures (Stick-breaking priors, MFM)

Limiting behavior of model-based clustering

- We studied the limiting clustering posterior as $p \to \infty$ with n fixed
- Consider a broad class of mixtures (Stick-breaking priors, MFM)
- Letting $c_i \in \{1, \ldots, \infty\}$ denote the cluster id for subject i, $y_i \mid c_i = h \sim \mathcal{K}(y_i; \theta_h)$, $\theta_h \sim P_0$

Limiting behavior of model-based clustering

- We studied the limiting clustering posterior as $p \to \infty$ with n fixed
- Consider a broad class of mixtures (Stick-breaking priors, MFM)
- Letting $c_i \in \{1, \ldots, \infty\}$ denote the cluster id for subject i, $y_i \mid c_i = h \sim \mathcal{K}(y_i; \theta_h)$, $\theta_h \sim P_0$
- Posterior probability of partition Ψ induced by clusters $c_1, \ldots, c_n$ conditionally on data $\mathbf{y} = (y_1, \ldots, y_n)^T$ is

$$\Pi(\Psi \mid \mathbf{y}) = \frac{\Pi(\Psi) \times \prod_{h \geq 1} \int \prod_{i:c_i=h} \mathcal{K}(y_i; \theta) dP_0(\theta)}{\sum_{\Psi' \in P} \Pi(\Psi') \times \prod_{h \geq 1} \int \prod_{i:c_i=h} \mathcal{K}(y_i; \theta) dP_0(\theta)}.$$

Limiting behavior

Theorem

Let $y_1, \ldots, y_n$ denote iid draws from p-variate continuous density f_0. Let Ψ denote the partition induced by the cluster labels $c_1, \ldots, c_n$, and let $c_1', \ldots, c_n'$ denote a new set of cluster labels obtained from $c_1, \ldots, c_n$ by merging an arbitrary pair of clusters, with Ψ' the related partition. If

$$\limsup_{p \to \infty} \frac{\prod_{h \geq 1} \int \prod_{i:c_i=h} \mathcal{K}(y_i; \theta) dP_0(\theta)}{\prod_{h \geq 1} \int \prod_{i:c_i'=h} \mathcal{K}(y_i; \theta) dP_0(\theta)} = 0 \text{ in } f_0\text{-probability},$$

then $\lim_{p \to \infty} \Pi(c_1 = \cdots = c_n \mid \mathbf{y}) = 1$. *Else if*

$$\liminf_{p \to \infty} \frac{\prod_{h \geq 1} \int \prod_{i:c_i=h} \mathcal{K}(y_i; \theta) dP_0(\theta)}{\prod_{h \geq 1} \int \prod_{i:c_i'=h} \mathcal{K}(y_i; \theta) dP_0(\theta)} = \infty \text{ in } f_0\text{-probability},$$

then $\lim_{p \to \infty} \Pi(c_1 \neq \cdots \neq c_n \mid \mathbf{y}) = 1$.

What does this Theorem mean?

- There are two different **bad limit posteriors**
- The 1st puts probability one on assigning all observations to **the same cluster**
- The 2nd puts probability one on assigning all observations to **their own cluster**
- Whether we converge to one of these very bad limits depends on a ratio of marginal likelihoods (MLs) and not on the specific EPPF

Examples

Consider the following important special cases

$$y_i \sim f, \quad f(y) = \sum_{h=1}^{\infty} \pi_h N_p(y; \xi_h, \Sigma), \quad \xi_h|\Sigma \sim P_{\xi|\Sigma}, \quad \Sigma \sim P_\Sigma.$$

Corollary 1

Assume $\xi_h = \mu_h$ with $\mu_h \in \mathbb{R}^p$. If $||y_i||^2 = O_p(p)$, $\mu_h|\Sigma \sim N_p(\mu_0, \kappa_0^{-1}\Sigma)$ and $\Sigma \sim IW(\nu_0, \Lambda_0)$, where $\mu_0, \kappa_0, \nu_0, \Lambda_0$ are the hyperparameters with $\nu_0 > p - 1$ and $\nu_0 = O(p)$, then $\Pi(c_1 = \cdots = c_n \mid \mathbf{y}) \to 1$.

Corollary 2

Assume $\Sigma = \sigma^2 I_p$ and $\xi_h = \mu_h 1_p$ where 1_p is a p vector of ones. If $||y_i||^2 = O_p(p)$, $\mu_h|\sigma^2 \sim N(\mu_0, \kappa_0^{-1}\sigma^2)$, and $\sigma^2 \sim IG(\nu_0, \lambda_0)$, where $\mu_0, \kappa_0, \nu_0, \lambda_0$ are fixed hyperparameters, then $\Pi(c_1 \neq \cdots \neq c_n \mid \mathbf{y}) \to 1$.

Comments

Poor MCMC mixing or an intrinsic property of the posterior in high dimensions?

It is not (just) an MCMC issue!

- too parsimonious model $\rightarrow$ all the observations to the same cluster for $p \rightarrow \infty$
- too flexible model $\rightarrow$ all the observations to n different clusters for $p \rightarrow \infty$
- This property holds regardless of the data or true data generating model!

So what?

The question is …

is it possible to define models that can circumvent this pitfall?

LAtent Mixtures for Bayesian (Lamb) clustering

LAtent Mixtures for Bayesian (Lamb) clustering

- We propose to cluster on the latent factor level as follows:

$$y_i \sim N_p(\Lambda\eta_i, \Sigma), \qquad \eta_i \sim \sum_{h=1}^{\infty} \pi_h N_d(\mu_h, \Delta_h)$$

LAtent Mixtures for Bayesian (Lamb) clustering

- We propose to cluster on the latent factor level as follows:

$$y_i \sim N_p(\Lambda\eta_i, \Sigma), \qquad \eta_i \sim \sum_{h=1}^{\infty} \pi_h N_d(\mu_h, \Delta_h)$$

- η_i's = d-dimensional latent factors with $d < n$

LAtent Mixtures for Bayesian (Lamb) clustering

- We propose to cluster on the latent factor level as follows:

$$y_i \sim N_p(\Lambda\eta_i, \Sigma), \qquad \eta_i \sim \sum_{h=1}^{\infty} \pi_h N_d(\mu_h, \Delta_h)$$

- η_i's = d-dimensional latent factors with $d < n$
- $\Lambda = p \times d$ factor loading matrix

LAtent Mixtures for Bayesian (Lamb) clustering

- We propose to cluster on the latent factor level as follows:

$$y_i \sim N_p(\Lambda\eta_i, \Sigma), \qquad \eta_i \sim \sum_{h=1}^{\infty} \pi_h N_d(\mu_h, \Delta_h)$$

- η_i's = d-dimensional latent factors with $d < n$
- $\Lambda = p \times d$ factor loading matrix
- $\Sigma = p \times p$ diagonal matrix

LAtent Mixtures for Bayesian (Lamb) clustering

- We propose to cluster on the latent factor level as follows:

$$y_i \sim N_p(\Lambda\eta_i, \Sigma), \qquad \eta_i \sim \sum_{h=1}^{\infty} \pi_h N_d(\mu_h, \Delta_h)$$

- η_i's = d-dimensional latent factors with $d < n$
- $\Lambda = p \times d$ factor loading matrix
- $\Sigma = p \times p$ diagonal matrix
- Similar models (but arising with different motivations) have been proposed by Galimberti et al. (2009); Baek et al. (2010); Montanari and Viroli (2010).

Definition

Let η_0 be the true values of the latent variables. We define the **Bayes oracle partition probability** as

$$\Pi(\Psi \mid \eta_0) = \frac{\Pi(\Psi) \times \int \prod_{h \geq 1} \prod_{i:c_i=h} \mathcal{K}(\eta_{0i}; \theta_h) dG_0(\theta)}{\sum_{\Psi' \in P} \Pi(\Psi') \times \int \prod_{h \geq 1} \prod_{i:c'_i=h} \mathcal{K}(\eta_{0i}; \theta_h) dG_0(\theta)}.$$

Consistency Properties

a) $y_i \sim N_p(\Lambda_0 \eta_{0i}, \sigma_0^2 I_p)$, for each $i = 1, \ldots, n$;

b) $\lim_{p\to\infty} \left|\left|\frac{1}{p}\Lambda_0^T\Lambda_0 - M\right|\right|_2 = 0$ with M some PD matrix

c) $\sigma_L < \sigma_0 < \sigma_U$ where σ_L and σ_U are known constants;

d) $||\eta_{0i}|| = O(1)$ for each $i = 1, \ldots, n$;

Theorem

Under $a) - d)$ for all $\Psi \in \mathcal{P}$

$$\lim_{p\to\infty} \Pi(\Psi \mid y) = \Pi(\Psi \mid \boldsymbol{\eta}_0).$$

Simulation study

Scenario 1 Mixture of factor analyzers

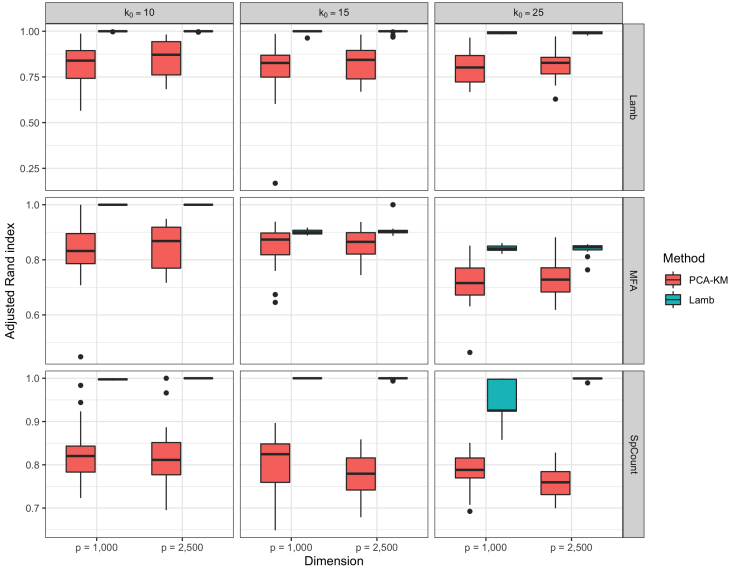
$$y_i \sim \sum_{h=1}^{k_0} \pi_h N_p(\mu_h, \Lambda_h \Lambda_h^T + \sigma^2 I_p).$$

Scenario 2 Lamb model

$$y_i \sim \sum_{h=1}^{k_0} \pi_h N_p(\Lambda \mu_h, \Lambda \Delta_h \Lambda^T + \sigma^2 I_p).$$

Scenario 3 Mixture of log transformed zero inflated Poisson (mimiking the sc-RNASeq data)

Simulation study's results



Application on scRNASeq data

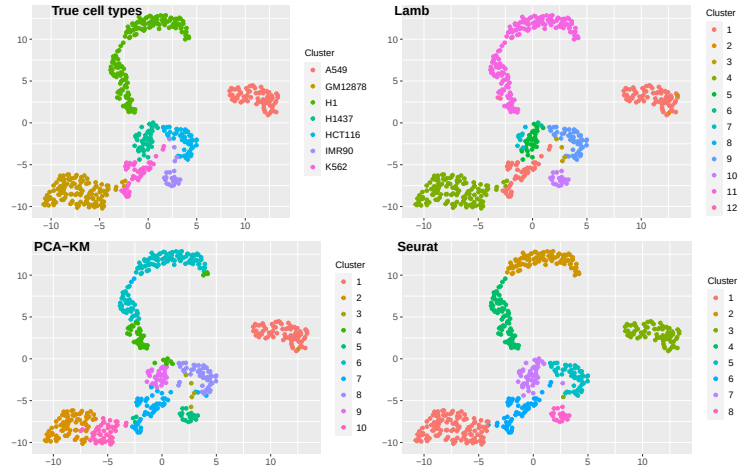


Figure: (a) true clustering; (b) Lamb estimate; (c) PCA-KM estimate; (d) Seurat estimate [Lamb has adjusted Rand index 0.956 vs 0.897 for Seurat & 0.831 for PCA-KM]

Take home messages

Take home messages

- In some (many?) statistical models high-dimensional data are problematic

Take home messages

- In some (many?) statistical models high-dimensional data are problematic
  - *"Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it"* (J. Friedman, 2001)

Take home messages

- In some (many?) statistical models high-dimensional data are problematic
  - *"Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it"* (J. Friedman, 2001)
  - Classical (or novel?) dimensionality reduction approaches are a possible solution

Take home messages

- In some (many?) statistical models high-dimensional data are problematic
  - *"Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it"* (J. Friedman, 2001)
  - Classical (or novel?) dimensionality reduction approaches are a possible solution
- Sparsity (e.g. in IFM) is beneficial also for providing useful insights

Take home messages

- In some (many?) statistical models high-dimensional data are problematic
  - "*Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it*" (J. Friedman, 2001)
  - Classical (or novel?) dimensionality reduction approaches are a possible solution
- Sparsity (e.g. in IFM) is beneficial also for providing useful insights
  - "*Statistical modeling: The two cultures*" (L. Breiman, 2001)

Take home messages

- In some (many?) statistical models high-dimensional data are problematic
  - "*Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it*" (J. Friedman, 2001)
  - Classical (or novel?) dimensionality reduction approaches are a possible solution
- Sparsity (e.g. in IFM) is beneficial also for providing useful insights
  - "*Statistical modeling: The two cultures*" (L. Breiman, 2001)
  - "*To exaplain or to predict?*" (G. Shmueli, 2010)

Take home messages

- In some (many?) statistical models high-dimensional data are problematic
  - "*Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it*" (J. Friedman, 2001)
  - Classical (or novel?) dimensionality reduction approaches are a possible solution
- Sparsity (e.g. in IFM) is beneficial also for providing useful insights
  - "*Statistical modeling: The two cultures*" (L. Breiman, 2001)
  - "*To exaplain or to predict?*" (G. Shmueli, 2010)
- Why Bayesian?

Take home messages

- In some (many?) statistical models high-dimensional data are problematic
  - "*Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it*" (J. Friedman, 2001)
  - Classical (or novel?) dimensionality reduction approaches are a possible solution
- Sparsity (e.g. in IFM) is beneficial also for providing useful insights
  - "*Statistical modeling: The two cultures*" (L. Breiman, 2001)
  - "*To exaplain or to predict?*" (G. Shmueli, 2010)
- Why Bayesian?
  - Uncertainty assessment

Take home messages

- In some (many?) statistical models high-dimensional data are problematic
  - "*Every time the amount of data increases by a factor of ten, we should totally rethink how we analyze it*" (J. Friedman, 2001)
  - Classical (or novel?) dimensionality reduction approaches are a possible solution
- Sparsity (e.g. in IFM) is beneficial also for providing useful insights
  - "*Statistical modeling: The two cultures*" (L. Breiman, 2001)
  - "*To exaplain or to predict?*" (G. Shmueli, 2010)
- Why Bayesian?
  - Uncertainty assessment
  - Learning of the latent dimension (in IFM, Envelope Models, Lamb)

Acknowledgements

This is a joint work with

Andrea Mascaretti
(University of Padova)

Lorenzo Schiavon
(University of Padova)

Noirrit K. Chandra
(University of Texas)

David B. Dunson
(Duke University)

References

- A. Bhattacharya and D. B. Dunson (2011). Sparse Bayesian infinite factor models. *Biometrika*
- N. K. Chandra, A. Canale, D.B. Dunson, (2020), Escaping the curse of dimensionality in Bayesian model based clustering *arxiv:2006.02700*
- Cook, R.D., Bing, L., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statistica Sinica*.
- D. Kowal and A. Canale, (2021), Semiparametric Functional Factor Models with Bayesian Rank Selection *arxiv:2108.02151*
- S. Legramanti, D. Durante, and D. B. Dunson (2020). Bayesian cumulative shrinkage for infinite factorizations. *Biometrika*.
- S. Montagna, S. T. Tokdar, B. Neelon, and D. B. Dunson (2012). Bayesian latent factor regression for functional and longitudinal data. *Biometrics*.
- J. S. Murray, D. B. Dunson, L. Carin, and J. E. Lucas (2013). Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association*.
- L. Schiavon, A. Canale, and D. B. Dunson (2022). Generalized infinite factorization models. *Biometrika*.

Thank you for your attention!

University of Padova …

…since 1222