Estimation in dyadic models based on estimation in i.i.d. sub-samples

Yassine Sbai Sassi

02/03/2022

Abstract

This paper proposes a general procedure to construct estimators for exchangeable network models. For any network model, consider an auxiliary *i.i.d.* model where each observation has the same distribution as any observation in the original model. The procedure returns estimators for the original model whenever valid estimators are known in the auxiliary *i.i.d.* model.

This paper then studies the asymptotic behavior of the *"the average MLE"*, the estimators returned by the procedure for parametric binomial network models. I show that the *average MLE* behaves asymptotically like the composite maximum likelihood estimator. Interestingly, the *average MLE* does not require the entire network to be observed. For instance, I show that for a balanced bipartite graph, observing almost any sub-graph with more than $N^{\frac{3}{2}+\epsilon}$ edges for some $\epsilon > 0$ (out of the total N^2 edges) is enough for the asymptotic result to hold. These results are readily extendable beyond the binomial model.

1 Introduction

Consider a the following general model:

$$Y_{ij} := h(X_i, X_j, U_i, U_j, V_{ij}; \beta) \tag{1}$$

for all $i, j \leq N$ and for some measurable and known measurable function h, *i.i.d.* variables X_i, U_i and V_{ij}, which are also mutually independent, and for some parameter β.[1] We are interested in estimating β using the observations $(Y_ij)_{i,j\leq N}$ and $(X_i)_{i\leq N}$.

[1]This model is in fact very general: any X-exchangeable random array would have a representation of the form (1). See Crane and Towsner (2018) for details.

Given that, in general, we know much more about *i.i.d.* models than about models with dyadic dependence such as the one in the equation (1), it would interesting to extract an *i.i.d.* sub-sample from a full sample (Y_{ij}) and (X_i). Observe that the set of edges $\{Y_{1,2}, Y_{3,4}, ..., Y_{N-1,N}\}$ (assuming N is even) are *i.i.d.*. Denoting $Y_{(i)} := Y_{2i-1,2i}$ and $X_{(i)} := (X_{2i-1}, X_{2i})$ for all $i = 1..N/2$, the observations $(Y_{(i)}, X_{(i)})_{i \leq N/2}$ become *i.i.d.* and follow:

$$Y_{(i)} = h(X_{(i)}, \epsilon_{(i)}, \beta) \tag{2}$$

with $\epsilon_{(i)}$ for all $i = 1..N/2$. Assuming the parameter β is identified under the model (2), it is also identified under (1). Moreover, any estimator for β with certain desirable properties in (2) would have those same properties under (1). In fact, there are many ways to extract *i.i.d.* sub-samples like the one used in (2): for any permutation $\sigma \in \mathbb{S}_N$, the observations $\{Y_{\sigma(2i-1),\sigma(2i)}, i = 1..N/2\}$ are *i.i.d.*.

This approach is too naive, it disregards most of the data. A more sensible estimator would be one that averages all - or a large number of - the estimators obtained through the *i.i.d.* sub-samples. This paper studies these *averaged* estimator for parametric binomial models (e.g. logit models). I show that if the set of permutations used to extract the *i.i.d.* samples is "diverse" enough, that is, if the sub-samples do not intersect too much (in a sense that I precise in the proposition 4), then the "average MLE" has the same asymptotic distribution as the composit maximum likelihood estimator (c.f. section 4.2. in Graham (2020) for details on the composite maximum likelihood). In the next section, I formally describe the procedure, the diversity and condition and show the asymptotic distribution of the "averaged" estimators. The third section discusses an interesting application: the procedure can also be useful when the network is not observable in its entirety. The last section concludes. All the proofs are relegated to the end of the paper.

2 The model and the main results

Consider the model:

$$Y_{ij} = \mathbb{1}(X_{ij}\beta_0 + U_i + U_j + V_{ij} \geq 0) \tag{3}$$

where: $X_{ij} = g(X_i, X_j)$ with (X_i) are i.i.d. random variable, U_i and V_{ij} are i.i.d random variables with mean 0 such that $\epsilon_{ij} = U_i + U_j + V_{ij}$ is distributed following CDF Φ and PDF ϕ. β_0 is the parameter of interest, β_0 is known to be in a set $K \subset \mathbb{R}^k$.

Assume we have an even number of observations $i, j = 1..N$, I am interested the following estimator: first, for every permutation $\sigma \in \mathbb{S}_N$ consider the i.i.d. observations

$\left(Y_{\sigma(2i-1),\sigma(2i)}, X_{\sigma(2i-1),\sigma(2i)}\right)_{i=1}^{\frac{N}{2}}$, to simplify, denote: $Y_{\sigma,i} := Y_{\sigma(2i-1),\sigma(2i)}$ and similarly for X. Define $\hat{\beta}_\sigma$ the maximum likelihood estimator of β_0 computed using the i.i.d. sample $(Y_\sigma, X_\sigma) = (Y_{\sigma,i}, X_{\sigma,i})$:

$$\hat{\beta}_\sigma := arg\max_{\beta} \sum_{i=1}^{\frac{N}{2}} Y_{\sigma,i} \log(\Phi(X_{\sigma,i}\beta)) + (1 - Y_{\sigma,i}) \log(1 - \Phi(X_{\sigma,i}\beta))$$

For every σ, denote: $\mathcal{L}_\sigma(X, Y; \beta) := \frac{2}{N} \sum_{i=1}^{\frac{N}{2}} Y_{\sigma,i} \log(\Phi(X_{\sigma,i}\beta)) + (1 - Y_{\sigma,i}) \log(1 - \Phi(X_{\sigma,i}\beta))$. Fix some set $S \subset \mathbb{S}_N$, define:

$$\hat{\beta}_S := \frac{1}{|S|} \sum_{\sigma \in S} \hat{\beta}_\sigma$$

the objective is to determine the asymptotic distribution of $\hat{\beta}_S$.

Note that for any $\sigma \in \mathbb{S}_N$, whenever $\hat{\beta}_\sigma$ is an interior point of the parameter space K:

$$0 = \frac{\partial \mathcal{L}_\sigma(X, Y; \hat{\beta}_\sigma)}{\partial \beta} = \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} + \frac{\partial^2 \mathcal{L}_\sigma(X, Y; \bar{\beta}_\sigma)}{\partial \beta^2} (\hat{\beta}_\sigma - \beta_0)$$

for some $\bar{\beta}_\sigma \in [\beta_0, \hat{\beta}_\sigma]$.[2] Therefore:

$$\begin{aligned}
\hat{\beta}_S - \beta_0 &= -\frac{1}{|S|} \sum_{\sigma \in S} \left(\frac{\partial^2 \mathcal{L}_\sigma(X, Y; \bar{\beta}_\sigma)}{\partial \beta^2} \right)^{-1} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} \\
&= -\Sigma(\beta_0)^{-1} \frac{1}{|S|} \sum_{\sigma \in S} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} \\
&\quad + \frac{1}{|S|} \sum_{\sigma \in S} \left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2 \mathcal{L}_\sigma(X, Y; \bar{\beta}_\sigma)}{\partial \beta^2} \right)^{-1} \right] \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta}
\end{aligned} \tag{4}$$

where $\Sigma(\beta_0) := E\left(\frac{\partial^2 \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta^2} \right)$.

Before discussing the asymptotic behavior of $\hat{\beta}_S$, a few technical comments are in order. First, these Taylor expansions are only valid if *all* the $\hat{\beta}_\sigma$'s are interior points. How can we be sure they are? Second, the equation (4) requires that $\frac{\partial^2 \mathcal{L}_\sigma(X, Y; \bar{\beta}_\sigma)}{\partial \beta^2}$ is invertible for any $\bar{\beta}_\sigma$ and for any σ.

The two following propositions and their corollary address these two concerns. I show that the $\hat{\beta}_\sigma$'s are not only all interior points with high probability (when the true parameter is itself an interior point), but that they are uniformly consistent as long as S does not grow too

[2] Throughout, as I will state in each proposition, I assume the parameter space to be convex.

fast in N. Moreover, I show that $\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}$ converge to their common expectation uniformly both in σ and in β.

Further, these uniform convergence results will allow me to neglect the second term of the equation (4): $\frac{1}{|S|}\sum_{\sigma \in S}\left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial \beta^2}\right)^{-1}\right]\frac{\partial \mathcal{L}_\sigma(X,Y;\beta_0)}{\partial \beta}$, relative to its first term. That is, the asymptotic distribution of $\hat{\beta}_S$ will be that of the first term of the equation (4): $-\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma \in S}\frac{\partial \mathcal{L}_\sigma(X,Y;\beta_0)}{\partial \beta}$.

Proposition 1. *Assume*

- *K is compact and convex;*
- *X has a compact support and*
- *the smallest eigenvalue of $\Sigma(\beta)$ is bounded away from 0 uniformly over $\beta \in K$, that is:*

$$\inf_{\beta \in K} \lambda_{\min}(\Sigma(\beta)) > 0$$

then, for any $\sigma \in \mathbb{S}_N$, for any $\epsilon \in \mathbb{R}_+^*$:

$$\mathbb{P}\left(\sup_{\beta \in K} ||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1}|| > \epsilon\right) \leq A\exp(-BN)$$

for some constants A and positive B that depend only on K, ϵ, $||.||$ the norm chosen on the matrix space and Σ.[3]

The second proposition shows that the $\hat{\beta}_\sigma'$s are close to the true parameter with a probability that grows exponentially to 1 with N:

Proposition 2. *Under the the assumptions of proposition 1, for all $\epsilon > 0$ there exist scalars A and $B > 0$ that do not depend on N such that:*

$$\mathbb{P}\left(||\hat{\beta}_{id} - \beta_0|| > \epsilon\right) \leq A\exp(-BN)$$

where id denotes the identity permutation, i.e. $id \in \mathbb{S}_N$ with $id(i) = i$ for all $i \in N$.

[3] In fact, what I show is that

$$\mathbb{P}\left([\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2} \text{ is not invertible for some } \beta] \text{ OR } [\text{it is invertible AND } \sup_{\beta \in K} ||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1}|| > \epsilon]\right) \leq A\exp(-BN)$$

I omit this detail in the statement of the proposition to simplify the exposition.

That each estimator $\hat{\beta}_\sigma$ is close to the true parameter with a probability that increases this fast (exponentially) has very strong implications: if the set S is small enough (with a cardinality that grows polynomial in N), then the $\hat{\beta}_\sigma$'s are uniformly consistent *almost surely*. The following corollary shows uniform consistency in probability for any set S that grows sub-exponentially but not necessarily polinomialy!) because convergence in probability is enough for our purposes. The claim on the uniform almost sure convergence follows by Borel-Cantelli.

Corollary 1. *In addition to the assumptions of proposition 1, assume that S grows sub-exponentially, that is: $|S| = o(\exp(AN))$ for all $A \in \mathbb{R}$. Then*

$$\sup_{\sigma \in S} |\hat{\beta}_\sigma - \beta_0| \rightarrow_p 0$$

If in addition β_0 is an interior point in K, then with probability approaching 1, β_σ is in the interior of K for all $\sigma \in S$.

Now that we dealt with the technical concerns regarding the validity of the Taylor expansion in (4), the two following propositions look at the asymptotic distribution of each of the terms in the final formula (4).

Proposition 3. *Fix some (sequence) $S \subset \mathbb{S}_N$. Define*

$$C_{S,ij} := |\{\sigma \in S : \exists k : \{\sigma(2k-1), \sigma(2k)\} = \{i, j\}\}|$$

the number of times the pair $\{i, j\}$ appears in the subset of edges in S. In addition to the assumptions of proposition 1, assume that $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$, then

$$\sqrt{N}\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma \in S}\frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta}$$
$$\rightarrow_d \mathcal{N}\left(0, 4 \times \Sigma(\beta_0)^{-1} Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)} - (1 - Y_{12})\frac{\phi(X_{12}\beta_0)}{1 - \Phi(X_{12}\beta_0)}|X_1, U_1\right)\right)\Sigma(\beta_0)^{-1}\right)$$

I will delay the discussion over the new condition in this theorem: $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ until we state the main result in this paper in proposition 4. Putting all the previous propositions together, we are now able to determine the asymptotic distribution of $\hat{\beta}_S$:

Proposition 4. *Assume that:*

- *K is compact and convex;*

- *X has a compact support and*

- *the smallest eigenvalue of* $\Sigma(\beta)$ *is bounded away from 0 uniformly over* $\beta \in K$*, that is:*

$$\inf_{\beta \in K} \lambda_{\min}(\Sigma(\beta)) > 0$$

- S *grows sub-exponentially, that is:* $|S| = o(\exp(AN))$ *for all* $A \in \mathbb{R}$

- $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$

Then:

$$\sqrt{N}(\hat{\beta}_S - \beta_0) \to_d \mathcal{N}\left(0, 4 \times \Sigma(\beta_0)^{-1} Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)} - (1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0)}|X_1, U_1\right)\right)\Sigma(\beta_0)^{-1}\right)$$

Remarks regarding the condition $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$

First, notice that:

$$\begin{aligned}\frac{\sum_{i<j} C_{S,ij}^2}{N|S|^2} &= \frac{\sum_{ij}\sum_{\sigma,\pi \in S} \mathbb{1}(i,j \in \sigma \cap \pi)}{N|S|^2} \\ &= \frac{\sum_{\sigma,\pi\in S} |\sigma \cap \pi|}{N|S|^2}\end{aligned}$$

where I notationaly identify permutations with perfect matchings (sets of edges), so that $\sigma \cap \pi := \{\{i,j\} : \exists k, k' \text{ s.t. } \{i,j\} = \{\sigma(2k-1), \sigma(2k)\} = \{\pi(2k'-1), \pi(2k')\}\}$. The condition $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ is then equivalent to $\sum_{\sigma,\pi\in S}|\sigma\cap\pi| = o(N|S|^2)$. This alternative formulation clarifies the need for the condition: it is a diversification requirement on the set S. S is not allowed to include permutations (perfect matchings) that share too many edges. Specifically, the average overlap between all the perfect matchings in S shouldn't grow faster than N.

This condition restricts the choice of the set of permutations $|S|$, for instance, $|S|$ can't be bounded (as a function of N), since for all i, j:

$$C_{S,ij}^2 \geq C_{S,ij}$$

therefore:

$$\sum_{i<j} C_{S,ij}^2 \geq \sum_{i<j} C_{S,ij} = \frac{N-1}{2}|S|$$

if $|S|$ does not go to infinity as $N \to \infty$, then the condition $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ can't be satisfied.

On the other side, any S such that $C_{S,ij} \in \{0,1\}$ for all i,j, i.e. where each pair appears at most once, and such that $|S| \to +\infty$ as $N \to \infty$, satisfies the condition. That is because in that case: $C_{S,ij}^2 = C_{S,ij}$ for all i,j, therefore $\sum_{i<j} C_{S,ij}^2 = \sum_{i<j} C_{S,ij} = \frac{N-1}{2}|S| = o(N|S|^2)$. Such an S is always guaranteed to exist. Fix some N (even) and consider the set of permutations where I first include the identity permutation, then I "rotate" the second elements in each pair (rotate the even indices). In other words, consider the following set of permutations:

$$
\begin{aligned}
S := \{&(1,2,3,4,...,N-3,N-2,N-1,N);\\
&(1,4,3,6,...,N-3,N,N-1,2);\\
&(1,6,3,8,...,N-3,2,N-1,4);\\
&.\\
&.\\
&.\\
&(1,N,3,2,...,N-3,N-4,N-1,N-2)\}
\end{aligned}
$$

where $\sigma = (i_1,...,i_N)$ denotes the permutation $\sigma(k) = i_k$. Notice that the odd indices (1,3, ...) do not change from one permutation to the other, whereas the even indices are rotated. In this example, $|S| = \frac{N}{2}$ and $C_{S,ij} \in \{0,1\}$ for all i,j.

For computational reasons, one would want to choose a set S that is as small as possible. Any subset of S defined above would work provided that its size explodes with N.

A weaker sufficient condition for $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ would be that each edge is allowed to be repeated in S at most $c_N = o(|S|)$. In which case, for any pair $\{i,j\}$:

$$C_{S,ij}^2 \le c_N C_{S,ij}$$

so:

$$\sum_{i<j} C_{S,ij}^2 \le c_N \sum_{i<j} C_{S,ij} = c_N \frac{N-1}{2}|S| = o(N|S|^2)$$

as desired.

Importantly, S can be random as long as it is independent from all other variables (X, U and V). In fact, picking a random S can relieve from the burden of verifying the condition $\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$ as discussed in the following corollary.

Corollary 2. *Assume that:*

- *K is compact convex;*

- *X has a compact support and*

- *the smallest eigenvalue of $\Sigma(\beta)$ is bounded away from 0 uniformly over $\beta \in K$, that is:*

$$\inf_{\beta \in K} \lambda_{\min}(\Sigma(\beta)) > 0$$

- *For all N, the perfect matchings in S are drawn uniformly with replacement from the set of perfect matchings and $|S|$ is a deterministic function of N with: $|S| = O(\log(N))$.*

Then:

$$\sqrt{N}(\hat{\beta}_S - \beta_0) \to_d \mathcal{N}\left(0, 4 \times \Sigma(\beta_0)^{-1} Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)} - (1 - Y_{12})\frac{\phi(X_{12}\beta_0)}{1 - \Phi(X_{12}\beta_0)}|X_1, U_1\right)\right)\Sigma(\beta_0)^{-1}\right)$$

An interesting application is one where the econometrician doesn't observe the complete N-node network, but can only observe a subgraph containing only a subset of $\mathbb{S}_N$. I discuss this application in the next section.

3 The average estimator for networks with missing data

Assume that $Y_N = (Y_{ij})_{i,j \leq N}$ is generate following equation (3) as before. However, assume that now the econometrician cannot observe the entire network, instead, the econometrician observes a subgraph of Y only. Specifically, assume there is a random graph $G_N = (G_{ij})_{i,j \leq N}$ with $G_{ij} \in \{0, 1\}$ for all i, j such that

1. for all $i, j \leq N$, Y_{ij} is observed if and only if $G_{ij} = 1$ and
2. G_N and Y_N are independent.

If G_N has a set of perfect matchings S_N that meets the conditions of proposition 4 above, that is such that S_N grows sub-exponentially and $\sum_{i<j} C^2_{S,ij} = o(N|S|^2)$, then S_N can be used to construct the estimator $\hat{\beta}_{S_N}$ and the proposition 4 can be readily applied. However, checking that such a set S_N exists is hard. To the best of my knowledge, using the fastest algorithms available, the enumeration of all the perfect matchings in G_N can performed at a time complexity $O(N) \times$ the number of perfect matchings in G_N (Uno (1997)). As in the corollary 2, we can get around this difficulty by sampling from S_N, as long as we know that $\sum_{i<j} C^2_{S,ij} = o_p(N|S|^2)$ (there is no need to assume that $|S_N|$ grows sub exponentially).

Proposition 5. *Assume that:*

- *K is compact convex.*

- *X has a compact support.*

- *the smallest eigenvalue of $\Sigma(\beta)$ is bounded away from 0 uniformly over $\beta \in K$, that is:*

$$\inf_{\beta \in K} \lambda_{\min}(\Sigma(\beta)) > 0$$

- *Assume the set S_N of all perfect matchings in G_N is such that $\sum_{i<j} C^2_{S_N,ij} = o_p(N|S_N|^2)$.*

For all N, construct $\tilde{S}_N$, a tuple of perfect matchings uniformly drawn (with replacement) from S_N with a deterministic cardinality and $|\tilde{S}_N| \to +\infty$. Then:

$$\begin{aligned}&\sqrt{N}(\hat{\beta}_{\tilde{S}} - \beta_0) \\ &\to_d \mathcal{N}\left(0, 4 \times \Sigma(\beta_0)^{-1} Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)} - (1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0)}|X_1, U_1\right)\right)\Sigma(\beta_0)^{-1}\right)\end{aligned}$$

Thanks to the proposition 5, we no longer need to check that $\sum_{i<j} C^2_{S,ij} = o(N|S|^2)$ conditional on G_N as suggested by the proposition 4. It is instead enough that the underlying model that generates G_N be such that $\sum_{i<j} C^2_{S,ij} = o_p(N|S|^2)$ in probability. But how large is the class of models for G_N satisfying $\sum_{i<j} C^2_{S,ij} = o_p(N|S|^2)$? The next proposition provides a partial answer for bipartite graphs.[4]

For the purposes of proposition 6 only, consider the following model instead of model (3):

$$Y_{ij} = \mathbb{1}(X_{ij}\beta_0 + U_i + W_j + V_{ij} \geq 0) \tag{5}$$

for $i \in N$ and $j \in M$, N and M being two sets of nodes. We assume $|N| = |M|$, a necessary condition for perfect matchings to exist. We will overload the notation N: when there is no ambiguity, it will also refer to the cardinality $|N|$.

Proposition 6. *For any $\epsilon > 0$, if G_N is drawn uniformly from the set of bipartite graphs with at least $N^{\frac{3}{2}+\epsilon}$ edges, then $\sum_{i<j} C^2_{S,ij} = o_p(N|S|^2)$.*

4 Concluding remarks

This paper offers a systematic procedure to translate what we know about *i.i.d.* models to exchangeable array models. It could be particularly useful for models where no other

[4]The model in equation (3) that is one for a non bi-partite graph Y_N. However, the propositions 1 to 5 would hold, under the same proofs, for the bipartite graph model (5).

estimators have been analysed. I am particularly thinking about semi-parametric models where the composite likelihood estimator is not available. The proofs for other models would be basically the same as the ones in this paper at the cost of adding some smoothness assumptions (that are satisfied by the binomial parametric model studied here and that I did not need to emphasize).

The estimator, however, is likely to be (very) inefficient. It does not exploit the dependence structure that dyadic models exhibit. That is clear in the parametric model studied in this paper: the average MLE cannot outperform the composite MLE, which also suffers from the same flaw. However, one question that I am leaving under the shadow is: what happens if we take exponentially many $i.i.d.$ samples? If we do, the average MLE -were it to be well defined - would be computationally infeasible, but what would its theoretical properties be? clearly, from the proofs in this paper (again, if the MLE's are all defined and are all interior points!), the asymptotic distribution of the average MLE would be nothing like the composite maximum likelihood anymore.

Perhaps related to the question of inefficiency, the use of this procedure for data sets with missing observations could be interesting. First, it intuitively illustrates how inefficient the estimators obtained are: in the last proposition, I show that in general, for balanced bi-partite models, around $1/\sqrt{N}$th of the total number of observations (edges) is in general enough to perform like the estimator returned by the procedure if every edge were used (or like the composite MLE)! Second, the result in the last section sheds only a very dim light over the question of what observations are allowed to be missing in general graphs: non bi-partite or unbalanced bi-partite. The same proof strategy does not seem to work for other settings and I am curious to know what other models for the observable graph (G_N in the last section) would guarantee that the diversity condition on the set of all perfect matching be satisfied with high probability. Of course, allowing the observables's graph G_N to be correlated with the actual graph Y_N is yet another interesting and probably much more challenging question. I have not put enough thought towards an answer to the last question yet.

Finally, this paper did not discuss how the standard errors could be estimated. That was not my focus so far. However, given that each "averaged estimator" is computed based on a set of $i.i.d.$ sub-samples that are themselves drawn $i.i.d.$ uniformly from the set of available $i.i.d.$ sub-samples, we end-up with a huge number of "averaged" estimators. The idea of computing the standard errors by computing multiple "averaged" estimators each on a different $i.i.d.$ sub-sample, then computing an "empirical standard error" based on all these "averaged" estimators, is an appealing place to start thinking about standard error estimation.

Beautiful stuff ahead!

References

Harry Crane and Henry Towsner. Relatively exchangeable structures. *The Journal of Symbolic Logic*, 83(2):416–442, 2018. doi: 10.1017/jsl.2017.61.

Bryan S. Graham. Chapter 2 - network data. In Steven N. Durlauf, Lars Peter Hansen, James J. Heckman, and Rosa L. Matzkin, editors, *Handbook of Econometrics, Volume 7A*, volume 7 of *Handbook of Econometrics*, pages 111–218. Elsevier, 2020. doi: https://doi.org/10.1016/bs.hoe.2020.05.001. URL https://www.sciencedirect.com/science/article/pii/S1573441220300015.

R.W. Keener. *Theoretical Statistics: Topics for a Core Course*. Springer Texts in Statistics. Springer New York, 2010. ISBN 9780387938394. URL https://books.google.co.in/books?id=aVJmcega44cC.

L. Lovász and M.D. Plummer. *Matching Theory*. AMS Chelsea Publishing Series. AMS Chelsea Pub., 2009. ISBN 9780821847596. URL https://books.google.com/books?id=OaoJBAAAQBAJ.

Kevin A. O'Neil and Richard A. Redner. Asymptotic Distributions of Weighted U-Statistics of Degree 2. *The Annals of Probability*, 21(2):1159 – 1169, 1993. doi: 10.1214/aop/1176989286. URL https://doi.org/10.1214/aop/1176989286.

Patrick Eugene O'Neil. Asymptotics in random (0, 1)-matrices. *Proceedings of the American Mathematical Society*, 25(1):39–45, 1970. ISSN 00029939, 10886826. URL http://www.jstor.org/stable/2036523.

Takeaki Uno. Algorithms for enumerating all perfect, maximum and maximal matchings in bipartite graphs. In *Proceedings of the 8th International Symposium on Algorithms and Computation*, ISAAC '97, page 92–101, Berlin, Heidelberg, 1997. Springer-Verlag. ISBN 3540638903.

5 Proofs

5.1 Proof of proposition 1

Proof. *(Proposition 1)* The proof follows in 3 steps:

Step 1: Prove that for any continuous function W from $support(X) \times support(Y) \times K$ into $\mathbb{R}$, with mean: $\mu(\beta) = E(W(X_{\sigma,i}, Y_{\sigma,i}, \beta))$, there are constants A' and B' such that for all $\epsilon > 0$:

$$\mathbb{P}\left(\sup_{\beta \in K} |\bar{W}(\beta) - \mu(\beta)| > 4\epsilon\right) \leq A' \exp(-B'N)$$

with $\bar{W}(\beta) := \frac{2}{N}\sum_{i=1}^{N/2} W(X_{\sigma,i}, Y_{\sigma,i}; \beta)$

Fix $\epsilon > 0$. For any $\beta \in K$, define:
$\lambda_\delta(\beta) := E\left(\sup_{\beta':||\beta'-\beta||\leq\delta} |W(X_{\sigma,i}, Y_{\sigma,i}, \beta) - W(X_{\sigma,i}, Y_{\sigma,i}, \beta')|\right)$ and $\delta > 0$ such that $\lambda_\delta(\beta) < \epsilon$ for all $\beta \in K$. Such δ exists because by theorem 9.1 in Keener (2010):

$$\sup_{\beta \in K} \lambda_\delta(\beta) \rightarrow_{\delta\rightarrow 0} 0$$

Since K is compact, let $(\beta_i)_{i=1..m}$ be a finite set of elements in K such that the open balls O_i centered at β_i with radius δ cover K. Following the proof of theorem 9.2 in Keener (2010), note that:

$$\begin{aligned}\sup_{\beta \in K} |\bar{W}(\beta) - \mu(\beta)| &= \max_{i=1..m} \sup_{\beta \in O_i} |\bar{W}(\beta) - \mu(\beta)| \\ &\leq \max_{i=1..m} \sup_{\beta \in O_i} |\bar{W}(\beta) - \bar{W}(\beta_i)| + |\bar{W}(\beta_i) - \mu(\beta_i)| + |\mu(\beta_i) - \mu(\beta)|\end{aligned}$$

Note that for all i and for all $\beta \in O_i$:

$$|\mu(\beta_i) - \mu(\beta)| \leq \lambda_\delta(beta_i) \leq \epsilon$$

second, observe:

$$\bar{M}_{\delta,N}(\beta) := \frac{2}{N}\sum_{i=1}^{\frac{N}{2}} \sup_{\beta':||\beta'-\beta||\leq\delta} |W(X_{\sigma,i}, Y_{\sigma,i}, \beta) - W(X_{\sigma,i}, Y_{\sigma,i}, \beta')|$$

and note that:

$$\begin{aligned}
\max_{i=1..m} \sup_{\beta \in O_i} |\bar{W}(\beta) - \bar{W}(\beta_i)| &\leq \max_{i=1..m} \bar{M}_{\delta,N}(\beta_i) \\
&\leq \max_{i=1..m} |\bar{M}_{\delta,N}(\beta_i) - \lambda_\delta(\beta_i)| + \max_{i=1..m} \lambda_\delta(\beta_i) \\
&\leq \max_{i=1..m} |\bar{M}_{\delta,N}(\beta_i) - \lambda_\delta(\beta_i)| + \epsilon
\end{aligned}$$

Therefore:

$$\sup_{\beta \in K} |\bar{W}(\beta) - \mu(\beta)| \leq \max_{i=1..m} |\bar{M}_{\delta,N}(\beta_i) - \lambda_\delta(\beta_i)| + \max_{i=1..m} |\bar{W}(\beta_i) - \mu(\beta_i)| + 2\epsilon$$

Hence:

$$\begin{aligned}
\mathbb{P}(\sup_{\beta \in K} |\bar{W}(\beta) - \mu(\beta)| \geq 4\epsilon) &\leq \mathbb{P}(\max_{i=1..m} |\bar{M}_{\delta,N}(\beta_i) - \lambda_\delta(\beta_i)| + \max_{i=1..m} |\bar{W}(\beta_i) - \mu(\beta_i)| + 2\epsilon \geq 4\epsilon) \\
&\leq \mathbb{P}(\max_{i=1..m} |\bar{M}_{\delta,N}(\beta_i) - \lambda_\delta(\beta_i)| + \max_{i=1..m} |\bar{W}(\beta_i) - \mu(\beta_i)| \geq 2\epsilon) \\
&\leq \mathbb{P}(\max_{i=1..m} |\bar{M}_{\delta,N}(\beta_i) - \lambda_\delta(\beta_i)|| \geq \epsilon) + \mathbb{P}(\max_{i=1..m} |\bar{W}(\beta_i) - \mu(\beta_i)| \geq \epsilon) \\
&\leq m \times \big(\mathbb{P}(|\bar{M}_{\delta,N}(\beta_1) - \lambda_\delta(\beta_1)|| \geq \epsilon) + \mathbb{P}(|\bar{W}(\beta_1) - \mu(\beta_1)| \geq \epsilon)\big)
\end{aligned}$$

By the compactness of $support(X) \times support(Y) \times K$ and the continuity of W, $(W(X_{\sigma,i}, Y_{\sigma,i}, \beta_1) - \mu(\beta_1))_i$ and $\sup_{\beta' : ||\beta' - \beta_1|| \leq \delta} |W(X_{\sigma,i}, Y_{\sigma,i}, \beta_1) - W(X_{\sigma,i}, Y_{\sigma,i}, \beta')| - \lambda_\delta(\beta_1))_i$ are i.i.d. and bounded, Hoeffding's inequality allows to conclude.

Step 2: Show that for any $\sigma \in \mathbb{S}_N$, for any $\epsilon \in \mathbb{R}_+^*$ there exist constants some constants A'' and B'' that depend only on K and ϵ

$$\mathbb{P}\left(\sup_{\beta \in K} ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| > 4\epsilon\right) \leq A'' \exp(-B''N)$$

To see that, it is enough to apply the result in step 1 element wise on the matrix $\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}$ then use a union bound to obtain the desired result for the max norm $||.||_\infty$.

Step 3: Show the final result.

Note that for any β, and any given σ (using a sub-multiplicative matrix norm this time):

$$\begin{aligned}
||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1}|| &\leq ||\Sigma(\beta)^{-1}|| \times ||\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1}|| \times ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \\
&\leq \frac{||\Sigma(\beta)^{-1}||}{\lambda_{\min}(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2})} \times ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}||
\end{aligned}$$

where $\lambda_{\min}(.)$ returns the smallest eigen value. Take some $x \in \mathbb{R}^k$ such that $||x|| = 1$ and $x' \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2} x = \lambda_{\min}(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2})$, then :

$$\lambda_{\min}(\Sigma(\beta)) - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \leq x'\Sigma(\beta)x - x'\left(\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}\right)x = x'\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}x$$

implying:[5]

$$\lambda_{\min}(\Sigma(\beta)) - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \leq \lambda_{\min}(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2})$$

under the event: $\sup_\beta ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| < \inf_\beta \lambda_{\min}(\Sigma(\beta))$ so:

$$\begin{aligned} ||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1}|| &\leq \frac{||\Sigma(\beta)^{-1}||}{\lambda_{\min}(\Sigma(\beta)) - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}||} \times ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \\ &\leq \frac{\sup_\beta\{||\Sigma(\beta)^{-1}||\}}{\inf_\beta\{\lambda_{\min}(\Sigma(\beta))\} - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}||} \times ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \end{aligned}$$

Therefore for any $\epsilon > 0$, there exists a function that only depends on epsilon $\gamma(\epsilon) > 0$ such that, under the event under the event $E_N := \sup_\beta ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| < \inf_\beta \lambda_{\min}(\Sigma(\beta))$ we have

$$\sup_\beta ||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1}|| \geq 4\epsilon \Rightarrow \sup_\beta ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \geq \gamma(\epsilon)$$

[5] Note here that we could similarly show:

$$\lambda_{\min}(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}) - ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}|| \leq \lambda_{\min}(\Sigma(\beta))$$

implying that:

$$||\lambda_{\min}(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}) - \lambda_{\min}(\Sigma(\beta))|| \leq ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}||$$

and leading to the result alluded to in a footnote to the proposition's statement:

$$\mathbb{P}\left(\frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2} \text{ is not invertible for some } \beta\right) \leq A'' \exp(-B''N)$$

for some generic A'', $B'' > 0$ that are independent of N.

then:

$$
\begin{aligned}
\mathbb{P}\left(\sup_{\beta} ||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1} || \geq 4\epsilon\right) &= \mathbb{P}\left(\sup_{\beta} ||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1} || \geq 4\epsilon; E_N\right) \\
&+ \mathbb{P}\left(\sup_{\beta} ||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1} || \geq 4\epsilon; not(E_N)\right) \\
&\leq \mathbb{P}\left(\sup_{\beta} ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2} || \geq \gamma(\epsilon); E_N\right) \\
&+ \mathbb{P}(not(E_N)) \\
&\leq \mathbb{P}\left(\sup_{\beta} ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2} || \geq \gamma(\epsilon)\right) \\
&+ \mathbb{P}\left(\sup_{\beta} ||\Sigma(\beta) - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2} || \geq \inf_{\beta} \lambda_{\min}(\Sigma(\beta))\right)
\end{aligned}
$$

which allows to conclude by step 2. □

5.2 Proof of proposition 2

Proof. Note that:

$$
\begin{aligned}
\hat{\beta}_{id} - \beta_0 &= \left(\frac{\partial^2 \mathcal{L}_{id}(X,Y;\bar{\beta}_{id})}{\partial \beta^2}\right)^{-1} \frac{\partial \mathcal{L}_{id}(X,Y;\beta_0)}{\partial \beta} \\
&= -\Sigma(\bar{\beta}_{id})^{-1} \frac{\partial \mathcal{L}_{id}(X,Y;\beta_0)}{\partial \beta} + \left(\Sigma(\bar{\beta}_{id})^{-1} - \left(\frac{\partial^2 \mathcal{L}_{id}(X,Y;\bar{\beta}_{id})}{\partial \beta^2}\right)^{-1}\right) \frac{\partial \mathcal{L}_{id}(X,Y;\beta_0)}{\partial \beta}
\end{aligned}
$$

Thanks to the compactness of the support of X and of K, $\frac{\partial \mathcal{L}_{id}(X,Y;\beta_0)}{\partial \beta}$ is bounded by some constant M and $\beta \to \Sigma(\beta)^{-1}$ is bounded by some constant L.

Hence:

$$
||\hat{\beta}_{id} - \beta_0|| \leq L|| \frac{\partial \mathcal{L}_{id}(X,Y;\beta_0)}{\partial \beta} || + M \sup_{\beta \in K} ||\Sigma(\beta)^{-1} - \frac{\partial^2 \mathcal{L}_\sigma(X,Y;\beta)}{\partial \beta^2}^{-1} ||
$$

Applying the Hoeffding bound to the first term and proposition 1 to the second, we obtain the desired result. □

5.3 Proof of corollary 1

Proof. Fix $\epsilon > 0$.

$$\mathbb{P}(\sup_{\sigma \in S} |\hat{\beta_\sigma} - \beta_0| > \epsilon) \leq |S| \mathbb{P}(|\hat{\beta_{id}} - \beta_0| > \epsilon)$$

and the proposition (2) completes the proof. $\square$

5.4 Proof of proposition 3

Proof. Fix some $S \subset \mathbb{S}_N$ and some $\lambda \in \mathbb{R}^k$. I want to determine the asymptotic distribution of:

$$\begin{aligned}
\lambda' \frac{1}{|S|} \Sigma(\beta_0)^{-1} \sum_{\sigma \in S} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} &= \lambda' \frac{1}{|S| N/2} \sum_{\sigma \in S} \sum_{i=1}^{\frac{N}{2}} \left(Y_{\sigma,i} \frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)} - (1 - Y_{\sigma,i}) \frac{\phi(X_{\sigma,i}\beta_0)}{1 - \Phi(X_{\sigma,i}\beta_0)} \right) \\
&\quad \times \Sigma(\beta_0)^{-1} X_{\sigma,i} \\
&=: \frac{1}{|S| N/2} \sum_{\sigma \in S} \sum_{i=1}^{\frac{N}{2}} f(X_{\sigma,i}, Y_{\sigma,i})
\end{aligned}$$

where $f(X_{\sigma,i}, Y_{\sigma,i}) := \left(Y_{\sigma,i} \frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)} - (1 - Y_{\sigma,i}) \frac{\phi(X_{\sigma,i}\beta_0)}{1 - \Phi(X_{\sigma,i}\beta_0)} \right) \lambda' \Sigma(\beta_0)^{-1} X_{\sigma,i}$. Although f depends on λ and β, they are omitted to simplify the notation. We can rearrange:

$$\lambda' \frac{1}{|S|} \sum_{\sigma \in S} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} = \frac{1}{|S| N/2} \sum_{i<j} C_{S,ij} f(X_{i,j}, Y_{i,j})$$

with $C_{S,ij} := |\{\sigma \in S : \exists k : \{\sigma(2k-1), \sigma(2k)\} = \{i, j\}\}|$, the number of times the pair $\{i, j\}$ appears in the subset of observations generated by S. Observe that for all i, j, by definition: $C_{S,ij} = C_{S,ji}$ and $C_{S,ii} = 0$. Also note that for all i:

$$\sum_{j=1}^{N} C_{S,ij} = |S|$$

since every pair appears exactly once per permutation $\sigma \in S$. Denote:

$$q(X_i, X_j, U_i, U_j) := E(f(X_{i,j}, Y_{i,j}) | X_i, X_j, U_i, U_j)$$

$$h(X_i, U_i) := E(q(X_i, X_j, U_i, U_j) | X_i, U_i)$$

and

$$\tilde{q}(X_i, X_j, U_i, U_j) := q(X_i, X_j, U_i, U_j) - h(X_i, U_i) - h(X_j, U_j)$$

where it is assumed that $C_{S,ii} = 0$ for all i and $C_{S,ij} = C_{S,ji}$ for all i and j. Observe, following O'Neil and Redner (1993), that

$$\begin{aligned}\sum_{i<j} C_{S,ij} q(X_i, X_j, U_i, U_j) &= \sum_{i<j} C_{S,ij} \tilde{q}(X_i, X_j, U_i, U_j) + \sum_i \left(\sum_{j=1}^{N} C_{S,ij} \right) h(X_i, U_i) \\ &= \sum_{i<j} C_{S,ij} \tilde{q}(X_i, X_j, U_i, U_j) + |S| \sum_i h(X_i, U_i)\end{aligned}$$

So:

$$\begin{aligned}\lambda' \Sigma(\beta_0)^{-1} \frac{1}{|S|} \sum_{\sigma \in S} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} &= \frac{1}{|S|N/2} \sum_{i<j} C_{S,ij} \left(f(X_{i,j}, Y_{i,j}) - q(X_i, X_j, U_i, U_j) \right) \\ &+ \frac{1}{|S|N/2} \sum_{i<j} C_{S,ij} \tilde{q}(X_i, X_j, U_i, U_j) + \frac{2}{N} \sum_i h(X_i, U_i)\end{aligned} \tag{6}$$

We have:

$$\begin{aligned}Var\left(\sum_{i<j} C_{S,ij} \left(f(X_{i,j}, Y_{i,j}) - q(X_i, X_j, U_i, U_j) \right) \right) &= Var\left(f(X_{1,2j}, Y_{1,2}) - q(X_1, X_2, U_1, U_2) \right) \sum_{i<j} C_{S,ij}^2 \\ Var\left(\sum_{i<j} C_{S,ij} \tilde{q}(X_i, X_j, U_i, U_j) \right) &= Var\left(\tilde{q}(X_1, X_2, U_1, U_2) \right) \sum_{i<j} C_{S,ij}^2\end{aligned} \tag{7}$$

Assuming :

$$\sum_{i<j} C_{S,ij}^2 = o(N|S|^2)$$

then:

$$\begin{aligned}\sqrt{N} \lambda' \Sigma(\beta_0)^{-1} \frac{1}{|S|} \sum_{\sigma \in S} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} &= \frac{2}{\sqrt{N}} \sum_i h(X_i, U_i) + o_p(1) \\ &\to_d \mathcal{N}(0, 4Var(h(X_1, U_1)))\end{aligned} \tag{8}$$

but

$$Var(h(X_1,U_1)) = \lambda'\Sigma(\beta_0)^{-1}Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0} - (1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0}|X_1,U_1\right)\right)\Sigma(\beta_0)^{-1}\lambda$$

therefore, using the wold device:

$$\sqrt{N}\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}$$
$$\to_d \mathcal{N}\left(0, 4\times\Sigma(\beta_0)^{-1}Var\left(E\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)} - (1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0)}|X_1,U_1\right)\right)\Sigma(\beta_0)^{-1}\right)$$

□

5.5 Proof of proposition 4

Proof. (Proposition 4)

Remember:

$$\hat{\beta}_S - \beta_0 = -\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta} + \frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}$$

First, I show that:

$$\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta} = o_p(1)$$

Note that because K is compact and $\beta \to \Sigma(\beta)^{-1}$ is conitnuously differentiable, then $\beta \to \Sigma(\beta)^{-1}$ is Lipschitz on K, let η be the Lipschitz constant.

$$\begin{aligned}
&\left|\left|\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\beta_0)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right| \\
&\leq \left|\left|\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\bar{\beta}_\sigma)^{-1} - \Sigma(\beta_0)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right| \\
&+ \left|\left|\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left[\Sigma(\bar{\beta}_\sigma)^{-1} - \left(\frac{\partial^2\mathcal{L}_\sigma(X,Y;\bar{\beta}_\sigma)}{\partial\beta^2}\right)^{-1}\right]\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right| \\
&\leq \left(\eta\sup_{\sigma\in S}||\hat{\beta}_\sigma - \beta_0|| + \sup_{\beta\in K}||\Sigma(\beta)^{-1} - \frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}||\right)\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left|\left|\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right|
\end{aligned}$$

First, note:

$$\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left|\left|\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right| = O_p(1)$$

because:

$$\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\left|\left|\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}\right|\right| \leq \sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}||\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}||$$

and

$$\begin{aligned}
&Var\left(\sqrt{N}\frac{1}{|S|}\sum_{\sigma\in S}\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}||\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}||\right)\\
&=\frac{N}{|S|^2}[|S|Var\left(\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}||\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}||\right)\\
&+|S|(|S|-1)Cov(\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}||\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}||,\\
&\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}||\left(Y_{\sigma',i}\frac{\phi(X_{\sigma',i}\beta_0)}{\Phi(X_{\sigma',i}\beta_0)}-(1-Y_{\sigma',i})\frac{\phi(X_{\sigma',i}\beta_0)}{1-\Phi(X_{\sigma',i}\beta_0)}\right)X_{\sigma',i}||)]\\
&\leq NVar\left(\frac{2}{N}\sum_{i=1}^{\frac{N}{2}}||\left(Y_{\sigma,i}\frac{\phi(X_{\sigma,i}\beta_0)}{\Phi(X_{\sigma,i}\beta_0)}-(1-Y_{\sigma,i})\frac{\phi(X_{\sigma,i}\beta_0)}{1-\Phi(X_{\sigma,i}\beta_0)}\right)X_{\sigma,i}||\right)\\
&=4Var\left(||\left(Y_{12}\frac{\phi(X_{12}\beta_0)}{\Phi(X_{12}\beta_0)}-(1-Y_{12})\frac{\phi(X_{12}\beta_0)}{1-\Phi(X_{12}\beta_0)}\right)X_{12}||\right)
\end{aligned}$$

By proposition 1:

$$\sup_{\beta\in K}||\Sigma(\beta)^{-1}-\frac{\partial^2\mathcal{L}_\sigma(X,Y;\beta)}{\partial\beta^2}^{-1}||=o_p(1)$$

and

$$\sup_{\sigma\in S}||\hat{\beta}_\sigma-\beta_0||=o_p(1)$$

since:

$$\mathbb{P}(\sup_{\sigma\in S}||\hat{\beta}_\sigma-\beta_0||>\epsilon)\leq|S|\mathbb{P}(||\hat{\beta}_{id}-\beta_0||>\epsilon)\to 0$$

where the limit is obtained by proposition 2 and by the assumption that $|S|=o(\exp(AN))$ for all $A\in\mathbb{R}$.

Finally:

$$\sqrt{N}(\hat{\beta}_S - \beta_0) = -\sqrt{N}\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta} + o_p(1) \tag{9}$$

and proposition 3 allows to conclude.

□

5.6 Proof of corollary 2

Proof. (Corollary 2.) Assume that for all N, S (in fact, S_N) is constructed by drawing permuations (or perfect matchings) with replacement from the set of perfect matchings. Denote $C_{ij,N}$ the number of perfect matchings in which the pair i,j appears in the set S_N. Let c_N be a deterministic sequence such that $c_N \to +\infty$ and $c_N = o(|S|)$.
Define the events $E_N := \{C_{ij,N} > c_N \text{ for some pair } i,j\}$. Then:

$$\begin{aligned}\mathbb{P}(E_N) &\leq \sum_{ij}\mathbb{P}(C_{ij,N} > c_N)\\ &= \frac{N(N-1)}{2}\mathbb{P}(C_{12,N} > c_N)\\ &= \frac{N(N-1)}{2}\sum_{k=c_N+1}^{|S|}\mathbb{P}(C_{12,N} = k)\end{aligned}$$

Let σ be the random variable corresponding to a single uniform draw from the set of all perfect matchings (i.e. permutations in $\mathbb{S}_N$). For any fixed pair i,j we have:[6]

$$\mathbb{P}(i,j\in\sigma) = \frac{\frac{(N-2)!}{(N/2-1)!2^{N/2-1}}}{\frac{N!}{(N/2)!2^{N/2}}} = \frac{2N/2}{N(N-1)} = \frac{1}{N-1}$$

so

$$\mathbb{P}(C_{12,N} = k) = \binom{|S|}{k}\left(\frac{1}{N-1}\right)^k\left(\frac{N-2}{N-1}\right)^{|S|-k} \leq \binom{|S|}{k}\left(\frac{1}{N-1}\right)^{c_N+1}$$

[6] I abuse notation here: $i,j\in\sigma$ means that the pair i,j forms an edge in the perfect matching σ, or in the language of permutations that there exists some k such that $\{\sigma(2k-1),\sigma(2k)\} = \{i,j\}$.

and

$$\begin{aligned}
\mathbb{P}(E_N) &\leq \frac{N(N-1)}{2}\left(\frac{1}{N-1}\right)^{c_N+1}\sum_{k=c_N+1}^{|S|}\binom{|S|}{k} \\
&\leq \frac{N(N-1)}{2}\left(\frac{1}{N-1}\right)^{c_N+1} \times 2^{|S|} \\
&= O\left(\left(\frac{1}{N}\right)^{C_N-2}\right) \\
&= o(\frac{1}{N^2})
\end{aligned}$$

Therefore

$$\sum_N \mathbb{P}(E_N) < +\infty$$

By the Borel–Cantelli lemma:

$$\mathbb{P}(\limsup E_N) = \mathbb{P}(\cap_{N\geq 1} \cup_{k=N}^{\infty} E_k) = 0$$

or equivalently:

$$\mathbb{P}(\exists N_0\ \forall N > N_0\ \forall i,j \leq N:\ C_{ij,N} \leq c_N) = 1$$

as shown earlier, if for all pairs i,j $C_{ij,N} \leq c_N = o(|S|)$, then:

$$\sum_{i<j} C_{ij,N}^2 \leq c_N \sum_{i<j} C_{ij,N} = c_N \times \frac{N-1}{2}|S| = o(N|S|^2)$$

hence with probability one, the condition: $\sum_{i<j} C_{ij,N}^2 = o(N|S|^2)$ is satisfied.[7] The rest of the proof for proposition 4 follows. □

5.7 Proof of proposition 5

Proof. Of proposition 5 Given the conditions of the proposition, denote $\tilde{\mathbb{P}}$ the probability conditional on S_N. By defintion:

$$\tilde{C}_{\tilde{S}_N,ij} := \sum_{\sigma\in\tilde{S}_N} \mathbb{1}(ij \in \sigma)$$

[7]Here I showed that $\frac{C_{ij,N}^2}{N|S|^2} \to 0$ almost surely. In fact, it was enough to show convergence in probability since that is enough to obtain equation 9 and conclude.

the terms in this sum are $i.i.d.$ conditional on S_N, because the perfect matchings in $\tilde{S}_N$ are $i.i.d.$, therefore:

$$\tilde{\mathbb{E}}(\tilde{C}_{\tilde{S}_N,ij}) = |\tilde{S}_N| \frac{C_{S_N,ij}}{|S_N|}$$

and

$$\tilde{\mathbb{V}}(\tilde{C}_{\tilde{S}_N,ij}) = |\tilde{S}_N| \frac{C_{S_N,ij}}{|S_N|} \left(1 - \frac{C_{S_N,ij}}{|S_N|}\right)$$

then

$$\tilde{\mathbb{E}}(\tilde{C}^2_{\tilde{S}_N,ij}) = |\tilde{S}_N| \frac{C_{S_N,ij}}{|S_N|} + (|\tilde{S}_N|^2 - |\tilde{S}_N|) \frac{C^2_{S_N,ij}}{|S_N|^2}$$

and:

$$\begin{aligned}
\tilde{\mathbb{E}}\left(\frac{\sum_{i<j} \tilde{C}^2_{\tilde{S}_N,ij}}{N|\tilde{S}_N|^2}\right) &= \frac{\sum_{i<j} C_{S_N,ij}}{N|\tilde{S}_N| \times |S_N|} + \left(1 - \frac{1}{|\tilde{S}_N|}\right) \frac{\sum_{i<j} C^2_{S_N,ij}}{N|S_N|^2} \\
&= \frac{1}{|\tilde{S}_N|} + \left(1 - \frac{1}{|\tilde{S}_N|}\right) \frac{\sum_{i<j} C^2_{S_N,ij}}{N|S_N|^2}
\end{aligned}$$

Remember equation (6):

$$\begin{aligned}
\lambda' \Sigma(\beta_0)^{-1} \frac{1}{|\tilde{S}_N|} \sum_{\sigma \in \tilde{S}_N} \frac{\partial \mathcal{L}_\sigma(X, Y; \beta_0)}{\partial \beta} &= \frac{1}{|\tilde{S}_N| N/2} \sum_{i<j} \tilde{C}_{\tilde{S}_N,ij} \left(f(X_{i,j}, Y_{i,j}) - q(X_i, X_j, U_i, U_j)\right) \\
&+ \frac{1}{|\tilde{S}_N| N/2} \sum_{i<j} \tilde{C}_{\tilde{S}_N,ij} \tilde{q}(X_i, X_j, U_i, U_j) + \frac{2}{N} \sum_i h(X_i, U_i)
\end{aligned}$$

equation (7) becomes:

$$Var\left(\sum_{i<j} \tilde{C}_{\tilde{S}_N,ij} \left(f(X_{i,j}, Y_{i,j}) - q(X_i, X_j, U_i, U_j)\right) |(S_N)_{N \geq 0}, (\tilde{S}_N)_{N \geq 0}\right) = Var\left(f(X_{1,2j}, Y_{1,2}) - q(X_1, X_2, U_1, U_2)\right) \sum_{i<j} \tilde{C}^2_{\tilde{S}_N,i}$$

$$Var\left(\sum_{i<j} \tilde{C}_{\tilde{S}_N,ij} \tilde{q}(X_i, X_j, U_i, U_j) |(S_N)_{N \geq 0}, (\tilde{S}_N)_{N \geq 0}\right) = Var\left(\tilde{q}(X_1, X_2, U_1, U_2)\right) \sum_{i<j} \tilde{C}^2_{\tilde{S}_N,ij}$$

then

$$Var[\frac{1}{|\tilde{S}_N|\sqrt{N}}\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}(f(X_{i,j},Y_{i,j})-q(X_i,X_j,U_i,U_j))|(S_N)_{N\geq 0}]$$

$$= Var\left(f(X_{1,2j},Y_{1,2})-q(X_1,X_2,U_1,U_2)\right)\tilde{\mathbb{E}}\left(\frac{1}{|\tilde{S}_N|\sqrt{N}}\sum_{i<j}\tilde{C}^2_{\tilde{S}_N,ij}\right)$$

$$+\tilde{\mathbb{V}}\left(E\left(\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}\left(f(X_{i,j},Y_{i,j})-q(X_i,X_j,U_i,U_j)\right)|(S_N)_{N\geq 0},(\tilde{S}_N)_{N\geq 0}\right)\right)$$

$$= Var\left(f(X_{1,2j},Y_{1,2})-q(X_1,X_2,U_1,U_2)\right)\left(\frac{1}{|\tilde{S}_N|}+\left(1-\frac{1}{|\tilde{S}_N|}\right)\frac{\sum_{i<j}C^2_{S_N,ij}}{N|S_N|^2}\right)$$

where the second equality results from the observation that for all i,j

$$E\left(\tilde{C}_{\tilde{S}_N,ij}\left(f(X_{i,j},Y_{i,j})-q(X_i,X_j,U_i,U_j)\right)|(S_N)_{N\geq 0},(\tilde{S}_N)_{N\geq 0}\right)=0$$

likewise:

$$Var[\frac{1}{|\tilde{S}_N|\sqrt{N}}\sum_{i<j}\tilde{C}_{\tilde{S}_N,ij}\tilde{q}(X_i,X_j,U_i,U_j)|(S_N)_{N\geq 0},(\tilde{S}_N)_{N\geq 0}]$$

$$= Var\left(\tilde{q}(X_1,X_2,U_1,U_2)\right)\left(\frac{1}{|\tilde{S}_N|}+\left(1-\frac{1}{|\tilde{S}_N|}\right)\frac{\sum_{i<j}C^2_{S_N,ij}}{N|S_N|^2}\right)$$

so equation (8) still holds, conditionally on $(S_N)_{N\geq 0}$ this time:

$$\sqrt{N}\lambda'\Sigma(\beta_0)^{-1}\frac{1}{|S|}\sum_{\sigma\in S}\frac{\partial\mathcal{L}_\sigma(X,Y;\beta_0)}{\partial\beta}=\frac{2}{\sqrt{N}}\sum_i h(X_i,U_i)+o_p(1)$$
$$\to_d \mathcal{N}(0,4Var(h(X_1,U_1)))$$

by dominated convergence, equation (8) also holds unconditionally. The rest of the argument for propositions 3 and 4 follows. □

5.8 Proof of proposition 6

Proof. Proposition 6.

The proof is for some fixed $\epsilon>0$.

First, without loss of generality, I assume that the sequence of graphs G_N are independent. Otherwise, I would work on another sequence (a coupling) $(G'_N)_N$ such that $G_N=_d G'_N$ and the G'_N are independent. In that case, $\frac{\sum_{i<j}C^2_{S',ij}}{N|S'|^2}=_d\frac{\sum_{i<j}C^2_{S,ij}}{N|S|^2}$ (the first ratio is computed on

G'_N and the second on G_N), therefore proving the proposition for G'_N implies that it also holds for G_N.

Let $e(G_N)$ denote the number of edges of the graph G_N. I will show the result conditionally on the sequence $(e(G_N)_N)_{N\geq 0}$, then proposition 6 will follow by dominated convergence. For the rest of the proof, the "ambient" probability is that conditional on $(e(G_N)_N)_{N\geq 0}$: I will omit to condition by $(e(G_N)_N)_{N\geq 0}$ in my notation. Further, I will use the notation e_N for $e(G_N)$.

Note that, conditional on $e(G_N)$, G_N is uniformly drawn from the set of graphs with exactly $e(G_N)$ edges.

First, I show that as $N \to +\infty$:

$$\begin{aligned} E(S_N) &\sim N! \left(\frac{e_N}{N^2}\right)^N \exp\left(-\frac{1}{2}\left(\frac{N^2}{e_N} - 1\right)\right) \\ E(S_N^2) &\sim E(S_N)^2 \\ E(C_{S_N,ij}^2) &\sim (N-1)!^2 \exp\left(1 - \frac{N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N-1} \end{aligned} \tag{10}$$

The first two statements result immediately from the theorems 1 and 2 in O'Neil (1970) (cf. the section 8.1 in Lovász and Plummer (2009) for details about the link between perfect matchings in a bipartite graph and the permanent of its bi-adjacency matrix). The proof for $E(C_{S_N,ij}^2)$ follows similar steps as those of the proofs for theorems 1 and 2 in O'Neil (1970). As in O'Neil (1970), denote, for any permutation $\sigma \in \mathbb{S}_N$,

$$x_\sigma := 1\left\{(i, \sigma(i)) \text{ is an edge in } G_N \text{ for all } i \in N\right\}$$

and for any integers M and $k \leq M$, define:

$$B_k^M := \{(\sigma, \pi) \in \mathbb{S}_M^2 : |\{i \in M : \pi(i) = \sigma(i)\}| = k\}$$

that is, if we identify every permutation in $\mathbb{S}_M$ to a perfect matching between two sets of carnality M each, then B_k^M is the set of perfect matching pairs that have exactly k edges in common.

By definition:

$$C_{S_N,ij} = \sum_{\sigma \in \mathbb{S}_N : \sigma(i) = j} x_\sigma$$

and:

$$C^2_{S_N,ij} = \sum_{\sigma,\pi\in\mathbb{S}_N:\sigma(i)=\pi(i)=j} x_\sigma x_\pi$$

so

$$E(C^2_{S_N,ij}) = \sum_{k=1}^{N} |B^{N-1}_{k-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma)\in B^N_k)$$

the equation (1.9) in O'Neil (1970) yields:

$$|B^N_k| = \frac{N!^2}{k!}e^{-1}\left(1 + O(\frac{1}{(N-k+1)!})\right)$$

and for $k \leq k_1 := [N^{5/8}]$, equation (1.14) in O'Neil (1970):

$$\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma)\in B^N_k) = \left(\frac{e_N}{N^2}\right)^{2N-k} \exp\left(-2\left(1-\frac{k}{N}\right)^2\left(\frac{N^2}{e_N}-1\right)\right)\left(1+O(N^{-1/4-\epsilon})+O(N^{-2\epsilon})\right)$$

hence

$$
\begin{aligned}
E(C^2_{S_N,ij}) &= (N-1)!^2 \exp\left(1 - \frac{2N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N} \sum_{k=1}^{k_1} \frac{1}{(k-1)!}\left[\frac{N^2}{e_N}\exp\big(4(N/e_N - 1/N)\big)\right]^k \\
&\times \Big(1 + O(N^{-1/4-\epsilon}) + O(N^{-2\epsilon})\Big) + \sum_{k=k_1+1}^{N} |B^{N-1}_{k-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma) \in B^N_k) \\
&= (N-1)!^2 \exp\left(1 - \frac{2N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N} \times \left[\frac{N^2}{e_N}\exp\big(4(N/e_N - 1/N)\big)\right] \\
&\times \sum_{k=0}^{\infty} \frac{1}{(k-1)!}\left[\frac{N^2}{e_N}\exp\big(4(N/e_N - 1/N)\big)\right]^k \times (1+o(1)) \times \Big(1 + O(N^{-1/4-\epsilon}) + O(N^{-2\epsilon})\Big) \\
&+ \sum_{k=k_1+1}^{N} |B^{N-1}_{k-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma) \in B^N_k) \\
&= (N-1)!^2 \exp\left(1 - \frac{2N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N} \times \left[\frac{N^2}{e_N}\exp\big(4(N/e_N - 1/N)\big)\right] \\
&\times \exp\left[\frac{N^2}{e_N}\exp\big(4(N/e_N - 1/N)\big)\right] \times (1+o(1)) \\
&+ \sum_{k=k_1+1}^{N} |B^{N-1}_{k-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma) \in B^N_k) \\
&= (N-1)!^2 \exp\left(1 - \frac{N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N-1} \times (1+o(1)) \\
&+ \sum_{k=k_1+1}^{N} |B^{N-1}_{k-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma) \in B^N_k)
\end{aligned}
$$

noting that $|B^N_k| \leq \frac{N!^2}{k!}$ (from equation 1.8 in O'Neil (1970)), we have:

$$
\begin{aligned}
\sum_{k=k_1+1}^{N} |B^{N-1}_{k-1}|\mathbb{P}(x_\sigma x_\pi = 1|(\pi,\sigma) \in B^N_k) &\leq \sum_{k=k_1+1}^{N} \frac{(N-1)!^2}{(k-1)!}\left(\frac{e_N}{N^2}\right)^{2N-k} \\
&= (N-1)!^2 \left(\frac{e_N}{N^2}\right)^{2N-1} \times O(N^{-(1/8)N^{5/8}})
\end{aligned}
$$

finally:

$$
E(C^2_{S_N,ij}) \sim (N-1)!^2 \exp\left(1 - \frac{N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N-1}
$$

Given the asymptotic results in (10), I can now show that:

$$\frac{|S_N|}{E(|S_N|)} \to_p 1 \tag{11}$$

indeed, for any $\epsilon > 0$:

$$\begin{aligned}
\mathbb{P}\left(|\frac{|S_N|}{E(|S_N|)} - 1| > \epsilon\right) &= \mathbb{P}\left(\frac{||S_N| - E(|S_N|)|}{E(|S_N|)} > \epsilon\right) \\
&\leq \frac{Var(|S_N|)}{\epsilon^2 E(|S_N|)^2} \\
&= \frac{E(|S_N|^2) - E(|S_N|)^2}{\epsilon^2 E(|S_N|)^2} \\
&\to 0
\end{aligned}$$

where the inequality is Markov's and where the limit is obtained thanks to the equation (10).

Observe that:

$$\begin{aligned}
\frac{E(\sum_{ij} C^2_{S_N,ij})}{NE(|S_N|^2)} &\sim \frac{N^2 \times (N-1)!^2 \exp\left(1 - \frac{N^2}{e_N}\right)\left(\frac{e_N}{N^2}\right)^{2N-1}}{N \times N!^2 \left(\frac{e_N}{N^2}\right)^{2N} \exp\left(1 - \frac{N^2}{e_N}\right)} \\
&= \frac{N}{e_N} \\
&\to 0
\end{aligned}$$

therefore:

$$\frac{\sum_{ij} C^2_{S_N,ij}}{NE(|S_N|^2)} \to_p 0$$

then the equation (11) gives:

$$\frac{\sum_{ij} C^2_{S_N,ij}}{N|S_N|^2} \to_p 0$$

as desired. □