Robust Inference for the Direct Average Treatment Effect with Treatment Assignment Interference

Matias D. Cattaneo*   Yihan He†   Ruiqi (Rae) Yu‡

June 26, 2025

Abstract

This paper develops methods for uncertainty quantification in causal inference settings with random network interference. We study the large-sample distributional properties of the classical difference-in-means Hájek treatment effect estimator, and propose a robust inference procedure for the (conditional) direct average treatment effect. Our framework allows for cross-unit interference in both the outcome equation and the treatment assignment mechanism. Drawing from statistical physics, we introduce a novel Ising model to capture complex dependencies in treatment assignment, and derive three results. First, we establish a Berry–Esseen-type distributional approximation that holds pointwise in the degree of interference induced by the Ising model. This approximation recovers existing results in the absence of treatment interference, and highlights the fragility of inference procedures that do not account for the presence of interference in treatment assignment. Second, we establish a uniform distributional approximation for the Hájek estimator and use it to develop robust inference procedures that remain valid uniformly over all interference regimes allowed by the model. Third, we propose a novel resampling method to implement the robust inference procedure and validate its performance through Monte Carlo simulations. A key technical innovation is the introduction of a conditional i.i.d. Gaussianization that may have broader applications. We also discuss extensions and generalizations of our results.

Keywords: causal inference under interferece; Ising Model; Distribution Theory; Robust Inference

---

*Department of Operations Research and Financial Engineering, Princeton University.
†Department of Operations Research and Financial Engineering, Princeton University.
‡Department of Operations Research and Financial Engineering, Princeton University.

1 Introduction

We study the large-sample distributional properties of the classical difference-in-means Hájek estimator for the direct average treatment effect, and propose a robust inference procedure for the (conditional) direct average treatment effect in the presence of cross-unit interference in both the outcome equation and treatment assignment mechanism. This problem arises in a variety of casusal inference settings, including social networks, medical trials, and socio-spatial studies, and has received renewed attention. Recent contributions include [1], [3], [10], [12], [13], [14], [19], [21], and references therein. See [9] for a modern textbook introduction to causal inference.

A central challenge in causal inference with interference is that units can affect one another in arbitrary ways, complicating identification and inference. To discipline these interactions, it is common to model units as vertices in a network, where edges encode potential interference relationships. While early research assumed the network was fixed and observed, recent advances allow for random and latent (unobserved) network structures (see Assumption 1 below). To manage this added complexity, [14] restricted the degree of interference, especially in the potential outcomes (e.g., by assuming *anonymity* or *exchangeability*, see Assumption 2), while typically assuming a random (independent) treatment assignment. Notable recent exceptions include [19] and [3], who allow for observed-covariate-based interference in treatment assignment. We contribute to this emerging literature by uncovering a rich family of limiting distributions, and establishing uniformly valid inference procedures across different interference regimes in the treatment assignment mechanism.

To model interference in treatment assignment, we introduce a class of Ising-based assignment mechanisms inspired by statistical physics [8]. Specifically, we assume that the vector of binary treatment assignments $\mathbf{T} = (T_1, \ldots, T_n)^\top \in \{0,1\}^n$ follows the distribution

$$\mathbb{P}_\beta(\mathbf{T} = \mathbf{t}) \propto \exp\left(\frac{\beta}{n} \sum_{i \neq j} (2t_i - 1)(2t_j - 1)\right), \tag{1}$$

where $\mathbf{t} = (t_1, \ldots, t_n)^\top \in \{0,1\}^n$, and the parameter $\beta \geq 0$ governs the degree of cross-unit dependence in treatment assignment (see Assumption 3). This model incorporates potential interference effects by allowing treatment assignments to interact globally, capturing the equilibrium distribution of treatment assignments under network influence, analogous to a system in thermal equilibrium with inverse temperature parameter β. When $\beta = 0$, the model reduces to the classical independent equiprobable treatment assignment. As β increases, treatment assignments become more dependent, modeling correlated behaviors such as peer effects or strategic interactions. This equiprobable Ising model allows us to study how interference in assignment affects inference procedures proposed in the literature, and to evaluate their validity and robustness as a function of the unknown dependence parameter β.

To streamline the analysis, and due to some technical issues, we focus on the moderate interference regime $\beta \in [0, 1]$; see Section 5 for extensions and generalizations. Our first contribution is to derive the large-sample distribution of the Hájek estimator in this setting. Theorem 3.1 provides a Berry-Esseen bound for the estimator, that is, a distributional approximation in Kolmogorov distance with explicit convergence rates. The closest antecedent is [14], who considered the same causal inference model under independent treatment assignment ($\beta = 0$), and established a Gaussian approximation for the Hájek estimator. Our results show that:

1. For $\beta \in [0, 1)$, the limiting distribution remains Gaussian, but the asymptotic variance includes an additional term capturing dependence in treatment assignment. The asymptotic variance is increasing and unbounded in β, encompassing the known special case $\beta = 0$.

2

2. At $\beta = 1$, the limiting distribution becomes non-Gaussian. Therefore, there is a discontinuity in the asymptotic distributional behavior of the Hájek estimator.

These results highlight a fragility in the inference methods proposed by [14], which are not robust to the presence of interference in the treatment assignment. Since their Gaussian approximation becomes invalid when $\beta > 0$, the associated inference procedures, while valid at $\beta = 0$, can perform poorly for moderate or large interference in the treatment assignment.

The lack of uniform validity (in β) poses a major challenge for developing robust inference procedures in the presence of interference in the treatment assignment. [18] showed that no consistent estimator exists for $\beta \in [0, 1)$, making plug-in inference procedures infeasible, even pointwise in $\beta \in [0, 1)$. To overcome these challenges, Theorem 3.2 establishes a uniform in $\beta \in [0, 1]$ distributional approximation for the Hájek estimator, along with a parallel result for the Maximum Pseudo-Likelihood estimator (MPLE) for β. Our uniform distributional approximation depends on a localization parameter that smoothly interpolates between the Gaussian and non-Gaussian regimes.

Building on these results, we construct uniformly valid confidence sets for the (conditional) direct average treatment effect τ_n. More precisely, using a Bonferroni-based correction that aggregates inference across β-regimes, we construct both infeasible (Theorem 4.1) and feasible (Theorem 4.2) prediction intervals, generically denoted by $\mathcal{C}_n(\alpha)$, such that

$$\liminf_{n \to \infty} \inf_{\beta \in [0,1]} \mathbb{P}_\beta[\tau_n \in \mathcal{C}_n(\alpha)] \geq 1 - \alpha,$$

for $\alpha \in [0, 1]$. The intervals $\mathcal{C}_n(\alpha)$ are based on the Hájek estimator and a resampling procedure designed to capturing capture uncertainty due to the latent network structure. To our knowledge, this is the first feasible inference procedure for this setting, and the first to provide robust uniform validity over $\beta \in [0, 1]$, including in the known special case $\beta = 0$. We also present a simulation study demonstrating the performance of our proposed methods.

Our methodological contributions are twofold. First, we introduce the Ising equiprobable treatment assignment model, which generalizes independent assignment to allow nonparametric, structured interference. Second, we develop a new robust and feasible inference procedure for the (conditional) direct average treatment effect, uniformly valid across a wide class of treatment dependence structures. This method combines novel uniform distributional approximations, a hierarchical confidence region construction, and a new resampling-based variance estimator. See [11], [20], [17], and references therein, for overviews of the literature.

On the technical side, our work also contributes to the applied probability and statistical mechanics literature. Interference in assignment induces strong dependence across units, complicating traditional asymptotic arguments. For example, as shown in Theorem 3.1, both concentration rates and limit distributions depend discontinuously on $\beta \in [0, 1]$. To address this technical challenge, we introduce a *De-Finetti Machine*, a new theoretical approach that leverages exchangeability in the Ising model to conditionally Gaussianize the assignment vector. This method transforms the problem into a conditional i.i.d. setting, enabling Berry-Esseen approximations even in the presence of strong dependence. Our main technical result is given in Lemma 3.1, which generalizes results from [5] and [7], and provides an alternative to Stein's method. This technique may be useful in other settings involving exchangeable or dependent random variables. Furthermore, our uniform distributional approximations also contribute to the growing literature on the probabilistic properties of the Ising model, and may be of independent interest in statistics, econometrics, and beyond.

2 Setup

We consider a random potential outcome framework under network interference. For each unit $i \in [n] = \{1, 2, \cdots, n\}$, let $Y_i(t; \mathbf{t}_{-i})$ denote its random potential outcome when assigned to treatment level $t \in \{0, 1\}$ while the other units are assigned to treatment levels $\mathbf{t}_{-i} \in \{0, 1\}^{n-1}$. The vector of observed random treatment assignments for the n units is $\mathbf{T} = (T_i : i \in [n])$, and $\mathbf{T}_{-i}$ denotes the associated random treatment assignment vector excluding T_i. Thus, the observed data is $(Y_i, T_i : i \in [n])$ with $Y_i = (1 - T_i)Y_i(0; \mathbf{T}_{-i}) + T_iY_i(1; \mathbf{T}_{-i})$ for each $i \in [n]$.

Interference among the n units is modelled via a latent network characterized by an undirected random graph $G(\mathbf{V}, \mathbf{E})$ with vertex set $\mathbf{V} = [n]$ and (random) adjacency matrix $\mathbf{E} = (E_{ij} : (i, j) \in [n] \times [n]) \in \{0, 1\}^{n \times n}$. The following assumption restricts this random graph structure.

Assumption 1 (Network Structure). *The random network $\mathbf{E}$ satisfies: For all $1 \leq i \leq j \leq n$ and $\rho_n \in (0, 1]$, $E_{ii} = 0$, $E_{ij} = E_{ji}$, and $E_{ij} = \mathbb{1}(\xi_{ij} \leq \min\{1, \rho_n G(U_i, U_j)\})$, where $G : [0, 1]^2 \mapsto \mathbb{R}_+$ is symmetric, continuous and positive on $[0, 1]^2$, $\mathbf{U} = (U_i : i \in [n])$ are i.i.d. $\mathsf{Uniform}[0, 1]$ random variables, $\mathbf{\Xi} = (\xi_{ij} : (i, j) \in [n] \times [n], i < j)$ are i.i.d $\mathsf{Uniform}[0, 1]$ random variables. Finally, $\mathbf{U}$ and $\mathbf{\Xi}$ are independent.*

This assumption corresponds to the *sparse graphon model* of [4]. The parameter ρ_n controls the sparsity of the network, and will play an important role in our theoretical results. The variable U_i is a *latent* heterogenous property of the ith unit, and $G(U_i, U_j)$ measures similarity between traits of U_i and U_j. This allows for a stochastic model for the edge formation.

Building on the underlying random graph structure, the following assumption imposes discipline on the interference entering the outcome equation.

Assumption 2 (Exchangeable Smooth Potential Outcomes Model). *For all $i \in [n]$, $Y_i(T_i; \mathbf{T}_{-i}) = f_i(T_i; M_i/N_i)$ where $M_i = \sum_{j \neq i} E_{ij} T_j$, $N_i = \sum_{j \neq i} T_j$, and $\mathbf{f} = (f_i : i \in [n])$ are i.i.d random functions. In addition, for all $i \in [n]$ and some integer $p \geq 4$, $\max_{1 \leq i \leq n} \max_{t \in \{0,1\}} |\partial_2^{(p)} f_i(t, \cdot)| \leq C$ for some C not depending on n and β. Finally, $\mathbf{f}$ is independent to $\mathbf{\Xi}$.*

This second assumption imposes two main restrictions on the potential outcomes. First, a dimension reduction is assumed via the underlying network structure (Assumption 1), making the potential outcomes for each unit $i \in [n]$ a function of only their own treatment assignment and the fraction of other treated units among their (connected) peers. Second, the potential outcomes are assumed to be smooth as a function of the fraction of treated peers, thereby ruling out certain types of outcome variables (e.g., binary or similarly limited dependent variable models). Assumption 2 explicitly parametrizes the smoothness level p because, together with the the sparsity parameter ρ_n in Assumption 1, it will play an important role in our theoretical results.

To complete the causal inference model, the following assumption restricts the treatment assignment distribution. We propose an Ising model from statistical physics [8].

Assumption 3 (Ising Equiprobable Treatment Assignment). *The treatment assignment mechanism follows a Curie-Weiss distribution:*

$$\mathbb{P}_\beta(\mathbf{T} = \mathbf{t}) = \frac{1}{C_\beta} \exp\left(\frac{\beta}{n} \sum_{1 \leq i \neq j \leq n} (2t_i - 1)(2t_j - 1)\right), \tag{2}$$

where $\mathbf{t} \in \{0, 1\}^n$, $\beta \in [0, 1]$, and C_β is determined by the condition $\sum_{\mathbf{t}} \mathbb{P}_\beta(\mathbf{T} = \mathbf{t}) = 1$.

This model encodes a nonparametric class of equiprobable, possibly dependent treatment assignment mechanisms. Assumption 3 implies $\mathbb{P}_\beta(T_i = 1) = 1/2$ for $i \in [n]$ and all $\beta \geq 0$, but allows for correlation in treatment assignment as controlled by β. When $\beta = 0$, treatment assignment becomes independent across units, and thus the assignment mechanism reduces to the canonical (equiprobable) randomized allocation. For $\beta \in [0, 1]$, the Ising mechanism induces positive pairwise correlations, capturing social interdependence phenomena like peer influence [16] characteristic of observational settings.

We study the distributional properties of, and proposed uncertainty quantification methods based on, the classical Hájek estimator

$$\widehat{\tau}_n = \frac{\sum_{i=1}^n T_i Y_i}{\sum_{i=1}^n T_i} - \frac{\sum_{i=1}^n (1 - T_i) Y_i}{\sum_{i=1}^n (1 - T_i)}, \tag{3}$$

which is commonly used in causal inference, both with and without interference. In particular, [14] studied the asymptotic properties of $\widehat{\tau}_n$ under Assumptions 1–3 with $\beta = 0$, and showed that

$$\sqrt{n}(\widehat{\tau}_n - \tau_n) \rightsquigarrow \mathsf{N}(0, \kappa_2), \qquad \kappa_s = \mathbb{E}[(R_i - \mathbb{E}[R_i] + Q_i)^s], \tag{4}$$

where $\rightsquigarrow$ denotes weak convergence as $n \to \infty$, $\mathsf{N}(0, \sigma^2)$ denotes a mean-zero Gaussian distribution with variance σ^2, the centering is the (conditional) direct average treatment effect

$$\tau_n = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[Y_i(1; \mathbf{T}_{-i}) - Y_i(0; \mathbf{T}_{-i}) | f_i(\cdot), \mathbf{E}], \tag{5}$$

and the random variables

$$R_i = f_i\left(1, \frac{1}{2}\right) + f_i\left(0, \frac{1}{2}\right) \qquad \text{and} \qquad Q_i = \mathbb{E}\left[\frac{G(U_i, U_j)}{\mathbb{E}[G(U_i, U_j) | U_j]}\left(f_j'\left(1, \frac{1}{2}\right) - f_j'\left(0, \frac{1}{2}\right)\right)\Big| U_i\right]$$

depend on the underlying stochastic potential outcomes and network structure. The (conditional) direct average treatment effect in (5) is a *predictand*, not an *estimand*, because it is a random variable that needs not to settle to a non-random probability limit under the assumptions imposed. Consequently, we propose prediction intervals for τ_n for uncertainty quantification.

3 Distribution Theory

We present pointwise and uniform in $\beta \in [0, 1]$ distributional approximations for $\widehat{\tau}_n$. We also study pointwise and uniform distributional approximations for the Maximum Pseudo-Likelihood Estimator (MPLE) for β. We employ a novel Berry-Esseen bound for the Ising model, which may be of independent interest. Proofs and technical details are given in the supplemental appendix.

3.1 Pointwise Distributional Approximation

Our first result is a Berry-Esseen bound for the Hájek estimator, pointwise in $\beta \in [0, 1]$. Let $\mathbb{P}_\beta$ denote a probability distribution under parametrization β.

Theorem 3.1 (Pointwise Distribution Theory). *Suppose Assumptions 1, 2, and 3 hold. Then,*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}_\beta[\widehat{\tau}_n - \tau_n \leq t] - L_n(t; \beta, \kappa_1, \kappa_2)| = O\left(\frac{\log n}{\sqrt{n \rho_n}} + \mathtt{r}_{n,\beta}\right),$$

where $L_n(\cdot; \beta, \kappa_1, \kappa_2)$ and $\mathtt{r}_{n,\beta}$ are as follows.

1. *High temperature:* If $\beta \in [0, 1)$, then

$$L_n(t; \beta, \kappa_1, \kappa_2) = \mathbb{P}\left[n^{-1/2}\left(\kappa_2 + \kappa_1^2 \frac{\beta}{1-\beta}\right)^{1/2} \mathsf{Z} \le t\right] \tag{6}$$

   with $\mathsf{Z} \sim \mathsf{N}(0, 1)$, *and* $\mathtt{r}_{n,\beta} = \frac{\sqrt{n \log n}}{(n\rho_n)^{(p+1)/2}}$.

2. *Critical temperature:* If $\beta = 1$, then

$$L_n(t; \beta, \kappa_1, \kappa_2) = \mathbb{P}[n^{-1/4}\kappa_1 \mathsf{W}_0 \le t] \tag{7}$$

   with

$$\mathbb{P}[\mathsf{W}_c \le w] = \frac{\int_{-\infty}^{w} \exp(-\frac{x^4}{12} - \frac{cx^2}{2})dx}{\int_{-\infty}^{\infty} \exp(-\frac{x^4}{12} - \frac{cx^2}{2})dx}, \qquad w \in \mathbb{R}, \quad c \in \mathbb{R}_+,$$

   and $\mathtt{r}_{n,\beta} = \frac{\log^3 n}{n^{1/4}} + \frac{n^{1/4}\sqrt{\log n}}{(n\rho_n)^{(p+1)/2}}$.

   In the high temperature regime, $\sqrt{n}(\widehat{\tau}_n - \tau_n)$ is asymptotically normal with variance $\kappa_2 + \kappa_1^2 \beta/(1 - \beta)$. Thus, our result recovers (4) when $\beta = 0$, but the asymptotic variance is strictly larger (increasing in $\beta \in [0, 1)$) unless $\kappa_1 = 0$ (no randomness from the underlying network) when $\beta \in (0, 1)$. In the critical temperature regime ($\beta = 1$), the limiting distribution is non-Gaussian. The distinct asymptotic behaviors of $\widehat{\tau}_n$ across the interference regimes mirror the phase transition phenomena observed in the Ising model's magnetization $m = \frac{1}{n}\sum_{i=1}^{n}(2T_i - 1)$. Theorem 3.1 highlights critical challenges in uncertainty quantification due to the fact that κ_1 and κ_2 are unknown quantities, and the interference regime parameter $\beta \in [0, 1]$ is also unknown.

   [2] established an impossibility result showing that no consistent estimator for β exists in the high-temperature regime. Due to the existence of the normalizing constant C_β in (2), maximum likelihood estimation is also computationally prohibitive. However, the conditional distribution of T_i given the rest of treatments adopts a closed form solution and can be optimized efficiently. Define $W_i = 2T_i - 1$, $\mathbf{W}_{-i} = \{W_j : j \in [n], j \ne i\}$, and $m_i = n^{-1}\sum_{j \ne i} W_j$. The MPLE for β is

$$\widehat{\beta}_n = \arg\max_{\beta \in [0,1]} \sum_{i \in [n]} \log \mathbb{P}_\beta[W_i | \mathbf{W}_{-i}] = \arg\max_{\beta \in [0,1]} \sum_{i \in [n]} -\log\left(\frac{1}{2}W_i \tanh(\beta m_i) + \frac{1}{2}\right).$$

We show in the supplementary appendix (Lemma SA-8) that the limiting distribution of $\widehat{\beta}_n$ also depends on the regime $\beta \in [0, 1]$. For $\beta \in [0, 1)$, $1 - \widehat{\beta}_n \rightsquigarrow (1 - \beta)\max\{(\chi_1^2)^{-1}, 0\}$, thereby ruling out consistent estimation, as already anticipated by [2]. For $\beta = 1$, $\sqrt{n}(\widehat{\beta}_n - 1) \rightsquigarrow \min\{\mathsf{W}_0^2/3 - 1/\mathsf{W}_0^2, 1\}$, where W_0 is given in Theorem 3.1.

3.2 Uniform Distributional Approximation

From Theorem 3.1, for all $\beta \in [0, 1)$, the limiting variance of $\sqrt{n}(\widehat{\tau}_n - \tau_n)$ is $\kappa_2 + \kappa_1^2\beta/(1 - \beta)$. Thus, when $\kappa_1 \ne 0$, the asymptotic variance diverges as β approaches the critical value $\beta = 1$. In contrast, Theorem 3.1 shows that when $\beta = 1$ the limiting variance of $n^{1/4}(\widehat{\tau}_n - \tau_n)$ is finite. This discrepancy indicates a lack of uniform (in β) validity for the two distributional approximations. The asymptotic distribution of $\widehat{\beta}_n - 1$ exhibits a discontinuity at $\beta = 1$, also highlighting the need for a distributional approximation that is uniform in β for valid inference across all regimes.

   We establish a uniform distributional approximation based on the drifting sequence $\beta_n = 1 + cn^{-1/2}$. This sequence follows the *knife-edge* rate, ensuring that the resulting approximating laws for $\widehat{\tau}_n$ and $\widehat{\beta}_n$ smoothly interpolate between the different pointwise distributional approximations indexed by $\beta \in [0, 1]$.

Theorem 3.2 (Uniform Distribution Theory)**.** *Suppose Assumptions 1, 2 and 3 hold, and define* $c_{\beta,n} = \sqrt{n}(1-\beta)$. *Then, the following distributional approximations hold.*
1. *Hájek Estimator:*

$$\lim_{n\to\infty} \sup_{0\leq\beta\leq 1} \sup_{t\in\mathbb{R}} |\mathbb{P}_\beta[\widehat{\tau}_n - \tau_n \leq t] - H_n(t; \kappa_1, \kappa_2, c_{\beta,n})| = 0,$$

where

$$H_n(t; \kappa_1, \kappa_2, c_{\beta,n}) = \mathbb{P}[n^{-1/2}\kappa_2^{1/2}\mathsf{Z} + \beta^{1/2}n^{-1/4}\kappa_1\mathsf{W}_{c_{\beta,n}} \leq t]$$

with $\mathsf{Z} \sim \mathsf{N}(0,1)$ *independent of* W_c.

2. *MPLE for* β:

$$\lim_{n\to\infty} \sup_{0\leq\beta\leq 1} \sup_{t\in\mathbb{R}} \left|\mathbb{P}_\beta[1 - \widehat{\beta}_n \leq t] - \mathbb{P}[\min\{\max\{\mathsf{T}_{c_{\beta,n},n}^{-2} - \mathsf{T}_{c_{\beta,n},n}^2/(3n), 0\}, 1\} \leq t]\right| = 0$$

where $\mathsf{T}_{c,n} = \mathsf{Z} + n^{1/4}\mathsf{W}_c$.

Theorem 3.2 establishes that $H_n(t; \kappa_1, \kappa_2, c_{\beta,n})$ uniformly approximates the distribution of $\widehat{\tau}_n - \tau_n$ in both the high-temperature and critical-temperature regimes. Under the *knife-edge* scaling, the leading term $n^{-1/2}\kappa_2^{1/2}\mathsf{Z}$ becomes negligible, and the typical *knife-edge* representation retains only the second term $\beta^{1/2}n^{-1/4}\kappa_1\mathsf{W}_c$. However, when $\beta \in [0,1)$ is fixed and $c_{\beta,n} = \sqrt{n}(1-\beta) \to \infty$, $\mathsf{W}_{c_{\beta,n}}$ approximates $n^{-1/4}\mathsf{N}(0, (1-\beta)^{-1})$, making both terms comparable in order. Consequently, we retain both terms in the distributional approximation. In the supplementary appendix (Lemma SA-5), we show that when β is fixed and $c_{\beta,n} = \sqrt{n}(1-\beta)$, we have $\sup_{t\in\mathbb{R}}|H_n(t; \kappa_1, \kappa_2, c_{\beta,n}) - L_n(t; \kappa_1, \kappa_2, \beta)| \to 0$. The same ideas apply to $1 - \widehat{\beta}_n$.

3.3 Main Technical Contribution

Our main results, Theorem 3.1 and Theorem 3.2, rely on a novel Berry-Esseen distributional approximation for Curie-Weiss magnetization with independent multipliers given in the following lemma. The lemma is self-contained because it may be of independent interest in other settings, and its proof and related technical details are given in the supplemental appendix.

Lemma 3.1 (Ising Berry-Esseen Bound)**.** *Let* $\mathbb{P}_\beta[\mathbf{W} = \mathbf{w}] \propto \exp(\frac{\beta}{n}\sum_{i\neq j} w_i w_j)$, *where* $\mathbf{W} = (W_1, \cdots, W_n)^\top \in \{-1,1\}^n$, $\mathbf{w} = (w_1, \cdots, w_n)^\top \in \{-1,1\}^n$, $\beta \geq 0$, *and* $(X_1, \cdots, X_n)$ *are i.i.d. with* $\mathbb{E}[|X_i|^3] < \infty$, *and independent of* $\mathbf{W}$. *Then, the following results hold.*
1. $\sup_{t\in\mathbb{R}} |\mathbb{P}_\beta(n^{-1}\sum_{i=1}^n X_i W_i \leq t) - L_n(t; \mathbb{E}[X_i], \mathbb{E}[X_i^2], \beta)| = O(n^{-1/2}(1 + \mathbb{1}(\beta = 1)\log^3 n))$, *where* L_n *is given in Theorem 3.1.*

2. $\sup_{\beta\in[0,1]} \sup_{t\in\mathbb{R}} |\mathbb{P}(n^{-1}\sum_{i=1}^n X_i W_i \leq t) - H_n(t; \mathbb{E}[X_i], \mathbb{E}[X_i^2], c_{\beta,n})| = O(n^{-1/2}\log^3 n)$, *where* $c_{\beta,n} = \sqrt{n}(\beta - 1)$, *and* H_n *is given in Theorem 3.2.*

This lemma generalizes the Berry-Esseen bounds for Curie-Weiss magnetization $n^{-1}\sum_{i=1}^n W_i$, $W_i = 2T_i - 1$, obtained by [5] and [7], to allow for multipliers $\mathbf{X}_i$'s independent to T_i's. Our rate differs from theirs only in a logarithmic term, while allowing for fairly general random weights X_i with bounded third moment.

Our proof strategy differs significantly from [5] and [7]. They rely on the existence of an exchangeable pair: for $A = n^{-1}\sum_{i=1}^n W_i$, the existence of another random variable A' such that (A, A') and (A', A) have the same distribution. The limiting distribution of A can be studied using

Stein's method characterized by g, if $\mathbb{E}[A - A'|A] = g(A) + r(A)$ where $g(A)$ is the leading term and $r(A)$ is negligible. See Equation (1.1) in [5], and Equation (1.3) in [7]. However, we have not succeed in finding such an exchangeable pair construction. The Glauber dynamics construction used in both previous works does not work here, due to the fact that $\mathbb{E}[X_i'(2T_i' - 1)|\mathbf{X}_{-i}, \mathbf{T}_{-i}] \approx \mathbb{E}[X_i] \tanh(2n^{-1} \sum_{i=1}^{n} T_i - 1)$, not preserving the multiplicative structure $n^{-1} \sum_{i=1}^{n} X_i(2T_i - 1)$, where X_i' and T_i' are random variables following the conditional distribution of X_i given $\mathbf{X}_{-i}$, and T_i given $\mathbf{T}_{-i}$, respectively. In contrast, our method is based on the *de Finetti's Lemma*, which ensures that the Ising spins $W_i = 2T_i - 1$ are i.i.d., conditional on a latent variable U_n. Hence, we can apply an i.i.d Berry-Esseen result condition on U_n, and characterize the conditional mean and variance based on distribution of U_n. The density of U_n at $u \in \mathbb{R}$ is proportional to $\exp(-1/2u^2 + n \log \cosh(\sqrt{\beta/n}u))$, and then show that U_n is close to $\mathsf{N}(0, (1-\beta)^{-1})$ when $\beta \in [0,1)$, and U_n is close to $n^{1/4}\mathsf{W}_0$ when $\beta = 1$. See the supplementary appendix (Section SA-2) for details.

4 Uniform Inference

Building on Theorem 3.2, we present prediction intervals for τ_n that are valid uniformly over $\beta \in [0, 1]$. First, we discuss an infeasible construction assuming κ_1 and κ_2 are known, thereby focusing on uniformity over β. Then, we present feasible prediction intervals for the (conditional) direct average treatment effect. Finally, we present a simulation study.

4.1 Known Random Graph Structure

This section addresses inference when the interference regime parameter β is unknown, but κ_1 and κ_2 are known. We propose a conservative prediction interval based on the following Bonferroni-correction procedure. In particular, in the first step, a uniform confidence interval for β is constructed under the *knife-edge* approximation, and in the second step, we choose the largest quantile for $\widehat{\tau}_n - \tau_n$ among all β's in the confidence interval. The quantile chosen is also based on the *knife-edge* approximation.

The following algorithm summarizes the infeasible, uniform-in-β inference procedure.

Algorithm 1: Infeasible Prediction Interval for τ_n

Input: Treatments and outcomes $(T_i, Y_i)_{i \in [n]}$, MPLE-estimator $\widehat{\beta}_n$, an upper bound K_n such that $\kappa_2 \leq K_n^2$, confidence level parameters $\alpha_1, \alpha_2 \in (0, 1)$.
Output: An $(1 - \alpha_1 - \alpha_2)$ prediction interval $\widetilde{\mathcal{C}}(\alpha_1, \alpha_2)$ for τ_n.

Get the maximum pseudo-likelihood estimator $\widehat{\beta}_n$ of β;

Define the $(1 - \alpha_1)$-confidence region given by $\mathcal{I}(\alpha_1) = \{\beta \in [0,1] : 1 - \widehat{\beta}_n \in [\mathsf{q}, \infty)\}$, where $\mathsf{q} = \inf\{q : \mathbb{P}[\min\{\max\{\mathsf{T}_{c_{\beta,n},n}^{-2} - \mathsf{T}_{c_{\beta,n},n}^2/(3n), 0\}, 1\} \leq q] \geq \alpha_1\}$;
Take $\mathsf{U} = \sup_{\beta \in \mathcal{I}(\alpha_1)} H_n(1 - \frac{\alpha_2}{2}; K_n, K_n, c_{\beta,n})$, $\mathsf{L} = \inf_{\beta \in \mathcal{I}(\alpha_1)} H_n(\frac{\alpha_2}{2}; K_n, K_n, c_{\beta,n})$.

return $\widetilde{\mathcal{C}}(\alpha_1, \alpha_2) = [\widehat{\tau}_n + \mathsf{L}, \widehat{\tau}_n + \mathsf{U}]$.

Validity of our proposed infeasible, uniform-in-β prediction interval for τ_n is established in the following theorem.

Theorem 4.1 (Infeasible Uniform Inference). *Suppose Assumptions 1, 2 and 3 hold, and let K_n be a sequence such that $\kappa_2 \leq K_n^2$. Then, the prediction interval $\widetilde{\mathcal{C}}(\alpha_1, \alpha_2)$ given in Algorithm 1 satisfies* $\liminf_{n \to \infty} \inf_{\beta \in [0,1]} \mathbb{P}_\beta[\tau_n \in \widetilde{\mathcal{C}}(\alpha_1, \alpha_2)] \geq 1 - \alpha_1 - \alpha_2.$

Algorithm 1 can be implemented without the knowledge of the degree of interference in the treatment assignment, but requires knowledge of of the underlying random graph structure.

4.2 Feasible Prediction Interval

The unknown parameters κ_1 and κ_2 capture moments of the underlying random graph structure. Building on [15], we propose a resampling method for consistent estimation of those parameters under an additional nonparametric assumption on the outcome equation.

Assumption 4. *Suppose $f_i(\cdot, \cdot) = f(\cdot, \cdot) + \varepsilon_i$, where $f(t, \cdot)$ is 4-times continuously differentiable on $[0, 1]$ for $t \in \{0, 1\}$, and $(\varepsilon_i : 1 \leq i \leq n)$ are i.i.d and independent of $\mathbf{E}$ and $\mathbf{T}$, with $\mathbb{E}[\varepsilon_i] = 0$ and $\mathbb{E}[|\varepsilon_i|^{2+\nu}] < \infty$ for some $\nu > 0$.*

This assumption allows for nonparametric learning the non-random regression function f. In the supplementary appendix (Section SA-5), we provide one example of such learner, but here we remain agnostic and thus present high-level conditions. This step aims to find a consistent estimate for both the function f and its derivative $\frac{\partial f(\cdot, x)}{\partial x}$.

We propose the following novel algorithm for estimating κ_2 based on resampling methods.

Algorithm 2: Estimation of κ_2

Input: Treatments and outcomes $(T_i, Y_i)_{i \in [n]}$, realized graph $\mathbf{E}$, non-parametric learner $\widehat{f}$ of f.

Output: An upper bound $\widehat{K}_n$ for κ_2.

Generate a new sample $(T_i^* : 1 \leq i \leq n)$ with $\beta = 0$;

Take $M_j^* = \sum_{l \neq j} E_{jl} T_l^*$, $N_j^* = \sum_{l \neq j} E_{jl}$, $M_{j,(i)}^* = \sum_{l \neq i, j} E_{jl} T_l^*$, $N_{j,(i)}^* = \sum_{l \neq i, j} E_{jl}$;

Take $\widehat{\varepsilon}_i = Y_i - T_i \widehat{f}(1, \frac{\sum_{j \neq i} E_{ij} T_j}{\sum_{j \neq i} T_j}) - (1 - T_i) \widehat{f}(0, \frac{\sum_{j \neq i} E_{ij} T_j}{\sum_{j \neq i} T_j})$;

Take $\tau_{(i)}^a = n^{-1} \sum_{j \neq i} 2 T_j^* (\widehat{f}(1, \frac{M_j^*}{N_j^*}) + \widehat{\varepsilon}_j) - 2(1 - T_j^*)(\widehat{f}(0, \frac{M_j^*}{N_j^*}) + \widehat{\varepsilon}_j)$, and

$\tau_{(i)}^b = n^{-1} \sum_{j \in [n]} 2 T_j^* (\widehat{f}(1, \frac{M_{j,(i)}^*}{N_{j,(i)}^*}) + \widehat{\varepsilon}_j) - 2(1 - T_i^*)(\widehat{f}(0, \frac{M_{j,(i)}^*}{N_{j,(i)}^*}) + \widehat{\varepsilon}_j)$;

Take $\overline{\tau}^a = n^{-1} \sum_{i \in [n]} \tau_{(i)}^a$, $\overline{\tau}^b = n^{-1} \sum_{i \in [n]} \tau_{(i)}^b$, and $\widehat{K}_n = n \sum_{i \in [n]} (\tau_{(i)}^a - \overline{\tau}^a + \tau_{(i)}^b - \overline{\tau}^b)^2$.

return $\widehat{K}_n$.

Our inference procedure consists of three steps. In step 1, we estimate f non-parametrically by $\widehat{f}$. In step 2, we construct *two* types of plug-in and leave-one-out estimators, denoted by $\{\tau_{(i)}^a\}_{i \in [n]}$ and $\{\tau_{(i)}^b\}_{i \in [n]}$ respectively. $\tau_{(i)}^a$ accounts for the randomness from flipping i-th unit's own treatment. $\tau_{(i)}^b$ accounts for randomness from flipping j-th unit's treatment, where j is a neighbor of i. In Step 3, we form our final variance estimator using the resampling based treatment effect estimators similar to the i.i.d. case. The supplementary appendix (Lemma SA-18, SA-19) presents formal results on the convergence guarantees for these estimators.

We outline the construction of our proposed feasible, robust prediction interval for τ_n in the following algorithm.

Algorithm 3: Feasible Prediction Interval for τ_n

Input: Treatments and outcomes $(T_i, Y_i)_{i \in [n]}$, realized graph **E**, non-parametric learner $\widehat{f}$ of f.

Output: A data-driven $(1 - \alpha_1 - \alpha_2)$ prediction interval $\widehat{\mathcal{C}}(\alpha_1, \alpha_2)$ for τ_n.

Get $\widehat{K}_n$ from Algorithm 2 using the treatments and outcomes $(T_i, Y_i)_{i \in [n]}$, the realized random graph **E**, a non-parametric learner $\widehat{f}$ for f;

Get $\widehat{\mathcal{C}}(\alpha_1, \alpha_2)$ from Algorithm 1 given $(T_i, Y_i)_{i \in [n]}$ and $\widehat{K}_n$.

return $\widehat{\mathcal{C}}(\alpha_1, \alpha_2)$.

The next theorem establishes the validity of our feasible, robust prediction interval for τ_n.

Theorem 4.2 (Feasible Robust Prediction Interval)**.** *Suppose Assumptions 1, 2, 3, and 4 hold. Suppose the non-parametric learner $\widehat{f}$ satisfies $\widehat{f}(\ell, \cdot) \in C_2([0,1])$, and $|\widehat{f}(\ell, \pi_*) - f(\ell, \pi_*)| = o_{\mathbb{P}}(1)$, $|\partial_2 \widehat{f}(\ell, \pi_*) - \partial_2 f(\ell, \pi_*)| = o_{\mathbb{P}}(1)$, for $\ell \in \{0, 1\}$. If $n\rho_n^3 \to \infty$, then the prediction interval $\widehat{\mathcal{C}}(\alpha_1, \alpha_2)$ given in Algorithm 3 satisfies*

$$\liminf_{n \to \infty} \sup_{\beta \in [0,1]} \mathbb{P}_\beta[\tau_n \in \widehat{\mathcal{C}}(\alpha_1, \alpha_2)] \geq 1 - \alpha_1 - \alpha_2.$$

4.3 Simulations

We exhibit the finite sample performance of our robust inference procedure. Take $(U_i : 1 \leq i \leq n)$ i.i.d Uniform$([0,1])$-distributed, graph function $G(\cdot, \cdot) \equiv 0.5$ and density $\rho_n = 0.5$. The Ising-treatments satisfy Assumption 3 with various n and β. Y_i has data generating process $Y_i = \mathbb{1}(T_i = 1)f(1, M_i/N_i) + \mathbb{1}(T_i = 0)f(0, M_i/N_i) + \varepsilon_i$, with $f(x_1, x_2) = x_1^2 + x_1(x_2 + 1)^2$, $(x_1, x_2) \in \mathbb{R}^2$ and $(\varepsilon_i : 1 \leq i \leq n)$ are i.i.d $N(0, 0.05)$ noise terms independent to $((U_i, T_i) : 1 \leq i \leq n)$. The Monte-Carlo simulations are repeated with 5000 iterations and look at the $1 - \alpha$ prediction interval with $\alpha = 0.1$.

Figure 1 (a) and (b) demonstrate the empirical coverage and interval length against β, while fixing $n = 500$. To compare multiple methods, `conserv` stands for Algorithm 3, "$\beta = 0$" stands for using the formula from Theorem 3.1, `Oracle` stands for using the law $n^{-1/2}\widehat{\kappa}_2^{1/2}\mathsf{Z} + n^{-1/4}\widehat{\kappa}_1 \mathsf{W}_{c_{\beta,n}}$ from Theorem 3.2 with $c_{\beta,n} = \sqrt{n}(1 - \beta)$ assumed to be known, and `Onestep` stands for Algorithm 1 but taking the first step confidence interval $\mathcal{I}(\alpha_1)$ to be the full range $[0,1]$ instead. For interval length, `Simulated` stands for the true interval length from Monte-Carlo simulations. `Conservative` and `Onestep` remain conservative except when β is close to 1, due to the second step in Algorithm 1 taking maximum quantile from $\beta \in \mathcal{I}(\alpha_1)$; `Oracle` has empirical coverage close to $1 - \alpha$ and interval length close to the true interval length from Monte-Carlo simulation; the approach of plugging in $\beta = 0$ becomes invalid as β deviates from zero. Figure 1 (c) and (d) demonstrate log-log plots of interval length against sample size, fixing $\beta = 0$. While the Monte-Carlo interval length `Simulated` interval length $\propto n^{-0.52}$, consistent with the $\sqrt{n}$-convergence with $\beta = 0$, `Conserv` has interval length $\propto n^{-0.34}$, an effect of taking the maximum quantile among $\beta \in \mathcal{I}(\alpha_1)$. Full replication codes and data are publicly available at the GitHub repository (https://github.com/ruiiiiqi/network_interference).

5 Extensions and Generalizations

We present pointwise distributional approximations under more general interference regimes in treatment assignment. These results demonstrate additional technical complications: developing
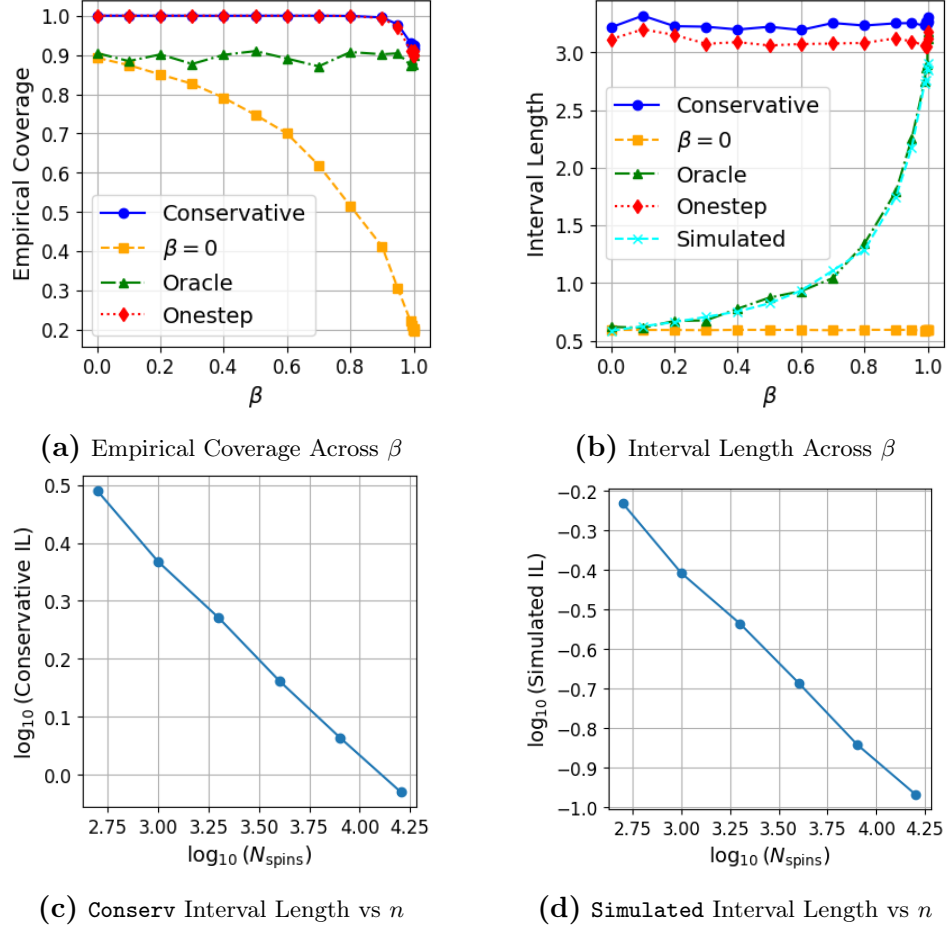
(a) Empirical Coverage Across β

(b) Interval Length Across β

(c) `Conserv` Interval Length vs n

(d) `Simulated` Interval Length vs n

Figure 1: (a) and (b) are empirical coverages and interval lengths of four methods across $\beta \in [0, 1]$: `Conservative` and `Onestep` remain conservative except when β is close to 1; `Oracle` has empirical coverage close to $1 - \alpha$ and interval length close to the true interval length from Monte-Carlo simulation; the approach of plugging in $\beta = 0$ becomes invalid as β deviates from zero. (c) shows `Conserv` interval length $\propto n^{-0.34}$. (d) shows `Simulated` interval length $\propto n^{-0.52}$.

valid (pointwise and uniform in β) inference procedures in the settings considered in this section is substantially more challenging, and thus beyond the scope of this paper.

5.1 Low Temperature Treatment Assignment

The low temperature regime corresponds to $\beta > 1$ in Assumption 3. In this case, the Hájek estimator concentrates in probability around a different (conditional) direct treatment effect, which now also depends on which side of the half line the random variable $\mathrm{sgn}(m) = \mathrm{sgn}(2n^{-1} \sum_{i=1}^{n} T_i - 1)$ lies on, where $\mathrm{sgn}(x) = +$ if $x \geq 0$, and $\mathrm{sgn}(x) = -$ if $x < 0$. This is due to the convergence of M_i/N_i to a two-point distribution depending on $\mathrm{sgn}(m)$. We thus define

$$\tau_{n,\ell} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[Y_i(1; \mathbf{T}_{-i}) - Y_i(0; \mathbf{T}_{-i}) | f_i(\cdot), \mathbf{E}, \mathrm{sgn}(m) = \ell], \qquad \ell \in \{-, +\},$$

which is a new causal predictand in the context of our causal inference model with interference.

The following lemma gives an analogue of Theorem 3.1 in the low temperature regime.

Lemma 5.1. *Under the Assumptions 1, 2 and 3 hold with $\beta > 1$, then*

$$\sup_{t \in \mathbb{R}} \max_{\ell \in \{-,+\}} |\mathbb{P}(\widehat{\tau}_n - \tau_{n,\ell} \leq t | \operatorname{sgn}(m) = \ell) - L_n(t; \beta, \kappa_{1,\ell}, \kappa_{2,\ell})| = O\left(\sqrt{\frac{n \log n}{(n\rho_n)^{p+1}}} + \frac{\log n}{\sqrt{n\rho_n}}\right),$$

where

$$L_n(t; \beta, \kappa_{1,\ell}, \kappa_{2,\ell}) = \mathbb{P}\left[n^{-1/2}\left(\kappa_{2,\ell}(1 - \pi_*^2) + \kappa_{1,\ell}^2 \frac{\beta(1 - \pi_*^2)}{1 - \beta(1 - \pi_*^2)}\right)^{1/2} \mathsf{Z} \leq t\right],$$

with $\kappa_{s,\ell} = \mathbb{E}[(R_{i,\ell} - \mathbb{E}[R_{i,\ell}] + Q_{i,\ell})^s]$,

$$R_{i,\ell} = (1 + \pi_\ell)^{-1} f_i\left(1, \frac{\pi_\ell + 1}{2}\right) + (1 - \pi_\ell)^{-1} f_i\left(0, \frac{\pi_\ell + 1}{2}\right),$$

$$Q_{i,\ell} = \mathbb{E}\left[\frac{G(U_i, U_j)}{\mathbb{E}[G(U_i, U_j)|U_j]} \frac{1}{2}\left(f_j'\left(1, \frac{\pi_\ell + 1}{2}\right) - f_j'\left(0, \frac{\pi_\ell + 1}{2}\right)\right)\Big|U_i\right],$$

π_* the positive root of $x = \tanh(\beta x)$, $\pi_+ = 1/2 + \pi_*/2$, $\pi_- = 1/2 - \pi_*/2$, and $\mathsf{Z} \sim \mathsf{N}(0,1)$ independent of m.

Due to the change of centering, establishing uniformly valid inference is substantially more challenging in this context.

5.2 Asymmetric Treatment Assignment

The following assumption allows for unequal treatment assignment probabilities, and thereby providing another (partial) generalization of Assumption 3.

Assumption 5 (Ising Asymmetric Treatment Assignment). *Let $\beta \in \mathbb{R}_+$ and let $h \in \mathbb{R} \setminus \{0\}$ denote, respectively, the interaction strength and the external field of a Curie–Weiss model. The treatment vector $\mathbf{T} = (T_1, \ldots, T_n)^\top \in \{0,1\}^n$ is assumed to follow the joint distribution*

$$\mathbb{P}_{\beta,h}(\mathbf{T} = \mathbf{t}) = \frac{1}{C_n(\beta, h)} \exp\left(\frac{\beta}{n} \sum_{1 \leq i < j \leq n} (2t_i - 1)(2t_j - 1) + h \sum_{i=1}^n (2t_i - 1)\right),$$

where $\mathbf{t} = (t_1, \ldots, t_n) \in \{0,1\}^n$, and $C_n(\beta, h)$ is the normalizing constant.

The external field h induces unequal marginal treatment probabilities, so that $\mathbb{P}_{\beta,h}(T_i = 1) = (1 + \pi)/2$ with $\pi \in (-1,1)$ satisfying $\pi = \tanh(\beta\pi + h)$. Setting $h = 0$ recovers the equiprobable mechanism of Assumption 3.

The following lemma gives an analogue of Theorem 3.1 for the generalized treatment assignment mechanism.

Lemma 5.2. *Suppose Assumptions 1, 2 and 5 hold. Then,*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}[\widehat{\tau}_n - \tau_n \leq t] - L_n(t; \beta, h, \kappa_1, \kappa_2)| = O\left(\frac{\log n}{\sqrt{n\rho_n}} + \sqrt{\frac{n \log n}{(n\rho_n)^{p+1}}}\right),$$

where

$$L_n(t; \beta, h, \kappa_1, \kappa_2) = \mathbb{P}\left[n^{-1/2}\left(\kappa_2(1 - \pi^2) + \kappa_1^2 \frac{\beta(1 - \pi^2)^2}{1 - \beta(1 - \pi^2)}\right)^{1/2} \mathsf{Z} \leq t\right]$$

12

with $Z \sim \mathsf{N}(0,1)$, π the unique solution to $x = \tanh(\beta x + h)$,

$$R_i = (1+\pi)^{-1} f_i\Big(1, \frac{\pi+1}{2}\Big) + (1-\pi)^{-1} f_i\Big(0, \frac{\pi+1}{2}\Big),$$

$$Q_i = \mathbb{E}\Big[\frac{G(U_i, U_j)}{\mathbb{E}[G(U_i, U_j)|U_j]} \frac{1}{2} \Big(f'_j\Big(1, \frac{\pi+1}{2}\Big) - f'_j\Big(0, \frac{\pi+1}{2}\Big) \Big) \Big| U_i \Big],$$

and $\kappa_s = \mathbb{E}[(R_i - \mathbb{E}[R_i] + Q_i)^s]$.

This results demonstrates that the Hájek estimator is $\sqrt{n}$-consistent and asymptotically Gaussian, with a limiting distribution generalizing our previous result for the high temperature case $h = 0$ and $\beta \in [0, 1)$. Unfortunately, establishing a general distributional approximation that is valid (poitwise and) uniform over $(h, \beta) \in \mathbb{R} \times \mathbb{R}_+$ is quite challenging.

5.3   Block Treatment Assignment

We have assumed that treatment assignments interfere with each other through a fully connected network, but this assumption can be relaxed by imposing a block structure in the underlying latent network.

Assumption 6 (Ising Block Treatment Assignment)**.** *For a fixed $K \in \mathbb{N}$, suppose $[n]$ has a partition $[n] = \sqcup_{k=1}^{K} \mathcal{C}_k$. For each partition $k = 1, \ldots, K$, assume the exists a positive vector $\mathbf{p} = (p_1, \cdots, p_K)$ such that the sample sizes $n_k = \sum_{i=1}^{n} \mathbb{1}(i \in \mathcal{C}_k)$ satisfy $n_k/n = p_k + O(n^{-1/2})$ as $n \to \infty$. With parameters $(\beta_k, h_k) \in \mathbb{R}_+ \times \mathbb{R}$ for $k = 1, \ldots, K$, the Ising block treatment assignment mechanism is*

$$\mathbb{P}_{\boldsymbol{\beta}, \boldsymbol{h}}(\mathbf{T} = \mathbf{t}) = \prod_{k=1}^{K} \mathbb{P}_{\boldsymbol{\beta}, \boldsymbol{h}}[(T_i, i \in \mathcal{C}_k) = (t_i, i \in \mathcal{C}_k)]$$

with $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)^{\top}$, $\boldsymbol{h} = (h_1, \ldots, h_K)^{\top}$,

$$\mathbb{P}_{\boldsymbol{\beta}, \boldsymbol{h}}[(T_i, i \in \mathcal{C}_k) = (t_i, i \in \mathcal{C}_k)] \propto \exp\Big(\frac{\beta_k}{n} \sum_{i,j \in \mathcal{C}_k, i<j} (2t_i - 1)(2t_j - 1) + h_k \sum_{i \in \mathcal{C}_k} (2t_i - 1) \Big).$$

Denote the collections of high temperature or nonzero external field blocks by $\mathcal{H} = \{k \in [K] : h_k = 0, \beta_k \in [0, 1)$ or $h_k \neq 0\}$, the collections of critical temperature blocks by $\mathcal{C} = \{k \in [K] : h_k = 0, \beta_k = 1\}$, and the collections of low temperature blocks by $\mathcal{L} = \{k \in [K] : h_k = 0, \beta_k > 1\}$. For block $k \in \mathcal{H} \cup \mathcal{C}$, π_k denotes the unique solution to $x = \tanh(\beta_k x + h_k)$. For block $k \in \mathcal{L}$, $\pi_{k,+}$ and $\pi_{k,-}$ denote the unique positive and negative solutions to $x = \tanh(\beta_k x + h_k)$, respectively. Due to the potential existence of low temperature blocks, we use $\boldsymbol{sgn}$ to collect the average spins in all low temperature blocks, and fill in the positions for high and critical temperature blocks with zeros:

$$\boldsymbol{sgn} = (\mathrm{sgn}(m_1)\mathbb{1}(1 \in \mathcal{L}), \cdots, \mathrm{sgn}(m_K)\mathbb{1}(K \in \mathcal{L})).$$

Let $\mathcal{S}$ denote the collection of all possible configurations of $\boldsymbol{sgn}$:

$$\mathcal{S} = \{(s_k)_{1 \le k \le K} : s_k = - \text{ or } + \text{ if } k \in \mathcal{L}, s_k = 0 \text{ otherwise}\}.$$

Finally, we denote the conditional fixed point of block k based on $\boldsymbol{sgn} = B_S$ by

$$\pi_{k,(B_S)} = \begin{cases} \pi_k, & \text{if } k \in \mathcal{H} \cup \mathcal{C}, \\ \pi_{k, s_k}, & \text{if } k \in \mathcal{L}, \end{cases} \qquad 1 \le k \le K, B_S \in \mathcal{S}.$$

In this setting, we consider the blockwise (conditional) average treatment effects $\boldsymbol{\tau}_n = (\tau_{n,1}, \cdots, \tau_{n,K})^\top$, where

$$\tau_{n,k} = \frac{1}{n_k} \sum_{i \in \mathcal{C}_k} \mathbb{E}[Y_i(1; \mathbf{T}_{-i}) - Y_i(0; \mathbf{T}_{-i})|f_i(\cdot), \mathbf{E}, \boldsymbol{sgn}], \qquad k = 1, \cdots, K,$$

and the associated K-dimensional Hájek estimator is $\widehat{\boldsymbol{\tau}}_n = (\widehat{\tau}_{n,1}, \cdots, \widehat{\tau}_{n,K})^\top$, where

$$\widehat{\tau}_{n,k} = \frac{\sum_{i \in \mathcal{C}_k} T_i Y_i}{\sum_{i \in \mathcal{C}_k} T_i} - \frac{\sum_{i \in \mathcal{C}_k} (1 - T_i) Y_i}{\sum_{i \in \mathcal{C}_k} (1 - T_i)}, \qquad k = 1, \cdots, K.$$

The following theorem gives an analogue of Theorem 3.1 in this generalized setting. Let $\mathcal{R}$ be the collection of all hyperrectangles in $\mathbb{R}^K$.

Lemma 5.3. *Suppose Assumptions 1, 2 and 6 hold. Then,*

$$\max_{\mathbf{s} \in \mathcal{S}} \sup_{A \in \mathcal{R}} |\mathbb{P}(\widehat{\boldsymbol{\tau}}_n - \boldsymbol{\tau}_n \in A|\boldsymbol{sgn} = \mathbf{s}) - L_n(A; \boldsymbol{\beta}, \boldsymbol{h})|$$

$$= O\Big(\frac{n^{1/2}}{(n\rho_n)^{(p+1)/2}} + \Big(\frac{\log^7 n}{n}\Big)^{1/6} + \mathbf{r}_{n,\boldsymbol{\beta},\boldsymbol{h}}\Big),$$

where

$$L_n(A; \boldsymbol{\beta}, \boldsymbol{h}) = \mathbb{P}\Big(n^{-1/2}\Big(\sum_{k=1}^K \mathbb{E}[\mathbf{S}_{k,i,(B_S)} \mathbf{S}_{k,i,(B_S)}^\top](1 - \pi_{k,(B_S)}^2) p_k^2\Big)^{1/2} \mathbf{Z}_K$$

$$+ n^{-1/2} \sum_{k \in \mathcal{H} \cup \mathcal{L}} p_k \sigma_{k,(B_S)} \mathbb{E}[\mathbf{S}_{k,i,(B_S)}] \mathsf{Z}_{(k)}$$

$$+ n^{-1/4} \sum_{k \in \mathcal{C}} p_k \mathbb{E}[\mathbf{S}_{k,i,(B_S)}] \mathsf{R}_{(k)} \in A\Big)$$

with $\mathbf{Z}_K \sim \mathsf{N}(\mathbf{0}, \mathbf{I})$, where $\mathbf{0}$ is a K-dimensional vector of zeros and $\mathbf{I}$ is the $K \times K$ identity matrix, $\mathsf{Z}_{(k)} \sim \mathsf{N}(0, 1)$ for $k \in \mathcal{H} \cup \mathcal{L}$ independent, $\mathsf{R}_{(k)}$ for $k \in \mathcal{C}$ independent random variables with distribution function

$$F_0(t) = \frac{\int_{-\infty}^t \exp(-z^4/12) dz}{\int_{-\infty}^\infty \exp(-z^4/12) dz}, \qquad t \in \mathbb{R},$$

and $\mathbf{Z}_K$, $(\mathsf{Z}_{(k)} : k \in \mathcal{H} \cup \mathcal{L})$ and $(\mathsf{R}_{(k)} : k \in \mathcal{C})$ mutually independent. In addition, $\mathbf{S}_{l,i,(B_S)} = (S_{1,l,i,(B_S)}, \cdots, S_{K,l,i,(B_S)})^\top$ with $S_{k,l,i,(B_S)} = Q_{i,(B_S)} + \mathbb{1}(k = l) p_k^{-1}(R_{i,l,(B_S)} - \mathbb{E}[R_{i,l,(B_S)}])$, $1 \leq k, l \leq K$ and $1 \leq i \leq n$,

$$R_{i,l,(B_S)} = \frac{g_i(1, \overline{\pi}_{(B_S)})}{1 + \pi_{l,(B_S)}} + \frac{g_i(-1, \overline{\pi}_{(B_S)})}{1 - \pi_{l,(B_S)}}, \qquad \overline{\pi}_{(B_S)} = \sum_{k=1}^K p_k \pi_{k,(B_S)},$$

$$Q_{i,(B_S)} = \mathbb{E}\Big[\frac{G(U_i, U_j)}{\mathbb{E}[G(U_i, U_j)|U_j]}(g_j'(1, \overline{\pi}_{(B_S)}) - g_j'(-1, \overline{\pi}_{(B_S)}))\Big|U_i\Big],$$

and

$$_{n,\boldsymbol{\beta},\boldsymbol{h}} = \begin{cases} \sqrt{\log n}(n\rho_n)^{-1/2} & \text{if } \sum_{k=1}^K \mathbb{1}(k \in \mathcal{C}) = 0 \\ \sqrt{\log n}(n\rho_n^2)^{-1/4} & \text{if } \sum_{k=1}^K \mathbb{1}(k \in \mathcal{C}) > 0 \end{cases}.$$

14

If there is no low temperature block, then $\boldsymbol{sgn} = (0, \cdots, 0)$ almost surely, and S is the singleton set containing $(0, \cdots, 0)$. Then, the result reduces to the unconditional distributional approximation. More generally, when comparing to the single-block results, Lemma 5.3 shows that an *overflow* behavior happens: as long as there is at least one low temperature block, the Hajek estimators for all blocks converge to the conditional estimand given low temperature block average Ising spin signs. Moreover, the distribution approximation error depends on whether there is at least one low temperature block. This is because unit i's potential outcome $Y_i(T_i; \mathbf{T}_{-i}) = f_i(T_i; M_i/N_i)$ depends on units from all blocks through M_i/N_i. Finally, the additional, slower rate $\log(n)^{7/6}n^{-1/6}$ comes from a multivariate conditional i.i.d Berry-Esseen result (see Section 3.3) with a degenerate covariance matrix, via [6, Theorem 2.1].

Supplementary material

The supplementary appendix presents more general theoretical results encompassing those discussed in the paper, and presents the proofs of those general results.

References

[1] Eric Auerbach, Hongchang Guo, and Max Tabord-Meehan. The local approach to causal inference under network interference. *arXiv preprint arXiv:2105.03810*, 2025.

[2] Bhaswar B Bhattacharya and Sumit Mukherjee. Inference in ising models. *Bernoulli*, 24(1):493–525, 2018.

[3] Sohom Bhattacharya and Subhabrata Sen. Causal effect estimation under network interference with mean-field methods. *Annals of Statistics*, 2025.

[4] Peter J Bickel and Aiyou Chen. A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences*, 106(50):21068–21073, 2009.

[5] Sourav Chatterjee and Qi-Man Shao. Nonnormal approximation by stein's method of exchangeable pairs with application to the curie–weiss model. *Annals of Applied Probability*, 21(2):464–483, 2011.

[6] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Central limit theorems and bootstrap in high dimensions. *Annals of Probability*, 45(4):2309 – 2352, 2017.

[7] Peter Eichelsbacher and Matthias Loewe. Stein's method for dependent random variables occuring in statistical mechanics. *Electronic Journal of Probability*, 15:962 – 988, 2010.

[8] Richard S Ellis. *Entropy, Large Deviations, and Statistical Mechanics*. Taylor & Francis, 2006.

[9] Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC, 2020.

[10] Yuchen Hu, Shuangning Li, and Stefan Wager. Average direct and indirect causal effects under interference. *Biometrika*, 109(4):1165–1172, 2022.

[11] Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

[12] Chanhwa Lee, Donglin Zeng, and Michael G Hudgens. Efficient nonparametric estimation of stochastic policy effects with clustered interference. *Journal of the American Statistical Association*, 120(549):382–39, 2025.

[13] Michael P Leung. Causal inference under approximate neighborhood interference. *Econometrica*, 90(1):267–293, 2022.

[14] Shuangning Li and Stefan Wager. Random graph asymptotics for treatment effect estimation under network interference. *Annals of Statistics*, 50(4):2334–2358, 2022.

[15] Qiaohui Lin, Robert Lunde, and Purnamrita Sarkar. On the theoretical properties of the network jackknife. In *International Conference on Machine Learning*, pages 6105–6115. PMLR, 2020.

[16] Adam Lipowski, Dorota Lipowska, and António Luis Ferreira. Phase transition and power-law coarsening in an ising-doped voter model. *Physical Review E*, 96(3):032145, 2017.

[17] Charles F Manski. Identification of treatment response with social interactions. *The Econometrics Journal*, 16(1):S1–S23, 2013.

[18] Rajarshi Mukherjee, Sumit Mukherjee, and Ming Yuan. Global testing against sparse alternatives under ising models. *Annals of Statistics*, 46(5):2062–2093, 2018.

[19] Elizabeth L Ogburn, Oleg Sofrygin, Ivan Diaz, and Mark J Van der Laan. Causal inference for social network data. *Journal of the American Statistical Association*, 119(545):597–611, 2024.

[20] Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical methods in medical research*, 21(1):55–75, 2012.

[21] Gonzalo Vazquez-Bare. Identification and estimation of spillover effects in randomized experiments. *Journal of Econometrics*, 237(1):105237, 2023.