On Posterior Consistency of Bayesian Factor Model

Coulet Maxime, Mactha Yassine, Stepien Léo

January 25, 2025

1. Introduction

Latent Factor models are a powerful technique for dimensionality reduction, extending Principal Component Analysis (PCA) by identifying latent factors that explain the covariance structure of the data. A Bayesian formulation of a latent factor model is as follows: given a set of N observations $Y = (y_1, \ldots, y_N)$, where $(y_i)_{i=1,\ldots,N}$ are G-dimensional vectors of variables, we assume the relationship

$$\overset{G\times 1}{y_i} = \overset{G\times K}{B}\overset{K\times 1}{\omega_i} + \overset{G\times 1}{\varepsilon_i}, \tag{1}$$

where latent factors ω_i and idiosyncratic terms ε_i are modeled by imposing specific densities. Similar to any Bayesian framework, the prior on the loadings B is crucial to the estimated model's properties. The Normal Bayesian factor model assumes the following:

$$\begin{aligned} \textbf{Likelihood:} \quad & y_i|\omega_i, B, \Sigma \overset{i.i.d.}{\sim} \mathcal{N}_G\left(B\omega_i, \Sigma\right), \\ \textbf{Latent Factors:} \quad & \omega_i \overset{i.i.d.}{\sim} \mathcal{N}_K\left(0_K, I_K\right) \end{aligned} \tag{2}$$

Similar to PCA, latent factor models are well-suited for handling large datasets with numerous observations. However, interpreting the coefficients of the loadings matrix presents two challenges. First, the loadings matrix is not identifiable because any right orthogonal transformation of the loadings matrix preserves the factorization. Specifically, for any orthogonal matrix O of dimension $K \times K$, the transformed expression $BO' \times O\omega_i = B\omega_i$ holds, and $O\omega_i$ retains the centered normal distribution with covariance $OO' = I_K$. Second, imposing sufficient restrictions on B to ensure identifiability becomes increasingly complex as the number of variables G increases, requiring many additional assumptions in both model design and inference.

To address these challenges, researchers have drawn inspiration from the LASSO technique, which suppresses the least explanatory variables by shrinking their coefficients to 0. They have applied similar methods to latent factor models, aiming to shrink to 0 the coefficients in B that capture non-significant effects of a factor on a variable. For instance, Ročková and George (2016) relies on empirical Bayes methods and a parameter-augmented version of the

Expectation-Maximization (EM) algorithm to estimate the posterior mode under a Spike-and-Slab LASSO prior imposed on the coefficients of B. Ma and Liu (2022) finds that, when factors are normally distributed the posterior distribution of the loadings exhibits "magnitude inflation" where all non-zero coefficients have absolute values larger than their true counterparts. Building on the insight that if the model is consistent under a flat prior for B, it should also be approximately consistent under a Spike-and-Slab prior, Ma and Liu (2022) propose replacing the normal factor assumption with an orthonormal factor assumption.

2. Magnitude Inflation Phenomenon

Ročková and George (2016)'s Spike-and-Slab LASSO prior is defined as:

$$
\begin{aligned}
B_{jk} \mid \gamma_{jk}, \lambda_0, \lambda_1 &\overset{\text{ind}}{\sim} (1-\gamma_{jk})\text{Laplace}(\lambda_0) + \gamma_{jk}\text{Laplace}(\lambda_1), \\
\gamma_{jk} &\overset{\text{ind}}{\sim} \text{Bernoulli}\left(\prod_{l=1}^{k} \nu_l\right), \\
\nu_l &\overset{\text{i.i.d.}}{\sim} \text{Beta}(\alpha, 1),
\end{aligned}
\tag{3}
$$

where $\lambda_0 \gg \lambda_1$ are scale parameters, and γ_{jk} is a latent binary indicator stored in the $G \times K$ feature allocation matrix Γ. Because $\lambda_0 \gg \lambda_1$, the first Laplace distribution has a significantly smaller variance than the second, making it correspond to the spike prior centered around 0, while the second represents the diffuse slab prior. To estimate the mode of the posterior distribution of loadings under the Spike-and-Slab LASSO prior, one recursively maximizes an objective function resembling either a LASSO or Ridge objective, depending on λ_0 and λ_1, hence the name.

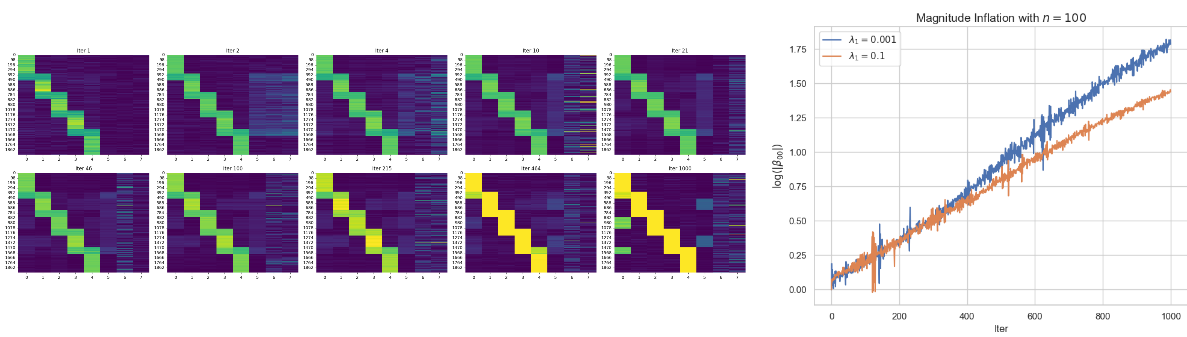

Figure 1: Factor loading matrix obtained with original Gibbs sampler (left) and the resulting inflation of the parameters (right)

To explore locally the posterior probability Ma and Liu (2022) propose a Gibbs sampler that is describe in algorithm 1. We parallelized it and implemented it and made several simulations to assess numerically the "magnitude inflation" phenomenon. In the first simulation, we generate 100 samples from a Bayesian factor model, with a specified loading matrix. This matrix is visualized in the heatmap located at the top-left corner of Figure 1, with the color scale ranging from blue to yellow. Here, blue represents a value of 0, while yellow corresponds to a value of 7. As the heatmap reveals, the sample remains relatively close to the true loading matrix in terms of which coefficients are zero and non-zero. However, as indicated by the color scale, the absolute value of the non-zero coefficients increases with more iterations, reflecting "magnitude inflation". By plotting the logarithm of the absolute value of a non-zero coefficient of the loading matrix, we observe a significant increase in absolute values. When examining different values of $\lambda_1 \in \{0.1, 0.001\}$, which correspond to slab priors with variances of 200 and 2×10^6 respectively, we find that "magnitude inflation" is more pronounced for the more diffuse slab prior ($\lambda_1 = 0.001$) consistent with Ma and Liu (2022) simulation and theoretical results. Nevertheless, it is important to note that the Gibbs Sampler may face numerical instability for two key reasons. First, sampling from truncated distributions involves trade-offs between computational time and precision. Second, sampling the feature sparsity coefficients, denoted as θ_k in Algorithm 1, requires computing the location parameter of the Beta distribution as the cardinality of the non-zero coefficients in column k of the loading matrix. However, this cardinality can be zero, and the Beta distribution is only defined for strictly positive location parameters. To address this issue, we propose setting the location parameter to an arbitrarily small positive value when the cardinality is zero. Nevertheless, this approach may not align perfectly with the true posterior in such cases. In approximately 30% of the simulations we conducted, we observed numerical instability and deviations in the Gibbs sampler from the true loading matrix. This is illustrated in Figure 2, where rows 6 and 7 become non-zero, in contrast to the structure of the true loading matrix.

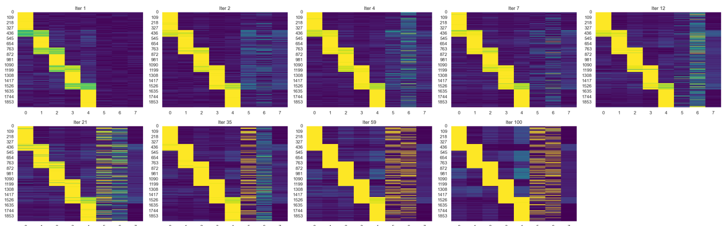

Figure 2: Sampling Instability with small scale

Theoretically Ma and Liu (2022) prove that, under the normal latent factor assumptions, as the variance of the Laplace slab priors tends to infinity (i.e., as the scale parameter approaches zero), the probability that a posterior sample B, given Y, Σ, Γ, has a matrix norm smaller than any constant converges to zero. An important result is that, for any density on the factors, the posterior distribution of the loading matrix under the spike-and-slab prior converges in distribution to that under a flat slab prior as the scale parameter of the slab prior approaches zero, becoming increasingly diffuse.

3. Orthonormal assumptions as a solution

Ma and Liu (2022) prove that consistency under either a flat prior or the spike-and-slab prior proposed by Ročková and George (2016) is achieved when the distribution of Ω ensures that the estimated $\frac{\Omega\Omega^\top}{n}$ from the factor model converges in probability to the identity matrix in the small-n, large-s paradigm as $n \to \infty$. As discussed in the previous section, under the assumption of normal factors, the estimated $\frac{\Omega\Omega^\top}{n}$ does not converge to the identity matrix when a flat prior is used. To address this, they propose the orthonormal factor assumption, which is consistent under a flat prior and provides improved control over the convergence of $\frac{\Omega\Omega^\top}{n}$ to the identity matrix.

Ensuring Consistency with Constraints on Ω

To impose stronger control over $\frac{\Omega\Omega^\top}{n}$, the model assumes a prior p_Ω that ensures:
1. All factors Ω are orthogonal, i.e., $\Omega^\top\Omega = I_K$, where K is the number of factors.
2. Each factor has equal norm such that:

$$\frac{\Omega}{\sqrt{n}} \in \mathrm{St}(K, n),$$

Here, $\mathrm{St}(K, n)$ denotes the Stiefel manifold, which represents the set of all ordered families of K vectors in $\mathbb{R}^n$.

Under this assumption, $\frac{\Omega\Omega^\top}{n} = I$, which guarantees that the posterior distribution of the loading matrix is consistent up to rotations. This avoids the inflation or deflation of the posterior samples.

Sampling from $\mathbf{St}(K, n)$

Let $\Omega_{k\cdot}$ denote the k-th row of the factor matrix Ω, and Ω_{-k} denote the remaining rows. The conditional distribution of $\Omega_{k\cdot} \mid Y, \Omega_{-k}, B, \Sigma$ is modified from a standard multivariate normal distribution to:

$$\pi(\mathrm{d}\Omega_{k\cdot} \mid Y, \Omega_{-k}, B, \Sigma) \propto f(\Omega_{k\cdot}; \bar{\Omega}_{k\cdot}, \sigma_k^2 I_n) \cdot p_{\Omega_{-k}}(\mathrm{d}\Omega_{k\cdot}),$$

where:
- $p_{\Omega_{-k}}$ is the uniform measure on the n-radius sphere centered in the orthogonal complement of Ω_{-k},
- $f(\Omega_{k\cdot}; \bar{\Omega}_{k\cdot}, \sigma_k^2 I_n)$ is the multivariate normal density with $\bar{\Omega}_{k\cdot}$ is a centered version of column k that is not affected by the projection in the orthogonal complement of Ω_{-k} and σ_k^2 is a scaling factor.

To sample from this distribution, we look at the intersections between the $\sqrt{n}$-radius centered sphere in the orthogonal complement of Ω_{-k} and the hyperplanes orthogonal to $\bar{\Omega}_{k\cdot}$. From the resulting $(n-K)$-dimensional spheres, we choose the one at a distance d from the center along $\bar{\Omega}_{k\cdot}$. Where d is sampled according to the distribution :

$$\pi(d \mid Y, \Omega_{-k}, B, \Sigma) \propto (n-d^2)^{(n-K-2)/2} \exp\left(\frac{\|P_{\Omega_{-k}^{\perp}}(\bar{\Omega}_{k\cdot})\|d}{\sigma_k^2}\right),$$

where $P_{\Omega_{-k}^{\perp}}$ is the projection operator onto the orthogonal complement of the space spanned by Ω_{-k}. This unimodal distribution is sampled using the Metropolis algorithm.

Our results

In our implementation we managed to have $\frac{\Omega\Omega^\top}{n}$ converge to I as shown in fig. 4. However we could not show this guarantees posterior consistency of the loading matrix (up to rotations) and avoids inflation or deflation in the posterior samples as our implementation was too unstable. We think it comes from our implementation of the sampling for which you can find the pseudo-code in the appendix algorithm 2 and the code on our `github`. As you may have noticed we say the spheres from which we sample Ω_k are $(n-K)$-dimensional when the paper says $(n-k)$. We assumed it was a typo in the paper since we consider spheres in the orthogonal complement of Ω_{-k} of dimension $n-K+1$ and then take the intersection with a hyperplane so it goes down another dimension. However there may just be something we misunderstood and it makes our implementation fail.

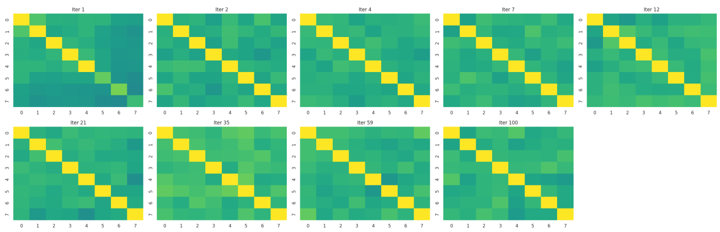

Figure 3: Orthonormal assumption of Ω

4. Conclusion

In this work, we successfully highlighted the limitations of the original Normal Sparse Factor Gibbs Sampler, demonstrating the inflation phenomenon in the posterior samples of the loading matrix. This issue, rooted in the normal latent factor assumptions, results in systematically inflated posterior coefficients, compromising the interpretability of the model.

We attempted to address this problem by replicating the orthonormal constraint on the latent factor matrix Ω, as proposed by Ma and Liu (2022). While we managed to enforce $\frac{\Omega\Omega^\top}{n}$ to converge to the identity matrix I, ensuring orthonormality, our implementation of the sampling procedure proved unstable. Consequently, we were unable to confirm whether this approach mitigates the inflation phenomenon or guarantees posterior consistency of the loading matrix.

Finally, we also explored an alternative implementation inspired by the method of Ghosh and Dunson (2009), we also provide the required Gibbs sampling framework in our Appendix. We managed to obtain results but did not have the time to compare.

Our findings underscore the challenges of implementing and evaluating advanced Bayesian factor models, particularly under orthonormal constraints, but it was a great learning opportunity.

References

Ghosh, J. and D. B. Dunson (Jan. 2009). “Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis”. In: *Journal of Computational and Graphical Statistics* 18.2, pp. 306–320. DOI: 10.1198/jcgs.2009.07145.

Ma, Y. and J. S. Liu (2022). “On Posterior Consistency of Bayesian Factor Models in High Dimensions”. In: *Bayesian Analysis* 17.3, pp. 901–929. DOI: 10.1214/21-BA1281. URL: https://doi.org/10.1214/21-BA1281.

Ročková, V. and E. I. George (2016). “Fast Bayesian Factor Analysis via Automatic Rotations to Sparsity”. In: *Journal of the American Statistical Association* 111.516, pp. 1608–1622. DOI: 10.1080/01621459.2015.1100620. eprint: https://doi.org/10.1080/01621459.2015.1100620. URL: https://doi.org/10.1080/01621459.2015.1100620.

5. Appendix

5.1. Replicating the results

Everything can be found on our `github`. To replicate the figures in this file:

- Go on the main branch and clone it
- Install the environment with poetry (Run poetry install in the command line)
- Then open notebooks/Replicate_results.ipynb and run all cells

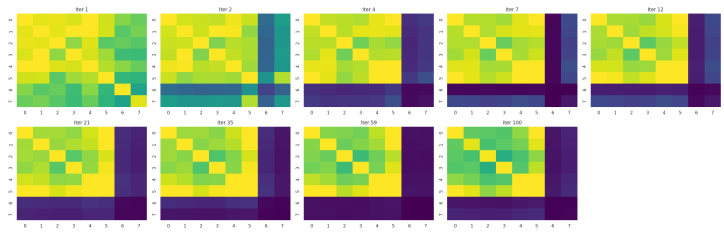

Figure 4: Orthonormal assumption of Ω

5.2. Sampling from the Normal Bayesian Factor under the Spike-and-slab LASSO prior

In the following algorithm description the superscript $^{(t-1)}$ indicates the use of the parameter value from the previous iteration.

Algorithm 1 Gibbs Sampler for SpSL Factor Model

Require: Data matrix $Y \in \mathbb{R}^{G\times n}$, initial parameters $B \in \mathbb{R}^{G\times K}, \Sigma = \mathrm{diag}(\sigma_1^2, \dots, \sigma_G^2) \in \mathbb{R}^{G\times G}, \Gamma \in \mathbb{R}^{G\times K}, \Theta \in \mathbb{R}^{K\times 1}, \alpha, \eta, \epsilon, \lambda_0, \lambda_1$, number of iterations T

Ensure: Updated parameters $B, \Sigma, \Gamma, \Theta$

Initialize $B, \Sigma, \Gamma, \Theta$

for $t = 1$ to T **do**

  for each $i = 1, \dots, N$ **do**

    Sample ω_i from a normal distribution:

$$\omega_i | B^{(t-1)}, \Sigma^{(t-1)} \overset{i.i.d.}{\sim} \mathcal{N}_K\big((I_K + B^T\Sigma^{-1}B)^{-1}B^T\Sigma^{-1}y_i,\, (I_K + B^T\Sigma^{-1}B)^{-1}\big)$$

  end for

  Update Ω with the new $(\omega_i)_i$.

  for each $j = 1, \dots, G$ and $k = 1, \dots, K$ **do**

    Compute a_{jk}, b_{jk}, c_{jk} given $B^{(t-1)}, \Omega^{(t-1)}, \Sigma^{(t-1)}, \Gamma^{(t-1)}$ and Y:

$$a_{jk} = \sum_{i=1}^{n} \frac{\omega_{ik}^2}{2\sigma_j^2}, \quad b_{jk} = \sum_{i=1}^{n} \frac{\omega_{ik}(y_{ij} - \sum_{l\neq k}\beta_{jl}\omega_{il})}{\sigma_j^2}, \quad c_{jk} = \lambda_1\gamma_{jk} + \lambda_0(1-\gamma_{jk})$$

    Sample β_{jk} from a mixture of truncated gaussian:

$$\beta_{jk} \overset{i.i.d.}{\sim} \mathcal{N}\left(\frac{b_{jk} - c_{jk}}{2a_{jk}}, \frac{1}{2a_{jk}}\right) \times \mathbf{1}\left[\beta_{jk} \geq 0\right] + \mathcal{N}\left(\frac{b_{jk} + c_{jk}}{2a_{jk}}, \frac{1}{2a_{jk}}\right) \times \mathbf{1}\left[\beta_{jk} < 0\right]$$

  end for

  Update B with the new $(\beta_{jk})_{jk}$.

  for each $j = 1, \dots, G$ and $k = 1, \dots, K$ **do**

    Sample γ_{jk} from a Bernoulli:

$$\gamma_{jk} | B, \theta_k^{(t-1)} \overset{i.i.d.}{\sim} \text{Bernoulli}\left(\frac{\lambda_1 \exp(-\lambda_1|\beta_{jk}|)\theta_k}{\lambda_0 \exp(-\lambda_0|\beta_{jk}|)(1-\theta_k) + \lambda_1 \exp(-\lambda_1|\beta_{jk}|)\theta_k}\right)$$

  end for

  Update Γ with the new $(\gamma_{jk})_{jk}$.

  for $k = K, K-1, \dots, 1$ **do**

    Compute α_k, β_k given Γ:

$$\alpha_k = \sum_{j=1}^{G} \gamma_{jk} + \alpha \times \mathbf{1}[k = K], \quad \beta_k = G - \sum_{j=1}^{G} \gamma_{jk} + 1$$

    Sample θ_k from a Truncated Beta with $\theta_{K+1} = 0$ and $\theta_0 = 1$:

$$\theta_k | \theta_{k+1}, \theta_{k-1}^{(t-1)} \sim \text{Beta}(\alpha_k, \beta_k) \times \mathbf{1}\left[\theta_{k+1} \leq \theta_k \leq \theta_{k-1}^{(t-1)}\right]$$

  end for

  Update Θ with the new $(\theta_k)_k$.

  for each $j = 1, \dots, G$ **do**

    Sample σ_j^2 from an Inverse-Gamma:

$$\sigma_j^2 \overset{i.i.d.}{\sim} \text{Inverse-Gamma}\left(\frac{\eta + n}{2}, \frac{1}{2}\left(\eta\epsilon + \sum_{i=1}^{n}\left(y_{ij} - B_{j.}^{\top}\omega_i\right)^2\right)\right)$$

  end for

  Update $\Sigma = \mathrm{diag}(\sigma_1^2, \dots, \sigma_G^2)$ with the new values.

end for

Return sampled paths of $B, \Sigma, \Gamma, \Theta$ =0

5.3. Sampling of Ω_k in the orthonormal case

Algorithm 2 Sampling from the intersection of a sphere and a hyperplane

Require: d (distance of the hyperplane from the origin), $\bar{\boldsymbol{\Omega}}_{\boldsymbol{k}}$ (normal vector defining the hyperplane), k (index of the row being sampled), ϵ (small value to avoid division by zero)

Ensure: Sampled point $\boldsymbol{x}$ from the intersection of the sphere and hyperplane.

1: Compute Ω_{-k} = Remove row k from Ω, with shape $(n, K-1)$

2: Randomly sample $\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \mathbb{I}_n)$

3: Normalize $\boldsymbol{X}$ to the unit sphere: $\boldsymbol{X}_{\text{unit}} = \frac{\boldsymbol{X}}{\|\boldsymbol{X}\|}$

4: Project $\boldsymbol{X}_{\text{unit}}$ orthogonally to the space spanned by Ω_{-k}:

$$\boldsymbol{X}_{\text{orthogonal}} = \text{Project}(\Omega_{-k}, \boldsymbol{X}_{\text{unit}})$$

5: Further project $\boldsymbol{X}_{\text{orthogonal}}$ orthogonally to $\bar{\boldsymbol{\Omega}}_{\boldsymbol{k}}$:

$$\boldsymbol{X}_{\text{hyperplane}} = \text{Project}(\bar{\boldsymbol{\Omega}}_{\boldsymbol{k}}, \boldsymbol{X}_{\text{orthogonal}})$$

6: Compute the norm of $\boldsymbol{X}_{\text{hyperplane}}$: $\lambda = \|\boldsymbol{X}_{\text{hyperplane}}\|$

7: Rescale $\boldsymbol{X}_{\text{hyperplane}}$ to lie on the sphere of radius $\sqrt{n - d^2}$:

$$\boldsymbol{X}_{\text{sphere}} = \sqrt{n - d^2} \cdot \frac{\boldsymbol{X}_{\text{hyperplane}}}{\lambda}$$

8: Shift the point $\boldsymbol{X}_{\text{sphere}}$ by d along $\bar{\boldsymbol{\Omega}}_{\boldsymbol{k}}$:

$$\boldsymbol{x} = \boldsymbol{X}_{\text{sphere}} + \frac{d \cdot \bar{\boldsymbol{\Omega}}_{\boldsymbol{k}}}{\max(\|\bar{\boldsymbol{\Omega}}_{\boldsymbol{k}}\|, \epsilon)}$$

9: **return** $\boldsymbol{x}$ =0

5.4. Gibbs Sampler proofs

5.4.1 Motivation

During our reimplementation of Ghosh and Dunson (2009) we noticed we had the model but no new update policy. Since it was a modified version with a diffuse parameter for the loading factors, the update policy from the original paper is no longer viable. The following is our endeavour at finding the update policy for the Ghosh Dunson model and the consequent results.

5.4.2 Base model for the base model

We first show the results on the original sampler to better understand what changes with Ghosh Dunson afterwards.

Model and Priors:

$$y_i|\omega_i, B, \Sigma \sim \mathrm{NG}(B\omega_i, \Sigma), \omega_i \sim \mathcal{N}_K(0, I_K)$$

$$\begin{aligned} p(\beta_{jk}|\gamma_{jk}, \lambda_0, \lambda_1) &= (1-\gamma_{jk})\psi(\beta_{jk}|\lambda_0) + \gamma_{jk}\psi(\beta_{jk}|\lambda_1), \qquad \lambda_0 \neq \lambda_1 \\ \gamma_{jk}|\theta_k &\sim \text{Bernoulli}(\theta_k), \\ \theta_k &= \prod_{l=1}^{k} \nu_l, \quad \nu_l \sim \text{Beta}(\alpha, 1), \\ \sigma_j^2 &\sim \text{Inverse-Gamma}\left(\frac{\eta}{2}, \frac{\eta\epsilon}{2}\right) \\ \psi(\beta_{jk}|\lambda) &\sim \text{Laplace}(0, 1/\lambda) \end{aligned}$$

---

posterior for β_{jk} At fixed j and k, to compute the likelihood term, we focus on the relationship between y_{ji} and β_{jk}. Substituting the model equation for y_{ji}, we have:

$$y_{ji} = \sum_{l=1}^{K} \beta_{jl}\omega_{il} + \epsilon_{ij}$$

$$y_{ji} = \beta_{jk}\omega_{ik} + \sum_{l\neq k} \beta_{jl}\omega_{il} + \epsilon_{ij}$$

$$\text{"Residual term" without } \beta_{jk}: \quad r_{ijk} = y_{ji} - \sum_{l\neq k} \beta_{jl}\omega_{il}$$

$$f(y_{ji} \mid \beta_{jk}, \omega_{ik}) \propto \exp\left(-\frac{1}{2\sigma_j^2}(r_{ijk} - \beta_{jk}\omega_{ik})^2\right)$$

$$\text{All rows that impact } \beta_{jk}: \prod_{i=1}^{n} f(y_{ji} \mid \beta_{jk}, \omega_{ik}) \propto \exp\left(-\frac{1}{2\sigma_j^2}\sum_{i=1}^{n}(r_{ijk} - \beta_{jk}\omega_{ik})^2\right)$$

$$\sum_{i=1}^{n}(r_{ijk} - \beta_{jk}\omega_{ik})^2 = \sum_{i=1}^{n} r_{ijk}^2 - 2\beta_{jk}\sum_{i=1}^{n} r_{ijk}\omega_{ik} + \beta_{jk}^2\sum_{i=1}^{n}\omega_{ik}^2$$

$$-\frac{1}{2\sigma_j^2}\sum_{i=1}^{n}(r_{ijk} - \beta_{jk}\omega_{ik})^2 = -\frac{1}{2\sigma_j^2}\left[\beta_{jk}^2\sum_{i=1}^{n}\omega_{ik}^2 - 2\beta_{jk}\sum_{i=1}^{n} r_{ijk}\omega_{ik}\right] + \text{constant}$$

The prior for β_{jk} is a mixture of Laplace distributions:

$$p(\beta_{jk} \mid \gamma_{jk}, \lambda_0, \lambda_1) = (1-\gamma_{jk})\psi(\beta_{jk} \mid \lambda_0) + \gamma_{jk}\psi(\beta_{jk} \mid \lambda_1)$$
$$\psi(\beta_{jk} \mid \lambda) \propto \exp(-\lambda|\beta_{jk}|)$$
$$p(\beta_{jk} \mid \gamma_{jk}, \lambda_0, \lambda_1) \propto \psi(\beta_{jk} \mid \lambda_1\gamma_{jk} + \lambda_0(1-\gamma_{jk}))$$

Since γ_{jk} is a bernouli variable

Combining the likelihood and prior terms, the conditional posterior for β_{jk} is proportional to:

$$\pi(\beta_{jk} \mid \beta_{-jk}, \Omega, \Gamma, \Sigma) \propto \exp\left(-\frac{1}{2\sigma_j^2}\beta_{jk}^2\sum_{i=1}^n \omega_{ik}^2 + \frac{1}{\sigma_j^2}\beta_{jk}\sum_{i=1}^n r_{ijk}\omega_{ik} - c_{jk}|\beta_{jk}|\right)$$
$$c_{jk} = \lambda_1\gamma_{jk} + \lambda_0(1-\gamma_{jk}) \mid a_{jk} = \frac{1}{2\sigma_j^2}\sum_{i=1}^n \omega_{ik}^2 \mid b_{jk} = \frac{1}{\sigma_j^2}\sum_{i=1}^n \omega_{ik}\left(y_{ji} - \sum_{l\neq k}\beta_{jl}\omega_{il}\right)$$

This is the way to get back the results from the paper's section 2.2. Where they actually had a typo by writing y_{ij} instead of y_{ji}

5.4.3 Modified Ghosh Dunson model

Model and Priors:

$$y_i|\omega_i, B, \Sigma \sim \text{NG}(B\omega_i, \Sigma), \quad \omega_i \sim \mathcal{N}_K(0, I_K)$$

$$\beta_{jk} = q_{jk}r_k,$$
$$p(r_k|\lambda) = \psi(r_k|\lambda),$$
$$p(q_{jk}|\gamma_{jk}, \lambda_0, \lambda_1) = (1-\gamma_{jk})\psi(q_{jk}|\lambda_0) + \gamma_{jk}\psi(q_{jk}|\lambda_1), \qquad \lambda_0 \neq \lambda_1$$
$$\gamma_{jk}|\theta_k \sim \text{Bernoulli}(\theta_k),$$
$$\theta_k = \prod_{l=1}^k \nu_l, \quad \nu_l \sim \text{Beta}(\alpha, 1),$$
$$\sigma_j^2 \sim \text{Inverse-Gamma}\left(\frac{\eta}{2}, \frac{\eta\epsilon}{2}\right)$$
$$\psi(r_k|\lambda) \sim \text{Laplace}(0, 1/\lambda)$$

posterior for q_{jk} At fixed j and k, we focus on the relationship between y_{ji} and q_{jk}. Substituting the model equation for y_{ji}, we have:

$$y_{ji} = \sum_{l=1}^{K} \beta_{jl}\omega_{il} + \epsilon_{ij}$$

$$y_{ji} = q_{jk}r_k\omega_{ik} + \sum_{l\neq k} q_{jl}r_l\omega_{il} + \epsilon_{ij}$$

$$\text{"Residual term" without } qjk: \quad res_{ijk} = y_{ji} - \sum_{l\neq k} q_{jl}r_l\omega_{il}$$

$$f(y_{ji} \mid q_{jk}, r_k, \omega_{ik}) \propto \exp\left(-\frac{1}{2\sigma_j^2}(res_{ijk} - q_{jk}r_k\omega_{ik})^2\right)$$

$$\text{All rows that impact } q_{jk}: \prod_{i=1}^{n} f(y_{ji} \mid q_{jk}, r_k, \omega_{ik}) \propto \exp\left(-\frac{1}{2\sigma_j^2}\sum_{i=1}^{n}(res_{ijk} - q_{jk}r_k\omega_{ik})^2\right)$$

$$\sum_{i=1}^{n}(res_{ijk} - q_{jk}r_k\omega_{ik})^2 = \sum_{i=1}^{n} res_{ijk}^2 - 2q_{jk}\sum_{i=1}^{n} res_{ijk}\omega_{ik}r_k + q_{jk}^2\sum_{i=1}^{n} r_k^2\omega_{ik}^2$$

$$-\frac{1}{2\sigma_j^2}\sum_{i=1}^{n}(res_{ijk} - q_{jk}r_k\omega_{ik})^2 = -\frac{1}{2\sigma_j^2}\left[q_{jk}^2\sum_{i=1}^{n}\omega_{ik}^2r_k^2 - 2q_{jk}\sum_{i=1}^{n} res_{ijk}\omega_{ik}r_k\right] + \text{constant}$$

The prior for qjk is the same as old β_{jk}:

$$p(q_{jk} \mid \gamma_{jk}, \lambda_0, \lambda_1) = (1-\gamma_{jk})\psi(q_{jk} \mid \lambda_0) + \gamma_{jk}\psi(q_{jk} \mid \lambda_1)$$

$$\psi(q_{jk} \mid \lambda) \propto \exp(-\lambda|q_{jk}|)$$

$$p(q_{jk} \mid \gamma_{jk}, \lambda_0, \lambda_1) \propto \psi(q_{jk} \mid \lambda_1\gamma_{jk} + \lambda_0(1-\gamma_{jk}))$$

Since γ_{jk} is a bernouli variable

Combining the likelihood and prior terms, the conditional posterior for q_{jk} is proportional to:

$$\pi(q_{jk} \mid q_{-jk}, \Omega, \Gamma, \Sigma) \propto \exp\left(-\frac{1}{2\sigma_j^2}q_{jk}^2\sum_{i=1}^{n}\omega_{ik}^2r_k^2 + \frac{1}{\sigma_j^2}q_{jk}\sum_{i=1}^{n} res_{ijk}\cdot\omega_{ik}r_k - c_{jk}|q_{jk}|\right)$$

$$c_{jk} = \lambda_1\gamma_{jk} + \lambda_0(1-\gamma_{jk}) \mid a_{jk} = \frac{1}{2\sigma_j^2}\sum_{i=1}^{n}\omega_{ik}^2r_k^2 \mid b_{jk} = \frac{1}{\sigma_j^2}\sum_{i=1}^{n}\omega_{ik}r_k\left(y_{ji} - \sum_{l\neq k} q_{jl}r_l\omega_{il}\right)$$

Where we have the diffuse parameter intervene in the coefficients this time.

Posterior for r_k By having k fixed we go back to the likelihood. The model equation for y_{ji} is:

$$y_{ji} = q_{jk} r_k \omega_{ik} + \sum_{l \neq k} q_{jl} r_l \omega_{il} + \epsilon_{ij}.$$

Let the residual term res_{ijk} be defined as:

$$res_{ijk} = y_{ji} - \sum_{l \neq k} q_{jl} r_l \omega_{il}.$$

The likelihood for y_{ji} given r_k, q_{jk}, and ω_{ik} is:

$$f(y_{ji} \mid r_k, q_{jk}, \omega_{ik}) \propto \exp\left(-\frac{1}{2\sigma_j^2} \left(res_{ijk} - q_{jk} r_k \omega_{ik} \right)^2 \right).$$

The total likelihood is the product of the likelihoods for all i and j:

$$\prod_{i=1}^{n} \prod_{j=1}^{J} f(y_{ji} \mid r_k, q_{jk}, \omega_{ik}) \propto \exp\left(-\sum_{i=1}^{n} \sum_{j=1}^{J} \frac{1}{2\sigma_j^2} \left(res_{jik} - q_{jk} r_k \omega_{ik} \right)^2 \right).$$

Expanding the quadratic term inside the sum and isolating r_k:

$$\sum_{i=1}^{n} \sum_{j=1}^{J} \frac{1}{2\sigma_j^2} \left(res_{jik} - q_{jk} r_k \omega_{ik} \right)^2 = \sum_{i=1}^{n} \sum_{j=1}^{J} res_{jik}^2 \frac{1}{2\sigma_j^2} - 2r_k \sum_{i=1}^{n} \sum_{j=1}^{J} res_{jik} \omega_{ik} \frac{q_{jk}}{2\sigma_j^2} + r_k^2 \sum_{i=1}^{n} \omega_{ik}^2 \sum_{j=1}^{J} \frac{q_{jk}^2}{2\sigma_j^2}.$$

Thus, the likelihood term becomes:

$$\sum_{i=1}^{n} \sum_{j=1}^{J} \frac{1}{2\sigma_j^2} \left(res_{jik} - q_{jk} r_k \omega_{ik} \right)^2 = \left[r_k^2 \sum_{i=1}^{n} \omega_{ik}^2 \sum_{j=1}^{J} \frac{q_{jk}^2}{2\sigma_j^2} - 2r_k \sum_{i=1}^{n} \sum_{j=1}^{J} res_{jik} \frac{q_{jk}}{2\sigma_j^2} \omega_{ik} \right] + \text{constant}.$$

We combine the prior for r_k and the likelihood term. The posterior for r_k is proportional

$$p(r_k \mid y_{ji}, q_{jk}, \omega_{ik}) \propto \exp\left(-\lambda |r_k| \right) \cdot \exp\left(-r_k^2 \sum_{i=1}^{n} \omega_{ik}^2 \sum_{j=1}^{J} \frac{q_{jk}^2}{2\sigma_j^2} + 2r_k \sum_{i=1}^{n} \sum_{j=1}^{J} res_{jik} \frac{q_{jk}}{2\sigma_j^2} \omega_{ik} \right).$$

$$p(r_k \mid \dots) \propto \exp\left(-a r_k^2 + b r_k - h|r_k| \right),$$

where:

- $a = \sum_{i=1}^{n} \sum_{j=1}^{J} \omega_{ik}^2 \frac{q_{jk}^2}{2\sigma_j^2}$
- $b = \sum_{i=1}^{n} \sum_{j=1}^{J} \frac{q_{jk}}{\sigma_j^2} \, res_{jik} \, \omega_{ik}$
- $h = \lambda$

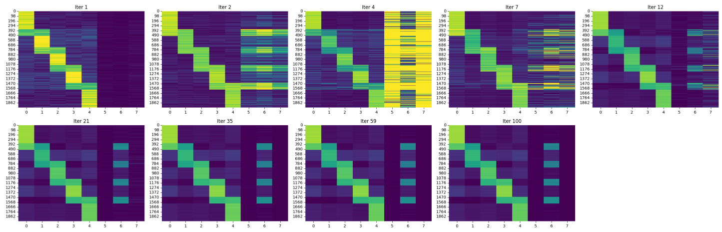

Figure 5: Factor loading matrix with Ghosh Dunson model

Results With these new updates and posteriors we are able to run the modified Ghosh dunson model's Gibbs sampler as all the other variables's updates stay the same. This enabled us to get this result