Consistency and Asymptotic Normality of Stochastic Block Model Estimators under sampling condition

1. Introduction

2. Model and assumptions

In SBM, nodes from a set $\mathcal{N} \triangleq \{1,\dots,n\}$ are distributed among a set $\mathcal{Q} \triangleq \{1,\dots,Q\}$ of hidden blocks that model the latent structure of the graph. The blocks are described by categorical variables $(Z_i, i \in \mathcal{N})$ with prior probabilities $\alpha = (\alpha_1,\dots,\alpha_Q)$, such that $\mathbb{P}(Z_i = q) = \alpha_q$, with $q \in \mathcal{Q}$. The probability of an edge between any dyad in $\mathcal{D} \triangleq \mathcal{N} \times \mathcal{N}$ only depends on the blocks the two nodes belong to. Hence, the presence of an edge between i and j, indicated by the binary variable Y_{ij}, is independent on the other edges conditionally on the latent blocks:

$$Y_{ij} \mid Z_i = q, Z_j = \ell \sim^{\text{ind}} \varphi(., \pi_{q\ell}), \qquad \forall (i,j) \in \mathcal{D}, \quad \forall (q,\ell) \in \mathcal{Q} \times \mathcal{Q}.$$

In the following, $Y = (Y_{ij})_{i,j \in \mathcal{D}}$ is the $n \times n$ adjacency matrix of the random graph, $Z = (Z_1,\dots,Z_n)$ the n-vector of the latent blocks. With a slight abuse of notation, we associate to Z_i a vector of indicator variables $(Z_{i1},\dots,Z_{iQ})$ such that $Z_i = q \Leftrightarrow Z_{iq} = 1, Z_{i\ell} = 0$, for all $\ell \neq q$. Notice that in the undirected binary case, $Y_{ij} = Y_{ji}$ for all $(i,j) \in \mathcal{D}$ and $Y_{ii} = 0$ for all $i \in \mathcal{N}$. Similarly, $\pi_{q\ell} = \pi_{\ell q}$ for all $(q,\ell) \in \mathcal{Q} \times \mathcal{Q}$.

Hence, the complete parameter set is $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\pi}) \in \boldsymbol{\Theta}$ and $\boldsymbol{\Theta}$ the parameter space.

When performing inference from data, we note $\boldsymbol{\theta}^\star = (\boldsymbol{\alpha}^\star, \boldsymbol{\pi}^\star)$ the true parameter set, *i.e.* the parameter values used to generate the data, and $\mathbf{z}^\star$ the true (and usually unobserved) assignment of rows and columns to their group. For given matrices of indicator variables $\mathbf{z}$, we also note:

- $z_{+q} = \sum_i z_{iq}$
- $z^\star_{+q}$ his counterpart for $\mathbf{z}^\star$.

The confusion matrix allows to compare the partition.

Definition 2.1 (confusion matrix). *For given assignments* $\mathbf{z}$ *and* $\mathbf{z}^\star$*, we define the* confusion matrix *between* $\mathbf{z}$ *and* $\mathbf{z}^\star$*, noted* $\mathbb{R}_{\mathcal{Q}}(\mathbf{z})$*, as follows:*

$$\mathbb{R}_{\mathcal{Q}}(\mathbf{z})_{qq'} = \frac{1}{n} \sum_i z^\star_{iq} z_{iq'} \tag{2.1}$$

2.1. Missing data for SBM

Regarding SBM inference, a missing value corresponds to a missing entry in the adjacency matrix Y, typically denoted by `NA`'s. Therefore, Y has three possible entries 0, 1 or `NA`. We rely on the $n \times n$ sampling matrix R to record the data sampled during this process:

$$(R_{ij}) = \begin{cases} 1 & \text{if } Y_{ij} \text{ is observed,} \\ 0 & \text{otherwise.} \end{cases} \tag{2.2}$$

As a shortcut, we use $Y^{\mathrm{o}} = \{Y_{ij} : R_{ij} = 1\}$ and $Y^{\mathrm{m}} = \{Y_{ij} : R_{ij} = 0\}$ to denote the sets of variables respectively associated with the *observed* and *missing* data. The *sampling design* is the description of the stochastic process that generates R. It is assumed that the network pre-exists this process, which is fully characterized by the conditional distribution $p_\psi(R|Y)$, the parameters of which are such that ψ and θ live in a product space $\Theta \times \Psi$. We then follow the framework of **?**] for missing data that we adapt to the presence of the latent variables Z_i: the joint probability density function of the observed data satisfies

$$p_{\theta,\psi}(Y^{\mathrm{o}}, R) = \int\int p_\theta(Y^{\mathrm{o}}, Y^{\mathrm{m}}, Z) p_\psi(R|Y^{\mathrm{o}}, Y^{\mathrm{m}}, Z) dY^{\mathrm{m}} dZ. \tag{2.3}$$

Simplifications may occur in (2.3) depending on the sampling design, leading to the three usual types of missingness: Missing completely at random (MCAR), Missing at random (MAR) or Not missing at random (NMAR). For SBM, this typology depends on the relations between the network Y, the latent structure Z and the sampling R, so that the missingness is characterized by four directed acyclic graphs (DAG) displayed in Figure **??**. Since the network pre-exists the sampling process, we do not consider DAG where R is a parent node.

The type of missingness for SBM can be defined as follows:

$$\text{Sampling design for SBM is} \begin{cases} \text{MCAR} & \text{if } R \perp\!\!\!\perp (Y^{\mathrm{m}}, Z, Y^{\mathrm{o}}), \\ \text{MAR} & \text{if } R \perp\!\!\!\perp (Y^{\mathrm{m}}, Z) \mid Y^{\mathrm{o}}, \\ \text{NMAR} & \text{otherwise.} \end{cases} \tag{2.4}$$

Property 2.2. *According to definition (2.4), if the sampling design is MAR, then maximising $p_{\theta,\psi}(Y^{\mathrm{o}}, R)$ in θ is equivalent to maximising $p_\theta(Y^{\mathrm{o}})$ in θ, this corresponds to ignorability notion defined in Rubin and recalled in Handcock and Gile.*

2.2. Sampling design examples

Definition 2.3 (Random dyad sampling)**.** *Each dyad $(i, j) \in \mathcal{D}$ has the same probability $\mathbb{P}(R_{ij} = 1) = \rho$ to be observed independently on the others.*

Definition 2.4 (Star sampling)**.** *Each node $i \in \{1, ...n\}$ has the same probability $\mathbb{P}(S_i = 1) = \rho$ to be observed independently on the others.*

These designs are trivially MCAR because each dyad/node is sampled with the same probability ρ which does not depend on Y.

Proposition 2.5. *Under random dyad sampling, defining* $N_i = \sum_{j\neq i} R_{ij}$ *and* $\Omega_{0,n} = \{\forall i \in \{1,...,n\},\ N_i \geqslant 1\}$*. Then*

$$\mathbb{P}\left(\lim_{n\to+\infty} \Omega_{0,n}\right) = 1.$$

Proof. Noticing that $N_i \sim \text{Bin}(n,\rho)$, then $\mathbb{P}(N_i \geqslant 1) = 1-\rho^n$. As a consequence $\mathbb{P}(\overline{\Omega_{0,n}}) \leqslant \sum_i \mathbb{P}(N_i = 0) = n\rho^n \underset{n\to+\infty}{\longrightarrow} 0$, and $\mathbb{P}(\Omega_{0,n}) \underset{n\to+\infty}{\longrightarrow} 1$. Then $\mathbb{P}(\limsup(\overline{\Omega_{0,n}})) = 0$ by Borel-Cantelli theorem (because $\sum_n \mathbb{P}(\overline{\Omega_{0,n}})$ does not converge), and as $\overline{\limsup \overline{\Omega_{0,n}}} = \overline{\bigcap_{n\geqslant 0}\bigcup_{q\geqslant n} \overline{\Omega_{0,n}}} = \bigcup_{n\geqslant 0}\bigcap_{q\geqslant n} \Omega_{0,n} = \liminf \Omega_{0,n}$, the result follow. □

Remark 2.6. Proposition 2.5 is trivially true for any node-centered sampling design.

In the following we will consider only MCAR samplings dyad-centered. Because of proposition 2.5, we will always consider that at least one dyad is sampled for each node of the network.

2.3. Likelihood

When the labels are known, the *complete log-likelihood* is given by:

$$\begin{aligned}
\mathcal{L}_c(\mathbf{z};\boldsymbol{\theta}) &= \log p(\mathbf{y^o},\mathbf{z};\boldsymbol{\theta}) \\
&= \log\left\{\left(\prod_{i,q} \alpha^{z_{iq}}\right)\left(\prod_{i,j,q,\ell} \varphi\left(y_{ij};\pi_{q\ell}\right)^{z_{iq}z_{j\ell}r_{ij}}\right)\right\} \\
&= \log\left\{\left(\prod_{i} \alpha_{z_i}\right)\left(\prod_{i,j} \varphi\left(y_{ij};\pi_{z_iz_j}\right)^{r_{ij}}\right)\right\}.
\end{aligned} \tag{2.5}$$

But the labels are usually unobserved, and the *observed log-likelihood* is obtained by marginalization over all the label configurations:

$$\mathcal{L}(\boldsymbol{\theta}) = \log p(\mathbf{y^o};\boldsymbol{\theta}) = \log\left(\sum_{\mathbf{z}\in\mathcal{Z}} p(\mathbf{y^o},\mathbf{z};\boldsymbol{\theta})\right). \tag{2.6}$$

2.4. Assumptions

We focus here on parametric models where φ belongs to a regular one-dimension exponential family in canonical form:

$$\varphi(x,\pi) = b(x)\exp(\pi x - \psi(\pi)), \tag{2.7}$$

where π belongs to the space $\mathcal{A}$, so that $\varphi(\cdot, \pi)$ is well defined for all $\pi \in \mathcal{A}$. Classical properties of exponential families insure that ψ is convex, infinitely differentiable on $\mathring{\mathcal{A}}$, that $(\psi')^{-1}$ is well defined on $\psi'(\mathring{\mathcal{A}})$. When $X_\pi \sim \varphi(., \pi)$, $\mathbb{E}[X_\pi] = \psi'(\pi)$ and $\mathbb{V}[X_\pi] = \psi''(\pi)$.

Moreover, we make the following assumptions on the parameter space :

H_1 : There exist a positive constant c, and a compact C_π such that

$$\mathbf{\Theta} \subset [c, 1-c]^{\mathcal{Q}} \times C_\pi^{\mathcal{Q}\times\mathcal{Q}} \quad \text{with} \quad C_\pi \subset \mathring{\mathcal{A}}.$$

H_2 : The true parameter $\boldsymbol{\theta}^\star = (\boldsymbol{\alpha}^\star, \boldsymbol{\pi}^\star)$ lies in the relative interior of $\mathbf{\Theta}$.
H_3 : The map $\pi \mapsto \varphi(\cdot, \pi)$ is injective.
H_4 : Each row and each column of $\boldsymbol{\pi}^\star$ is unique.

The previous assumptions are standard. Assumption H_1 ensure that the group proportions are bounded away from 0 and 1 so that no group disappears when n goes to infinity. It also ensures that π is bounded away from the boundaries of the $\mathcal{A}$ and that there exists a $\kappa > 0$, such that $[\pi_{q\ell} - \kappa, \pi_{q\ell} + \kappa] \subset \mathring{\mathcal{A}}$ for all parameters $\pi_{q\ell}$ of $\boldsymbol{\theta} \in \mathbf{\Theta}$. Assumptions H_3 and H_4 are necessary to ensure that the model is identifiable. If the map $\pi \mapsto \varphi(., \pi)$ is not injective, the model is trivially not identifiable. Similarly, if rows q and q' are identical, we can build a more parsimonious model that induces the same distribution of $\mathbf{y}$ by merging groups q and q'. In the following, we consider that $\mathcal{Q}$, the number of classes (or groups) is known.

Moreover, we define the $\delta(\boldsymbol{\pi})$, that captures the differences between groups: lower values means that there are two classes that are very similar.

Definition 2.7 (class distinctness)**.** *For $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\pi}) \in \mathbf{\Theta}$. We define:*

$$\delta(\boldsymbol{\pi}) = \min_{\ell,\ell'} \max_q \mathrm{KL}(\pi_{q\ell}, \pi_{q\ell'})$$

with $\mathrm{KL}(\pi, \pi') = \mathbb{E}_\pi[\log(\varphi(X, \pi)/\varphi(X, \pi'))] = \psi'(\pi)(\pi - \pi') + \psi(\pi') - \psi(\pi)$ *the Kullback divergence between $\varphi(., \pi)$ and $\varphi(., \pi')$, when φ comes from an exponential family.*

Remark 2.8. Since all $\boldsymbol{\pi}$ have distinct rows and columns, $\delta(\boldsymbol{\alpha}) > 0$.

Remark 2.9. Since we restricted π in a bounded subset of $\mathring{\mathcal{A}}$, there exists two positive values M_π and κ such that $C_\pi + (-\kappa, \kappa) \subset [-M_\pi, M_\pi] \subset \mathring{\mathcal{A}}$. Moreover, the variance of X_π is bounded away from 0 and $+\infty$. We note

$$\sup_{\pi \in [-M_\pi, M_\pi]} \mathbb{V}(X_\pi) = \bar{\sigma}^2 < +\infty \quad \text{and} \quad \inf_{\pi \in [-M_\pi, M_\pi]} \mathbb{V}(X_\pi) = \underline{\sigma}^2 > 0. \tag{2.8}$$

Proposition 2.10. *With the previous notations, if $\pi \in C_\pi$ and $X_\pi \sim \varphi(., \pi)$, then X_π is subexponential with parameters $(\bar{\sigma}^2, \kappa^{-1})$.*

Remark 2.11. These assumptions are satisfied for many distributions, including but not limited to:

- Bernoulli, when the proportion p is bounded away from 0 and 1, or natural parameter $\pi = \log(p/(1-p))$ bounded away from $\pm\infty$;
- Poisson, when the mean λ is bounded away from 0 and $+\infty$, or natural parameter $\pi = \log(\lambda)$ bounded away from $\pm\infty$;
- Gaussian with known variance when the mean μ, which is also the natural parameter, is bounded away from $\pm\infty$.

In particular, the conditions stating that ψ is twice differentiable and that $(\psi')^{-1}$ exists are equivalent to assuming that X_π has positive and finite variance for all values of π in the parameter space.

2.5. Symmetry

The study of the asymptotic properties of the MLE will lead to take into account symmetry properties on the parameter set. We first recall the definition of a permutation, then define equivalence relationships for assignments and parameter, and precise symmetry.

Definition 2.12 (permutation)**.** *Let s be a permutation on $\{1,\dots,g\}$ and t a permutation on $\{1,\dots,m\}$. If $\boldsymbol{A}$ is a matrix with g columns, we define $\boldsymbol{A}^s$ as the matrix obtained by permuting the columns of $\boldsymbol{A}$ according to s,* i.e. *for any row i and column q of $\boldsymbol{A}$, $A^s_{iq} = A_{is(q)}$. If $\boldsymbol{B}$ is a matrix with m columns and $\boldsymbol{C}$ is a matrix with g rows and m columns, $\boldsymbol{B}^t$ and $\boldsymbol{C}^{s,t}$ are defined similarly:*

$$\boldsymbol{A}^s = \left(A_{is(q)}\right)_{i,q} \quad \boldsymbol{B}^t = \left(B_{jt(\ell)}\right)_{j,\ell} \quad \boldsymbol{C}^{s,t} = \left(C_{s(q)t(\ell)}\right)_{q,\ell}$$

Definition 2.13 (equivalence)**.** *We define the following equivalence relationships:*

- *Two assignments $\mathbf{z}$ and $\mathbf{z}'$ are* equivalent*, noted $\sim$, if they are equal up to label permutation,* i.e. *it exists a permutation s such that $\mathbf{z}' = \mathbf{z}^s$.*
- *Two parameters are $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$ are* equivalent*, noted $\sim$, if they are equal up to label permutation,* i.e. *it exists a permutation s such that $(\boldsymbol{\alpha}^s, \boldsymbol{\pi}^s) = (\boldsymbol{\alpha}', \boldsymbol{\pi}')$. This is* label-switching.
- *$(\boldsymbol{\theta}, \mathbf{z})$ and $(\boldsymbol{\theta}', \mathbf{z}')$ are* equivalent*, noted $\sim$, if they are equal up to label permutation on $\boldsymbol{\pi}$,* i.e. *it exists a permutation s such that $(\boldsymbol{\pi}^s, \mathbf{z}^s) = (\boldsymbol{\pi}', \mathbf{z}')$.*

Definition 2.14 (distance)**.** *We define the following distance, up to equivalence, between configurations $\mathbf{z}$ and $\mathbf{z}^\star$:*

$$\|\mathbf{z} - \mathbf{z}^\star\|_{0,\sim} = \inf_{\mathbf{z}'\sim\mathbf{z}} \|\mathbf{z}' - \mathbf{z}^\star\|_0$$

where, for all matrix $\mathbf{z}$, we use the Hamming norm $\|\cdot\|_0$ defined by

$$\|\mathbf{z}\|_0 = \sum_{i,q} \mathbb{1}\{z_{iq} \neq 0\}.$$

The last equivalence relationship is not concerned with $\boldsymbol{\pi}$. It is useful when dealing with the conditional likelihood $p(\mathbf{x}|\mathbf{z};\boldsymbol{\theta})$ which does not depend on $\boldsymbol{\pi}$: in fact, if $(\boldsymbol{\theta},\mathbf{z}) \sim (\boldsymbol{\theta}',\mathbf{z}')$, then for all $\mathbf{x}$, we have $p(\mathbf{x}|\mathbf{z};\boldsymbol{\theta}) = p(\mathbf{x}|\mathbf{z}';\boldsymbol{\theta}')$. Note also that $\mathbf{z} \sim \mathbf{z}^\star$ if and only if the confusion matrix $\mathbb{R}_{\mathcal{Q}}(\mathbf{z})$ is equivalent to a diagonal matrix.

Definition 2.15 (symmetry). *We say that the parameter* $\boldsymbol{\theta}$ exhibits symmetry for the permutation s *if*

$$(\boldsymbol{\alpha}^s, \boldsymbol{\pi}^s) = (\boldsymbol{\alpha}, \boldsymbol{\pi}).$$

$\boldsymbol{\theta}$ exhibits symmetry *if it exhibits symmetry for any non trivial pair of permutations* (s,t)*. Finally the set of pairs* (s,t) *for which* $\boldsymbol{\theta}$ *exhibits symmetry is noted* $\mathrm{Sym}(\boldsymbol{\theta})$*.*

Remark 2.16. The set of parameters that exhibit symmetry is a manifold of null Lebesgue measure in $\boldsymbol{\Theta}$. The notion of symmetry allows us to deal with a notion of non-identifiability of the class labels that is subtler than and different from label switching.

3. Asymptotic properties in the complete data model

As stated in the introduction, we first study the asymptotic properties of the complete data model. Let $\widehat{\boldsymbol{\theta}}_c = (\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\pi}})$ be the MLE of $\boldsymbol{\theta}$ in the complete data model, where the real assignments $\mathbf{z} = \mathbf{z}^\star$ are known. We can derive the following general estimates from Equation (2.5):

$$\begin{aligned} \hat{\alpha}_q = \widehat{\alpha}_q(\mathbf{z}) &= \frac{z_{+q}}{n} \\ \widehat{y}_{q\ell}(\mathbf{z}) = \frac{\sum_{i<j} y_{ij} r_{ij} z_{iq} z_{j\ell}}{\sum_{i<j} r_{ij} z_{iq} z_{j\ell}} & \quad \hat{\alpha}_{q\ell} = \widehat{\pi}_{q\ell}(\mathbf{z}) = (\psi')^{-1}\left(\widehat{y}_{q\ell}(\mathbf{z})\right) \end{aligned} \tag{3.1}$$

Lemma 3.1.

$$U_n = \frac{1}{\binom{n}{2}} \sum_{i<j} r_{ij} z_{iq} z_{j\ell} \xrightarrow[n\to+\infty]{\mathbb{P}} \rho\alpha_q\alpha_l$$

Proof. Noticing that $\mathbb{E}[r_{ij}z_{iq}z_{j\ell}] = \rho\alpha_q\alpha_l$ and defining $q^{q,\ell}_{i,j} = r_{ij}z_{iq}z_{j\ell} - \rho\alpha_q\alpha_l$. By Hoeffding decomposition for U-statistics (see [**?**])

$$U_n = \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n} \frac{1}{\lfloor\frac{n}{2}\rfloor} \sum_{i=1}^{\lfloor\frac{n}{2}\rfloor} q^{q,\ell}_{\sigma(i),\sigma(i+\lfloor\frac{n}{2}\rfloor)}, \tag{3.2}$$

where for each permutation $\sigma \in \mathfrak{S}$, $\sum_{i=1}^{\lfloor\frac{n}{2}\rfloor} q^{q,\ell}_{\sigma(i),\sigma(i+\lfloor\frac{n}{2}\rfloor)}$ is a sum of independant r.v. Then, for $\gamma > 0$ by Jensen's inequality and Hoeffding's lemma about bounded r.v.

$$\begin{aligned}\mathbb{E}\left[\exp(\gamma U_n)\right] &\leq \frac{1}{n!}\sum_{\sigma\in\mathfrak{S}_n}\mathbb{E}\exp\left(\frac{\gamma}{\lfloor\frac{n}{2}\rfloor}\sum_{i=1}^{\lfloor\frac{n}{2}\rfloor} q^{q,\ell}_{\sigma(i),\sigma(i+\lfloor\frac{n}{2}\rfloor)}\right)\\ &\leq \exp\left(\frac{\gamma^2}{8\lfloor\frac{n}{2}\rfloor}\right).\end{aligned}$$

Finally, using the same proof than Hoeffding's allows us to conclude. □

Proposition 3.2. *The matrix* $\Sigma_{\boldsymbol{\alpha}^\star} = \mathrm{Diag}(\boldsymbol{\alpha}^\star) - \boldsymbol{\alpha}^\star(\boldsymbol{\alpha}^\star)^T$ *is semi-definite positive, of rank* $\mathcal{Q}-1$, *and* $\widehat{\boldsymbol{\alpha}}$ *is asymptotically normal:*

$$\sqrt{n}\left(\hat{\boldsymbol{\alpha}}\left(\mathbf{z}^\star\right)-\boldsymbol{\alpha}^\star\right)\xrightarrow[n\to\infty]{\mathcal{D}}\mathcal{N}(0,\Sigma_{\boldsymbol{\alpha}^\star}) \tag{3.3}$$

Similarly, let $V(\boldsymbol{\pi}^\star)$ *be the matrix defined by* $[V(\boldsymbol{\pi}^\star)]_{q\ell} = 1/\psi''(\pi^\star_{q\ell})$ *and* $\Sigma_{\boldsymbol{\pi}^\star} = \rho^{-1}\,\mathrm{Diag}^{-1}(\boldsymbol{\alpha}^\star)V(\boldsymbol{\pi}^\star)\,\mathrm{Diag}^{-1}(\boldsymbol{\alpha}^\star)$. *Then:*

$$\sqrt{n(n-1)/2}\left(\widehat{\pi}_{q\ell}\left(\mathbf{z}^\star\right)-\pi^\star_{q\ell}\right)\xrightarrow[n,\to\infty]{\mathcal{D}}\mathcal{N}(0,\Sigma_{\boldsymbol{\pi}^\star,q\ell}) \quad \textit{for all } q,\ell \tag{3.4}$$

where the components are independent.

Proof: Since $\hat{\boldsymbol{\alpha}}\left(\mathbf{z}^\star\right) = \left(\hat{\alpha}_1\left(\mathbf{z}^\star\right),\ldots,\hat{\alpha}_g\left(\mathbf{z}^\star\right)\right)$ is the sample mean of n i.i.d. multinomial random variables with parameters 1 and $\boldsymbol{\alpha}^\star$, a simple application of the central limit theorem (CLT) gives:

$$\Sigma_{\boldsymbol{\alpha}^\star,qq'} = \begin{cases}\alpha^\star_q(1-\alpha^\star_q) & \text{if} \quad q=q'\\ -\alpha^\star_q\alpha^\star_{q'} & \text{if} \quad q\neq q'\end{cases}$$

which proves Equation (3.3) where $\Sigma_{\boldsymbol{\alpha}^\star}$ is semi-definite positive of rank $\mathcal{Q}-1$.

Similarly, $\psi'\left(\widehat{\pi}_{q\ell}\left(\mathbf{z}^\star\right)\right)$ is the average of $\sum_{i<j} r_{ij}z^\star_{iq}z^\star_{j\ell}$ i.i.d. random variables with mean $\psi'\left(\pi^\star_{q\ell}\right)$ and variance $\psi''\left(\pi^\star_{q\ell}\right)$. $\sum_{i<j} r_{ij}z^\star_{iq}z^\star_{j\ell}$ is itself random but thanks to lemma 3.1 : $\frac{1}{\binom{n}{2}}\sum_{i<j} r_{ij}z^\star_{iq}z^\star_{j\ell}\xrightarrow[n\to+\infty]{\mathbb{P}}p\alpha^\star_q\alpha^\star_l$. Therefore, by Slutsky's lemma and the CLT for random sums of random variables [**?**], we have:

$$\begin{aligned}\sqrt{\frac{n(n-1)}{2}\rho\alpha^\star_q\alpha^\star_\ell}\left(\psi'\left(\widehat{\pi}_{q\ell}\left(\mathbf{z}^\star\right)\right)-\psi'(\pi^\star_{q\ell})\right) &= \sqrt{\frac{n(n-1)}{2}\rho\alpha^\star_q\alpha^\star_\ell}\left(\frac{\sum_{i<j}X_{ij}R_{ij}z^\star_{iq}z^\star_{j\ell}}{\sum_{i<j}R_{ij}z^\star_{iq}z^\star_{j\ell}}-\psi'(\pi^\star_{q\ell})\right)\\ &\xrightarrow[n,d\to+\infty]{\mathcal{D}}\mathcal{N}\left(0,\psi''(\pi^\star_{q\ell})\right)\end{aligned}$$

The differentiability of $(\psi')^{-1}$ and the delta method then gives:

$$\sqrt{\frac{n(n-1)}{2}}\left(\widehat{\pi}_{q\ell}\left(\mathbf{z}^\star\right)-\pi^\star_{q\ell}\right)\xrightarrow[n,d\to+\infty]{\mathcal{D}}\mathcal{N}\left(0,\frac{1}{\rho\alpha^\star_q\rho^\star_\ell\psi''(\pi^\star_{q\ell})}\right)$$

and the independence results from the independence of $\widehat{\pi}_{q\ell}\left(\mathbf{z}^\star\right)$ and $\widehat{\pi}_{q'\ell'}\left(\mathbf{z}^\star\right)$ as soon as $q\neq q'$ or $\ell\neq\ell'$, as they involve different sets of i.i.d. variables.

□

Proposition 3.3 (Local asymptotic normality). *Let $\mathcal{L}_c^\star$ the function defined on $\mathbf{\Theta}$ by $\mathcal{L}_c^\star(\boldsymbol{\alpha},\boldsymbol{\pi}) = \log p(\mathbf{y}^o, \mathbf{z}^\star; \boldsymbol{\theta})$. For any s, t and u in a compact set, we have:*

$$\begin{aligned}\mathcal{L}_c^\star\left(\boldsymbol{\alpha}^\star + \frac{s}{\sqrt{n}}, \boldsymbol{\pi}^\star + \frac{u}{\sqrt{\frac{n(n-1)}{2}}}\right) &= \mathcal{L}_c^\star(\boldsymbol{\theta}^\star) + s^T\mathbf{Y}_{\boldsymbol{\alpha}^\star} + Tr(u^T\mathbf{Y}_{\boldsymbol{\pi}^\star}) \\ &- \left(\frac{1}{2}s^T\Sigma_{\boldsymbol{\alpha}^\star}s + \frac{1}{2}Tr\left((u\odot u)^T\Sigma_{\boldsymbol{\pi}^\star}\right)\right) \\ &+ o_P(1)\end{aligned}$$

where $\odot$ denote the Hadamard product of two matrices (element-wise product) and $\Sigma_{\boldsymbol{\alpha}^\star}$, $\Sigma_{\boldsymbol{\rho}^\star}$ and $\Sigma_{\boldsymbol{\pi}^\star}$ are defined in Proposition 3.2. $\mathbf{Y}_{\boldsymbol{\alpha}^\star}$, $\mathbf{Y}_{\boldsymbol{\rho}^\star}$ are asymptotically Gaussian with zero mean and respective variance matrices $\Sigma_{\boldsymbol{\alpha}^\star}$, $\Sigma_{\boldsymbol{\rho}^\star}$ and $\mathbf{Y}_{\boldsymbol{\pi}^\star}$ is a matrix of asymptotically independent Gaussian components with zero mean and variance matrix $\Sigma_{\boldsymbol{\pi}^\star}$.

Proof.
By Taylor expansion,

$$\begin{aligned}&\mathcal{L}_c^\star\left(\boldsymbol{\alpha}^\star + \frac{s}{\sqrt{n}}, \boldsymbol{\pi}^\star + \frac{u}{\sqrt{\frac{n(n-1)}{2}}}\right) \\ = \;& \mathcal{L}_c^\star(\boldsymbol{\theta}^\star) + \frac{1}{\sqrt{n}}s^T\nabla\mathcal{L}_{c\boldsymbol{\alpha}}^\star(\boldsymbol{\theta}^\star) + \frac{1}{\sqrt{\frac{n(n-1)}{2}}}\mathrm{Tr}\left(u^T\nabla\mathcal{L}_{c\boldsymbol{\pi}}^\star(\boldsymbol{\theta}^\star)\right) \\ &+\frac{1}{n}s^T\mathbf{H}_{\boldsymbol{\alpha}}(\boldsymbol{\theta}^\star)s + \frac{1}{\frac{n(n-1)}{2}}\mathrm{Tr}\left((u\odot u)^T\mathbf{H}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^\star)\right) + o_P(1)\end{aligned}$$

where $\nabla\mathcal{L}_{c\boldsymbol{\alpha}}^\star(\boldsymbol{\theta}^\star)$ and $\nabla\mathcal{L}_{c\boldsymbol{\pi}}^\star(\boldsymbol{\theta}^\star)$ denote the respective components of the gradient of $\mathcal{L}_c^\star$ evaluated at $\boldsymbol{\theta}^\star$ and $\mathbf{H}_{\boldsymbol{\alpha}}$ and $\mathbf{H}_{\boldsymbol{\pi}}$ denote the conditional hessian of $\mathcal{L}_c^\star$ evaluated at $\boldsymbol{\theta}^\star$. By inspection, $\mathbf{H}_{\boldsymbol{\alpha}}/n$ and $\mathbf{H}_{\boldsymbol{\pi}}/\frac{n(n-1)}{2}$ converge in probability to constant matrices and the random vectors $\nabla\mathcal{L}_{c\boldsymbol{\alpha}}^\star(\boldsymbol{\theta}^\star)/\sqrt{n}$ and $\nabla\mathcal{L}_{c\boldsymbol{\pi}}^\star(\boldsymbol{\theta}^\star)/\sqrt{\frac{n(n-1)}{2}}$ converge in distribution by central limit theorem.

□

4. Profile Likelihood

To study the likelihood behaviors, we shall work conditionally to the real configurations $\mathbf{z}^\star$ that have enough observations in each group. We therefore define regular configurations which occur with high probability, then introduce conditional and profile log-likelihood ratio.

4.1. Regular assignments

Definition 4.1 (c-regular assignments)**.** *Let* $\mathbf{z} \in \mathcal{Z}$*. For any* $c > 0$*, we say that* $\mathbf{z}$ *is* c-regular *if*

$$\min_q z_{+q} \geq cn. \tag{4.1}$$

In regular configurations, each group has $\Omega(n)$ members, where $u_n = \Omega(n)$ if there exists two constant $a, b > 0$ such that for n enough large $an \leq u_n \leq bn$. $c/2$-regular assignments, with c defined in Assumption H_1, have high $\mathbb{P}_{\boldsymbol{\theta}^\star}$-probability in the space of all assignments, uniformly over all $\boldsymbol{\theta}^\star \in \boldsymbol{\Theta}$.

Each z_{+q} is a sum of n i.i.d Bernoulli r.v. with parameter $\alpha_q \geq \alpha_{\min} \geq c$. A simple Hoeffding bound shows that

$$\mathbb{P}_{\boldsymbol{\theta}^\star}\left(z_{+q} \leq n\frac{c}{2}\right) \leq \mathbb{P}_{\boldsymbol{\theta}^\star}\left(z_{+q} \leq n\frac{\alpha_q}{2}\right) \leq \exp\left(-2n\left(\frac{\alpha_q}{2}\right)^2\right) \leq \exp\left(-\frac{nc^2}{2}\right)$$

taking a union bound over $\mathcal{Q}$ values of q lead to Proposition 4.2.

Proposition 4.2. *Define* $\mathcal{Z}_1$ *as the subset of* $\mathcal{Z}$ *made of* $c/2$*-regular assignments, with* c *defined in assumption* H_1*. Note* Ω_1 *the event* $\{\mathbf{z}^\star \in \mathcal{Z}_1\}$*, then:*

$$\mathbb{P}_{\boldsymbol{\theta}^\star}\left(\bar{\Omega}_1\right) \leq \mathcal{Q} \exp\left(-\frac{nc^2}{2}\right).$$

We define now balls of configurations taking into account equivalent assignments classes.

Definition 4.3 (Set of local assignments)**.** *We note* $S(\mathbf{z}^\star, r)$ *the set of configurations that have a representative (for* $\sim$*) within relative radius* r *of* $\mathbf{z}^\star$*:*

$$S(\mathbf{z}^\star, r) = \{\mathbf{z} : \|\mathbf{z} - \mathbf{z}^\star\|_{0,\sim} \leq rn\}$$

4.2. Conditional and profile log-likelihoods

We first introduce few notations.

Definition 4.4. *We define the conditional log-likelihood ratio* F_n *and its expectation* G *as:*

$$\begin{aligned} F_n(\boldsymbol{\theta}, \mathbf{z}) &= \log \frac{p(\mathbf{y^o}|\mathbf{z}; \boldsymbol{\theta})}{p(\mathbf{y^o}|\mathbf{z}^\star; \boldsymbol{\theta}^\star)} \\ G(\boldsymbol{\theta}, \mathbf{z}) &= \mathbb{E}_{\boldsymbol{\theta}^\star}\left[\log \frac{p(\mathbf{y^o}|\mathbf{z}; \boldsymbol{\theta})}{p(\mathbf{y^o}|\mathbf{z}^\star; \boldsymbol{\theta}^\star)} \middle| \mathbf{z}^\star\right] \end{aligned} \tag{4.2}$$

We also define the profile log-likelihood ratio Λ *and its expectation* $\tilde{\Lambda}$ *as:*

$$\begin{aligned} \Lambda(\mathbf{z}) &= \max_{\boldsymbol{\theta}} F_n(\boldsymbol{\theta}, \mathbf{z}) \\ \tilde{\Lambda}(\mathbf{z}) &= \max_{\boldsymbol{\theta}} G(\boldsymbol{\theta}, \mathbf{z}). \end{aligned} \tag{4.3}$$

Remark 4.5. As F_n and G only depend on $\boldsymbol{\theta}$ through $\boldsymbol{\pi}$, we will sometimes replace $\boldsymbol{\theta}$ with $\boldsymbol{\pi}$ in the expressions of F_n and G. Replacing F_n and G by their profiled version Λ and $\tilde{\Lambda}$ allows us to get rid of the continuous argument of F_n and to effectively use discrete contrasts Λ and $\tilde{\Lambda}$.

The following proposition shows which values of $\boldsymbol{\pi}$ maximize F_n and G to attain Λ and $\tilde{\Lambda}$.

Proposition 4.6 (maximum of G and $\tilde{\Lambda}$ in θ)**.** *Conditionally on $\mathbf{z}^\star$, define the following quantities:*

$$\boldsymbol{S}^\star = (S^\star_{q\ell})_{q\ell} = \left(\psi'(\pi^\star_{q\ell})\right)_{q\ell}$$
$$\bar{y}_{q\ell}(\mathbf{z}) = \mathbb{E}_{\boldsymbol{\theta}^\star}[\widehat{y}_{q\ell}(\mathbf{z})|\mathbf{z}^\star] = \frac{\left[\mathrm{I\!R}_{\mathcal{Q}}(\mathbf{z})^T \boldsymbol{S}^\star \mathrm{I\!R}_{\mathcal{Q}}(\mathbf{z})\right]_{q\ell}}{\widehat{\alpha}_q(\mathbf{z})\widehat{\alpha}_\ell(\mathbf{z})} \tag{4.4}$$

with $\bar{y}_{q\ell}(\mathbf{z}) = 0$ for $\mathbf{z}$ such that $\widehat{\alpha}_q(\mathbf{z}) = 0$ or $\widehat{\alpha}_\ell(\mathbf{z}) = 0$. Then $F_{nd}(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w})$ and $G(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w})$ are maximum in $\boldsymbol{\pi}$ for $\widehat{\boldsymbol{\pi}}(\mathbf{z}, \mathbf{w})$ and $\bar{\boldsymbol{\pi}}(\mathbf{z}, \mathbf{w})$ defined by:

$$\widehat{\pi}(\mathbf{z}, \mathbf{w})_{q\ell} = (\psi')^{-1}(\widehat{x}_{q\ell}(\mathbf{z}, \mathbf{w})) \quad \textit{and} \quad \bar{\pi}(\mathbf{z}, \mathbf{w})_{q\ell} = (\psi')^{-1}(\bar{x}_{q\ell}(\mathbf{z}, \mathbf{w}))$$

so that

$$\Lambda(\mathbf{z}, \mathbf{w}) = F_{nd}(\widehat{\boldsymbol{\pi}}(\mathbf{z}, \mathbf{w}), \mathbf{z}, \mathbf{w})$$
$$\tilde{\Lambda}(\mathbf{z}, \mathbf{w}) = G(\bar{\boldsymbol{\pi}}(\mathbf{z}, \mathbf{w}), \mathbf{z}, \mathbf{w})$$

Note that although $\bar{x}_{q\ell} = \mathbb{E}_{\boldsymbol{\theta}^\star}\left[\hat{x}_{q\ell}|\, \mathbf{z}^\star, \mathbf{w}^\star\right]$, in general $\bar{\pi}_{q\ell} \neq \mathbb{E}_{\boldsymbol{\theta}^\star}\left[\hat{\pi}_{q\ell}|\, \mathbf{z}^\star, \mathbf{w}^\star\right]$ by non linearity of $(\psi')^{-1}$. Nevertheless, $(\psi')^{-1}$ is Lipschitz over compact subsets of $\psi'(\mathring{\mathcal{A}})$ and therefore, with high probability, $|\bar{\pi}_{q\ell} - \hat{\pi}_{q\ell}|$ and $|\widehat{x}_{q\ell} - \bar{x}_{q\ell}|$ are of the same order of magnitude.

The maximum and argmax of G and $\tilde{\Lambda}$ are characterized by the following propositions.

Proposition 4.7 (maximum of G and $\tilde{\Lambda}$ in $(\theta, \mathbf{z}, \mathbf{w})$)**.** *Let $\mathrm{KL}(\pi, \pi') = \psi'(\pi)(\pi - \pi') + \psi(\pi') - \psi(\pi)$ be the Kullback divergence between $\varphi(., \pi)$ and $\varphi(., \pi')$ then:*

$$G(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}) = -nd \sum_{q,q'} \sum_{\ell,\ell'} \mathrm{I\!R}_g(\mathbf{z})_{q,q'} \mathrm{I\!R}_m(\mathbf{w})_{\ell,\ell'} \,\mathrm{KL}(\pi^\star_{q\ell}, \pi_{q'\ell'}) \leq 0. \tag{4.5}$$

Conditionally on the set Ω_1 of regular assignments and for $n, d > 2/c$,

(i) G is maximized at $(\boldsymbol{\pi}^\star, \mathbf{z}^\star, \mathbf{w}^\star)$ and its equivalence class.
(ii) $\tilde{\Lambda}$ is maximized at $(\mathbf{z}^\star, \mathbf{w}^\star)$ and its equivalence class; moreover, $\tilde{\Lambda}(\mathbf{z}^\star, \mathbf{w}^\star) = 0$.
(iii) The maximum of $\tilde{\Lambda}$ (and hence the maximum of G) is well separated.

Property *(iii)* of Proposition 4.7 is a direct consequence of the local upperbound for $\tilde{\Lambda}$ as stated as follows:

Proposition 4.8 (Local upperbound for $\tilde{\Lambda}$)**.** *Conditionally upon Ω_1, there exists a positive constant C such that for all $(\mathbf{z}, \mathbf{w}) \in S(\mathbf{z}^\star, \mathbf{w}^\star, C)$:*

$$\tilde{\Lambda}(\mathbf{z}, \mathbf{w}) \leq -\frac{c\delta(\boldsymbol{\pi}^\star)}{4}\left(d\|\mathbf{z} - \mathbf{z}^\star\|_{0,\sim} + n\|\mathbf{w} - \mathbf{w}^\star\|_{0,\sim}\right) \tag{4.6}$$

Proofs of Propositions 4.6, 4.7 and 4.8 are reported in Appendix A.

5. Main Result

We are now ready to present our main result stated in Theorem 5.1.

Theorem 5.1 (complete-observed)**.** *Consider that assumptions H_1 to H_4 hold for the Latent Block Model of known order with $n \times d$ observations coming from an univariate exponential family and define $\# \operatorname{Sym}(\boldsymbol{\theta})$ as the set of pairs of permutation (s,t) for which $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\rho}, \boldsymbol{\pi})$ exhibits symmetry. Then, for n and d tending to infinity with asymptotic rates $\log(d)/n \to 0$ and $\log(n)/d \to 0$, the observed likelihood ratio behaves like the complete likelihood ratio, up to a bounded multiplicative factor:*

$$\frac{p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^\star)} = \frac{\# \operatorname{Sym}(\boldsymbol{\theta})}{\# \operatorname{Sym}(\boldsymbol{\theta}^\star)} \max_{\boldsymbol{\theta}' \sim \boldsymbol{\theta}} \frac{p(\mathbf{x}, \mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}')}{p(\mathbf{x}, \mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star)} (1 + o_P(1)) + o_P(1)$$

where the o_P is uniform over all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

The maximum over all $\boldsymbol{\theta}'$ that are equivalent to $\boldsymbol{\theta}$ stems from the fact that because of label-switching, $\boldsymbol{\theta}$ is only identifiable up to its $\sim$-equivalence class from the observed likelihood, whereas it is completely identifiable from the complete likelihood.
As already pointed out, if $\boldsymbol{\Theta}$ exhibits symmetry, the maximum likelihood assignment is not unique under the true model, and $\# \operatorname{Sym}(\boldsymbol{\theta})$ terms contribute with the same weight. This was not taken into account by [**?**]. The next corollary is deduced immediately :

Corollary 5.2. *If $\boldsymbol{\Theta}$ contains only parameters that do not exhibit symmetry:*

$$\frac{p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^\star)} = \max_{\boldsymbol{\theta}' \sim \boldsymbol{\theta}} \frac{p(\mathbf{x}, \mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}')}{p(\mathbf{x}, \mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star)} (1 + o_P(1)) + o_P(1)$$

where the o_P is uniform over all $\boldsymbol{\Theta}$.

Using the conditional log-likelihood, the observed likelihood can be written as

$$\begin{aligned} p(\mathbf{x};\boldsymbol{\theta}) &= \sum_{(\mathbf{z},\mathbf{w})} p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta}) \\ &= p(\mathbf{x}|\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star) \sum_{(\mathbf{z},\mathbf{w})} p(\mathbf{z},\mathbf{w};\boldsymbol{\theta}) \exp(F_{nd}(\boldsymbol{\theta},\mathbf{z},\mathbf{w})). \end{aligned} \tag{5.1}$$

The proof proceeds with an examination of the asymptotic behavior of F_{nd} on three types of configurations that partition $\mathcal{Z} \times \mathcal{W}$:

1. *global control*: for $(\mathbf{z},\mathbf{w})$ such that $\tilde{\Lambda}(\mathbf{z},\mathbf{w}) = \Omega(-nd)$, Proposition 5.3 proves a large deviation behavior for $F_{nd} = -\Omega_P(nd)$ and in turn those assignments contribute a o_P of $p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star))$ to the sum (Proposition 5.4).

2. *local control*: a small deviation result (Proposition 5.5) is needed to show that the combined contribution of assignments close to but not equivalent to $(\mathbf{z}^\star, \mathbf{w}^\star)$ is also a o_P of $p(\mathbf{x}, \mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star)$ (Proposition 5.6).
3. *equivalent assignments*: Proposition 5.7 examines which of the remaining assignments, all equivalent to $(\mathbf{z}^\star, \mathbf{w}^\star)$, contribute to the sum.

These results are presented in next section 5.1 and their proofs reported in Appendix A. They are then put together in section 5.2 to achieve the proof of our main result. The remainder of the section is devoted to the asymptotics of the ML and variational estimators as a consequence of the main result.

5.1. Different asymptotic behaviors

We begin with a large deviations inequality for configurations $(\mathbf{z}, \mathbf{w})$ far from $(\mathbf{z}^\star, \mathbf{w}^\star)$ and leverage it to prove that far away configurations make a small contribution to $p(\mathbf{x}; \boldsymbol{\theta})$.

5.1.1. Global Control

Proposition 5.3 (large deviations of F_{nd}). *Let* $\mathrm{Diam}(\boldsymbol{\Theta}) = \sup_{\boldsymbol{\theta}, \boldsymbol{\theta}'} \|\boldsymbol{\theta} - \boldsymbol{\theta}'\|_\infty$. *For all* $\varepsilon_{n,d} < \kappa/(2\sqrt{2}\,\mathrm{Diam}(\boldsymbol{\Theta}))$ *and* n, d *large enough that*

$$\begin{aligned} &\Delta^1_{nd}(\varepsilon_{nd}) \\ &= \mathbb{P}\left(\sup_{\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}} \left\{ F_{nd}(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}) - \tilde{\Lambda}(\mathbf{z}, \mathbf{w}) \right\} \geq \bar{\sigma} nd\, \mathrm{Diam}(\boldsymbol{\Theta}) 2\sqrt{2} \varepsilon_{nd} \left[1 + \frac{gm}{2\sqrt{2nd}\varepsilon_{nd}} \right] \right) \\ &\leq g^n m^d \exp\left(-\frac{nd\varepsilon_{nd}^2}{2} \right) \end{aligned} \tag{5.2}$$

Proposition 5.4 (contribution of global assignments). *Assume* $\log(d)/n \to 0$, $\log(n)/d \to 0$ *when* n *and* d *tend to infinity, and choose* t_{nd} *decreasing to* 0 *such that* $t_{nd} \gg \max(\frac{n+d}{nd}, \frac{\log(nd)}{\sqrt{nd}})$. *Then conditionally on* Ω_1 *and for* n, d *large enough that* $2\sqrt{2nd}t_{nd} \geq gm$, *we have:*

$$\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \sum_{(\mathbf{z}, \mathbf{w}) \notin S(\mathbf{z}^\star, \mathbf{w}^\star, t_{nd})} p(\mathbf{z}, \mathbf{w}, \mathbf{x}; \boldsymbol{\theta}) = o_P(p(\mathbf{z}^\star, \mathbf{w}^\star, \mathbf{x}; \boldsymbol{\theta}^\star))$$

5.1.2. Local Control

Proposition 5.3 gives deviations of order $\mathcal{O}_P(\sqrt{nd})$, which are only useful for $(\mathbf{z}, \mathbf{w})$ such that G and $\tilde{\Lambda}$ are large compared to $\sqrt{nd}$. For $(\mathbf{z}, \mathbf{w})$ close to $(\mathbf{z}^\star, \mathbf{w}^\star)$, we need tighter concentration inequalities, of order $o_P(-(n+d))$, as follows:

Proposition 5.5 (small deviations F_{nd})**.** *Conditionally upon Ω_1, there exists three positive constant c_1, c_2 and C such that for all $\varepsilon \leq \kappa\underline{\sigma}^2$, for all $(\mathbf{z},\mathbf{w}) \nsim (\mathbf{z}^\star,\mathbf{w}^\star)$ such that $(\mathbf{z},\mathbf{w}) \in S(\mathbf{z}^\star,\mathbf{w}^\star,C)$:*

$$\Delta^2_{nd}(\varepsilon) = \mathbb{P}_{\boldsymbol{\theta}^\star}\left(\sup_{\boldsymbol{\theta}} \frac{F_{nd}(\boldsymbol{\theta},\mathbf{z},\mathbf{w}) - \tilde{\Lambda}(\mathbf{z},\mathbf{w})}{d\|\mathbf{z}-\mathbf{z}^\star\|_{0,\sim} + n\|\mathbf{w}-\mathbf{w}^\star\|_{0,\sim}} \geq \varepsilon\right) \leq \exp\left(-\frac{ndc^2\varepsilon^2}{128(c_1\bar{\sigma}^2 + c_2\kappa^{-1}\varepsilon)}\right) \tag{5.3}$$

The next propositions builds on Proposition 5.5 and 4.7 to show that the combined contributions of assignments close to $(\mathbf{z}^\star,\mathbf{w}^\star)$ to the observed likelihood is also a o_P of $p(\mathbf{z}^\star,\mathbf{w}^\star,\mathbf{x};\boldsymbol{\theta}^\star)$

Proposition 5.6 (contribution of local assignments)**.** *With the previous notations*

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{\substack{(\mathbf{z},\mathbf{w})\in S(\mathbf{z}^\star,\mathbf{w}^\star,C)\\(\mathbf{z},\mathbf{w})\nsim(\mathbf{z}^\star,\mathbf{w}^\star)}} p(\mathbf{z},\mathbf{w},\mathbf{x};\boldsymbol{\theta}) = o_P(p(\mathbf{z}^\star,\mathbf{w}^\star,\mathbf{x};\boldsymbol{\theta}^\star))$$

5.1.3. Equivalent assignments

It remains to study the contribution of equivalent assignments.

Proposition 5.7 (contribution of equivalent assignments)**.** *For all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, we have*

$$\sum_{(\mathbf{z},\mathbf{w})\sim(\mathbf{z}^\star,\mathbf{w}^\star)} \frac{p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta})}{p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star)} = \#\,\mathrm{Sym}(\boldsymbol{\theta}) \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} \frac{p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}')}{p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star)}(1+o_P(1))$$

where the o_P is uniform in $\boldsymbol{\theta}$.

5.2. Proof of the main result

Proof.
We work conditionally on Ω_1. Choose $(\mathbf{z}^\star,\mathbf{w}^\star) \in \mathcal{Z}_1 \times \mathcal{W}_1$ and a sequence t_{nd} decreasing to 0 but satisfying $t_{nd} \gg \max\left(\frac{n+d}{nd}, \frac{\log(nd)}{\sqrt{nd}}\right)$ (this is possible since $\log(d)/n \to 0$ and $\log(n)/d \to 0$). According to Proposition 5.4,

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{(\mathbf{z},\mathbf{w})\notin S(\mathbf{z}^\star,\mathbf{w}^\star,t_{nd})} p(\mathbf{z},\mathbf{w},\mathbf{x};\boldsymbol{\theta}) = o_P(p(\mathbf{z}^\star,\mathbf{w}^\star,\mathbf{x};\boldsymbol{\theta}^\star))$$

Since t_{nd} decreases to 0, it gets smaller than C (used in proposition 5.6) for n, d large enough. As this point, Proposition 5.6 ensures that:

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \sum_{\substack{(\mathbf{z},\mathbf{w})\in S(\mathbf{z}^\star,\mathbf{w}^\star,t_{nd})\\(\mathbf{z},\mathbf{w})\nsim(\mathbf{z}^\star,\mathbf{w}^\star)}} p(\mathbf{z},\mathbf{w},\mathbf{x};\boldsymbol{\theta}) = o_P(p(\mathbf{z}^\star,\mathbf{w}^\star,\mathbf{x};\boldsymbol{\theta}^\star))$$

And therefore the observed likelihood ratio reduces as:

$$\frac{p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^\star)} = \frac{\displaystyle\sum_{(\mathbf{z},\mathbf{w})\sim(\mathbf{z}^\star,\mathbf{w}^\star)} p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta}) + \sum_{(\mathbf{z},\mathbf{w})\nsim(\mathbf{z}^\star,\mathbf{w}^\star)} p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta})}{\displaystyle\sum_{(\mathbf{z},\mathbf{w})\sim(\mathbf{z}^\star,\mathbf{w}^\star)} p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta}^\star) + \sum_{(\mathbf{z},\mathbf{w})\nsim(\mathbf{z}^\star,\mathbf{w}^\star)} p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta}^\star)}$$
$$= \frac{\displaystyle\sum_{(\mathbf{z},\mathbf{w})\sim(\mathbf{z}^\star,\mathbf{w}^\star)} p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta}) + p(\mathbf{x};\mathbf{z}^\star,\mathbf{w}^\star,\boldsymbol{\theta}^\star)o_P(1)}{\displaystyle\sum_{(\mathbf{z},\mathbf{w})\sim(\mathbf{z}^\star,\mathbf{w}^\star)} p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta}^\star) + p(\mathbf{x};\mathbf{z}^\star,\mathbf{w}^\star,\boldsymbol{\theta}^\star)o_P(1)}$$

And Proposition 5.7 allows us to conclude

$$\frac{p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^\star)} = \frac{\#\,\mathrm{Sym}(\boldsymbol{\theta})}{\#\,\mathrm{Sym}(\boldsymbol{\theta}^\star)} \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} \frac{p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}')}{p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star)}(1+o_P(1)) + o_P(1).$$

□

5.3. *Asymptotics for the MLE of $\boldsymbol{\theta}$*

The asymptotic behavior of the maximum likelihood estimator in the incomplete data model is a direct consequence of Theorem 5.1.

Corollary 5.8 (Asymptotic behavior of $\widehat{\boldsymbol{\theta}}_{MLE}$)**.** *Denote $\widehat{\boldsymbol{\theta}}_{MLE}$ the maximum likelihood estimator and use the notations of Proposition 3.2. If $\#\,\mathrm{Sym}(\boldsymbol{\theta}) = 1$, there exist permutations s of $\{1,\dots,g\}$ and t of $\{1,\dots,m\}$ such that*

$$\hat{\boldsymbol{\alpha}}(\mathbf{z}^\star) - \widehat{\boldsymbol{\alpha}}^s_{MLE} = o_P\left(n^{-1/2}\right), \qquad \hat{\boldsymbol{\rho}}(\mathbf{w}^\star) - \widehat{\boldsymbol{\rho}}^t_{MLE} = o_P\left(d^{-1/2}\right),$$
$$\hat{\boldsymbol{\pi}}(\mathbf{z}^\star,\mathbf{w}^\star) - \widehat{\boldsymbol{\pi}}^{s,t}_{MLE} = o_P\left((nd)^{-1/2}\right).$$

If $\#\,\mathrm{Sym}(\boldsymbol{\theta}) \neq 1$, $\widehat{\boldsymbol{\theta}}_{MLE}$ is still consistent: there exist permutations s of $\{1,\dots,g\}$ and t of $\{1,\dots,m\}$ such that

$$\hat{\boldsymbol{\alpha}}(\mathbf{z}^\star) - \widehat{\boldsymbol{\alpha}}^s_{MLE} = o_P(1), \qquad \hat{\boldsymbol{\rho}}(\mathbf{w}^\star) - \widehat{\boldsymbol{\rho}}^t_{MLE} = o_P(1),$$
$$\hat{\boldsymbol{\pi}}(\mathbf{z}^\star,\mathbf{w}^\star) - \widehat{\boldsymbol{\pi}}^{s,t}_{MLE} = o_P(1).$$

Hence, the maximum likelihood estimator for the LBM is consistent and asymptotically normal, with the same behavior as the maximum likelihood estimator in the complete data model when θ does not exhibit any symmetry. The proof in appendix A.9 relies on the local asymptotic normality of the MLE in the complete model, as stated in Proposition 3.3 and on our main Theorem.

Appendix A: Proofs

A.1. Proof of Proposition 4.6

Proof.
Define $\nu(y,\pi) = y\pi - \psi(\pi)$. For y fixed, $\nu(y,\pi)$ is maximized at $\pi = (\psi')^{-1}(y)$. Define $n_{q\ell}(z) = \left(\sum_{i<j} r_{ij} z_{iq} z_{j\ell}\right)$, manipulations yield

$$\begin{aligned}
F_n(\boldsymbol{\pi},\mathbf{z}) &= \log p(\mathbf{y^o};\mathbf{z},\boldsymbol{\theta}) - \log p(\mathbf{y^o};\mathbf{z}^\star,\boldsymbol{\theta}^\star) \\
&= \left[\sum_q \sum_\ell n_{q\ell}(z)\nu(\widehat{y}_{q\ell}(\mathbf{z}),\pi_{q\ell}) - \sum_q \sum_\ell n_{q\ell}(z^\star)\nu(\widehat{y}_{q\ell}(\mathbf{z}^\star),\pi^\star_{q\ell})\right]
\end{aligned}$$

which is maximized at $\pi_{q\ell} = (\psi')^{-1}(\widehat{y}_{q\ell}(\mathbf{z}))$. Similarly

$$\begin{aligned}
G(\boldsymbol{\pi},\mathbf{z}) &= \mathbb{E}_{\boldsymbol{\theta}^\star}[\log p(\mathbf{y^o};\mathbf{z},\boldsymbol{\theta}) - \log p(\mathbf{y^o};\mathbf{z}^\star,\boldsymbol{\theta}^\star)] \\
&= \left[\sum_q \sum_\ell n_{q\ell}(z)\nu(\bar{y}_{q\ell}(\mathbf{z}),\pi_{q\ell}) - \sum_q \sum_\ell n_{q\ell}(z^\star)\nu(\psi'(\pi^\star_{q\ell}),\pi^\star_{q\ell})\right]
\end{aligned}$$

is maximized at $\pi_{q\ell} = (\psi')^{-1}(\bar{y}_{q\ell}(\mathbf{z}))$

□

A.2. Proof of Proposition 4.7 (maximum of G and $\tilde{\Lambda}$)

Proof.
We condition on $\mathbf{z}^\star$ and prove Equation (4.5):

$$\begin{aligned}
G(\boldsymbol{\theta},\mathbf{z}) &= \mathbb{E}_{\boldsymbol{\theta}^\star}\left[\frac{p(\mathbf{y^o};\mathbf{z},\boldsymbol{\theta})}{p(\mathbf{y^o};\mathbf{z}^\star,\boldsymbol{\theta}^\star)}\middle|\,\mathbf{z}^\star\right] \\
&= \sum_i \sum_j \sum_{q,q'} \sum_{\ell,\ell'} \mathbb{E}_{\boldsymbol{\theta}^\star}\left[y_{ij}(\pi_{q'\ell'} - \pi^\star_{q\ell}) - (\psi(\pi_{q'\ell'}) - \psi(\pi^\star_{q\ell}))\right]\rho z^\star_{iq} z_{iq'} z^\star_{j\ell} z_{j\ell'} \\
&= n^2\rho \sum_{q,q'} \sum_{\ell,\ell'} \mathrm{IR}_{\mathcal{Q}}(\mathbf{z})_{q,q'} \mathrm{IR}_{\mathcal{Q}}(\mathbf{z})_{\ell,\ell'} \left[\psi'(\pi^\star_{q\ell})(\pi_{q'\ell'} - \pi^\star_{q\ell}) + \psi(\pi^\star_{q\ell}) - \psi(\pi_{q'\ell'})\right] \\
&= -n^2\rho \sum_{q,q'} \sum_{\ell,\ell'} \mathrm{IR}_{\mathcal{Q}}(\mathbf{z})_{q,q'} \mathrm{IR}_{\mathcal{Q}}(\mathbf{z})_{\ell,\ell'}\, \mathrm{KL}(\pi^\star_{q\ell},\pi_{q'\ell'})
\end{aligned}$$

If $\mathbf{z}^\star$ is regular, and for $n > 2/c$, all the rows of $\mathrm{IR}_{\mathcal{Q}}(\mathbf{z})$ have at least one positive element and we can apply lemma B.4 (which is an adaptation for LBM of Lemma 3.2 of [**?**] for SBM) to characterize the maximum for G.

The maximality of $\tilde{\Lambda}(\mathbf{z}^\star)$ results from the fact that $\tilde{\Lambda}(\mathbf{z}) = G(\bar{\boldsymbol{\pi}}(\mathbf{z}),\mathbf{z})$ where $\bar{\boldsymbol{\pi}}(\mathbf{z})$ is a particular value of $\boldsymbol{\pi}$, $\tilde{\Lambda}$ is immediately maximum at $\mathbf{z} \sim \mathbf{z}^\star$, and for those, we have $\bar{\boldsymbol{\pi}}(\mathbf{z}) \sim \boldsymbol{\pi}^\star$.

The separation and local behavior of G around $\mathbf{z}^\star$ is a direct consequence of the proposition 4.8.

□

A.3. Proof of Proposition 4.8 (Local upper bound for $\tilde{\Lambda}$)

Proof.

We work conditionally on $(\mathbf{z}^\star, \mathbf{w}^\star)$. The principle of the proof relies on the extension of $\tilde{\Lambda}$ to a continuous subspace of$M_g([0,1]) \times M_m([0,1])$, in which confusion matrices are naturally embedded. The regularity assumption allows us to work on a subspace that is bounded away from the borders of $M_g([0,1]) \times M_m([0,1])$. The proof then proceeds by computing the gradient of $\tilde{\Lambda}$ at and around its argmax and using those gradients to control the local behavior of $\tilde{\Lambda}$ around its argmax. The local behavior allows us in turn to show that $\tilde{\Lambda}$ is well-separated.

Note that $\tilde{\Lambda}$ only depends on $\mathbf{z}$ and $\mathbf{w}$ through $\mathrm{I\!R}_g(\mathbf{z})$ and $\mathrm{I\!R}_m(\mathbf{w})$. We can therefore extend it to matrices $(U,V) \in \mathcal{U}_c \times \mathcal{V}_c$ where $\mathcal{U}$ is the subset of matrices $\mathcal{M}_g([0,1])$ with each row sum higher than $c/2$ and $\mathcal{V}$ is a similar subset of $\mathcal{M}_m([0,1])$.

$$\tilde{\Lambda}(U,V) = -nd \sum_{q,q'} \sum_{\ell,\ell'} U_{qq'} V_{\ell\ell'} \,\mathrm{KL}\left(\pi^\star_{q\ell}, \bar{\pi}_{q'\ell'}\right)$$

where

$$\bar{\pi}_{q\ell} = \bar{\pi}_{q\ell}(U,V) = (\psi')^{-1}\left(\frac{\left[U^T \boldsymbol{S}^\star V\right]_{q\ell}}{\left[U^T \mathbf{1} V\right]_{q\ell}}\right)$$

and $\mathbf{1}$ is the $g \times m$ matrix filled with 1. Confusion matrices $\mathrm{I\!R}_g(\mathbf{z})$ and $\mathrm{I\!R}_m(\mathbf{w})$ satisfy $\mathrm{I\!R}_g(\mathbf{z})\mathbb{1} = \boldsymbol{\alpha}(\mathbf{z}^\star)$ and $\mathrm{I\!R}_m(\mathbf{w})\mathbb{1} = \boldsymbol{\rho}(\mathbf{w}^\star)$, with $\mathbb{1} = (1,\dots,1)^T$ a vector only containing 1 values, and are obviously in $\mathcal{U}_c$ and $\mathcal{V}_c$ as soon as $(\mathbf{z}^\star, \mathbf{w}^\star)$ is $c/2$ regular.

The maps $f_{q,\ell} : (U,V) \mapsto KL(\pi^\star_{q\ell}, \bar{\pi}_{q\ell}(U,V))$ are twice differentiable with second derivatives bounded over $\mathcal{U}_c \times \mathcal{V}_c$ and therefore so is $\tilde{\Lambda}(U,V)$. Tedious but straightforward computations show that the derivative of $\tilde{\Lambda}$ at $(D_\alpha, D_\rho) \coloneqq (\mathrm{Diag}(\boldsymbol{\alpha}(\mathbf{z}^\star)), \mathrm{Diag}(\boldsymbol{\rho}(\mathbf{w}^\star)))$ is:

$$A_{qq'}(\mathbf{w}^\star) \coloneqq \frac{\partial \tilde{\Lambda}}{\partial U_{qq'}}(D_\alpha, D_\rho) = \sum_\ell \rho_\ell(\mathbf{w}^\star)\,\mathrm{KL}\left(\pi^\star_{q\ell}, \pi^\star_{q'\ell}\right)$$

$$B_{\ell\ell'}(\mathbf{z}^\star) \coloneqq \frac{\partial \tilde{\Lambda}}{\partial V_{\ell\ell'}}(D_\alpha, D_\rho) = \sum_q \rho_\ell(\mathbf{z}^\star)\,\mathrm{KL}\left(\pi^\star_{q\ell}, \pi^\star_{q\ell'}\right)$$

$A(\mathbf{w}^\star)$ and $B(\mathbf{z}^\star)$ are the matrix-derivative of $-\tilde{\Lambda}/nd$ at (D_α, D_ρ). Since $(\mathbf{z}^\star, \mathbf{w}^\star)$ is $c/2$-regular and by definition of $\delta(\boldsymbol{\pi}^\star)$, $A(\mathbf{w}^\star)_{qq'} \geq c\delta(\boldsymbol{\pi}^\star)/2$ (resp. $B(\mathbf{w}^\star)_{\ell\ell'} \geq c\delta(\boldsymbol{\pi}^\star)/2$) if $q \neq q'$ (resp. $\ell \neq \ell'$) and $A(\mathbf{w}^\star)_{qq} = 0$ (resp. $B(\mathbf{z}^\star)_{\ell\ell} = 0$) for all q (resp. ℓ). By boundedness of the second derivative, there exists $C > 0$ such that for all (D_α, D_ρ) and all $(H,G) \in B(D_\alpha, D_\rho, C)$, we have:

$$\frac{-1}{nd}\frac{\partial \tilde{\Lambda}}{\partial U_{qq'}}(H,G) \begin{cases} \geq \frac{3c\delta(\boldsymbol{\pi}^\star)}{8} & \text{if } q \neq q' \\ \leq \frac{c\delta(\boldsymbol{\pi}^\star)}{8} & \text{if } q = q' \end{cases} \quad \text{and} \quad \frac{-1}{nd}\frac{\partial \tilde{\Lambda}}{\partial V_{\ell\ell'}}(H,G) \begin{cases} \geq \frac{3c\delta(\boldsymbol{\pi}^\star)}{8} & \text{if } \ell \neq \ell' \\ \leq \frac{c\delta(\boldsymbol{\pi}^\star)}{8} & \text{if } \ell = \ell' \end{cases}$$

Choose U and V in $(\mathcal{U}_c \times \mathcal{V}_c) \cap B(D_\alpha, D_\rho, C)$ satisfying $U\mathbb{1} = \boldsymbol{\alpha}(\mathbf{z}^\star)$ and $V\mathbb{1} = \boldsymbol{\rho}(\mathbf{w}^\star)$. $U - D_\alpha$ and $V - D_\rho$ have nonnegative off diagonal coefficients and negative diagonal coefficients. Furthermore, the coefficients of U, V, D_α, D_ρ sum up to 1 and $\mathrm{Tr}(D_\alpha) = \mathrm{Tr}(D_\rho) = 1$. By Taylor expansion, there exists a couple (H, G) also in $(\mathcal{U}_c \times \mathcal{V}_c) \cap B(D_\alpha, D_\rho, C)$ such that

$$\frac{-1}{nd}\tilde{\Lambda}\,(U,V) = \frac{-1}{nd}\tilde{\Lambda}\,(D_\alpha, D_\rho) + \mathrm{Tr}\left((U - D_\alpha)\frac{\partial\tilde{\Lambda}}{\partial U}(H,G)\right) + \mathrm{Tr}\left((V - D_\rho)\frac{\partial\tilde{\Lambda}}{\partial V}(H,G)\right)$$

$$\geq \frac{c\delta(\boldsymbol{\pi}^\star)}{8}[3\sum_{q\neq q'}(U-D_\alpha)_{qq'} + 3\sum_{\ell\neq\ell'}(V-D_\rho)_{\ell\ell'} - \sum_{q}(U-D_\alpha)_{qq} - \sum_{\ell}(V-D_\rho)_{\ell\ell}]$$

$$= \frac{c\delta(\boldsymbol{\pi}^\star)}{4}[(1 - \mathrm{Tr}(U)) + (1 - \mathrm{Tr}(V))]$$

To conclude the proof, assume without loss of generality that $(\mathbf{z}, \mathbf{w}) \in S(\mathbf{z}^\star, \mathbf{w}^\star, C)$ achieves the $\|.\|_{0,\sim}$ norm (i.e. it is the closest to $(\mathbf{z}^\star, \mathbf{w}^\star)$ in its representative class). Then $(U, V) = (\mathbb{R}_g(\mathbf{z}), \mathbb{R}_m(\mathbf{w}))$ is in $(\mathcal{U}_c \times \mathcal{V}_c) \cap B(D_\alpha, D_\rho, C)$ and satisfy $U\mathbb{1} = \boldsymbol{\alpha}(\mathbf{z}^\star)$ (resp. $V\mathbb{1} = \boldsymbol{\rho}(\mathbf{w}^\star)$). We just need to note $n(1 - \mathrm{Tr}(\mathbb{R}_g(\mathbf{z}))) = \|\mathbf{z} - \mathbf{z}^\star\|_{0,\sim}$ (resp. $d(1 - \mathrm{Tr}(\mathbb{R}_m(\mathbf{w}))) = \|\mathbf{w} - \mathbf{w}^\star\|_{0,\sim}$) to end the proof.

□

The maps $f_{q,\ell} : x \mapsto KL(\pi^\star_{q\ell}, (\psi')^{-1}(x))$ are twice differentiable with a continuous second derivative bounded by $\underline{\sigma}^{-2}$ on $\psi'(C_\pi)$. All terms $\left[U^T \boldsymbol{S}^\star V\right]_{q\ell} \left[U^T \mathbf{1} V\right]^{-1}_{q\ell}$ are convex combinations of the $\psi'(\pi^\star_{q\ell})$ and therefore in $\psi'(C_\pi)$. Furthermore, their first and second order derivative are also bounded as soon as each row sum of U and V is bounded away from 0. By composition, all second order partial derivatives of $\tilde{\Lambda}$ are therefore continuous and bounded on $\mathcal{U} \times \mathcal{V}$.

We now compute the first derivative of $\tilde{\Lambda}$ at $(D_\alpha, D_\rho) := (\mathrm{Diag}(\boldsymbol{\alpha}(\mathbf{z}^\star)), \mathrm{Diag}(\boldsymbol{\rho}(\mathbf{w}^\star)))$ by doing a first-order Taylor expansion of $\tilde{\Lambda}\,(D_\alpha + U, D_\rho + V)$ for small U and V.

Tedious but straightforward manipulations show:

$$\begin{aligned}
\bar{\pi}_{q\ell}(D_\alpha + U, D_\rho + V) =& \pi^\star_{q\ell} + \frac{1}{\alpha_q(\mathbf{z}^\star)}\sum_{q'} U_{qq'}(S_{q'\ell} - 1) \\
& + \frac{1}{\rho_\ell(\mathbf{w}^\star)}\sum_{\ell'} V_{\ell\ell'}(S_{q\ell'} - 1) + o(\|U\|_1, \|V\|_1) \\
\mathrm{KL}\left(\pi^\star_{q\ell}, \bar{\pi}_{q'\ell'}\right) =& \mathrm{KL}\left(\pi^\star_{q\ell}, \pi^\star_{q',\ell'}\right) + \begin{cases} \mathcal{O}(\|U\|_1, \|V\|_1) & \text{if } (q',\ell') \neq (q,\ell) \\ o(\|U\|_1, \|V\|_1) & \text{if } (q',\ell') = (q,\ell) \end{cases}
\end{aligned}$$

where the second line comes from the fact that $f'_{q,\ell}(\psi'(\pi^\star_{q\ell})) = 0$. Keeping only the first order term in U and V in $\tilde{\Lambda}$ and noting that $\tilde{\Lambda}\,(D_\alpha, D_\rho) = 0$ yields:

$$
\begin{aligned}
&\frac{-1}{nd}[\tilde{\Lambda}\left(D_\alpha+U,D_\rho+V\right)-\tilde{\Lambda}\left(D_\alpha,D_\rho\right)]=\frac{-1}{nd}\tilde{\Lambda}\left(D_\alpha+U,D_\rho+V\right)\\
&=\sum_q D_{\alpha,qq}\sum_{\ell,\ell'}V_{\ell\ell'}\operatorname{KL}\left(\pi^\star_{q\ell},\bar{\pi}_{q\ell'}\right)+\sum_\ell D_{\rho,\ell\ell}\sum_{q,q'}U_{qq'}\operatorname{KL}\left(\pi^\star_{q\ell},\bar{\pi}_{q'\ell}\right)+o(\|U\|_1,\|V\|_1)\\
&=\sum_q \alpha_q(\mathbf{z}^\star)\sum_{\ell,\ell'}V_{\ell\ell'}\operatorname{KL}\left(\pi^\star_{q\ell},\pi^\star_{q\ell'}\right)+\sum_\ell \rho_\ell(\mathbf{w}^\star)\sum_{q,q'}U_{qq'}\operatorname{KL}\left(\pi^\star_{q\ell},\pi^\star_{q'\ell}\right)+o(\|U\|_1,\|V\|_1)\\
&\qquad=\operatorname{Tr}(UA(\mathbf{w}^\star))+\operatorname{Tr}(VB(\mathbf{z}^\star))+o(\|U\|_1,\|V\|_1)
\end{aligned}
$$

where $A_{qq'}(\mathbf{w}^\star):=\sum_\ell \rho_\ell(\mathbf{w}^\star)\operatorname{KL}\left(\pi^\star_{q\ell},\pi^\star_{q'\ell}\right)$ and $B_{\ell\ell'}(\mathbf{z}^\star):=\sum_q \rho_\ell(\mathbf{z}^\star)\operatorname{KL}\left(\pi^\star_{q\ell},\pi^\star_{q\ell'}\right)$. A and B are the matrix-derivative of $-\tilde{\Lambda}/nd$ at (D_α,D_ρ). Since $(\mathbf{z}^\star,\mathbf{w}^\star)$ is $c/2$-regular and by definition of $\delta(\boldsymbol{\pi}^\star)$, $A_{qq'}\geq c\delta(\boldsymbol{\pi}^\star)/2$ for $q\neq q'$ and $B_{\ell\ell'}\geq c\delta(\boldsymbol{\pi}^\star)/2$ for $\ell\neq\ell'$ and the diagonal terms of A and B are null. By boundedness of the lower second derivative of $\tilde{\Lambda}$, there exists a constant $C>0$ such that for all $(H,G)\in B(D_\alpha,D_\rho,C)$, we have:

$$
\frac{-1}{nd}\frac{\partial\tilde{\Lambda}}{\partial U_{qq'}}(H,G)\begin{cases}\geq\frac{3c\delta(\boldsymbol{\pi}^\star)}{8} & \text{if } q\neq q'\\ \leq\frac{c\delta(\boldsymbol{\pi}^\star)}{8} & \text{if } q=q'\end{cases}\quad\text{and}\quad\frac{-1}{nd}\frac{\partial\tilde{\Lambda}}{\partial V_{\ell\ell'}}(H,G)\begin{cases}\geq\frac{3c\delta(\boldsymbol{\pi}^\star)}{8} & \text{if } \ell\neq \ell'\\ \leq\frac{c\delta(\boldsymbol{\pi}^\star)}{8} & \text{if } \ell=\ell'\end{cases}
$$

In particular, if U and V have nonnegative non diagonal coefficients and negative diagonal coefficients.

$$
\begin{aligned}
&\frac{-1}{nd}\left[\operatorname{Tr}\left(U\frac{\partial\tilde{\Lambda}}{\partial U}(H,G)\right)+\operatorname{Tr}\left(V\frac{\partial\tilde{\Lambda}}{\partial V}(H,G)\right)\right]\\
&\qquad\qquad\geq\frac{c\delta(\boldsymbol{\pi}^\star)}{4}\left[\sum_{qq'}U_{qq'}+\sum_{\ell\ell'}V_{\ell\ell'}-\operatorname{Tr}(U)-\operatorname{Tr}(V)\right]
\end{aligned}
$$

Choose U and V in $(\mathcal{U}\times\mathcal{V})\cap B(D_\alpha,D_\rho,c_3)$ satisfying $U\mathbb{1}=\boldsymbol{\alpha}(\mathbf{z}^\star)$ and $V\mathbb{1}=\boldsymbol{\rho}(\mathbf{w}^\star)$. Note that $U-D_\alpha$ and $V-D_\rho$ have nonnegative non diagonal coefficients, negative diagonal coefficients, that their coefficients sum up to 1 and that $\operatorname{Tr}(D_\alpha)=\operatorname{Tr}(D_\rho)=1$. By Taylor expansion, there exists a couple (H,G) also in $(\mathcal{U}\times\mathcal{V})\cap B(D_\alpha,D_\rho,C)$ such that

$$
\begin{aligned}
&\frac{-1}{nd}\tilde{\Lambda}\left(U,V\right)=\frac{-1}{nd}\tilde{\Lambda}\left(D_\alpha+(U-D_\alpha),D_\rho+(V-D_\rho)\right)\\
&\qquad=\operatorname{Tr}\left((U-D_\alpha)\frac{\partial\tilde{\Lambda}}{\partial U}(H,G)\right)+\operatorname{Tr}\left((V-D_\rho)\frac{\partial\tilde{\Lambda}}{\partial V}(H,G)\right)\\
&\geq\frac{c\delta(\boldsymbol{\pi}^\star)}{4}[\sum_{q,q'}(U-D_\alpha)_{qq'}+\sum_{\ell,\ell'}(V-D_\rho)_{\ell\ell'}-\operatorname{Tr}(U-D_\alpha)-\operatorname{Tr}(V-D_\rho)]\\
&\qquad\qquad=\frac{c\delta(\boldsymbol{\pi}^\star)}{4}[(1-\operatorname{Tr}(U))+(1-\operatorname{Tr}(V))]
\end{aligned}
$$

To conclude the proof, choose any assignment $(\mathbf{z}, \mathbf{w})$ and without loss of generality assume that $(\mathbf{z}, \mathbf{w})$ are closest to $(\mathbf{z}^\star, \mathbf{w}^\star)$ in their equivalence class. Then $\mathbb{R}_g(\mathbf{z})$ is in $\mathcal{U}$ and additionally satifies $\mathbb{R}_g(\mathbf{z})\mathbb{1} = \boldsymbol{\alpha}(\mathbf{z}^\star)$ and $\|\mathbf{z} - \mathbf{z}^\star\|_{0,\sim} = n\|\mathbb{R}_g(\mathbf{z}) - D_\alpha\|_1/2 = n(1 - \mathrm{Tr}(\mathbb{R}_g(\mathbf{z})))$. Similar equalities hold for $\mathbb{R}_m(\mathbf{w})$ and $\|\mathbf{w} - \mathbf{w}^\star\|_0$.

□

A.4. Proof of Proposition 5.3 (global convergence F_{nd})

Proof.
Conditionally upon $(\mathbf{z}^\star, \mathbf{w}^\star)$,

$$\begin{aligned}
F_{nd}(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}) - \tilde{\Lambda}(\mathbf{z}, \mathbf{w}) &\leq F_{nd}(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}) - G(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}) \\
&= \sum_i \sum_j (\pi_{z_i w_j} - \pi^\star_{z_i^\star w_j^\star}) \left(x_{ij} - \psi'(\pi^\star_{z_i^\star w_j^\star}) \right) \\
&= \sum_{qq'} \sum_{\ell\ell'} \left(\pi_{q'\ell'} - \pi^\star_{q\ell} \right) W_{qq'\ell\ell'} \\
&\leq \sup_{\substack{\Gamma \in \mathbb{R}^{g^2 \times m^2} \\ \|\Gamma\|_\infty \leq \mathrm{Diam}(\boldsymbol{\Theta})}} \sum_{qq'} \sum_{\ell\ell'} \Gamma_{qq'\ell\ell'} W_{qq'\ell\ell'} \coloneqq Z
\end{aligned}$$

uniformly in $\boldsymbol{\theta}$, where the $W_{qq'\ell\ell'}$ are independent and defined by:

$$W_{qq'\ell\ell'} = \sum_i \sum_j z^\star_{iq} w^\star_{j\ell} z_{i,q'} w_{j\ell'} \left(x_{ij} - \psi'(\pi^\star_{q\ell}) \right)$$

is the sum of $nd\mathbb{R}_g(\mathbf{z})_{qq'}\mathbb{R}_m(\mathbf{w})_{\ell\ell'}$ sub-exponential variables with parameters $(\bar{\sigma}^2, 1/\kappa)$ and is therefore itself sub-exponential with parameters $(nd\mathbb{R}_g(\mathbf{z})_{qq'}\mathbb{R}_m(\mathbf{w})_{\ell\ell'}\bar{\sigma}^2, 1/\kappa)$. According to Proposition B.3, $\mathbb{E}_{\boldsymbol{\theta}^\star}[Z|\mathbf{z}^\star, \mathbf{w}^\star] \leq gm\,\mathrm{Diam}(\boldsymbol{\Theta})\sqrt{nd\bar{\sigma}^2}$ and Z is sub-exponential with parameters $(nd\,\mathrm{Diam}(\boldsymbol{\Theta})^2(2\sqrt{2})^2\bar{\sigma}^2, 2\sqrt{2}\,\mathrm{Diam}(\boldsymbol{\Theta})/\kappa)$. In particular, for all $\varepsilon_{n,d} < \kappa/2\sqrt{2}\,\mathrm{Diam}(\boldsymbol{\Theta})$

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\theta}^\star}\left(Z \geq \bar{\sigma} gm\,\mathrm{Diam}(\boldsymbol{\Theta})\sqrt{nd}\left\{ 1 + \frac{\sqrt{8nd}\varepsilon_{n,d}}{gm} \right\} \middle| \mathbf{z}^\star, \mathbf{w}^\star \right) \\
&\qquad \leq \mathbb{P}_{\boldsymbol{\theta}^\star}\left(Z \geq \mathbb{E}_{\boldsymbol{\theta}^\star}[Z|\mathbf{z}^\star, \mathbf{w}^\star] + \bar{\sigma}\,\mathrm{Diam}(\boldsymbol{\Theta}) nd 2\sqrt{2}\varepsilon_{n,d} \middle| \mathbf{z}^\star, \mathbf{w}^\star \right) \\
&\qquad\qquad \leq \exp\left(-\frac{nd\varepsilon_{n,d}^2}{2} \right)
\end{aligned}$$

We can then remove the conditioning and take a union bound to prove Equation (5.2).

□

A.5. Proof of Proposition 5.4 (contribution of far away assignments)

Proof.
Conditionally on $(\mathbf{z}^\star, \mathbf{w}^\star)$, we know from proposition 4.7 that $\tilde{\Lambda}$ is maximal in $(\mathbf{z}^\star, \mathbf{w}^\star)$ and its equivalence class. Choose $0 < t_{nd}$ decreasing to 0 but satisfying $t_{nd} \gg \max\left(\frac{n+d}{nd}, \frac{\log(nd)}{\sqrt{nd}}\right)$. This is always possible because we assume that $\log(d)/n \to 0$ and $\log(n)/d \to 0$. According to 4.7 (iii), for all $(\mathbf{z}, \mathbf{w}) \notin (\mathbf{z}^\star, \mathbf{w}^\star, t_{nd})$

$$\tilde{\Lambda}(\mathbf{z}, \mathbf{w}) \leq -\frac{c\delta(\boldsymbol{\pi}^\star)}{4}(n\|\mathbf{w}-\mathbf{w}^\star\|_{0,\sim} + d\|\mathbf{z}-\mathbf{z}^\star\|_{0,\sim}) \leq -\frac{c\delta(\boldsymbol{\pi}^\star)}{4} ndt_{nd} \quad \text{(A.1)}$$

since either $\|\mathbf{z}-\mathbf{z}^\star\|_{0,\sim} \geq nt_{nd}$ or $\|\mathbf{w}-\mathbf{w}^\star\|_{0,\sim} \geq dt_{nd}$.

Set $\varepsilon_{nd} = \frac{\inf(c\delta(\boldsymbol{\pi}^\star)t_{nd}/16\bar{\sigma}, \kappa)}{\mathrm{Diam}(\boldsymbol{\Theta})}$. By proposition 5.3, and with our choice of ε_{nd}, with probability higher than $1 - \Delta^1_{nd}(\varepsilon_{nd})$,

$$\begin{aligned}
&\sum_{(\mathbf{z},\mathbf{w})\notin S(\mathbf{z}^\star,\mathbf{w}^\star,t_{nd})} p(\mathbf{x}, \mathbf{z}, \mathbf{w}; \boldsymbol{\theta}) \\
&= p(\mathbf{x}|\mathbf{z}^\star, \mathbf{w}^\star, \boldsymbol{\theta}^\star) \sum_{(\mathbf{z},\mathbf{w})\notin S(\mathbf{z}^\star,\mathbf{w}^\star,t_{nd})} p(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta}) e^{F_{nd}(\boldsymbol{\theta},\mathbf{z},\mathbf{w}) - \tilde{\Lambda}(\mathbf{z},\mathbf{w}) + \tilde{\Lambda}(\mathbf{z},\mathbf{w})} \\
&\leq p(\mathbf{x}|\mathbf{z}^\star, \mathbf{w}^\star, \boldsymbol{\theta}^\star) \sum_{\mathbf{z},\mathbf{w}} p(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta}) e^{F_{nd}(\boldsymbol{\theta},\mathbf{z},\mathbf{w}) - \tilde{\Lambda}(\mathbf{z},\mathbf{w}) - ndt_{nd}c\delta(\boldsymbol{\pi}^\star)/4} \\
&\leq p(\mathbf{x}|\mathbf{z}^\star, \mathbf{w}^\star, \boldsymbol{\theta}^\star) \sum_{\mathbf{z},\mathbf{w}} p(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta}) e^{ndt_{nd}c\delta(\boldsymbol{\pi}^\star)/8} \\
&= \frac{p(\mathbf{x}, \mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star)}{p(\mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star)} e^{-ndt_{nd}c\delta(\boldsymbol{\pi}^\star)/8} \\
&\leq p(\mathbf{x}, \mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star) \exp\left(-ndt_{nd}\frac{c\delta(\boldsymbol{\pi}^\star)}{8} + (n+d)\log\frac{1-c}{c}\right) \\
&= p(\mathbf{x}, \mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star) o(1)
\end{aligned}$$

where the second line comes from inequality (A.1), the third from the global control studied in Proposition 5.3 and the definition of ε_{nd}, the fourth from the definition of $p(\mathbf{x}, \mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star)$, the fifth from the bounds on $\boldsymbol{\alpha}^\star$ and $\boldsymbol{\rho}^\star$ and the last from $t_{nd} \gg (n+d)/nd$.

In addition, we have $\varepsilon_{nd} \gg \log(nd)/\sqrt{nd}$ so that the series $\sum_{n,d} \Delta^1_{nd}(\varepsilon_{nd})$ converges and:

$$\sum_{(\mathbf{z},\mathbf{w})\notin S(\mathbf{z}^\star,\mathbf{w}^\star,t_{nd})} p(\mathbf{x}, \mathbf{z}, \mathbf{w}; \boldsymbol{\theta}) = p(\mathbf{x}; \mathbf{z}^\star, \mathbf{w}^\star, \boldsymbol{\theta}^\star) o_P(1)$$

□

A.6. Proof of Proposition 5.5 (local convergence F_{nd})

Proof.

We work conditionally on $(\mathbf{z}^\star, \mathbf{w}^\star) \in \mathcal{Z}_1 \times \mathcal{W}_1$. Choose $\varepsilon \leq \kappa \underline{\sigma}^2$ small. Assignments $(\mathbf{z}, \mathbf{w})$ at $\|.\|_{0,\sim}$-distance less than $c/4$ of $(\mathbf{z}^\star, \mathbf{w}^\star)$ are $c/4$-regular. According to Proposition B.1, $\widehat{x}_{q\ell}$ and $\bar{x}_{q\ell}$ are at distance at most ε with probability higher than $1 - \exp\left(-\frac{ndc^2\varepsilon^2}{128(\bar{\sigma}^2+\kappa^{-1}\varepsilon)}\right)$. Manipulation of Λ and $\tilde{\Lambda}$ yield

$$
\begin{aligned}
\frac{F_{nd}(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}) - \tilde{\Lambda}(\mathbf{z}, \mathbf{w})}{nd} &\leq \frac{\Lambda(\mathbf{z}, \mathbf{w}) - \tilde{\Lambda}(\mathbf{z}, \mathbf{w})}{nd} \\
&= \sum_{q,q''} \sum_{\ell,\ell'} \mathbb{R}_g(\mathbf{z})_{qq'} \mathbb{R}_m(\mathbf{w})_{\ell\ell'} \left[f_{q\ell}(\widehat{x}_{q'\ell'}) - f_{q\ell}(\bar{x}_{q'\ell'}) \right]
\end{aligned}
$$

where $f_{q\ell}(x) = -S^\star_{q\ell}(\psi')^{-1}(x) + \psi \circ (\psi')^{-1}(x)$. The functions $f_{q\ell}$ are twice differentiable on $\mathring{\mathcal{A}}$ with bounded first and second derivatives over $I = \psi'([-M_\pi, M_\pi])$ so that:

$$
f_{q\ell}(y) - f_{q\ell}(x) = f'_{q\ell}(x)\,(y - x) + o\,(y - x)
$$

where the o is uniform over pairs $(x, y) \in I^2$ at distance less than ε and does not depend on $(\mathbf{z}^\star, \mathbf{w}^\star)$. $\bar{x}_{q\ell}$ is a convex combination of the $S^\star_{q\ell} = \psi'(\pi^\star_{q\ell}) \in \psi'(C_\pi)$. Since ψ' is monotonic, $\bar{x}_{q\ell} \in \psi'(C_\pi) \subset I$. Similarly, $|\widehat{x}_{q\ell} - \bar{x}_{q\ell}| \leq \kappa \underline{\sigma}^2$ and $|\psi''| \geq \underline{\sigma}^2$ over I therefore $\widehat{x}_{q\ell} \in I$. We now bound $f'_{q\ell}$:

$$
\begin{aligned}
|f'_{q\ell}(\bar{x}_{q'\ell'})| = \left| \frac{\bar{x}_{q'\ell'} - S^\star_{q\ell}}{\psi'' \circ (\psi')^{-1}(\bar{x}_{q'\ell'})} \right| &= \left| \frac{\frac{\left[\mathbb{R}_g(\mathbf{z})^T \boldsymbol{S}^\star \mathbb{R}_m(\mathbf{w})\right]_{q'\ell'}}{\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})} - S^\star_{q\ell}}{\psi'' \circ (\psi')^{-1}(\bar{x}_{q'\ell'})} \right| \\
&\leq \left(1 - \frac{\mathbb{R}_g(\mathbf{z})_{qq'} \mathbb{R}_m(\mathbf{w})_{\ell\ell'}}{\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})}\right) \frac{S^\star_{\max} - S^\star_{\min}}{\underline{\sigma}^2}
\end{aligned}
$$

where $S^\star_{\max} = \max_{q,\ell} S^\star_{q\ell}$ and $S^\star_{\min} = \min_{q,\ell} S^\star_{q\ell}$. In particular,

$$
\begin{aligned}
\mathbb{R}_g(\mathbf{z})_{qq'} \mathbb{R}_m(\mathbf{w})_{\ell\ell'} |f'_{q\ell}(\bar{x}_{q'\ell'})| &\leq \mathbb{R}_g(\mathbf{z})_{qq'} \mathbb{R}_m(\mathbf{w})_{\ell\ell'} \left(1 - \frac{\mathbb{R}_g(\mathbf{z})_{qq'} \mathbb{R}_m(\mathbf{w})_{\ell\ell'}}{\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})}\right) \frac{S^\star_{\max} - S^\star_{\min}}{\underline{\sigma}^2} \\
&\leq \begin{cases} \mathbb{R}_g(\mathbf{z})_{qq'} \mathbb{R}_m(\mathbf{w})_{\ell\ell'} \frac{S^\star_{\max} - S^\star_{\min}}{\underline{\sigma}^2} & \text{if } (q', \ell') \neq (q, \ell) \\ \left[\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w}) - \mathbb{R}_g(\mathbf{z})_{qq} \mathbb{R}_m(\mathbf{w})_{\ell\ell}\right] \frac{S^\star_{\max} - S^\star_{\min}}{\underline{\sigma}^2} & \text{if } (q, \ell) = (q, \ell) \end{cases}
\end{aligned}
$$

Wrapping everything,

$$\begin{aligned}
\frac{|\Lambda(\mathbf{z},\mathbf{w})-\tilde{\Lambda}(\mathbf{z},\mathbf{w})|}{nd} &= \left|\sum_{q,q'}\sum_{\ell,\ell'} \mathbb{R}_g(\mathbf{z})_{qq'}\mathbb{R}_m(\mathbf{w})_{\ell\ell'}\left[f'_{q\ell}(\bar{x}_{q'\ell'})(\widehat{x}_{q\ell}-\bar{x}_{q\ell})+o(\widehat{x}_{q\ell}-\bar{x}_{q\ell})\right]\right| \\
&\leq \left[\sum_{(q',\ell')\neq(q,\ell)} \mathbb{R}_g(\mathbf{z})_{qq'}\mathbb{R}_m(\mathbf{w})_{\ell\ell'} + \sum_{q,\ell}(\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})-\mathbb{R}_g(\mathbf{z})_{qq}\mathbb{R}_m(\mathbf{w})_{\ell\ell})\right] \\
&\times \frac{S^\star_{\max}-S^\star_{\min}}{\underline{\sigma}^2}\max_{q,\ell}|\widehat{x}_{q\ell}-\bar{x}_{q\ell}|(1+o(1)) \\
&= 2\left[\sum_{(q',\ell')\neq(q,\ell)} \mathbb{R}_g(\mathbf{z})_{qq'}\mathbb{R}_m(\mathbf{w})_{\ell\ell'}\right]\frac{S^\star_{\max}-S^\star_{\min}}{\underline{\sigma}^2}\max_{q,\ell}|\widehat{x}_{q\ell}-\bar{x}_{q\ell}|(1+o(1)) \\
&= 2\left[1-\mathrm{Tr}(\mathbb{R}_g(\mathbf{z}))\,\mathrm{Tr}(\mathbb{R}_m(\mathbf{w}))\right]\frac{S^\star_{\max}-S^\star_{\min}}{\underline{\sigma}^2}\max_{q,\ell}|\widehat{x}_{q\ell}-\bar{x}_{q\ell}|(1+o(1)) \\
&\leq 2\left(\frac{\|\mathbf{z}-\mathbf{z}^\star\|}{n}+\frac{\|\mathbf{w}-\mathbf{w}^\star\|}{d}\right)\frac{S^\star_{\max}-S^\star_{\min}}{\underline{\sigma}^2}\max_{q,\ell}|\widehat{x}_{q\ell}-\bar{x}_{q\ell}|(1+o(1)) \\
&\leq 2\left(\frac{\|\mathbf{z}-\mathbf{z}^\star\|}{n}+\frac{\|\mathbf{w}-\mathbf{w}^\star\|}{d}\right)\frac{S^\star_{\max}-S^\star_{\min}}{\underline{\sigma}^2}\varepsilon(1+o(1))
\end{aligned}$$

We can remove the conditioning on $(\mathbf{z}^\star,\mathbf{w}^\star)$ to prove Equation (5.3) with $c_2 = 2(S^\star_{\max}-S^\star_{\min})/\underline{\sigma}^2$ and $c_1 = c_2^2$.

□

A.7. Proof of Proposition 5.6 (contribution of local assignments)

Proof.
By Proposition 4.2, it is enough to prove that the sum is small compared to $p(\mathbf{z}^\star,\mathbf{w}^\star,\mathbf{x};\boldsymbol{\theta}^\star)$ on Ω_1. We work conditionally on $(\mathbf{z}^\star,\mathbf{w}^\star)\in\mathcal{Z}_1\times\mathcal{W}_1$. Choose $(\mathbf{z},\mathbf{w})$ in $S(\mathbf{z}^\star,\mathbf{w}^\star,C)$ with C defined in proposition 5.4.

$$\log\left(\frac{p(\mathbf{z},\mathbf{w},\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{z}^\star,\mathbf{w}^\star,\mathbf{x};\boldsymbol{\theta}^\star)}\right) = \log\left(\frac{p(\mathbf{z},\mathbf{w};\boldsymbol{\theta})}{p(\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star)}\right)+F_{nd}(\boldsymbol{\theta},\mathbf{z},\mathbf{w})$$

For C small enough, we can assume without loss of generality that $(\mathbf{z},\mathbf{w})$ is the representative closest to $(\mathbf{z}^\star,\mathbf{w}^\star)$ and note $r_1 = \|\mathbf{z}-\mathbf{z}^\star\|_0$ and $r_2 = \|\mathbf{w}-\mathbf{w}^\star\|_0$. We choose $\varepsilon_{nd} \leq \min(\kappa\underline{\sigma}^2, c\delta(\boldsymbol{\pi}^\star)/8)$. Then with probability at least

$1 - \exp\left(-\frac{nd\tilde{c}^2\varepsilon_{nd}^2}{8(c_1\bar{\sigma}^2+c_2\kappa^{-1}\varepsilon_{nd})}\right)$:

$$\begin{aligned}
F_{nd}(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}) &\leq \Lambda(\mathbf{z}, \mathbf{w}) - \tilde{\Lambda}(\mathbf{z}, \mathbf{w}) + \tilde{\Lambda}(\mathbf{z}, \mathbf{w}) \\
&\leq \Lambda(\mathbf{z}, \mathbf{w}) - \tilde{\Lambda}(\mathbf{z}, \mathbf{w}) - \frac{c\delta(\boldsymbol{\pi}^\star)}{4}(dr_1 + nr_2) \\
&\leq \varepsilon_{nd}(dr_1 + nr_2) - \frac{c\delta(\boldsymbol{\pi}^\star)}{4}(dr_1 + nr_2) \\
&\leq -\frac{c\delta(\boldsymbol{\pi}^\star)}{8}(dr_1 + nr_2)
\end{aligned}$$

where the first line comes from the definition of Λ, the second line from Proposition 4.7, the third from Proposition 5.5 and the last from $\varepsilon_{nd} \leq c\delta(\boldsymbol{\pi}^\star)/8$. A union bound shows that

$$\begin{aligned}
\Delta_{nd}(\varepsilon_{nd}) = \mathbb{P}_{\boldsymbol{\theta}^\star}&\left(\sup_{\substack{(\mathbf{z},\mathbf{w})\in(\mathbf{z}^\star,\mathbf{w}^\star,\tilde{c})\\ \boldsymbol{\theta}\in\boldsymbol{\Theta}}} F_{nd}(\boldsymbol{\theta}, \mathbf{z}, \mathbf{w}) \geq -\frac{c\delta(\boldsymbol{\pi}^\star)}{8}(d\|\mathbf{z} - \mathbf{z}^\star\|_{0,\sim} + n\|\mathbf{w} - \mathbf{w}^\star\|_{0,\sim})\right) \\
&\leq g^n m^d \exp\left(-\frac{nd\tilde{c}^2\varepsilon_{nd}^2}{8(c_1\bar{\sigma}^2 + c_2\kappa^{-1}\varepsilon_{nd})}\right)
\end{aligned}$$

Thanks to corollary B.6, we also know that:

$$\log\left(\frac{p(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta})}{p(\mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star)}\right) \leq \mathcal{O}_P(1)\exp\left\{M_{c/4}(r_1 + r_2)\right\}$$

There are at most $\binom{n}{r_1}\binom{n}{r_2}g^{r_1}m^{r_2}$ assignments $(\mathbf{z}, \mathbf{w})$ at distance r_1 and r_2 of $(\mathbf{z}^\star, \mathbf{w}^\star)$ and each of them has at most $g^g m^m$ equivalent configurations. Therefore, with probability $1 - \Delta_{nd}(\varepsilon_{nd})$,

$$\begin{aligned}
&\frac{\sum_{\substack{(\mathbf{z},\mathbf{w})\in S(\mathbf{z}^\star,\mathbf{w}^\star,\tilde{c})\\ (\mathbf{z},\mathbf{w})\nsim(\mathbf{z}^\star,\mathbf{w}^\star)}} p(\mathbf{z}, \mathbf{w}, \mathbf{x}; \boldsymbol{\theta})}{p(\mathbf{z}^\star, \mathbf{w}^\star, \mathbf{x}; \boldsymbol{\theta}^\star)} \\
&\leq \mathcal{O}_P(1)\sum_{r_1+r_2\geq 1}\binom{n}{r_1}\binom{n}{r_2}g^{g+r_1}m^{m+r_2}\exp\left((r_1 + r_2)M_{c/4} - \frac{c\delta(\boldsymbol{\pi}^\star)}{8}(dr_1 + nr_2)\right) \\
&= \mathcal{O}_P(1)\left(1 + e^{(g+1)\log g + M_{c/4} - d\frac{c\delta(\boldsymbol{\pi}^\star)}{8}}\right)^n\left(1 + e^{(m+1)\log m + M_{c/4} - n\frac{c\delta(\boldsymbol{\pi}^\star)}{8}}\right)^d - 1 \\
&\leq \mathcal{O}_P(1)a_{nd}\exp(a_{nd})
\end{aligned}$$

where $a_{nd} = ne^{(g+1)\log g + M_{c/4} - d\frac{c\delta(\boldsymbol{\pi}^\star)}{8}} + de^{(m+1)\log m + M_{c/4} - n\frac{c\delta(\boldsymbol{\pi}^\star)}{8}} = o(1)$ as soon as $n \gg \log d$ and $d \gg \log n$. If we take $\varepsilon_{nd} \gg \log(nd)/\sqrt{nd}$, the series $\sum_{n,d}\Delta_{nd}(\varepsilon_{nd})$ converges which proves the results.

□

A.8. Proof of Proposition 5.7 (contribution of equivalent assignments)

Proof.
Choose (s,t) permutations of $\{1,\dots,g\}$ and $\{1,\dots,m\}$ and assume that $\mathbf{z}=\mathbf{z}^{\star,s}$ and $\mathbf{w}=\mathbf{w}^{\star,t}$. Then $p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta})=p(\mathbf{x},\mathbf{z}^{\star,s},\mathbf{w}^{\star,t};\boldsymbol{\theta})=p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}^{s,t})$. If furthermore $(s,t)\in \mathrm{Sym}(\boldsymbol{\theta})$, $\boldsymbol{\theta}^{s,t}=\boldsymbol{\theta}$ and immediately $p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta})=p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta})$. We can therefore partition the sum as

$$\begin{aligned}
\sum_{(\mathbf{z},\mathbf{w})\sim(\mathbf{z},\mathbf{w})} p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta}) &= \sum_{s,t} p(\mathbf{x},\mathbf{z}^{\star,s},\mathbf{w}^{\star,t};\boldsymbol{\theta}) \\
&= \sum_{s,t} p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}^{s,t}) \\
&= \sum_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} \#\,\mathrm{Sym}(\boldsymbol{\theta}') p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}') \\
&= \#\,\mathrm{Sym}(\boldsymbol{\theta}) \sum_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}')
\end{aligned}$$

$p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta})$ unimodal in $\boldsymbol{\theta}$, with a mode in $\widehat{\boldsymbol{\theta}}_{MC}$. By consistency of $\widehat{\boldsymbol{\theta}}_{MC}$, either $p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta})=o_P(p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}^{\star}))$ or $p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta})=\mathcal{O}_P(p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}^{\star}))$ and $\boldsymbol{\theta}\to\boldsymbol{\theta}^{\star}$. In the latter case, any $\boldsymbol{\theta}'\sim\boldsymbol{\theta}$ other than $\boldsymbol{\theta}$ is bounded away from $\boldsymbol{\theta}^{\star}$ and thus $p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}')=o_P(p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}^{\star}))$. In summary,

$$\sum_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} \frac{p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}')}{p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}^{\star})} = \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} \frac{p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}')}{p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}^{\star})}(1+o_P(1))$$

□

A.9. Proof of Corollary 5.8: Behavior of $\widehat{\theta}_{MLE}$

Theorem 5.1, states that:

$$\frac{p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^{\star})} = \frac{\#\,\mathrm{Sym}(\boldsymbol{\theta})}{\#\,\mathrm{Sym}(\boldsymbol{\theta}^{\star})} \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} \frac{p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}')}{p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}^{\star})}\,(1+o_P(1))+o_P(1)$$

Then,

$$\begin{aligned}
p(\mathbf{x};\boldsymbol{\theta}) &= \#\,\mathrm{Sym}(\boldsymbol{\theta}) \frac{p(\mathbf{x};\boldsymbol{\theta}^{\star})}{\#\,\mathrm{Sym}(\boldsymbol{\theta}^{\star}) p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}^{\star})} \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}')\,(1+o_P(1))+o_P(1) \\
&= \#\,\mathrm{Sym}(\boldsymbol{\theta}) \frac{1}{\#\,\mathrm{Sym}(\boldsymbol{\theta}^{\star}) p(\mathbf{z}^{\star},\mathbf{w}^{\star}|\mathbf{x};\boldsymbol{\theta}^{\star})} \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} p(\mathbf{x},\mathbf{z}^{\star},\mathbf{w}^{\star};\boldsymbol{\theta}')\,(1+o_P(1))+o_P(1).
\end{aligned}$$

Now, using Corollary 3 p. 553 of Mariadassou and Matias [?]

$$p(\cdot,\cdot|\mathbf{x};\boldsymbol{\theta}^\star) \xrightarrow[n,d\to+\infty]{(\mathcal{D})} \frac{1}{\#\operatorname{Sym}(\boldsymbol{\theta}^\star)} \sum_{(\mathbf{z},\mathbf{w})\overset{\boldsymbol{\theta}^\star}{\sim}(\mathbf{z}^\star,\mathbf{w}^\star)} \delta_{(\mathbf{z},\mathbf{w})}(\cdot,\cdot),$$

we can deduce that

$$\begin{aligned} p(\mathbf{x};\boldsymbol{\theta}) &= \#\operatorname{Sym}(\boldsymbol{\theta})\frac{1}{\#\operatorname{Sym}(\boldsymbol{\theta}^\star)p(\mathbf{z}^\star,\mathbf{w}^\star|\mathbf{x};\boldsymbol{\theta}^\star)}\max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}')\,(1+o_P(1))+o_P(1)\\ &= \#\operatorname{Sym}(\boldsymbol{\theta})\frac{1}{1+o_P(1)}\max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}')\,(1+o_P(1))+o_P(1)\\ &= \#\operatorname{Sym}(\boldsymbol{\theta})\max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}')\,(1+o_P(1))+o_P(1). \end{aligned} \tag{A.2}$$

Finaly, we conclude with the proposition 3.3.

A.10. Proof of Corollary ??: Behavior of $J(\mathbb{Q},\boldsymbol{\theta})$

Remark first that for every $\boldsymbol{\theta}$ and for every $(\mathbf{z},\mathbf{w})$,

$$p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta}) \le \exp\left[J(\delta_{\mathbf{z}}\times\delta_{\mathbf{w}},\boldsymbol{\theta})\right] \le \max_{\mathbb{Q}\in\mathcal{Q}}\exp\left[J(\mathbb{Q},\boldsymbol{\theta})\right] \le p(\mathbf{x};\boldsymbol{\theta})$$

where $\delta_{\mathbf{z}}$ denotes the dirac mass on $\mathbf{z}$. By dividing by $p(\mathbf{x};\boldsymbol{\theta}^\star)$, we obtain

$$\frac{p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^\star)} \le \frac{\max_{\mathbb{Q}\in\mathcal{Q}}\exp\left[J(\mathbb{Q},\boldsymbol{\theta})\right]}{p(\mathbf{x};\boldsymbol{\theta}^\star)} \le \frac{p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^\star)}.$$

As this inequality is true for every couple $(\mathbf{z},\mathbf{w})$, we have:

$$\max_{(\mathbf{z},\mathbf{w})\in\mathcal{Z}\times\mathcal{W}}\frac{p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^\star)} \le \frac{\max_{\mathbb{Q}\in\mathcal{Q}}\exp\left[J(\mathbb{Q},\boldsymbol{\theta})\right]}{p(\mathbf{x};\boldsymbol{\theta}^\star)}.$$

Moreover, using Equation A.2, we get a lower bound:

$$\begin{aligned} \max_{(\mathbf{z},\mathbf{w})\in\mathcal{Z}\times\mathcal{W}}\frac{p(\mathbf{x},\mathbf{z},\mathbf{w};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^\star)} &= \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}}\frac{p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}'\right)(1+o_p(1))}{p(\mathbf{x};\boldsymbol{\theta}^\star)}+o_p(1)\\ &= \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}}\frac{p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}'\right)(1+o_p(1))}{\#\operatorname{Sym}(\boldsymbol{\theta}^\star)p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star)(1+o_p(1))}+o_p(1)\\ &= \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}}\frac{p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}'\right)(1+o_p(1))}{\#\operatorname{Sym}(\boldsymbol{\theta}^\star)p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star)}+o_p(1). \end{aligned}$$

Now, Theorem 5.1 leads to the following upper bound:

$$\begin{aligned} \frac{\max_{\mathbb{Q}\in\mathcal{Q}}\exp\left[J(\mathbb{Q},\boldsymbol{\theta})\right]}{p(\mathbf{x};\boldsymbol{\theta}^\star)} &\le \frac{p(\mathbf{x};\boldsymbol{\theta})}{p(\mathbf{x};\boldsymbol{\theta}^\star)}\\ &\le \frac{\#\operatorname{Sym}(\boldsymbol{\theta})}{\#\operatorname{Sym}(\boldsymbol{\theta}^\star)}\max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}}\frac{p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}'\right)(1+o_p(1))}{p(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star)}+o_p(1) \end{aligned}$$

so that we have the following control

$$\max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} \frac{p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}'\right)(1+o_p(1))}{\#\,\mathrm{Sym}(\boldsymbol{\theta}^\star)p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star\right)} + o_p(1) \leq \frac{\max_{\mathbb{Q}\in\mathcal{Q}} \exp\left[J\left(\mathbb{Q},\boldsymbol{\theta}\right)\right]}{p\left(\mathbf{x};\boldsymbol{\theta}^\star\right)}$$
$$\leq \frac{\#\,\mathrm{Sym}(\boldsymbol{\theta})}{\#\,\mathrm{Sym}(\boldsymbol{\theta}^\star)} \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} \frac{p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}'\right)(1+o_p(1))}{p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star\right)} + o_p(1).$$

In the particular case where $\#\,\mathrm{Sym}(\boldsymbol{\theta}) = 1$, we have

$$\frac{\max_{\mathbb{Q}\in\mathcal{Q}} \exp\left[J\left(\mathbb{Q},\boldsymbol{\theta}\right)\right]}{p\left(\mathbf{x};\boldsymbol{\theta}^\star\right)} = \frac{1}{\#\,\mathrm{Sym}(\boldsymbol{\theta}^\star)} \max_{\boldsymbol{\theta}'\sim\boldsymbol{\theta}} \frac{p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}'\right)(1+o_p(1))}{p\left(\mathbf{x},\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star\right)} + o_p(1)$$

and, following the same reasoning as the appendix A.9, we have the result.

Appendix B: Technical Lemma

B.1. Sub-exponential variables

We now prove two propositions regarding subexponential variables. Recall first that a random variable X is sub-exponential with parameters (τ^2, b) if for all λ such that $|\lambda| \leq 1/b$,

$$\mathbb{E}[e^{\lambda(X-\mathbb{E}(X))}] \leq \exp\left(\frac{\lambda^2\tau^2}{2}\right).$$

In particular, all distributions coming from a natural exponential family are sub-exponential. Sub-exponential variables satisfy a large deviation Bernstein-type inequality:

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \begin{cases} \exp\left(-\frac{t^2}{2\tau^2}\right) & \text{if} \quad 0 \leq t \leq \frac{\tau^2}{b} \\ \exp\left(-\frac{t}{2b}\right) & \text{if} \quad t \geq \frac{\tau^2}{b} \end{cases} \tag{B.1}$$

So that

$$\mathbb{P}(X - \mathbb{E}[X] \geq t) \leq \exp\left(-\frac{t^2}{2(\tau^2 + bt)}\right)$$

The subexponential property is preserved by summation and multiplication.

- If X is sub-exponential with parameters (τ^2, b) and $\alpha \in \mathbb{R}$, then so is αX with parameters $(\alpha^2\tau^2, \alpha b)$
- If the X_i, $i = 1, \ldots, n$ are sub-exponential with parameters (τ_i^2, b_i) and independent, then so is $X = X_1 + \cdots + X_n$ with parameters $(\sum_i \tau_i^2, \max_i b_i)$

Proposition B.1 (Maximum in $(\mathbf{z},\mathbf{w})$). *Let $(\mathbf{z},\mathbf{w})$ be a configuration and $\hat{x}_{q,\ell}(\mathbf{z},\mathbf{w})$ resp. $\bar{x}_{q\ell}(\mathbf{z},\mathbf{w})$ be as defined in Equations (3.1) and (4.4). Under* *the assumptions of the section 2.4, for all* $\varepsilon > 0$

$$\mathbb{P}\left(\max_{\mathbf{z},\mathbf{w}} \max_{k,l} \widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})|\hat{x}_{q,\ell} - \bar{x}_{q\ell}| > \varepsilon\right) \leq g^{n+1}m^{d+1}\exp\left(-\frac{nd\varepsilon^2}{2(\bar{\sigma}^2 + \kappa^{-1}\varepsilon)}\right). \tag{B.2}$$

Additionally, the suprema over all $c/2$*-regular assignments satisfies:*

$$\mathbb{P}\left(\max_{\mathbf{z}\in\mathcal{Z}_1,\mathbf{w}\in\mathcal{W}_1} \max_{k,l} |\hat{x}_{q,\ell} - \bar{x}_{q\ell}| > \varepsilon\right) \leq g^{n+1}m^{d+1}\exp\left(-\frac{ndc^2\varepsilon^2}{8(\bar{\sigma}^2 + \kappa^{-1}\varepsilon)}\right). \tag{B.3}$$

Note that equations B.2 and B.3 remain valid when replacing $c/2$ by any $\tilde{c} < c/2$.

Proof.

The random variables X_{ij} are subexponential with parameters $(\bar{\sigma}^2, 1/\kappa)$. Conditionally to $(\mathbf{z}^\star, \mathbf{w}^\star)$, $z_{+q}w_{+\ell}(\hat{x}_{q,\ell} - \bar{x}_{q\ell})$ is a sum of $z_{+q}w_{+\ell}$ centered subexponential random variables. By Bernstein's inequality [**?**], we therefore have for all $t > 0$

$$\mathbb{P}(z_{+q}w_{+\ell}|\hat{x}_{q,\ell} - \bar{x}_{q\ell}| \geq t) \leq 2\exp\left(-\frac{t^2}{2(z_{+q}w_{+\ell}\bar{\sigma}^2 + \kappa^{-1}t)}\right)$$

In particular, if $t = ndx$,

$$\mathbb{P}\left(\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})|\hat{x}_{q,\ell} - \bar{x}_{q\ell}| \geq x\right) \leq 2\exp\left(-\frac{ndx^2}{2(\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})\bar{\sigma}^2 + \kappa^{-1}x)}\right) \leq 2\exp\left(-\frac{ndx^2}{2(\bar{\sigma}^2 + \kappa^{-1}x)}\right)$$

uniformly over $(\mathbf{z}, \mathbf{w})$. Equation (B.2) then results from a union bound. Similarly,

$$\begin{aligned}\mathbb{P}\left(|\hat{x}_{q,\ell} - \bar{x}_{q\ell}| \geq x\right) &= \mathbb{P}\left(\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})|\hat{x}_{q,\ell} - \bar{x}_{q\ell}| \geq \widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})x\right) \\ &\leq 2\exp\left(-\frac{ndx^2\widehat{\alpha}_q(\mathbf{z})^2\widehat{\rho}_\ell(\mathbf{w})^2}{2(\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w})\bar{\sigma}^2 + \kappa^{-1}x\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w}))}\right) \\ &\leq 2\exp\left(-\frac{ndc^2x^2}{8(\bar{\sigma}^2 + \kappa^{-1}x)}\right)\end{aligned}$$

Where the last inequality comes from the fact that $c/2$-regular assignments satisfy $\widehat{\alpha}_q(\mathbf{z})\widehat{\rho}_\ell(\mathbf{w}) \geq c^2/4$. Equation (B.3) then results from a union bound over $\mathcal{Z}_1 \times \mathcal{W}_1 \subset \mathcal{Z} \times \mathcal{W}$.

□

Lemma B.2. *If* X *is a zero mean random variable, subexponential with parameters* (σ^2, b)*, then* $|X|$ *is subexponential with parameters* $(8\sigma^2, 2\sqrt{2}b)$*.*

Proof.
Note $\mu = \mathbb{E}|X|$ and consider $Y = |X| - \mu$. Choose λ such that $|\lambda| < (2\sqrt{2}b)^{-1}$. We need to bound $\mathbb{E}[e^{\lambda Y}]$. Note first that $\mathbb{E}[e^{\lambda Y}] \leq \mathbb{E}[e^{\lambda X}] + \mathbb{E}[e^{-\lambda X}] < +\infty$ is properly defined by subexponential property of X and we have

$$\mathbb{E}[e^{\lambda Y}] \leq 1 + \sum_{k=2} \frac{|\lambda|^k \mathbb{E}[|Y|^k]}{k!}$$

where we used the fact that $\mathbb{E}[Y] = 0$. We know bound odd moments of $|\lambda Y|$.

$$\mathbb{E}[|\lambda Y|^{2k+1}] \leq (\mathbb{E}[|\lambda Y|^{2k}]\mathbb{E}[|\lambda Y|^{2k+2}])^{1/2} \leq \frac{1}{2}(\lambda^{2k}\mathbb{E}[Y^{2k}] + \lambda^{2k+2}\mathbb{E}[Y^{2k+2}])$$

where we used first Cauchy-Schwarz and then the arithmetic-geometric mean inequality. The Taylor series expansion can thus be reduced to

$$\begin{aligned}
\mathbb{E}[e^{\lambda Y}] &\leq 1 + \left(\frac{1}{2} + \frac{1}{2.3!}\right)\mathbb{E}[Y^2]\lambda^2 + \sum_{k=2}^{+\infty}\left(\frac{1}{(2k)!} + \frac{1}{2}\left[\frac{1}{(2k-1)!} + \frac{1}{(2k+1)!}\right]\right)\lambda^{2k}\mathbb{E}[Y^{2k}] \\
&\leq \sum_{k=0}^{+\infty} 2^k \frac{\lambda^{2k}\mathbb{E}[Y^{2k}]}{(2k)!} \\
&\leq \sum_{k=0}^{+\infty} 2^{3k} \frac{\lambda^{2k}\mathbb{E}[X^{2k}]}{(2k)!} = \cosh\left(2\sqrt{2}\lambda X\right) = \mathbb{E}\left[\frac{e^{2\sqrt{2}\lambda X} + e^{-2\sqrt{2}\lambda X}}{2}\right] \\
&\leq e^{\frac{8\lambda^2\sigma^2}{2}}
\end{aligned}$$

where we used the well-known inequality $\mathbb{E}[|X - \mathbb{E}[X]|^k] \leq 2^k \mathbb{E}[|X|^k]$ to substitute $2^{2k}\mathbb{E}[X^{2k}]$ to $\mathbb{E}[Y^{2k}]$.

□

Proposition B.3 (concentration for subexponential)**.** *Let $X_1, \ldots, X_n$ be independent zero mean random variables, subexponential with parameters (σ_i^2, b_i). Note $V_0^2 = \sum_i \sigma_i^2$ and $b = \max_i b_i$. Then the random variable Z defined by:*

$$Z = \sup_{\substack{\Gamma \in \mathbb{R}^n \\ \|\Gamma\|_\infty \leq M}} \sum_i \Gamma_i X_i$$

is also subexponential with parameters $(8M^2V_0^2, 2\sqrt{2}Mb)$. Moreover $\mathbb{E}[Z] \leq MV_0\sqrt{n}$ so that for all $t > 0$,

$$\mathbb{P}(Z - MV_0\sqrt{n} \geq t) \leq \exp\left(-\frac{t^2}{2(8M^2V_0^2 + 2\sqrt{2}Mbt)}\right) \tag{B.4}$$

Proof.
Note first that Z can be simplified to $Z = M \sum_i |X_i|$. We just need to bound bound $\mathbb{E}[Z]$. The rest of the proposition results from the fact that the $|X_i|$ are subexponential $(8\sigma_i^2, 2\sqrt{2}b_i)$ by Lemma B.2 and standard properties of sums of independent rescaled subexponential variables.

$$\begin{aligned}\mathbb{E}[Z] = \mathbb{E}\left[\sup_{\substack{\Gamma\in\mathbb{R}^n\\ \|\Gamma\|_\infty\le M}} \sum_i \Gamma_i X_i\right] &= \mathbb{E}\left[\sum_i M|X_i|\right] \le M\sum_i \sqrt{\mathbb{E}[X_i^2]} \\ = M\sum_i \sigma_i \le M\left(\sum_i 1\right)^{1/2}\left(\sum_i \sigma_i^2\right)^{1/2} &= MV_0\sqrt{n}\end{aligned}$$

using Cauchy-Schwarz.

□

The final lemma is the working horse for proving Proposition 4.7.

Lemma B.4.
Let η and $\bar\eta$ be two matrices from $M_{g\times m}(\Theta)$ and $f : \Theta\times\Theta \to \mathbb{R}_+$ a positive function, A a (squared) confusion matrix of size g and B a (squared) confusion matrix of size m. We denote $D_{q\ell q'\ell'} = f(\eta_{q\ell}, \bar\eta_{q'\ell'})$. Assume that

- *all the rows of η are distinct;*
- *all the columns η are distinct;*
- *$f(x, y) = 0 \Leftrightarrow x = y$;*
- *each row of A has a non zero element;*
- *each row of B has a non zero element;*

and note

$$\Sigma = \sum_{qq'}\sum_{\ell\ell'} A_{qq'}B_{\ell\ell'}d_{q\ell q'\ell'} \tag{B.5}$$

Then,

$$\Sigma = 0 \Leftrightarrow \begin{cases} A, B \text{ are permutation matrices } s, t \\ \bar\eta = \eta^{s,t} \text{ cad } \forall(q,\ell), \bar\eta_{q\ell} = \eta_{s(q)t(\ell)}\end{cases}$$

Proof.
If A and B are the permutation matrices corresponding to the permutations s et t: $A_{ij} = 0$ if $i \neq s(j)$ and $B_{ij} = 0$ if $i \neq t(j)$. As each row of A contains a non zero element and as $A_{s(q)q} > 0$ (resp. $B_{s(\ell)\ell} > 0$) for all q (resp. ℓ), the following sum Σ reduces to

$$\Sigma = \sum_{qq'}\sum_{\ell\ell'} A_{qq'}B_{\ell\ell'}d_{q\ell q'\ell'} = \sum_q\sum_\ell A_{s(q)q}B_{t(\ell)\ell}d_{s(q)t(\ell)q\ell}$$

Σ is null and sum of positive components, each component is null. However, all $A_{s(q)q}$ and $B_{t(\ell)\ell}$ are not null, so that for all (q,ℓ), $d_{s(q)t(\ell)q\ell} = 0$ and $\bar\eta_{q\ell} =$

$\eta_{s(q)t(\ell)}$.
Now, if A is not a permutation matrix while $\Sigma = 0$ (the same reasoning holds for B or both). Then A owns a column k that contains two non zero elements, say $A_{q_1 q}$ and $A_{q_2 q}$. Let $\ell \in \{1 \dots m\}$, there exists by assumption ℓ' such that $B_{\ell\ell'} \neq 0$. As $\Sigma = 0$, both products $A_{q_1 q} B_{\ell\ell'} d_{q_1 \ell q \ell'}$ and $A_{q_2 q} B_{\ell\ell'} d_{q_2 \ell q \ell'}$ are zero.

$$\begin{cases} A_{q_1 q} B_{\ell\ell'} d_{q_1 \ell q \ell'} &= 0 \\ A_{q_2 q} B_{\ell\ell'} d_{q_2 \ell q \ell'} &= 0 \end{cases} \Leftrightarrow \begin{cases} d_{q_1 \ell q \ell'} &= 0 \\ d_{q_2 \ell q \ell'} &= 0 \end{cases} \Leftrightarrow \begin{cases} \eta_{q_1 \ell} = \bar{\eta}_{q \ell'} \\ \eta_{q_2 \ell} = \bar{\eta}_{q \ell'} \end{cases} \Leftrightarrow \eta_{q_1 \ell} = \eta_{q_2 \ell}$$

The previous equality is true for all ℓ, thus rows q_1 and q_2 of η are identical, and contradict the assumptions.

□

B.2. Likelihood ratio of assignments

Lemma B.5.
Let $\mathcal{Z}_1$ be the subset of $\mathcal{Z}$ of c-regular configurations, as defined in Definition 4.1. Let $\mathbb{S}^g = \{\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_g) \in [0,1]^g : \sum_{k=1}^g \alpha_k = 1\}$ be the g-dimensional simplex and note $\mathbb{S}^g_c = \mathbb{S}^g \cap [c, 1-c]^g$. Then there exists two positive constants M_c and M'_c such that for all $\mathbf{z}$, $\mathbf{z}^\star$ in $\mathcal{Z}_1$ and all $\boldsymbol{\alpha} \in \mathbb{S}^g_c$

$$|\log p(\mathbf{z}; \hat{\boldsymbol{\alpha}}(\mathbf{z})) - \log p(\mathbf{z}^\star; \hat{\boldsymbol{\alpha}}(\mathbf{z}^\star))| \quad \leq \quad M_c \|\mathbf{z} - \mathbf{z}^\star\|_0$$

Proof.
Consider the entropy map $H : \mathbb{S}^g \to \mathbb{R}$ defined as $H(\boldsymbol{\alpha}) = -\sum_{k=1}^g \alpha_k \log(\alpha_k)$. The gradient ∇H is uniformly bounded by $\frac{M_c}{2} = \log \frac{1-c}{c}$ in $\|.\|_\infty$-norm over $\mathbb{S}^g \cap [c, 1-c]^g$. Therefore, for all $\boldsymbol{\alpha}$, $\boldsymbol{\alpha}^\star \in \mathbb{S}^g \cap [c, 1-c]^g$, we have

$$|H(\boldsymbol{\alpha}) - H(\boldsymbol{\alpha}^\star)| \leq \frac{M_c}{2} \|\boldsymbol{\alpha} - \boldsymbol{\alpha}^\star\|_1$$

To prove the inequality, we remark that $\mathbf{z} \in \mathcal{Z}_1$ translates to $\hat{\boldsymbol{\alpha}}(\mathbf{z}) \in \mathbb{S}^g \cap [c, 1-c]^g$, that $\log p(\mathbf{z}; \hat{\boldsymbol{\alpha}}(\mathbf{z})) - \log p(\mathbf{z}^\star; \hat{\boldsymbol{\alpha}}(\mathbf{z}^\star)) = n[H(\hat{\boldsymbol{\alpha}}(\mathbf{z})) - H(\hat{\boldsymbol{\alpha}}(\mathbf{z}^\star))]$ and finally that $\|\hat{\boldsymbol{\alpha}}(\mathbf{z}) - \hat{\boldsymbol{\alpha}}(\mathbf{z}^\star)\|_1 \leq \frac{2}{n} \|\mathbf{z} - \mathbf{z}^\star\|_0$.

□

Corollary B.6. *Let $\mathbf{z}^\star$ (resp. $\mathbf{w}^\star$) be $c/2$-regular and $\mathbf{z}$ (resp. $\mathbf{w}$) at $\|.\|_0$-distance $c/4$ of $\mathbf{z}^\star$ (resp. $\mathbf{w}^\star$). Then, for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$*

$$\log \frac{p(\mathbf{z}, \mathbf{w}; \boldsymbol{\theta})}{p(\mathbf{z}^\star, \mathbf{w}^\star; \boldsymbol{\theta}^\star)} \leq \mathcal{O}_P(1) \exp\left\{M_{c/4}(\|\mathbf{z} - \mathbf{z}^\star\|_0 + \|\mathbf{w} - \mathbf{w}^\star\|_0)\right\}$$

Proof.
Note then that:

$$
\begin{aligned}
\frac{p(\mathbf{z},\mathbf{w};\boldsymbol{\theta})}{p(\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\theta}^\star)} &= \frac{p(\mathbf{z},\mathbf{w};\boldsymbol{\alpha},\boldsymbol{\rho})}{p(\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\alpha}^\star,\boldsymbol{\rho}^\star)} = \frac{p(\mathbf{z},\mathbf{w};\boldsymbol{\alpha},\boldsymbol{\rho})}{p(\mathbf{z}^\star,\mathbf{w}^\star;\hat{\boldsymbol{\alpha}}(\mathbf{z}^\star),\hat{\boldsymbol{\rho}}(\mathbf{w}^\star))} \frac{p(\mathbf{z}^\star,\mathbf{w}^\star;\hat{\boldsymbol{\alpha}}(\mathbf{z}^\star),\hat{\boldsymbol{\rho}}(\mathbf{w}^\star))}{p(\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\alpha}^\star,\boldsymbol{\rho}^\star)} \\
&\leq \frac{p(\mathbf{z},\mathbf{w};\hat{\boldsymbol{\alpha}}(\mathbf{z}),\hat{\boldsymbol{\rho}}(\mathbf{w}))}{p(\mathbf{z}^\star,\mathbf{w}^\star;\hat{\boldsymbol{\alpha}}(\mathbf{z}^\star),\hat{\boldsymbol{\rho}}(\mathbf{w}^\star))} \frac{p(\mathbf{z}^\star,\mathbf{w}^\star;\hat{\boldsymbol{\alpha}}(\mathbf{z}^\star),\hat{\boldsymbol{\rho}}(\mathbf{w}^\star))}{p(\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\alpha}^\star,\boldsymbol{\rho}^\star)} \\
&\leq \exp\left\{M_{c/4}(\|\mathbf{z}-\mathbf{z}^\star\|_0 + \|\mathbf{w}-\mathbf{w}^\star\|_0)\right\} \times \frac{p(\mathbf{z}^\star,\mathbf{w}^\star;\hat{\boldsymbol{\alpha}}(\mathbf{z}^\star),\hat{\boldsymbol{\rho}}(\mathbf{w}^\star))}{p(\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\alpha}^\star,\boldsymbol{\rho}^\star)} \\
&\leq \mathcal{O}_P(1)\exp\left\{M_{c/4}(\|\mathbf{z}-\mathbf{z}^\star\|_0 + \|\mathbf{w}-\mathbf{w}^\star\|_0)\right\}
\end{aligned}
$$

where the first inequality comes from the definition of $\hat{\boldsymbol{\alpha}}(\mathbf{z})$ and $\hat{\boldsymbol{\rho}}(\mathbf{w})$ and the second from Lemma B.5 and the fact that $\mathbf{z}^\star$ and $\mathbf{z}$ (resp. $\mathbf{w}^\star$ and $\mathbf{w}$) are $c/4$-regular. Finally, local asymptotic normality of the MLE for multinomial proportions ensures that $\frac{p(\mathbf{z}^\star,\mathbf{w}^\star;\hat{\boldsymbol{\alpha}}(\mathbf{z}^\star),\hat{\boldsymbol{\rho}}(\mathbf{w}^\star))}{p(\mathbf{z}^\star,\mathbf{w}^\star;\boldsymbol{\alpha}^\star,\boldsymbol{\rho}^\star)} = \mathcal{O}_P(1)$.

□

References