GENERAL COVARIANCE-BASED CONDITIONS FOR CENTRAL LIMIT THEOREMS WITH DEPENDENT TRIANGULAR ARRAYS

ARUN G. CHANDRASEKHAR[‡], MATTHEW O. JACKSON[‡,⋆], TYLER H. MCCORMICK[⋄,§], AND VYDHOURIE THIYAGESWARAN[⋄]

Abstract. We present a general central limit theorem with simple, easy-to-check covariance-based sufficient conditions for triangular arrays of random vectors when all variables could potentially be interdependent. The result is constructed from Stein's method, but the bounds are distinct from related work. We show that these covariance conditions nest standard assumptions studied in the literature such as M-dependence, mixing random fields, non-mixing autoregressive processes, and dependency graphs, which themselves need not necessarily imply each other. This permits researchers to work with high-level but intuitive conditions based on overall correlation instead of more complicated and restrictive conditions such as strong mixing in random fields that may not have any obvious micro-foundation. As examples of the implications, the theorem implies asymptotic normality in estimating treatment effects with spillovers in more settings than previously admitted, as well as processes with global dependencies such as epidemic spread and information diffusion.

1. Introduction

In many contexts researchers have a set of random vectors that exhibit interdependence and use the data to develop estimators and therefore, need to study whether they are asymptotically normal. Often, dependency is modeled in idiosyncratic ways, with perhaps unintuitive if not unappealing conditions to describe otherwise straightforward assumptions on correlation. It is not clear, for instance, if the α- or ϕ-mixing random field structures aptly captures rainfall shocks in an empirical analysis, though the researcher is willing to assume that there is decaying covariance in some sense. Further, in settings such as treatment effects with spillovers, epidemic spread, information diffusion, general equilibrium effects, and so on, the correlation can be non-zero for any finite set of observations across all random vectors of interest, which is not allowed for in the literature.

To address these concerns, we present simple covariance-based sufficient conditions for a central limit theorem to be applied to a triangular array of dependent random vectors, a special case of which was first developed in a companion paper Chandrasekhar and Jackson (2021). Specifically, we use Stein's method to write three high-level, easy-to-interpret conditions Stein (1986). In this paper, we prove that our conditions are implied by (but

Date: Revision: May 24, 2023.
We thank Paul Goldsmith-Pinkham, Stéphane Bonhomme, and Samuel Higbee for helpful comments.
[‡]Stanford, Department of Economics.
[⋆]Santa Fe Institute.
[⋄]University of Washington, Department of Statistics.
[§]University of Washington, Department of Sociology.

do not imply) the assumptions used in a wide array of specific dependence models such as M-dependence, mixing random fields, non-mixing autoregressive processes, and dependency graphs, which themselves do not nest each other. Unlike the proofs for asymptotic normality in many of the aforementioned cases, the covariance-based arguments are compact.

It is useful to contextualize the literature on central limit theorems for dependent data before returning to the contributions. There are several approaches to establishing sufficient conditions for central limit theorems. The first can be thought of as Lindeberg approaches, the second more generally characteristic function based approaches, and the third—our focus here—Stein method approaches.

The Stein method observes that

$$\mathrm{E}[Yf(Y)] = \mathrm{E}[f'(Y)]$$

for all continuously differentiable functions if and only if Y has a standard normal distribution. So, when by considering normalized sums taking the role of Y, it is enough to show that this equality holds for all such functions asymptotically.

This method has been utilized in various forms. For instance Bolthausen (1982) used this construction in establishing conditions for dependence in time series data; namely, the author establishes how the probability of a joint set of events differ from treating them as independent decays in temporal distance. So as the mixing coefficients decay fast enough in distance, the proof proceeds by checking that the Stein argument follows leveraging the mixing structure. Of course, this argument in time can be generalized to space, and, further, to random fields. That is, random variables carry indices in a Euclidean space and an analogous mixing condition is made, with decay based on Euclidean distance. Together with higher moment conditions, a Stein-based argument can be shown. Again, a literature (e.g., Jenish and Prucha (2009)) developed and refined such conditions such as using near epoch dependence.

A peculiar consequence of this when taken to applied literature is that it organized itself, at least in part, around a number of such conditions despite not having any obvious microfoundations for the assumed dependence structure. For example, spatial standard errors are often used, as in Conley (1999) when conducting Z-estimation (or GMM). However, in actual applications, for instance agricultural shocks such as rainfall or pests or soil, it is not clear that they should follow a specific form of interdependence satisfy ϕ-mixing with a certain decay rate as invoked in Conley (1999). Surely these shocks correlate over space, but it is hard to say much beyond that. To take a different example, models of network formation often orient themselves by embedding nodes in a random field to deliver central limit theorems. However, this is not without consequence as it forces a specific, and sometimes undesirable, pattern of link formation. For example, a certain clustering pattern or whether and when tree-like structures can be generated in large graphs can be restricted with such modeling techniques (Hoff et al., 2002; Lubold et al., 2020). If nodes i and j have a distance that inversely relates to the probability of a link between them forming, by the triangle inequality, the distances between i and k and i and j determine possible distances between j and k and therefore the probabilities for the jk link forming. But that pattern then emerges due to the spatial embedding rather than a fundamental modeling of human behavior. So in both of these cases, it is less restrictive and more appropriate to impose high level conditions on correlations if asymptotic normality still follows.

In contrast to embedding the index of variables in a metric space and toggling dependency by distance, a literature on dependency graphs emerged (Baldi and Rinott, 1989; Goldstein and Rinott, 1996; Ross, 2011). There, observations have indices in a graph where those that are not edge-adjacent are independent. This provides a different strategy to apply Stein's method, by creating for each observation a dependency neighborhood. Sufficient sparsity in the graph structure allows for a central limit theorem to apply, despite not forcing time-like nor space-like structure. Examples include dependency graphs (Ross, 2011; Goldstein and Rinott, 1996; Chen and Shao, 2004) and more general treatment in Chen and Shao (2004).

Both cases of embedding indices in a metric space or using a more unstructured dependency structure are similar in the sense that they constrain the total amount of correlation between the n random variables. In principle there are $\binom{n}{2}$-order components to this sum, but be it mixing conditions or sparsity conditions, they force this total sum to be order n. All strategies, therefore, present some attempt to reduce the size of the covariance sum.

We consider a triangular array of n random vectors $X^n_{1:n}$, which are neither necessarily independent nor identically distributed. We study when their appropriately normalized sample mean is asymptotically normally disturbed. In principle at any n, all X^n_i and X^n_j can be correlated. The proof follows the well-known Stein's method, though we develop and apply specific bounds for our purpose. To apply Stein's method we must first associate each random variable with a set of other random variables with whom it carries higher level of correlation, still keeping in mind that it could in principle have some non-zero correlation with all variables. We call these *affinity sets*, denoted $\mathcal{A}^n_{(i,d)}$ in dimension d.[1] We provide sufficient conditions for asymptotic normality in terms of the total amount of covariance within an affinity set and the total amount of covariance across affinity sets. as long as in the limit, the amount of interdependence in the overall mean is coming from the covariances within affinity sets, asymptotic normality follows. We work with weaker conditions based on bounds on sums of covariances, differently from conditions in the previous literature.[2]

Our contributions are several-fold. First we do not require a sparse dependency structure in finite sample at any n. That is, there can be non-zero correlation between any pair

[1]We note that we use the term *affinity set* rather than "dependency neighborhood" to emphasize the possibility of high and low non-zero covariance structures with arbitrary groupings that satisfy summation conditions. Thus, for the sake of convenience and precision, we avoid using the terms "high dependency neighborhood", and "low dependency neighborhood".

[2]In Arratia et al. (1989), the authors present Chen's method Chen (1975) for Poisson approximation rather than normality, which has a similar approach to ours in collecting random variables into dependencies. While this results in nice finite sample bounds, these bounds consist of almost three separate pieces, making it less friendly to understanding these bounds together with growing sums of covariance of the samples. In fact, all five of the examples studied in Arratia et al. (1989) deal only with cases where at least one of these pieces being identically zero in the cases where Chen's method succeeds. Many examples require no exact zeros, which our approach focuses on.

X_i, X_j. Much of the dependency graph literature leverages the independence structure in constructing their bounds and, therefore, the bounds we build are different.[3]

Second, because of this possibility of non-zero covariance across all random vectors, we choose to organize our bounds through covariance conditions. We are reminded of a discussion in Chen (1975) in the context of Poisson approximation. Covariance conditions are easy-to-interpret and check, and from an applied perspective often easier to justify from a microfoundation.

Third, the result is for random vectors and while the application of the Cramér-Wold device is simple in our setting—by the nature of how indexing works—it is useful to have and instructive for a practitioner.

Fourth, our setup nests many of the previous literature's examples, most of which do not nest each other. We illustrate the utility of our central limit theorem through several distinct applications. We show that it is easy to prove asymptotic normality for M-dependent processes, non-mixing auto regressive processes, random fields, and dependency graphs. In each, it is straightforward to examine the covariance structure of the random variables in question and check that our sufficient conditions are met.

Fifth, we show that our generalizations permit a wider and more practical set of analyses that were otherwise ruled out or limited in the literature. This includes treatment effects with spillover models and things like epidemic and diffusion models. Specifically, we extend the treatment effects with spillovers analysis, as in Aronow and Samii (2017), to allow every individual's exposure to treatment to possibly be increasing in every other node's treatment assignment, and nonetheless the relevant estimator is still asymptotically normally distributed. Of course this case, which is ubiquitous in practice, is assumed away in applied work because conventional central limit theorems do not cover such a case.

The other applied example we provide concerns a sub-critical epidemic process with a number of periods longer than the graph's diameter. So, whether an individual is infected is correlated with the infection status of any other individual (assuming a connected, unweighted graph). Again this practical situation is excluded by the previous central limit theorems in the literature.

2. The Theorem

We consider a triangular array of n random variables $X^n_{1:n} \in \mathbb{R}^p$, with entries $X^n_{i,d}$ and $d \in \{1, \ldots, p\}$, such that each have finite variance (possibly varying with n). We let $Z^n_i \in \mathbb{R}^p$ denote the de-meaned variable, $Z^n_i = X^n_i - \mathrm{E}\left[X^n_i\right]$. The sum, $S^n \in \mathbb{R}^p$, is given by $S^n := \sum_{i=1}^n Z^n_i$. We suppress the dependency on n for clarity, e.g., writing $X_{i,d}$, unless otherwise needed.

[3]In Chatterjee (2008), the author uses a decoupling approach to break apart the dependency into various independent components. This gives a slightly different perspective on the problem, treating any condition on the variance term as a separate problem. The condition one would need to satisfy in this setting is to have a bounded sum of third moments of the coordinate-wise derivative of the function of interest. Our result considers variance from the beginning as we deal with affinity sets, and as a result leads to conditions of a more approachable form for more applied researchers.

2.1. **Affinity Set.** Every real-valued random variable $X_{i,d}$ has an *affinity set*, denoted $\mathcal{A}^n_{(i,d)}$, which can depend on n. We require $(i,d) \in \mathcal{A}^n_{(i,d)}$.

Heuristically, $\mathcal{A}^n_{(i,d)}$ includes the indices j, d' if the covariance between $X_{j,d'}$ and $X_{i,d}$ is relatively high in magnitude, but not if the covariance is low. Note that there is no independence requirement at any n and, in fact, our sufficient conditions for the central limit theorem involve the total mass of covariances both within and outside of $\mathcal{A}^n_{(i,d)}$ to satisfy appropriate conditions. Given these requirements, the precise construction of affinity sets is rather flexible.

2.2. **The Central Limit Theorem.** We present a multidimensional central limit theorem. Let Ω_n be a $p \times p$ matrix which houses the bulk of covariance across observations and dimensions in a manner described below. It has entries

$$\Omega_{n,dd'} := \sum_{i=1}^{n} \sum_{(j,d') \in \mathcal{A}^n_{(i,d)}} \operatorname{cov}\left(Z_{i,d}, Z_{j,d'}\right),$$

of the total sum of variance-covariances across all the pairs of variables and dimensions in each other's affinity sets.[4] In what follows, we maintain the assumption that $\|\Omega_n\|_F \to \infty$, where $\|\cdot\|_F$ is the Frobenius norm.

Our first assumption is that the the total mass of the variance-covariance is not driven by the covariance between members of a given affinity set neither of which are the reference random variables themselves. That is, given reference variable $X_{i,d}$, the covariance of some $X_{j,d'}$ and $X_{k,d''}$ where both are in the reference variable's affinity set is small in total across all such triples of variables.

Assumption **1** (Bound on total weighted-covariance within affinity set)**.**

$$\sum_{(i,d);(j,d'),(k,d'') \in \mathcal{A}^n_{(i,d)}} \mathrm{E}\left[|Z_{i,d}| Z_{j,d'} Z_{k,d''}\right] = o\left(\left(\|\Omega_n\|_F\right)^{3/2}\right).$$

The second assumption is that the total mass of the variance-covariance is not driven by random variables across affinity sets relative to two distinct reference variables. That is, given two random variables $X_{i,d}$ and $X_{j,d'}$, the aggregate amount of weighted covariance between two other random variables (each within one of the reference variables' affinity sets) is small.

Assumption **2** (Bound on total weighted-covariance across affinity sets)**.**

$$\sum_{(i,d),(j,d');(k,d'') \in \mathcal{A}^n_{(i,d)},(l,\hat{d}) \in \mathcal{A}^n_{(j,d')}} \operatorname{cov}\left(Z_{i,d} Z_{k,d''}, Z_{j,d'} Z_{l',\hat{d}}\right) = o\left(\left(\|\Omega_n\|_F\right)^{2}\right),$$

The third assumption is that the total mass of variance-covariance is not driven by reference random variables and the variables outside of its affinity sets.

[4] Notice that this is distinct from a total variance-covariance matrix $\Sigma_{n,dd'} := \sum_{i=1}^{n} \sum_{j=1}^{n} \operatorname{cov}\left(Z_{i,d}, Z_{j,d'}\right)$, which includes terms outside of $\mathcal{A}^n_{(i,d)}$.

ASSUMPTION **3** (Bound on total weighted-covariance across affinity sets)**.**

$$\sum_{(i,d);(j,d')\notin\mathcal{A}^n_{(i,d)}} \mathrm{E}\left(Z_{i,d}Z_{j,d'}\cdot \mathrm{sign}\left(\mathrm{E}[Z_{i,d}Z_{j,d'}|Z_{j,d'}]\right)\right) = o\left(\|\Omega_n\|_F\right).$$

With these three assumptions established, a central limit theorem follows.

THEOREM **1.** *If (1)-(3) are satisfied, then* $\Omega_n^{-1/2}S^n \rightsquigarrow \mathcal{N}(0, I_{p\times p})$.

The proof is provided in Appendix A. The argument follows by applying the Cramér-Wold device to the arguments following Stein's method in as Chandrasekhar and Jackson (2021) for the univariate case. Specifically, since the Cramér-Wold device requires for all $c\in\mathbb{R}^p$ fixed in n that the c-weighted sum satisfies a central limit theorem (Biscio et al., 2018) — i.e., $(c'\Omega_n c)^{-1/2}c'S^n \rightsquigarrow \mathcal{N}(0,1)$ — we can consider a problem of np random variables with affinity sets. Then, by checking Assumptions 1-3 for the case of $c = 1_p$ the result follows.

An important special case is $\mathcal{A}^n_{(i,d)} = \{(i,d)\}$ and when there is positive correlation throughout in a manner described below. It nests a variety of cases in practice. We provide an illustration in Example 2.

COROLLARY **1.** *If* $\mathcal{A}^n_{(i,d)} = \{(i,d)\}$, $\mathrm{E}[Z_{i,d}Z_{j,d'}|Z_{j,d'}]\geq 0$ *for every* $(j,d')\neq(i,d)$*, and*[5]

(i) $\sum_{(i,d),(j,d')}\mathrm{cov}(Z^2_{i,d}, Z^2_{j,d'}) = o\left((\|\Omega_n\|_F)^2\right)$*, and*
(ii) $\sum_{(i,d)\neq(j,d')}\mathrm{cov}(Z_{i,d}, Z_{j,d'}) = o\left(\|\Omega_n\|_F\right)$,

then $\Omega_n^{-1/2}S^n \rightsquigarrow \mathcal{N}(0, I_{p\times p})$.

It is also useful to note that, for instance if we take $p = 1$ and the X_is are Bernoulli random variables with $\mathrm{E}[X_i]\to 0$ (uniformly), then condition (ii) implies condition (i) (Chandrasekhar and Jackson, 2021).

3. APPLICATIONS

We present four distinct examples from the literature that focus on proving asymptotic normality: (i) M-dependence, (ii) non-mixing autoregressive processes, (iii) mixing random fields, and (iv) dependency graphs These examples do not necessarily nest each other, as will become clear. In each case we provide a sketch of the core assumptions made in the relevant papers to be self-contained for the reader. We both show how these assumptions imply our covariance restrictions and also the relative complexity of these setups.

3.1. M-dependence.

[5]If $\mathrm{E}[Z_{i,d}Z_{j,d'}|Z_{j,d'}]\geq 0$ does not hold, then (ii) can just be substituted by Assumption 3.

3.1.1. *Environment.* We take the example of Theorem 2.1 of Romano and Wolf (2000). In this example, we have real-valued time series data so $p = 1$ (so we can drop the index d) and Ω_n is a scalar and Z_i and Z_j are independent if $|i-j| > M$. For the sake of convenience of the reader, we include the assumptions made in their paper: for some $\delta > 0$, and $-1 \leq \gamma < 1$,

(1) $\mathrm{E}|X_{n,i}|^{2+\delta} \leq \Delta_n$ for all i
(2) $\mathrm{var}\left(\sum_{i=a}^{a+k-1} X_{n,i}\right) k^{-1-\gamma} \leq K_n$ for all a and $k \geq m$
(3) $\mathrm{var}\left(\sum_{i=1}^{r} X_{n,i}\right) r^{-1} m^{-\gamma} \geq L_n$
(4) $K_n = O\left(L_n\right)$
(5) $\Delta_n = O\left({L_n}^{1+\delta/2}\right)$
(6) $m^{1+(1-\gamma)(1+2/\delta)} = o(r)$

3.1.2. *Application of Theorem 1.* We consider the M-ball, ($\mathcal{A}_i^n = \{j : |j-i| \leq M\}$. In this case, for all j with $|i-j| > M$, we have that the covariance $\mathrm{cov}(Z_i, Z_j) = 0$ by independence, so Assumption 3 is satisfied. Under bounded third and fourth moments, we check the remaining assumptions. Assumption 1 is easily verified:

$$\sum_{i;j,k \in \mathcal{A}_i^n} \mathrm{E}[|Z_i| Z_j Z_k] = O(\sum_i \mathrm{E}[|Z_i|^3]) = o\left(\Omega_n^{3/2}\right),$$

following their assumption (6).

Our Assumption 2 is satisfied similarly following their assumption(6):

$$\begin{aligned}\sum_{i,j;k \in \mathcal{A}_i^n, l \in \mathcal{A}_j^n} \mathrm{cov}(Z_i Z_k, Z_j Z_l) &= O(\sum_{\substack{i,j:|i-j|\leq M,\\ k:|k-i|\leq M,\\ l:|l-i|\leq 2M}} \mathrm{var}(Z_i^2)) \\ &= O(n \cdot M^3 \cdot \mathrm{var}(Z_i^2)) = o(\Omega_n^2).\end{aligned}$$

This is from the fact that if they are not within that distance then automatically it is impossible for the Z_k and Z_l to induce any correlation as well.

Note that our conditions generate bounded fourth moment requirements, which is not necessarily a condition invoked in every analysis of M-dependent processes in the literature, which sometimes have slightly lower moment requirements. In all other examples in the paper, we need not add any moment assumptions to the examples. Nonetheless, our results are not intended to provide the tightest bounds, but rather general conditions, spanning various types of dependence, that are easily checkable for most applied settings.

3.2. **Andrews' Non-Mixing Autoregressive Processes.**

3.2.1. *Environment.* This example is from Andrews (1984). The goal of that paper is to present interdependence in a time series, but one that does not satisfy strong (α-) mixing in order to clarify the distinction between dependence and mixing. Again, $p = 1$. The setting is where ϵ_t is Bernoulli(q) and $X_t = \sum_{l=0}^{\infty} \rho^l \epsilon_{t-l}$, where $\rho \in (0, 1/2]$. We show that, for a constant C depending only on ρ, asymptotic normality of the normalized, mean zero Z_t: $\frac{1}{\sqrt{Cnq(1-q)}} \sum_t Z_t \rightsquigarrow \mathcal{N}(0,1)$.

Assume, without loss of generality, that $s > t$, so

$$\text{cov}(Z_t, Z_s) = \text{cov}(\sum_{l=0}^{\infty} \rho^l \epsilon_{t-l}, \sum_{k=0}^{\infty} \rho^k \epsilon_{s-k}) = \rho^M \sum_{l=0}^{\infty} \rho^{2l} \text{var}(\epsilon_{t-l}) = c \cdot q(1-q)\rho^M$$

where $M = s - t$, and c is a constant.

3.2.2. *Application of Theorem 1.* We can take $\mathcal{A}_t^n = \{t\}$ and apply Corollary 1. Here $\Omega_n = Cnq(1-q)$ for some constant C that depends only on the choice of ρ. As M grows, $\text{cov}(Z_t, Z_{t+M}) \to 0$ as $M \to \infty$.

It is easy to see that $\mathrm{E}[Z_t Z_s | Z_s] \geq 0$ for all $s \neq t$. Condition (ii) follows immediately since $\sum_{t \neq s} \text{cov}(Z_t, Z_s) = \sum_{t \neq s} c \cdot q(1-q)\rho^{|t-s|} < nc \cdot q(1-q)$ where c is a constant.

We check Condition (i)

$$\begin{aligned}
\sum_{t \neq s} \text{cov}(Z_t^2, Z_s^2) \leq & \sum_{t \neq s} \{\rho^{4|t-s|} \text{var}((\sum_{l=0}^{\infty} \rho^l \epsilon_{t \wedge s - l})^2) + 2\rho^{|t-s|} \text{cov}((\sum_{l=0}^{\infty} \rho^l \epsilon_{t \wedge s - l})^2, (\sum_{k=0}^{|t-s|-1} \rho^k \epsilon_{t \vee s - k}) \cdot (\sum_{l=0}^{\infty} \rho^l \epsilon_{t \wedge s - l})) \\
& + 4c^2 \cdot q^2 (1-q) \rho^{|t-s|}\} \\
\leq & \sum_{t \neq s} \{\rho^{4|t-s|} \text{var}((\sum_{l=0}^{\infty} \rho^l \epsilon_{t \wedge s - l})^2) + 2C\rho^{|t-s|} \text{var}((\sum_{l=0}^{\infty} \rho^l \epsilon_{t \wedge s - l})^{3/2}) + 4c^2 \cdot q^2 (1-q) \rho^{|t-s|}\} \\
= & \, o((\Omega_n)^2),
\end{aligned}$$

and the proof is complete.

3.3. **Random Fields.**

3.3.1. *Environment.* The next example looks at random fields. Note that this setting effectively nests many time series and spatial mixing models as special cases. Specifically, we take the setting of Jenish and Prucha (2009), Theorem 1. They have a setting with either ϕ- or α- mixing in random fields allowing for non-stationarity and asymptotically unbounded second moments. They treat real mean-zero random field arrays $\{Z_{i,n}; i \in D_n \subseteq \mathbb{R}^d, n \in \mathbb{N}\}$, where each pair of elements i, j have some minimum distance $\rho(i,j) > 0$, where $\rho(i,j) := \max_{1 \leq l \leq d} |i_l - j_l|$, between them. At each point on the lattice, there is a real-valued random variable drawn, so $p = 1$. For their central limit theorem results, the authors assume (see their Assumptions 2 and 5 restated below for convenience) a version of uniform integrability that allows for asymptotically unbounded second moments, while maintaining that no single variance summand dominates by scaling $X_{i,n} := Z_{i,n} / \max_{i \in D_n} c_{i,n}$ so that $X_{i,n}$ is uniformly integrable in L_2. The authors also assume (see their Assumption 3 restated below) some conditions on the inverse function α_{inv} on mixing coefficients α (and analogously for ϕ, see their Assumption 4) together with the tail quantile functions $Q_{i,n}$ (where $Q_X(u) := \inf\{x : F_X(x) \geq 1 - u\}$ where F_X is the cumulative distribution function for the random variable X), essentially requiring nice trade-off conditions between the two, such that under α-mixing decaying at a rate $O(\rho^{d+\delta})$ for some $\delta > 0$, $\sum_{m=1}^{\infty} m^{d-1} \sup_n \alpha_{k,l,n}(\rho) < \infty$ for all $k + l \leq 4$, and $\sup_n \sup_{i \in D_n} \int_0^1 \alpha_{inv}^d(u) Q_{i,n}(u) du$ tends to zero in the limit of upper quantiles. Again, for the sake of convenience of the reader, we include the assumptions made in their paper:

(1) Assumption 2: $\lim_{k\to\infty}\sup_n \sup_{i\in D_n} \mathrm{E}[|Z_{i,n}/c_{i,n}|^2\mathbf{1}\{|Z_{i,n}/c_{i,n}|>k\}] = 0$ for $c_{i,n}\in\mathbb{R}^+$
(2) Assumption 3: The following conditions must be satisfied by the α-mixing coefficients:
   (a) $\lim_{k\to\infty}\sup_n\sup_{i\in D_n}\int_0^1 \alpha_{inv}^d(u)\left(Q_{|Z_{i,n}/c_{i,n}|\mathbf{1}\{Z_{i,n}/c_{i,n}>k\}}\right)^2 du = 0$
   (b) $\sum_{m=1}^{\infty} m^{d-1}\sup_n \alpha_{k,l,n}(r) < \infty$ for $k+l\le 4$ where
$$\alpha_{k,l,n}(r) = \sup(\alpha_n(U,V), |U|\le k, |V|\le l, \rho(U,V)\ge r)$$
   (c) $\sup_n \alpha_{1,\infty,n}(m) = O\left(m^{-d-\epsilon}\right)$
(3) Assumption 5: $\liminf_{n\to\infty}|D_n|^{-1}M_n{}^{-2}\sigma_n^2 > 0$

3.3.2. *Application of Theorem 1.* In the following, we assume that $Z_{i,n}$s have bounded second moments (otherwise, we can replace them with their scaled versions (see above), and the results should go through under bounded third and fourth moments). Here, for any $\epsilon>0$, we can take $\mathcal{A}_i^n = \{j : \rho(i,j)\le K(\epsilon)\}$, and K is a non-increasing function. That is, we pick $K(\epsilon)$ to be large enough, and this can be decided by understanding the cumulative distribution function of the random variables. From the first part of Lemma B.1 (see also (B.10)) in Jenish and Prucha (2009), together with their assumptions 3,4 (under $\phi-$mixing), 5 and Rio's covariance inequality, we know that for any $\epsilon>0$, there exists infinitely many $k\neq i$ such that $\rho(i,k)\ge K(\epsilon)$, such that
$$|\mathrm{cov}(Z_{i,n},Z_{k,n})| \le 4\int_0^{\overline{\alpha}_{1,1}(K(\epsilon))} Q_{i,n}(u)Q_{k,n}(u)du \le \epsilon.$$

We first restate Lemma B.1 in their paper for the sake of convenience of the reader:

Lemma **1** (Lemma B.1 Jenish and Prucha (2009), Bradley (2007))**.** *Let $\alpha(m)$, $m=1,2,\ldots$ be a non-increasing sequence such that $0\le\alpha(m)\le 1$, and $\alpha(m)\to 0$ as $m\to\infty$. Set $\alpha(0)=1$, and define $\alpha^{-1}(u):(0,1)\to\mathbb{N}\cup\{0\}$ such that*
$$\alpha^{-1}(u) = \max\{m\ge 0 : \alpha(m) > u\}$$
for $u\in(0,1)$.

Let $f:(0,1)\to[0,\infty)$ be a Borel function, then $q\ge 1$:
(1) $\sum_{m=1}^{\infty} m^{q-1}\int_0^{\alpha(m)} f(u)du \le \int_0^1[\alpha^{-1}(u)]^q f(u)du$

(2) $\int_0^1[\alpha^{-1}(u)]^q du \le q\sum_{m-1}^{\infty}\alpha(m)m^{q-1}$*, for any $q\ge 1$.*

Now we restate result B.10 in their paper:

$$\sup_n \sup_{i\in D_n}\int_0^1 \alpha_{inv}^d Q_{i,n}^2(u)du = K_1 < \infty \tag{3.1}$$

where $\alpha_{inv}(u) := \max\{m\ge 0 : \sup_n \alpha_{1,1,n}(m) > u\}$ for $u\in(0,1)$.

Therefore, we can pick ϵ arbitrarily small so that
$$4\int_0^{\overline{\alpha}_{1,1}(K(\epsilon))} Q_{i,n}(u)Q_{k,n}(u)du \le \epsilon \ll \int_0^1 (Q_{i,n})^2(u)du = \mathrm{var}(Z_{i,n}).$$

Now, we verify that our key conditions are satisfied in this setting. We write $K := K(\epsilon)$, and take ϵ to be small. First, we check Condition 1:

$$\sum_{i;j,k\in\mathcal{A}_i^n} \mathrm{E}[|Z_{i,n}|Z_{j,n}Z_{k,n}] = O(\sum_i \mathrm{E}[|Z_{i,n}|^3]) = o\left(\Omega_n^{3/2}\right),$$

where the last equality follows from their Assumption 3.

Next, we check that Condition 2 is satisfied:

$$\begin{aligned}\sum_{i,j;k\in\mathcal{A}_i^n,l\in\mathcal{A}_j^n} \mathrm{cov}(Z_{i,n}Z_{k,n}, Z_{j,n}Z_{l,n}) &= O(\sum_{\substack{i,j:|i-j|\le K,\\ k:|k-i|\le K,\\ l:|l-i|\le 2K}} \mathrm{var}(Z_{i,n}^2))\\ &= O(n\cdot K^3\cdot \mathrm{var}(Z_{i,n}^2)) = o(\Omega_n^2).\end{aligned}$$

Finally, we check Condition 3:

$$\sum_{i;j\notin\mathcal{A}_i^n} \mathrm{E}(Z_{i,n}Z_{j,n}\cdot \mathrm{sign}\ (\mathrm{E}[Z_{i,n}Z_{j,n}|Z_{j,n}])) = O(\epsilon) = o\left(\Omega_n\right)$$

by construction.

We note that our conditions also provide consistent estimators for covariance matrices of moment conditions for parameters of interest in the GMM setting, under full-rank conditions of expected derivatives Conley (1999), since the author uses the CLT from Bolthausen (1982) under stationary random fields which is generalized in the setting above. Leveraging this, we can treat more general metrics as well. In Conley and Topa (2002), the authors develop consistent covariance estimators, using these conditions, combining different "distance" metrics including physical distance as well as ethnicity (or occupation, for another example) distance in L_2 at an aggregate level (using census tracts data).

3.4. **Dependency Graphs and Chen and Shao (2004).** Next we turn to dependency graphs. There is an undirected, unweighted graph G with dependency neighborhoods $N_i := \{j:\ G_{ij}=1\}$ such that Z_i is independent of all Z_j for $j\notin N_i$ (Baldi and Rinott, 1989; Chen and Shao, 2004; Ross, 2011). Let $\mathcal{A}_i^n = \{j:\ G_{ij}=1\}$. Denote the maximum cardinality of these to be D_n. In Ross (2011) (see Theorem 3.6), together with a bounded fourth moment assumption, we see that the conditions there imply the conditions here. Indeed, we see that

$$\sum_{i;j,k\in\mathcal{A}_i^n} \mathrm{E}\left[|Z_i|Z_jZ_k\right] \le D^2\sum_{i=1}^n \mathrm{E}\left[Z_i^3\right],$$

and for $\Omega_n^{-1/2}S_n \rightsquigarrow N(0,1)$, we need $D^2\sum_{i=1}^n \mathrm{E}\left[Z_i^3\right] \le o\left(\Omega_n^{3/2}\right)$ in Ross (2011) (Theorem 3.6), and hence Assumption 1 is satisfied. Similarly,

$$\sum_{i,i';j\in\mathcal{A}(i,n),j'\in\mathcal{A}_i^n} \mathrm{cov}\left(Z_iZ_j, Z_{i'}Z_{j'}\right) = O(D^3\sum_{i=1}^n \mathrm{E}[Z_i^4]),$$

and for $\Omega_n^{-1/2}S_n \rightsquigarrow N(0,1)$, we need $D^3\sum_{i=1}^n \mathrm{E}[Z_i^4] \le o\left((\Omega_n)^2\right)$ in Ross (2011) (Theorem 3.6), and hence Assumption 2 is satisfied. Condition 3 holds trivially, by definition of the dependency neighborhoods.

Now, we consider Chen and Shao (2004). In particular, we consider their weakest assumption, LD1: Given index set $\mathcal{I}$, for any $i \in \mathcal{I}$, there exists an $A_i \in \mathcal{I}$ such that X_i and $X_{A_i^C}$ are independent. The affinity sets can be defined by the complement of the independence sets. So $\mathcal{A}_i^n = \{j : Z_j \text{ is not independent of } Z_i\}$, which is similar to the dependency graphs setting. The goals of their paper are different. They develop finite-sample Berry-Esseen bounds, with bounded p-th moments, where $2 < p \leq 3$. This is different from our approach. In our paper, we focus on covariance conditions in the asymptotics, and collect relatively more dependent sets along the triangular array.

3.5. **Peer Effects Models.**

3.5.1. *Environment.* We now turn to an example of treatment effects with spillovers. We look at Aronow and Samii (2017) who study a setting in which nodes in a network are assigned treatment status $T_i \in \{0, 1\}$. For now, we consider the i.i.d. case for simplicity. There are spillovers determined by the topology of the network where treatment status within one's network neighborhood may influence one's own outcome.

Their simplification is as follows. Rather than being arbitrary, the exposure function f takes on one of K finite values; i.e., $f(T_i; T_{1:n}, G) \in \{d_1, \ldots, d_K\}$. To see this, let $N_{i,:}$ denote a dummy vector of whether j is in i's neighborhood: let $N_{ij} = 1$ when $G_{ij} = 1$ with the convention $N_{ii} = 0$. With this simplification, an estimand of the average causal effect is of the form

$$\tau(d_k, d_l) = \frac{1}{n} \sum_i y_i(d_k) - \frac{1}{n} \sum y_i(d_l)$$

where d_k and d_l are the induced exposures under the treatment vectors.

The Horvitz and Thompson estimator from Horvitz and Thompson (1952) provides:

$$\hat{\tau}_{HT}(d_k, d_l) = \frac{1}{n} \sum_i \mathbf{1}\{D_i = d_k\} \frac{y_i(d_k)}{\pi_i(d_k)} - \frac{1}{n} \sum \mathbf{1}\{D_i = d_l\} \frac{y_i(d_l)}{\pi_i(d_l)}$$

where $\pi_i(d_k)$ is the probability that that node i receives exposure d_k over all treatments.

The authors of Aronow and Samii (2017) consider an empirical study in their paper where they have a finite set $K = 4$ of (i) only i is treated in their neighborhood ($d_1 = T_i \cdot 1\{T'_{1:n} N_{i,:} = 0\}$), (ii) at least one is treated in i's neighborhood and i is treated ($d_2 = T_i \cdot 1\{T'_{1:n} N_{i,:} > 0\}$), (iii) i is not treated but some member of the neighborhood is ($d_3 = (1 - T_i) \cdot 1\{T'_{1:n} N_{i,:} > 0\}$), and (iv) neither i nor any neighbor is treated ($d_4 = (1 - T_i) \cdot \prod_{j:\ N_{ij}=1} (1 - T_j)$). We will see that our result allows for a more generalized setting.

3.5.2. *Application of Theorem 1.* To obtain consistency and asymptotic normality they have a covariance restriction of local dependence (Condition 5) and apply which uses Chen and Shao (2004) to prove the result. Namely, it is a high level condition on the model developed above (e.g., the exposures determined by the topology of the graph). Specifically, there is a dependency graph H (with entries $H_{ij} \in \{0, 1\}$) with uniformly bounded degree by some integer m independent of n. That is, $\sum_j H_{ij} \leq m$ for every i. This setting is much more restrictive than our conditions. We can work with larger real exposure values, and in settings concentrating the mass of influence in a neighborhood while allowing from spillovers from everywhere. This is important to allow for in centrality-based diffusion models, SIR models, and financial flow networks, since the the spillovers in these settings are less restricted than the sparse dependency graph in their Condition 5.

Indeed, we allow the dependency graph to be a complete graph, as long as the correlations between the nodes in this dependency graph satisfy our Assumptions 1-3. That is, we can handle cases where, for a given treatment assignment, each node has n real exposure conditions for which the exposure conditions of the whole graph can be well approximated by simple functions, where small perturbations to any node in large regions of small correlations may not perturb the outcomes in these regions (i.e., across affinity sets) while perturbations of the same size in any region of larger correlations (i.e., within affinity sets) may effectively cause significant changes in the outcomes in that region. One can think of the "shorter" monotonic regions in the simple function to be over affinity sets, and "longer" monotonic regions to be across different affinity sets. One can think about monotonically non-decreasing functions, for instance, in epidemic spread settings where any increase in the "treatment" cannot decrease the number of infected nodes.

To take an example, let the true exposure of i be given by $e_i(T_{1:n})$. Then consider a case where $e_i(T^+_{1:n}) - e_i(T_{1:n}) \geq 0$, where $T^+_{1:n}$ indicates an increase in any element $j \in [n]$ from $T_{1:n}$. This is a structure that would happen naturally in a setting with diffusion. The potential outcome for i given treatment assignment is assumed to be $y_i(e_i)$.

In practice, for parsimony and ease exposures are often binned. So consider the problem where the 2^n possible exposures $e_i(T_{1:n})$ can be approximated by K well-separated "effective" exposures $\{d_1, d_2, ..., d_K\}$ where $|d_i - d_j| > \delta$ for any $i, j \in \{1, 2, ..., K\}$ and some $\delta > 0$, and for any $r \in \{1, 2, ..., K\}$, $i, j \in \{1, 2, ..., 2^n\}$, we have, $e_i(T_{1:n}), e_j(T_{1:n}) \in d_r$ if and only if $|e_i(T_{1:n}) - e_j(T_{1:n})| < \delta$ and we have $y_i(e_i)$ smooth in its argument for every i.

Then, following the above, the researcher's target estimand is the average causal effect switching between two exposure bins,

$$\tau(d_k, d_l) = \frac{1}{n}\sum_i 1\{e_i(T_{1:n}) \in d_k\} y_i(e_i) - \frac{1}{n}\sum_i 1\{e_i(T_{1:n}) \in d_l\} y_i(e_i).$$

The estimator for this estimand cannot directly be shown to be asymptotically normally distributed using the prior literature. It is ruled out by Condition 5 in Aronow and Samii (2017) which uses Chen and Shao (2004). However, it is straightforward to apply our result.

An example of this is a sub-critical diffusion process with randomly selected set of nodes M_n being assigned some treatment, and every other node is subsequently infected with some probability. The following example in the SIR model setting illustrates this.

3.6. **Diffusion Models.**

3.6.1. *Environment.* We begin by describing the environment before checking the assumptions of our central limit theorem. Consider a setting of a sequence of graphs G_n on which a simple, finite-time SIR diffusion process occurs. A set M_n, of size m_n, of random nodes with treatment indicated $W_i \in \{0, 1\}$, are seeded at $t = 0$ and in period $t = 1$ each infects each of its network neighbors, $\{j : G_{ij,n} = 1\}$ i.i.d. with probability q_n. The seed is henceforth inactive. In period $t = 2$ each of the nodes infected at period $t = 1$ infect each of its network neighbors who were previously never infected i.i.d. with probability q_n. The process continues for T_n periods. Let $X_i^n \in \{0,1\}$ be a binary indicator of whether i was ever infected throughout the process.

Assume that the sequence of SIR models under study, (G_n, q_n, T_n, W_n) with G_n an unweighted and undirected connected graph, $m_n \to \infty$ (with $\alpha_n := m_n/n = o(1)$), $q_n \to 0$, and $T_n \to \infty$ (with $T_n \geq \mathrm{diam}(G_n)$ at each n), are such that the process is sub-critical. Since the number of periods is at least as large as the diameter, it guarantees that for a connected G_n, $\mathrm{cov}(X_i^n, X_j^n) > 0$ for each i, j.

The statistician may be interested in a number of quantities. For instance, the unknown parameter q_n may be of interest. Suppose $\mathrm{E}[\Psi_i(X_i; q_n, W_{1:n})] = 0$ is a (scalar) moment condition satisfied only at the true parameter q_n given known seeding $W_{1:n}$. The Z-estimator (or GMM) derives from the empirical analog, setting $\sum_i \Psi_i(X_i; \hat{q}, W_{1:n}) = 0$.

By a standard expansion argument

$$(\hat{q} - q_n) = -\left\{\sum_i \nabla_q \Psi_i(X_i; \tilde{q}, W_{1:n})\right\}^{-1} \times \sum_i \Psi_i(X_i; q_n, W_{1:n})$$

To study the asymptotic normality of the estimator we need to study

$$\frac{1}{\sqrt{\mathrm{var}\left(\sum_i \Psi_i\right)}} \sum_i \Psi_i(X_i; q_n, W_{1:n})$$

which involves developing affinity sets for each Ψ_i.[6]

In the example below, for simplicity we consider the case where the estimator may directly work with $X_{1:n}^n$. Letting $Z_i = X_i^n - \mathrm{E}(X_i^n)$ be the de-meaned outcome and we want to show that

$$\frac{1}{\sqrt{\sum_i \sum_{j \in \mathcal{A}_i^n} \mathrm{cov}(Z_i^n, Z_j^n)}} \sum_i Z_i^n \rightsquigarrow \mathcal{N}(0,1).$$

Under sub-criticality, we have that a non-zero but vanishing share of nodes are infected from a single seed. Without sub-criticality, obviously most of the graph is enveloped by a single diffusion and no meaningful inference can be made with a single network.

Let us define

$$\mathcal{B}_j^n := \{i : \mathrm{P}(X_i^n = 1 \mid j \in M_n, m_n = 1) > \epsilon_n\}.$$

Then $\mathcal{B}_j^n$ is the set of nodes for which, if j is the only seed, the probability of being infected in the process is at least ϵ_n. As noted above, in a sub-critical process $\left|\mathcal{B}_j^n\right| = o(n)$ for every

[6]This is a general setup. To see a related but different example, consider a diffusion model with re-infection (SIS). Here X_i denotes the number of times i was infected. Then one can use the expected number of infections for each node given the seeded set as moments:

$$\mathrm{E}\left\{\Psi_i(X_i; q_n, W_{1:n})\right\} = \left[\sum_{t=1}^{T} q_n^t G_n^t \cdot W_{1:n}\right]_i .$$

Again, the main difficulty is that it is not clear what the structure of the correlations of $\Psi_i(X_i)$ and $\Psi_j(X_j)$. Here the moments correspond to sums of expected walk counts from the set of seeds to a given node i, which can be complicated and in a connected graph non-zero for every pair of nodes.

j, not necessarily uniformly in j. For simplicity assume that there is a sequence $\epsilon_n \to 0$ such that this holds uniformly (otherwise, one can simply consider sums). Let $\beta_n := \sup_j \left|\mathcal{B}_j^n\right|/n$ which tends to zero, such that $\alpha_n\beta_n = o(n)$.

Next, we assume that the rate at which infections happen within the affinity set is higher than outside of it, and the share of seeds is sufficiently high and affinity sets are large enough but not too large. That is, there exists some $\mathcal{B}_j^{n\prime} \subset \mathcal{B}_j^n$ such that $|\mathcal{B}_j^{n\prime}| = \Theta(|\mathcal{B}_j^n|)$ and $\mathrm{P}(X_i = 1 | j \in M_n, m_n = 1) \geq \gamma_n$ with $\gamma_n/\epsilon_n \to \infty$, such that $\alpha_n^3 = O(\gamma_n)$, and $\beta_n = O(\gamma_n)$.

It is clear that none of the prior examples such as random fields and dependency graphs cover this case since since all X_i^n are correlated. We now show that Theorem 1 applies to this case.

3.6.2. *Application of Theorem 1.* Let us define the affinity sets $\mathcal{A}_i^n = \mathcal{B}_i^n$. Next, consider a random seed k. Let $\mathcal{E}_{i,j,k} := \{a \notin \mathcal{B}_b^n : \ a, b \in \{i,j,k\},\ a \neq b\}$ denote the event that none of the nodes are in each others' affinity sets. It is clear that $\mathrm{P}(\mathcal{E}_{i,j,k}) \to 1$, since $|\mathcal{B}_a^n| = o(n)$ for $a \in \{i,j,k\}$ and that seeds are uniformly randomly chosen.

If we look at an affinity set, it is sufficient to just look at the variance components and check it is a higher order of magnitude

$$\sum_i \mathrm{var}(Z_i^n) = n \times (1 - (1-\beta_n)^{2m_n})\gamma_n^2 = O(n^2 \times \alpha_n\beta_n \times \gamma_n^2).$$

Now to check Assumption 1, we compute:

$$\begin{aligned}\sum_{i;j,k\in\mathcal{A}_i^n} \mathrm{E}[|Z_i|Z_jZ_k] &= \sum_{i;j,k\in\mathcal{A}_i^n, j\notin\mathcal{A}_k^n \text{ or } k\notin\mathcal{A}_j^n} \mathrm{E}[|Z_i|Z_jZ_k] + \sum_{i;j,k\in\mathcal{A}_i^n,\mathcal{A}_j^n,\mathcal{A}_k^n} \mathrm{E}[|Z_i|Z_jZ_k] \\ &\approx n^3\beta_n\epsilon_n^2 + n\alpha_n\beta_n\epsilon_n.\end{aligned}$$

Thus, we have

$$\frac{n^6\alpha_n^3\beta_n^3\gamma_n^6}{n^6\beta_n^3\epsilon_n^6} = \frac{\alpha_n^3\gamma_n^6}{\epsilon_n^6} \to \infty,$$

which is satisfied.

We note that the probability that no seed is in any other affinity set

$$\mathrm{P}(\cap_{k\in M_n}\mathcal{E}_{i,j,k}) = (1-\beta_n)^{2m_n} \approx 1 - 2m_n\beta_n = 1 - 2n \times \alpha_n\beta_n.$$

This puts an intuitive restriction on the number of seeds and percolation size as a function of n.

Next, we verify Assumption 2. We have,

$$\begin{aligned}\sum_{i;j,k\in\mathcal{A}_i^n, r\in\mathcal{A}_k^n} \mathrm{cov}(Z_i^nZ_j^n, Z_k^nZ_r^n) &= O(\sum_{i;j,k\in\mathcal{A}_i^n; r\in A_k^n, \notin A_i^n} \mathrm{cov}(Z_i^nZ_j^nZ_k^nZ_r^n)) \\ &= O(n^2 \times (n-1)(1-\beta_n)\beta_n\epsilon_n^4) \\ &= O(n^3\beta_n\epsilon_n^4).\end{aligned}$$

Therefore, we have

$$\frac{n^4\alpha_n^2\beta_n^2\gamma_n^4}{n^3\beta_n\epsilon_n^4} = n\alpha_n\beta_n(\frac{\gamma_n}{\epsilon_n})^4 \to \infty,$$

which is satisfied. Next, we verify Assumption 3. Given the event $\mathcal{E}_{i,j,k}$, we can bound the conditional covariance

$$\operatorname{cov}(Z_i^n, Z_j^n \mid \mathcal{E}_{i,j,k}) = O(\epsilon_n^2)$$

by bounding the probabilities of two contagions. So then

$$\sum_{i,j:\ j\notin\mathcal{A}_i^n} \sum_{k\in M_n} \operatorname{cov}(Z_i^n, Z_j^n \mid \mathcal{E}_{i,j,k})\mathrm{P}(\mathcal{E}_{i,j,k}) = C \sum_{i,j:\ j\notin\mathcal{A}_i^n} (1-2n\alpha_n\beta_n)\cdot \operatorname{cov}(Z_i^n, Z_j^n \mid \mathcal{E}_{i,j,k})$$
$$\approx (1-\alpha_n)\cdot n \times ((1-\alpha_n)\cdot n - 1)(1-\beta_n) \times (1-2n\alpha_n\beta_n)\cdot \epsilon_n^2$$

for some constant C. Keeping orders we have

$$\sum_{i,j:\ i\notin\mathcal{A}_j^n} \sum_{k\in M_n} \operatorname{cov}(Z_i^n, Z_j^n \mid \mathcal{E}_{i,j,k})\mathrm{P}(\mathcal{E}_{i,j,k}) = O((n\epsilon_n)^2).$$

Therefore, we have,

$$\frac{n^2\gamma_n^2\alpha_n\beta_n}{n^2\epsilon_n^2} = \frac{\gamma_n^2\alpha_n\beta_n}{\epsilon_n^2} \to \infty,$$

which is also satisfied.

4. Discussion

References

Andrews, D. W. (1984): "Non-strong mixing autoregressive processes," *Journal of Applied Probability*, 21, 930–934. 3.2.1

Aronow, P. M. and C. Samii (2017): "Estimating average causal effects under general interference, with application to a social network experiment," . 1, 3.5.1, 3.5.2

Arratia, R., L. Goldstein, and L. Gordon (1989): "Two moments suffice for Poisson approximations: the Chen-Stein method," *The Annals of Probability*, 9–25. 2

Baldi, P. and Y. Rinott (1989): "On normal approximations of distributions in terms of dependency graphs," *The Annals of Probability*, 1646–1650. 1, 3.4

Biscio, C. A. N., A. Poinas, and R. Waagepetersen (2018): "A note on gaps in proofs of central limit theorems," *Statistics & Probability Letters*, 135, 7–10. 2.2

Bolthausen, E. (1982): "On the central limit theorem for stationary mixing random fields," *Annals of Probability*, 10, 1047–1050. 1, 3.3.2

Bradley, R. C. (2007): "Introduction to strong mixing conditions," *(No Title)*. 1

Chandrasekhar, A. G. and M. O. Jackson (2021): "A network formation model based on subgraphs," . 1, 2.2, 2.2, A, A.2, A.3, A, A

Chatterjee, S. (2008): "A new method of normal approximation," . 3

Chen, L. H. (1975): "Poisson approximation for dependent trials," *The Annals of Probability*, 3, 534–545. 2, 1

Chen, L. H. and Q.-M. Shao (2004): "Normal approximation under local dependence," *The Annals of Probability*, 32, 1985–2028. 1, 3.4, 3.5.2

Conley, T. G. (1999): "GMM estimation with cross sectional dependence," *Journal of econometrics*, 92, 1–45. 1, 3.3.2

Conley, T. G. and G. Topa (2002): "Socio-economic distance and spatial patterns in unemployment," *Journal of Applied Econometrics*, 17, 303–327. 3.3.2

Cramér, H. and H. Wold (1936): "Some theorems on distribution functions," *Journal of the London Mathematical Society*, 1, 290–294. A.4

Goldstein, L. and Y. Rinott (1996): "Multivariate normal approximations by Stein's method and size bias couplings," *Journal of Applied Probability*, 1–17. 1

Hoff, P. D., A. E. Raftery, and M. S. Handcock (2002): "Latent space approaches to social network analysis," *Journal of the american Statistical association*, 97, 1090–1098. 1

Horvitz, D. G. and D. J. Thompson (1952): "A generalization of sampling without replacement from a finite universe," *Journal of the American statistical Association*, 47, 663–685. 3.5.1

Jenish, N. and I. R. Prucha (2009): "Central limit theorems and uniform laws of large numbers for arrays of random fields," *Journal of econometrics*, 150, 86–98. 1, 3.3.1, 3.3.2, 1

Lubold, S., A. G. Chandrasekhar, and T. H. McCormick (2020): "Identifying the latent space geometry of network models through analysis of curvature," Tech. rep., National Bureau of Economic Research. 1

Romano, J. P. and M. Wolf (2000): "A more general central limit theorem for m-dependent random variables with unbounded m," *Statistics & probability letters*, 47, 115–124. 3.1.1

Ross, N. (2011): "Fundamentals of Stein's method," *Probab. Surv*, 8, 210–293. 1, 3.4, A.1, A

Stein, C. (1986): "Approximate computation of expectations," *Lecture Notes-Monograph Series*, 7, i–164. 1, A, A.1

Appendix A. Proof of Central Limit Theorem 1 and Corollary 1

Recall that Ω_n be a $p \times p$ matrix has entries

$$\Omega_{n,dd'} := \sum_{i=1}^{n} \sum_{(j,d') \in \mathcal{A}^n_{(i,d)}} \operatorname{cov}\left(Z_{i,d}, Z_{j,d'}\right).$$

We start with the case $p = 1$, reproducing elements of the proof to Theorem 2 in Chandrasekhar and Jackson (2021). Let $\Omega_n := \sum_{i=1}^{n} \sum_{j \in \mathcal{A}^n_i} \operatorname{cov}\left(Z_i, Z_j\right).$

The proof uses Stein's lemma from Stein (1986).

Lemma A.1 (Stein (1986); Ross (2011)).
If Y is a random variable and Z has the standard normal distribution, then

$$d_W(Y, Z) \leq \sup_{\{f:||f||,||f''||\leq 2,||f'||\leq\sqrt{2/\pi}\}} \left|\mathrm{E}[f'(Y) - Yf(Y)]\right|.$$

Further $d_K(Y, Z) \leq (2/\pi)^{1/4}(d_W(Y, Z))^{1/2}$.

By this lemma, if we show that a normalized sum of random variables satisfies

$$\sup_{\{f:||f||,||f''||\leq 2,||f'||\leq\sqrt{2/\pi}\}} \left|\mathrm{E}[f'(\overline{S}^n) - \overline{S}^n f(\overline{S}^n)]\right| \to 0,$$

then $d_W(\overline{S}^n, Z) \to 0$, and so it must be asymptotically normally distributed.

The following lemmas are useful in the proof.

Lemma **A.2** (Chandrasekhar and Jackson (2021), Lemma B.2)**.**
A solution to $\max_h \mathrm{E}[Zh(Y)]$ *s.t.* $|h| \leq 1$ *(where* h *is measurable) is* $h(Y) = \mathrm{sign}(\mathrm{E}[Z|Y])$, *where we break ties, setting* $\mathrm{sign}(\mathrm{E}[Z|Y]) = 1$ *when* $\mathrm{E}[Z|Y] = 0$.

Lemma **A.3** (Chandrasekhar and Jackson (2021), Lemma B.3)**.**
$\mathrm{E}[XYh(Y)]$ *when* $h(\cdot)$ *is measurable and bounded by* $\sqrt{\frac{2}{\pi}}$ *satisfies*

$$\mathrm{E}[XYh(Y)] \leq \sqrt{\frac{2}{\pi}}\mathrm{E}\left[XY \cdot \mathrm{sign}(\mathrm{E}[X|Y]Y)\right].$$

Lemma **A.4** (Cramér-Wold Device, Cramér and Wold (1936))**.** *The sequence* $\{X_n\}_n$ *of random vectors of dimension* $p \in \mathbb{N}$ *weakly converges to the random vector* $X \in \mathbb{R}^p$*, as* $n \to \infty$*, if and only if for any* $c \in \mathbb{R}^p$*,*

$$c^T X_n \rightsquigarrow c^T X$$

as $n \to \infty$.

Proof of Theorem 1. The $p = 1$ case is Theorem 2 in Chandrasekhar and Jackson (2021). So we reproduce a sketch to provide intuition. By Lemma A.1, it is sufficient to show that the appropriate sequence of random variables $\overline{S}^n$ satisfies

$$\sup_{\{f:||f||,||f''||\leq 2,||f'||\leq\sqrt{2/\pi}\}} \left|\mathrm{E}[f'(\overline{S}^n) - \overline{S}^n f(\overline{S}^n)]\right| \to 0.$$

Let

$$S_i := \sum_{j \notin \mathcal{A}_i^n} Z_j \text{ and } \overline{S}_i := S_i/\Omega_n^{1/2}.$$

Let $\overline{S}^n = S^n/\Omega_n^{1/2}$.

We consider

$$\mathrm{E}[f'(\overline{S}^n) - \overline{S}^n f(\overline{S}^n)]| \tag{A.1}$$

for all f such that $||f||, ||f''|| \leq 2, ||f'|| \leq \sqrt{2/\pi}$. Observe that

$$\mathrm{E}\left[\overline{S}f\left(\overline{S}\right)\right] = \mathrm{E}\left[\frac{1}{\Omega_n^{1/2}}\sum_i^n Z_i \cdot f\left(\overline{S}\right)\right] = \mathrm{E}\left[\frac{1}{\Omega_n^{1/2}}\sum_i^n Z_i\left(f\left(\overline{S}\right) - f\left(\overline{S}_i\right)\right)\right] + \mathrm{E}\left[\frac{1}{\Omega_n^{1/2}}\sum_i^n Z_i \cdot f\left(\overline{S}_i\right)\right].$$

We first consider the second term above. By a first-order Taylor approximation of $f\left(\overline{S}\right)$ about $f(0)$, and observing $\mathrm{E}[Z_i]=0$ and the triangle inequality, we get an upper bound:

$$\mathrm{E}\left[\Omega_n^{1/2}\sum_i^n Z_i\cdot f\left(\overline{S}_i\right)\right]=\mathrm{E}\left[\Omega_n^{1/2}\sum_i^n Z_i\cdot f\left(0\right)\right]+\mathrm{E}\left[\Omega_n^{1/2}\sum_i^n Z_i\cdot(\overline{S}_i)f'\left(\tilde{S}_i\right)\right].$$

The first term is zero since $\mathrm{E}[Z_i]=0$. We upper bound the second term by applying Lemma A.3. To see this, we expand $\overline{S}_i$,

$$\mathrm{E}\left[\Omega_n^{1/2}\sum_i^n Z_i\cdot\left(\Omega_n^{1/2}\sum_{j\notin\mathcal{A}_i^n}Z_j\right)f'\left(\tilde{S}_i\right)\right]\leq\Omega_n^{-1}\sqrt{2/\pi}\cdot\mathrm{E}\left[\sum_i\sum_{j\notin\mathcal{A}_i^n}Z_iZ_j\cdot\mathrm{sign}(\mathrm{E}[Z_i\mid Z_j]Z_j)\right]$$

By Assumption 3, we have that the upper bound is $o(1)$. This kind of bound is generated by the fact that in principle any two random variables can be correlated for a given n.

Now, plugging this back into (A.1), following a similar reasoning as in Ross (2011), we have once again by using the triangle inequality an upper bound. Followed by a second-order Taylor series approximation and a bound on the derivatives of f and a use of Cauchy-Schwarz inequality, together with Assumptions 1, and 2. In both of these pieces, rather than relying on conditional independence and applying the arithmetic-geometric mean inequality to write conditions in terms of moment restrictions, which cannot be used with the general dependency structure, we collect covariance terms.

Therefore, we have shown that the convergence $\overline{S}^n\rightsquigarrow N(0,1)$ in distribution in each dimension. Now, we consider the multidimensional setting, and let $Y=(Y_1,Y_2,...,Y_p)$ be a mean-zero normally distributed random vector with covariance the p-dimensional identity matrix. By the Cramér-Wold device (Lemma A.4), it is sufficient to show that

$$\sum_{u=1}^p c_u\sum_{i=1}^n\sum_{k=1}^p Z_{ik}(\Omega_n{}^{-1/2})_{ku}\rightsquigarrow\sum_{u=1}^p c_u\sum_{i=1}^n Y_{iu}\tag{A.2}$$

for all $c\in\mathbb{R}^p$.

But, from the proof above, we see that for each $u\in[p]$,

$$\sum_{i=1}^n\sum_{k=1}^p Z_{ik}(\Omega_n{}^{-1/2})_{ku}\rightsquigarrow\sum_{i=1}^n Y_{iu}.$$

It immediately follows that A.2 is satisfied so we have shown $(\Omega_n)^{-1/2}S^n\rightsquigarrow\mathcal{N}(0,I_{p\times p})$. ■

Proof sketch of Corollary 1. We refer the reader to the proof to Corollary 2 in Chandrasekhar and Jackson (2021). Applying the Cramér-Wold device just as above gives us the result. ■