
Supplementary Material

Bayesian Sparse Factor Models and DAGs

Inference and Comparison

Ricardo Henao
 DTU Informatics
 Technical University of Denmark
 2800 Lyngby, Denmark
 Bioinformatics Centre
 University of Copenhagen
 2200 Copenhagen, Denmark
 rhenao@binf.ku.dk

Ole Winther
 DTU Informatics
 Technical University of Denmark
 2800 Lyngby, Denmark
 Bioinformatics Centre
 University of Copenhagen
 2200 Copenhagen, Denmark
 owi@imm.dtu.dk

1 Inference

Given a set of N observations in d dimensions, the data $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$ and a number of factors equal to the number of dimensions $m = d$, MCMC analysis is rather standard and can be implemented through Gibbs sampling. Note that in the following, $\mathbf{X}_{i\cdot}$ and $\mathbf{X}_{\cdot i}$ are rows and columns of \mathbf{X} respectively, i, j, n are indexes for dimensions, factors and observations respectively. In the following we describe the conditional distributions needed to sample from the factor model hierarchy. To sample from the DAG model we only have to replace \mathbf{Z} by \mathbf{X} , Ψ by \mathbf{Z} , \mathbf{A} by \mathbf{B} . Note also that \mathbf{B} is strictly lower triangular, thus we only need to sample for the elements in its lower part.

Noise variance We can sample each element of Ψ independently one at the time using

$$\psi_i^{-1} | \mathbf{X}_{i\cdot}, \mathbf{A}_{i\cdot}, \mathbf{Z}, \mathbf{D}_i, s_s, s_r \sim \text{Gamma} \left(\psi_i^{-1} \left| s_s + \frac{N+d}{2}, s_r + \mathbf{C} \right. \right),$$

where \mathbf{D}_i is a diagonal matrix with entries τ_{ij}^{-1} and

$$\mathbf{C} = \frac{1}{2} (\mathbf{X}_{i\cdot} - \mathbf{A}_{i\cdot} \mathbf{Z}) (\mathbf{X}_{i\cdot} - \mathbf{A}_{i\cdot} \mathbf{Z})^\top + \frac{1}{2} \mathbf{A}_{i\cdot} \mathbf{D}_i^{-1} \mathbf{A}_{i\cdot}^\top.$$

Factors The conditional distribution of \mathbf{Z} depends on $\pi(z_{jn} | \cdot)$ and is written independently for each element z_{jn} by

$$z_{jn} | \mathbf{X}_{\cdot n}, \mathbf{A}, \Psi, v_{jn} \sim \mathcal{N}(z_{jn} | \mathbf{A}_{\cdot j}^\top \Psi^{-1} \boldsymbol{\epsilon}_{\setminus jn}, c), \quad c = (\mathbf{A}_{\cdot j}^\top \Psi^{-1} \mathbf{A}_{\cdot j} + v_{jn}^{-1})^{-1},$$

where $\boldsymbol{\epsilon}_{\setminus jn} = \mathbf{X}_{\cdot n} - \mathbf{A} \mathbf{Z}_{\cdot n} |_{z_{jn}=0}$. In addition, we need to sample from the conditional of v_{jn} using

$$v_{jn}^{-1} | z_{jn}, \lambda \sim \text{IG} \left(v_{jn}^{-1} \left| \frac{\lambda}{z_{jn}}, \lambda^2 \right. \right),$$

where $\text{IG}(\cdot | \mu, \lambda)$ is the inverse Gaussian distribution with mean μ and scale parameter λ [1]. Furthermore, the conditional of λ^2 is given by

$$\lambda^2 | \nu_{ij}, \ell_s, \ell_r \sim \text{Gamma} \left(\lambda^2 \left| \ell_s + mN, \ell_s + \frac{1}{2} \sum_{j=1}^m \sum_{n=1}^N v_{ij} \right. \right).$$

Mixing matrix In order to sample each element a_{ij} from the conditional distribution of \mathbf{A} we use

$$a_{ij} | \mathbf{X}_{i:}, \mathbf{A}_{\setminus ij}, \mathbf{Z}_{j:}, \psi_i, \tau_{ij} \sim \mathcal{N}(a_{ij} | c \epsilon_{\setminus ij} \mathbf{Z}_{j:}^\top, c \psi_i), \quad c = (\mathbf{Z}_{j:} \mathbf{Z}_{j:}^\top + \tau_{ij}^{-1})^{-1},$$

where $\epsilon_{\setminus ij} = \mathbf{X}_{i:} - \mathbf{A}_{i:} \mathbf{Z}_{j:} |_{\mathbf{A}_{ij}=0}$ and noting that we only need to sample those elements a_{ij} for which $r_{ij} = 1$, i.e. just the slab distribution. Sampling from the conditional distributions for τ_{ij} can be done using

$$\tau_{ij}^{-1} | a_{jn}, t_s, t_r \sim \text{Gamma} \left(\tau_{ij}^{-1} \left| t_s + \frac{1}{2}, t_r + \frac{a_{ij}^2}{2\psi_i} \right. \right).$$

The conditional distributions for the remaining parameters in the slab and spike prior can be written as

$$r_{ij} | \mathbf{X}_{i:}, \mathbf{A}_{\setminus ij}, \mathbf{Z}_{j:}, \psi_i, \tau_{ij}, \eta_{ij} \sim \text{Bernoulli} \left(r_{ij} \left| \frac{\xi_{\eta_{ij}}}{1 + \xi_{\eta_{ij}}} \right. \right),$$

where

$$\xi_{\eta_{ij}} = \frac{\eta_{ij}}{1 - \eta_{ij}} \frac{\psi_i^{1/2}}{(\mathbf{Z}_{j:} \mathbf{Z}_{j:}^\top + \tau_{ij}^{-1})^{1/2}} \exp \left(\frac{(\epsilon_{\setminus ij} \mathbf{Z}_{j:}^\top)^2}{2\psi_i (\mathbf{Z}_{j:} \mathbf{Z}_{j:}^\top + \tau_{ij}^{-1})} \right),$$

and

$$\eta_{ij} | r_{ij}, \alpha_p, \alpha_m \sim \text{Beta}(\eta_{ij} | \alpha_p \alpha_m + r_{ij}, \alpha_p (1 - \alpha_m) + 1 - r_{ij}),$$

just if $q_{ij} = 1$, otherwise $\eta_{ij} = 0$.

Finally, for the shared sparsity rate as

$$q_{ij} | \alpha_m, \nu_j \sim \text{Bernoulli} \left(q_{ij} \left| 1 - \frac{\nu_j (1 - \alpha_m)}{1 + \nu_j \alpha_m} \right. \right),$$

and

$$\nu_j | q_{ij}, \beta_p, \beta_m \sim \text{Beta}(\nu_j | \beta_p \beta_m + \sum_{i=1}^d q_{ij}, \beta_p (1 - \beta_m) + d - \sum_{i=1}^d q_{ij}).$$

References

- [chhikara89](#)[1] R. S. Chhikara and L. Folks. *The inverse Gaussian distribution: theory, methodology, and applications*. M. Dekker, New York, 1989.