A Low-rank Bayesian Nonparametric Model for Binary Matrices

Adrien.Todeschini@inria.fr and François.Caron@stats.ox.ac.uk

Introduction

Binary matrices Describe relations between 2 types of entities:

- **readers / books**
- users / movies
- customers / products
- authors / publications
- ...

Aims

- **Interpretability** via a **low-rank** representation
- Capture **power-law** degree distributions of real world datasets
- Develop efficient computational procedure for posterior simulation

BNP approach

Elegant and useful

- Potential number of books may be very large and considered **infinite**
- Captures **power-law** properties

Previous work (rank one)

- Indian buffet process [GG05]
- Beta-Bernoulli process [TJ07]
- Stable Beta process [TG09]
- BNP models for bipartite graphs [Car12]

Formulation Represent the set of books read by all readers by a collection of atomic measures $(Z_1, ..., Z_n)$

$$Z_i = \sum_{j=1}^{\infty} z_{ij}\delta_{\theta_j}$$

where Z_i represents the set of books read by reader i by a point process

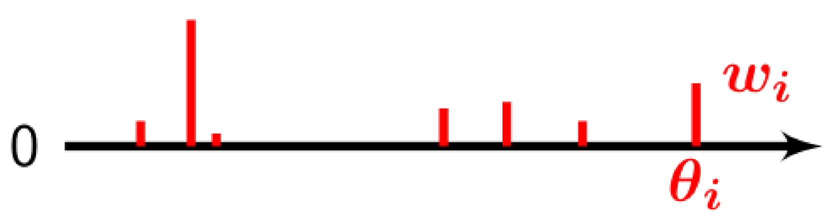
- $z_{ij} = 1$ if reader i has read book j, 0 otherwise
- $\{\theta_j\}$ is the set of books

Completely random measures (CRM) [Kin67]

Random masses $w_j > 0$ at random locations $\theta_j \in \Theta$ characterized by a Poisson process over $\mathbb{R}^+ \times \Theta$

$$W = \sum_{j=1}^{\infty} w_j\delta_{\theta_j}$$

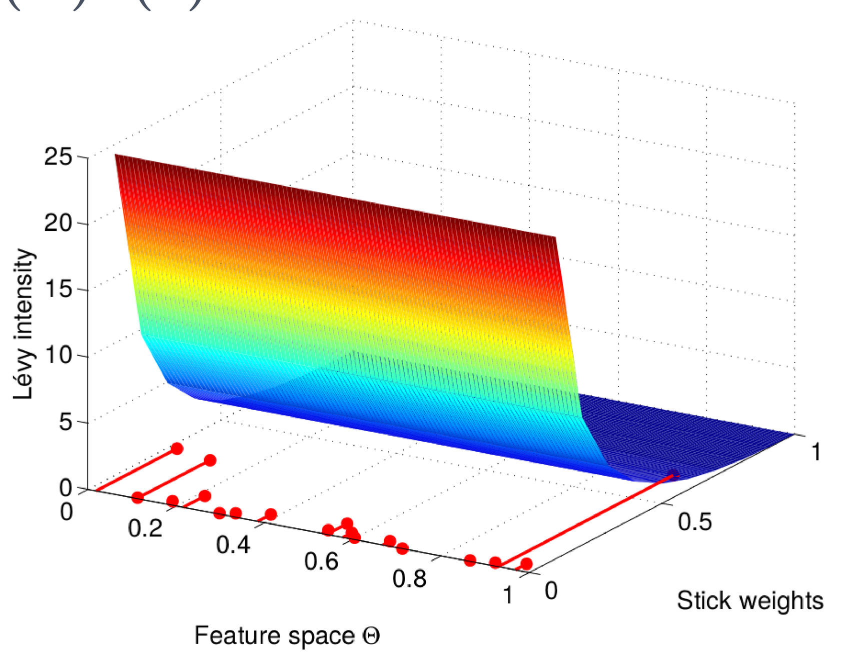
$$\{(w_j, \theta_j)\} \sim \text{PP}(\nu)$$



Homogeneous CRM $\nu(dw, d\theta) = \rho(w)h(\theta)dwd\theta$

$$W \sim \text{CRM}(\rho, h)$$

$$w_j \sim \text{PP}(\rho) \quad \perp\!\!\!\perp \quad \theta_j \overset{\text{iid}}{\sim} H$$



with finite total mass ⇒

$$\int_0^{\infty} (1 - e^{-w})\rho(dw) < \infty$$

Generalized gamma process (GGP) Lévy intensity:

$$\rho(dw) = \frac{\alpha}{\Gamma(1-\sigma)} w^{-1-\sigma} \exp(-w\tau)dw$$

where $\alpha > 0$ and $\{\sigma \in (0, 1), \tau \geq 0\}$ or $\{\sigma = 0, \tau > 0\}$

- Infinitely many atoms: $\int_0^{\infty} \rho(dw) = \infty$
- Interests: **generality**, **interpretability**, attractive **conjugacy** properties
- Admits as special cases: gamma process ($\sigma = 0$), inverse Gaussian process ($\sigma = \frac{1}{2}$), stable process ($\tau = 0$)
- Exhibits **power-law** behavior when $\sigma > 0$

References

Car12 Francois Caron. Bayesian nonparametric models for bipartite graphs. In F. Pereira, C.J.C. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 2051–2059. Curran Associates, Inc., 2012.

FT12 S. Favaro and Y.W. Teh. MCMC for normalized random measure mixture models. Technical report, University of Turin, 2012.

GG05 T Griffiths and Z. Ghahramani. Infinite latent feature models and the Indian buffet process. In *NIPS*, 2005.

GL14 J. E. Griffin and F. Leisen. Compound random measures and their use in Bayesian nonparametrics. *arXiv preprint arXiv:1410.0611*, 2014.

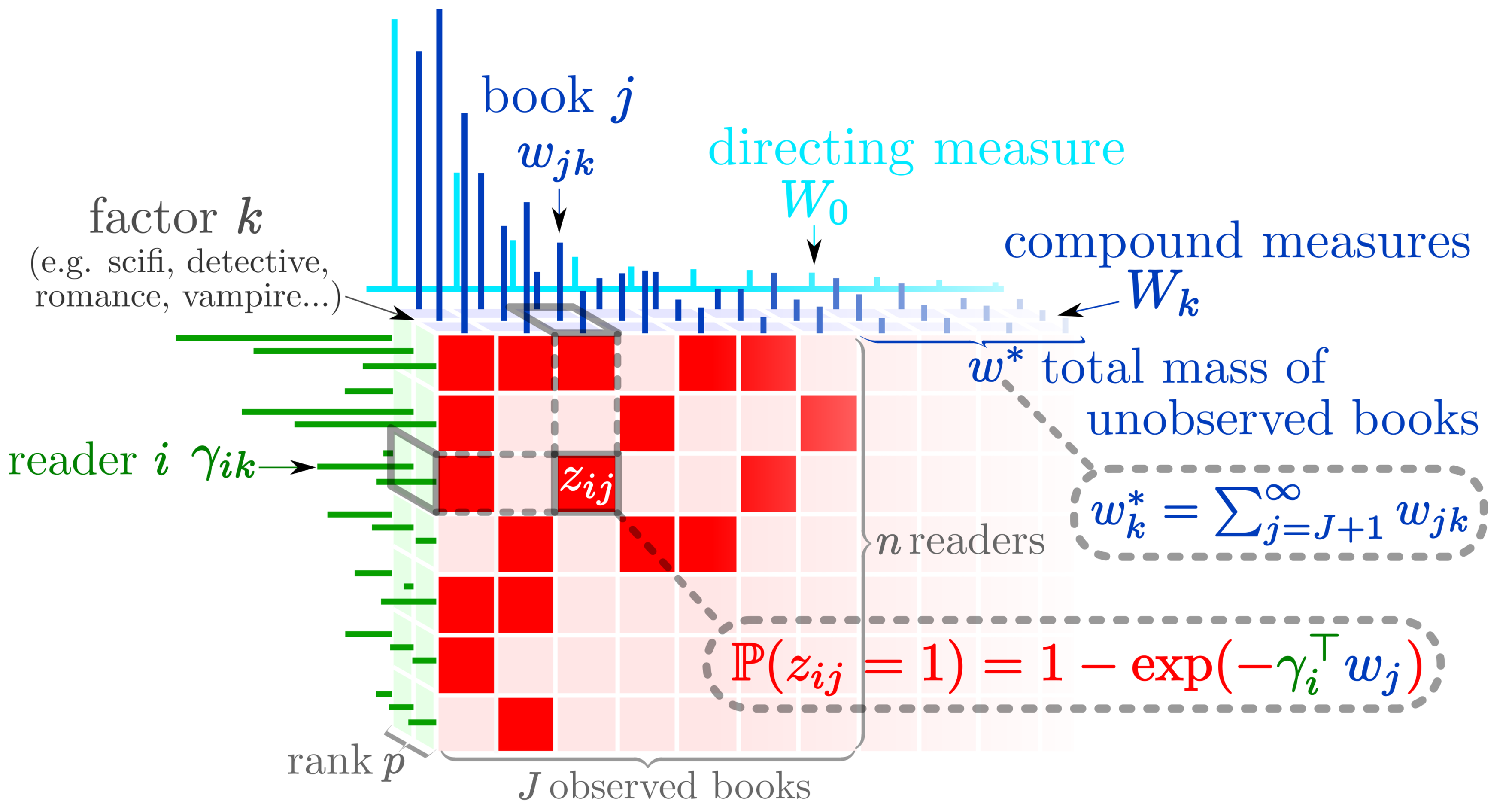
Kin67 J.F.C. Kingman. Completely random measures. *Pacific Journal of Mathematics*, 21(1):59–78, 1967.

TG09 Y.W. Teh and D. Görür. Indian buffet processes with power-law behavior. In *NIPS*, 2009.

TJ07 R. Thibaux and M. Jordan. Hierarchical beta processes and the Indian buffet process. In *International Conference on Artificial Intelligence and Statistics*, volume 11, pages 564–571, 2007.

VDP08 David A Van Dyk and Taeyoung Park. Partially collapsed gibbs samplers: Theory and methods. *Journal of the American Statistical Association*, 103(482):790–796, 2008.

Probabilistic model



Likelihood

$$z_{ij}|\gamma_i, w_j \sim \text{Ber}\left(1 - \exp\left(-\sum_{k=1}^{p} \gamma_{ik} w_{jk}\right)\right)$$

Prior for readers parameters

$$\gamma_{ik} \sim \text{Gamma}(a_k, b_k)$$

Nonparametric prior for books

Compound random measure [GL14] The weights w_{jk} come from a multivariate random measure

$$(W_1, ..., W_p) \sim \text{CCRM}(\rho_0, \lambda_{1:p}, h)$$

$$W_k = \sum_{j=1}^{\infty} w_{jk}\delta_{\theta_j}$$

where ρ_0 is the Lévy intensity of a GGP(α, σ, τ)

Hierarchical construction

$$w_{jk} = w_{j0}\beta_{jk} \quad \text{where}$$

w_{j0} come from the directing measure W_0

$$W_0 \sim \text{CRM}(\rho_0, h)$$

$$W_0 = \sum_{j=1}^{\infty} w_{j0}\delta_{\theta_j}$$

β_{jk} are gamma distributed

$$\beta_{jk} \sim \text{Gamma}(\lambda_k, \lambda_k)$$

so that λ tunes the correlation between the measures W_k.

Inference

Goal Approximate $p(\gamma_{1:n,1:p}, w_{1:J,1:p}, w^*_{1:p}|Z_{1:n})$

Gibbs sampler We introduce a set of latent variables to have conjugacy properties. At each MCMC iteration:

- Update latent variables
- Update γ_{ik}|rest $\sim$ Gamma, $i = 1, ..., n$, $k = 1, ..., p$
- Update w_{jk}|rest $\sim$ Gamma, $j = 1, ..., J$, $k = 1, ..., p$
- Update w_k^*|rest, $k = 1, ..., p$ using adaptive thinning strategy [FT12]

Hyperparameters $\{\alpha, \sigma, \tau, \lambda_{1:p}, a_{1:p}, b_{1:p}\}$ updated using partially collapsed Gibbs [VDP08] for good mixing

Future work

- Scalable inference
- Experiments on real world datasets
- Nonparametric prior over the parameters of readers γ
- Low-rank BNP models for symmetric matrices (adjacency in simple graphs)
- Binary tensor data