Supplementary Material to "Treatment Network Effect Estimation under Dependence"

Matias D. Cattaneo[1], Yihan He[1], and Ruiqi (Rae) Yu[1]

[1]Department of Operations Research and Financial Engineering, Princeton University

February 18, 2025

Abstract

This Supplemental Material contains general theoretical results encompassing those discussed in the main paper, includes proofs of those general results, and discusses additional methodological and technical results.

Contents

SA-1 Notation

For a sequence of real-valued random variables X_n, we say $X_n = O_{\psi_p}(r_n)$ if there exists $N \in \mathbb{N}$ and $M > 0$ such that $\|X_n\|_{\psi_p} \leq M r_n$ for all $n \geq N$, where $\|\cdot\|_{\psi_p}$ is the Orlicz norm w.r.p $\psi_p(x) = \exp(x^p) - 1$. We say $X_n = O_{\psi_p, tc}(r_n)$, tc stands for tail control, if there exists $N \in \mathbb{N}$ and $M > 0$ such that for all $n \geq N$ and $t > 0$, $\mathbb{P}(|X_n| \geq t) \leq 2n \exp(-(t/(M r_n))^p) + M n^{-1/2}$.

SA-2 Berry-Esseen Results for Curie-Weiss magnetization with Independent Multipliers

For $\beta \geq 0$, the *Curie-Weiss model of ferromagnetic interaction* at inverse temperature β and zero external field is given by the following Gibbs measure on $\{-1, +1\}^n$:

$$P_\beta(\mathbf{w}) = \frac{1}{Z_\beta} \exp\left(\frac{\beta}{n} \sum_{i<j} w_i w_j\right), \quad \mathbf{w} = (w_1, \cdots, w_n) \in \{-1, 1\}^n, \tag{SA-1}$$

where Z_β is the normalizing constant.

Suppose $\mathbf{W} = (W_1, \cdots, W_n)$ is a random vector with law P_β. Then $\mathbb{E}[W_i] = 0$ and $m = n^{-1} \sum_{i=1}^n W_i$. The Curie-Weiss model has a phase transition phenomena between regimes. The case $0 \leq \beta \leq 1$ is called the *high temperature* regime, where m concentrates around 0. The case $\beta > 1$ is called the *low temperature* regime, where m concentrates on the set $\{-\pi_*, \pi_*\}$, π_* being the unique positive solution to $x = \tanh(\beta x)$. The case $\beta = 1$ is called the *critical temperature* regime.

Suppose $\mathbf{X} = (X_1, \cdots, X_n)$ has i.i.d components such that $\mathbb{E}\left[|X_1|^3\right] < \infty$ independent to $\mathbf{W}$. The goal is to study the limiting distribution and the rate of convergence for

$$g_n = n^{-1} \sum_{i=1}^n W_i X_i.$$

The magnetization $n^{-1} \sum_{i=1}^n W_i$ has been studied using Stein's method [5], [3]. Due to the multipliers, the Stein's method can not be directly applied for g_n. We use a novel strategy based on the following de Finetti's lemma to show Berry Essseen results.

Lemma SA-1 (de Finetti's Lemma). *There exists a latent variable* U_n *with density*

$$f_{\mathsf{U}_n}(u) = I_{\mathsf{U}_n}^{-1} \exp\left(-\frac{1}{2}u^2 + n \log\cosh\left(\sqrt{\frac{\beta}{n}} u\right)\right),$$

where $I_{\mathsf{U}_n} = \int_{-\infty}^{\infty} \exp(-\frac{1}{2}u^2 + n \log\cosh(\sqrt{\frac{\beta}{n}}u))du$*, such that* $W_1, \cdots, W_n$ *are i.i.d condition on* U_n.

Lemma SA-2. *Take* U_n *to be a random variable with density function* $f_{\mathsf{U}_n}(u) = I_{\mathsf{U}_n}^{-1} \exp(-\frac{1}{2}u^2 + n \log\cosh(\sqrt{\frac{\beta}{n}}u + h))$ *where* $I_{\mathsf{U}_n} = \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}u^2 + n \log\cosh\left(\sqrt{\beta/n}u + h\right)\right) du$*. Take* $\mathsf{W}_n = n^{-\frac{1}{4}} \mathsf{U}_n$*. Then*

1. *High-temparature case: Suppose* $h \neq 0$ *or* $h = 0, \beta < 1$*. Then* $\|\mathsf{U}_n - \mathbb{E}[\mathsf{U}_n]\|_{\psi_2} \lesssim 1$.
2. *Critical-temparature case: Suppose* $h = 0$ *and* $\beta = 1$*. Then* $\|\mathsf{U}_n\|_{\psi_2} \lesssim n^{1/4}$.

3. *Low-temparature case: Suppose $h = 0$ and $\beta > 1$. Then condition on $\mathsf{U}_n \in \mathcal{C}_l$, $\|\mathsf{U}_n - \mathbb{E}[\mathsf{U}_n|\mathsf{U}_n \in \mathcal{C}_l]\|_{\psi_2} \lesssim 1$.*

4. *Drifting sequence case: Suppose $h = 0$, $\beta = 1 - cn^{-\frac{1}{2}}, c \in \mathbb{R}^+$. Then $\|\mathsf{U}_n\|_{\psi_2} \leq \mathsf{C}n^{1/4}$ for large enough n with C not depending on β.*

Fix $\beta > 0$. We characterize the limiting distribution of $n^{-1}\sum_{i=1}^n W_i X_i$ and the rate of convergence as $n \to \infty$ in the following lemma. In particular, we will see that the limiting distribution changes from a Gaussian distribution under high temperature, to a non-Gaussian distribution under critical temperature, to a Gaussian mixture under low temperature.

Lemma SA-3 (Fixed Temperature Berry-Esseen)**.** *Then*

1. *When $\beta < 1$,*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(n^{\frac{1}{2}}(\mathbb{E}[X_i^2] + \mathbb{E}[X_i]^2 \frac{\beta}{1-\beta})^{-\frac{1}{2}} g_n \leq t) - \Phi_{N(0,1)}(t)| = O(n^{-\frac{1}{2}}).$$

2. *When $\beta = 1$, denote $F_0(t) = \frac{\int_{-\infty}^t \exp(-z^4/12)dz}{\int_{-\infty}^\infty \exp(-z^4/12)dz}, t \in \mathbb{R}$, then*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(n^{\frac{1}{4}} \mathbb{E}[X_i]^{-1} g_n \leq t) - F_0(t)| = O((\log n)^3 n^{-\frac{1}{2}}).$$

3. *When $\beta > 1$, denote $g_{n,\ell} = \frac{1}{n}\sum_{i=1}^n X_i(W_i - \pi_\ell)$, $\mathcal{C}_+ = [0, \infty)$ and $\mathcal{C}_- = (-\infty, 0)$, then*

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(n^{\frac{1}{2}} \left(\mathbb{E}[X_i^2](1-\pi_\ell^2) + \mathbb{E}[X_i]^2 \frac{\beta(1-\pi_\ell^2)}{1-\beta(1-\pi_\ell^2)}\right)^{-\frac{1}{2}} g_{n,\ell} \leq t | m \in \mathcal{C}_\ell) - \Phi_{N(0,1)}(t)|$$
$$= O(n^{-\frac{1}{2}}), \quad t \in \{-, +\}.$$

Lemma SA-4 (Size-Dependent Temperature Berry-Esseen)**.** *Suppose Z is a standard Gaussian random variable. (1) Suppose $\beta_n = 1 + cn^{-\frac{1}{2}}$, where $c < 0$ does not depend on n. Then*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(n^{\frac{1}{4}} g_n \leq t) - \mathbb{P}(n^{-\frac{1}{4}} \mathbb{E}[X_i^2]^{\frac{1}{2}} \mathsf{Z} + \beta_n^{\frac{1}{2}} \mathbb{E}[X_i] \mathsf{W}_c \leq t) \right| = O((\log n)^3 n^{-\frac{1}{2}}),$$

where $O(\cdot)$ is up to a universal constant.

(2) Suppose $\beta_n = 1 + cn^{-\frac{1}{2}}$, where $c > 0$ does not depend on n. Then

$$\sup_{c \in \mathbb{R}^+} \sup_{t \in \mathbb{R}} \left| \mathbb{P}(n^{\frac{1}{4}} g_n \leq t | m \in \mathcal{I}_{c,\ell}) - \mathbb{P}(n^{-\frac{1}{4}} \mathbb{E}[X_i^2]^{\frac{1}{2}} \mathsf{Z} + \beta_n^{\frac{1}{2}} \mathbb{E}[X_i] \mathsf{W}_{c,n} \leq t | \mathsf{W}_{c,n} \in \mathcal{I}_{c,\ell}) \right|$$
$$= O((\log n)^3 n^{-\frac{1}{2}}),$$

with $\mathcal{I}_{c,n,-} = (-\infty, K_{c,n,-})$ and $\mathcal{I}_{c,n,+} = (K_{c,n,+}, \infty)$ such that $\mathbb{E}[\mathsf{W}_{c,n}|\mathsf{W}_{c,n} \in \mathcal{I}_{c,n,\ell}] = w_{c,n,\ell}$ for $\ell \in \{-, +\}$.

Lemma SA-5 ($\sqrt{n}$-sequence is knife-edge)**.** *(1) Suppose $|\beta_n - 1| = o(n^{-\frac{1}{2}})$, then*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}(n^{\frac{1}{4}} g_n \leq t) - \mathbb{P}(\mathbb{E}[X_i] \mathsf{W}_0 \leq t) \right| = o(1).$$

(2) Suppose $1-\beta_n \gg n^{-\frac{1}{2}}$*, then*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}(\mathbb{V}[g_n]^{-\frac{1}{2}}g_n \leq t)-\Phi(t)\right| = o(1).$$

(3) Suppose $\beta_n - 1 \gg n^{-\frac{1}{2}}$*, then for* $\ell \in \{-,+\}$*,*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\Big(\mathbb{V}[g_n|m\in\mathcal{I}_\ell])^{-\frac{1}{2}}(g_n-\mathbb{E}[g_n|m\in\mathcal{I}_\ell]) \leq t\Big)-\Phi(t)\right| = o(1),$$

where $\mathcal{I}_+ = [0,\infty)$ *and* $\mathcal{I}_- = (-\infty, 0)$*.*

SA-3 Pseudo-Likelihood Estimator for Curie-Weiss Regimes

Lemma SA-1 (No Consistent Variance Estimator)**.** *Suppose Assumptions 1,2,3 hold. Then there is no consistent estimator of* $n\mathbb{V}[\widehat{\tau}_n - \tau_n]$*.*

The pseudo-likelihood estimator for Curie-Weiss regime with no external field is given by

$$\begin{aligned}\widehat{\beta} &= \arg\max_{\beta}\sum_{i\in[n]}\log\mathbb{P}_\beta\left(W_i|W_{-i}\right)\\ &= \arg\max_{\beta}\sum_{i\in[n]}-\log\left(\frac{W_i\tanh(\beta n^{-1}\sum_{j\neq i}W_j)+1}{2}\right).\end{aligned}$$

Lemma SA-2 (Fixed Temperature Distribution Approximation)**.** *(1) If* $\beta\in[0,1)$*, then*

$$\widehat{\beta}\xrightarrow{d}\max\left\{1-\frac{1-\beta}{\chi^2(1)},0\right\}.$$

(2) If $\beta = 1$*, then*

$$n^{\frac{1}{2}}(1-\widehat{\beta})\xrightarrow{d}\max\left\{\frac{1}{\mathsf{W}_0^2}-\frac{\mathsf{W}_0^2}{3},0\right\}.$$

(3) If $\beta > 1$*, we define an unrestricted pseud-likelihood estimator,*

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

Then

$$\sup_{t\in\mathbb{R}}|\mathbb{P}(n^{1/2}(\widehat{\beta}_{UR}-\beta)\leq t|m\in\mathcal{I}_\ell)-\mathbb{P}((\frac{1-\beta(1-\pi_\ell^2)}{1-\pi_\ell^2})^{1/2}\mathsf{Z}\leq t)| = o(1).$$

Lemma SA-3 (Drifting Temperature Distribution Approximation)**.** *For any* $\beta\in[0,1]$*, define* $c_{\beta,n} = \sqrt{n}(1-\beta)$*, and suppose*

$$\mathbb{P}(z_{\beta,n}\leq t) = \mathbb{P}(\mathsf{Z}+n^{\frac{1}{4}}\mathsf{W}_{c_{\beta,n}}\leq t), \qquad t\in\mathbb{R}.$$

then

$$\sup_{\beta\in[0,1]}\sup_{t\in\mathbb{R}}|\mathbb{P}(1-\widehat{\beta}\leq t)-\mathbb{P}(\min\{\max\{z_{\beta,n}^{-2}-\frac{1}{3n}z_{\beta,n}^2,0\},1\}\leq t)| = o(1).$$

SA-4 Stochastic Linearization

Throughout this section, we prove under a more generic setting. We assume $W_i = 2T_i - 1$, and $(W_i)_{i\in[n]}$ satisfies a Curie-Weiss model with a possibly non-zero external field, that is,

Assumption 1 (Curie-Weiss)**.** *Suppose* $\mathbf{W} = (W_i)_{1\le i\le n}$ *are such that for some* $C_{\beta,h} \in \mathbb{R}$*,*

$$\mathbb{P}(\mathbf{W} = \mathbf{w}) = C_{\beta,h}^{-1} \exp\left(\frac{\beta}{n}\sum_{1\le i<j\le n} W_i W_j + h\sum_{i=1}^{n} W_i\right),$$

where $C_{\beta,h}$ *is a normalizing constant.*

Morever, for the ease of proof, we let g_i to be the function such that

$$g_i(x,y) = f_i(\frac{1}{2}x + \frac{1}{2}, \frac{1}{2}y + \frac{1}{2}), \qquad x \in \{-1,1\}, y \in [-1,1].$$

We denote $M_i = \sum_{j\neq i} E_{ij} W_i$, $N_i = \sum_{j\neq i} E_{ij}$. Then

$$g_i(T_i, \mathbf{T}_{-i}) = f_i(T_i, \frac{\sum_{j\neq i} E_{ij} T_i}{\sum_{j\neq i} E_{ij}}) = g_i(W_i, \frac{M_i}{N_i}).$$

Define $\pi = \mathbb{E}[W_i]$, $m = n^{-1}\sum_{i=1}^{n} W_i$ and for $1 \le i \le n$, $m_i = n^{-1}\sum_{j\neq i} W_j$. Define the following rates that will be used in the convergence analysis:

$$\mathtt{a}_{\beta,h} = \begin{cases} n^{1/2}, & \text{if } \beta \neq 1 \text{ or } \beta = 1, h \neq 0, \\ n^{3/4}, & \text{if } \beta = 1, h = 0, \end{cases} \qquad \mathtt{r}_{\beta,h} = \begin{cases} n^{1/2}, & \text{if } \beta \neq 1 \text{ or } \beta = 1, h \neq 0, \\ n^{1/4}, & \text{if } \beta = 1, h = 0. \end{cases}$$

and

$$\mathtt{p}_{\beta,h} = \begin{cases} 1/2, & \text{if } \beta \neq 1 \text{ or } \beta = 1, h \neq 0, \\ 1/4, & \text{if } \beta = 1, h = 0, \end{cases} \quad \psi_{\beta,h}(x) = \begin{cases} \exp(x^2) - 1, & \text{if } \beta \neq 1 \text{ or } \beta = 1, h \neq 0, \\ \exp(x^4) - 1, & \text{if } \beta = 1, h = 0. \end{cases}$$

SA-4.1 The Unbiased Estimator

Denote $p_i = \mathbb{P}(W_i = 1; \mathbf{W}_{-i}) = (\exp(-2\beta m_i - 2h) + 1)^{-1}$. We propose an unbiased estimator given by

$$\widehat{\tau}_{n,\mathrm{UB}} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{T_i Y_i}{p_i} - \frac{(1-T_i)Y_i}{1-p_i}\right].$$

Lemma SA-1 (Unbiased Estimator)**.** $\widehat{\tau}_{n,UB}$ *is an unbiased estimator for* τ_n *in the sense that,*

$$\mathbb{E}[\widehat{\tau}_{n,UB} | \mathbf{E}, (f_i)_{i\in[n]}] = \tau_n.$$

We will show the followings have weak limits:

$$n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\left[\frac{T_i Y_i}{p_i} - \frac{(1-T_i)Y_i}{1-p_i} - \tau_n\right].$$

W.l.o.g, we analyse the error for treated data, the error for control data follows in the same way. First, decompose by

$$n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \left[\frac{T_i Y_i}{p_i} - \frac{(1-T_i)Y_i}{1-p_i} \right] = \Delta_1 + \Delta_2,$$

$$\Delta_1 = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \left[\frac{T_i}{p_i} Y_i(1,\pi) - \frac{1-T_i}{1-p_i} g_i(-1,\pi) \right],$$

$$\Delta_2 = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \left[\frac{T_i}{p_i} \Big(g_i\Big(1, \frac{M_i}{N_i}\Big) - g_i\Big(1,\pi\Big) \Big) - \frac{1-T_i}{1-p_i} \Big(g_i\Big(-1, \frac{M_i}{N_i}\Big) - g_i\Big(-1,\pi\Big) \Big) \right].$$

Lemma SA-2. *Suppose Assumption 1,2, and 3 hold. Then*

$$\Delta_1 - \mathbb{E}[\Delta_1 | \mathbf{E}, (f_i)_{i \in [n]}] = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \Big(\frac{g_i(1,\pi)}{1+\pi} + \frac{g_i(-1,\pi)}{1-\pi} - \beta \mathtt{d} \Big) (W_i - \pi) + O_{\psi_2, tc}(\sqrt{\log n} n^{-\mathtt{r}_{\beta,h}}),$$

where $\mathtt{d} = (1-\pi)\mathbb{E}[g_i(1,\pi)] + (1+\pi)\mathbb{E}[g_i(-1,\pi)]$.

Now consider Δ_2. Since $\frac{T_i}{p_i} = \frac{T_i - p_i}{p_i} + 1$, we have the decomposition,

$$\Delta_2 = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \frac{T_i}{p_i} \left[g_i\left(1, \frac{M_i}{N_i}\right) - g_i\left(1,\pi\right) \right] = \Delta_{2,1} + \Delta_{2,2} + \Delta_{2,3} \tag{SA-2}$$

where

$$\Delta_{2,1} = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} g_i'(1,\pi) \left(\frac{M_i}{N_i} - \pi \right),$$

$$\Delta_{2,2} = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \frac{T_i - p_i}{p_i} g_i'\left(1,\pi\right) \left(\frac{M_i}{N_i} - \pi \right),$$

$$\Delta_{2,3} = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \frac{T_i Y_i''\left(1, \eta_i^*\right)}{2p_i} \left(\frac{M_i}{N_i} - \pi \right)^2$$

where η_i^* is some random quantity between $\frac{M_i}{N_i}$ and π. Define $b_i = \sum_{j \neq i} \frac{E_{ij}}{N_j} Y_j'\left(1,\pi\right)$. Then by reordering the terms,

$$\Delta_{2,1} = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} b_i \left(W_i - \pi\right).$$

Lemma SA-3. *Suppose Assumption 1,2,3 hold. Then condition on* $\mathbf{U}$ *such that* $A(\mathbf{U}) \in \mathscr{A} = \{A \in \mathbb{R}^{n \times n} : \min_{i \in [n]} \sum_{j \neq i} A_{ij} \geq 32 \log n\}$,

$$\Delta_{2,2} = O_{\psi_2, tc}\left(\log n \max_{i \in [n]} \mathbb{E}[N_i | \mathbf{U}]^{-1/2} \right) + O_{\psi_{\beta,\gamma}, tc}(\sqrt{\log n} n^{-\mathtt{r}_{\beta,h}}).$$

For the term $\Delta_{2,3}$, we further decompose it into two parts:

$$\Delta_{2,3} = \Delta_{2,3,1} + \Delta_{2,3,2},$$

where

$$\Delta_{2,3,1} = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \left[g_i\left(1, \frac{M_i}{N_i}\right) - g_i(1,\pi) - g_i'(1,\pi)\left(\frac{M_i}{N_i} - \pi\right)\right],$$

$$\Delta_{2,3,2} = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \frac{1}{2}\frac{W_i - \mathbb{E}[W_i|\mathbf{W}_{-i}]}{p_i} \left[g_i\left(1, \frac{M_i}{N_i}\right) - g_i(1,\pi) - g_i'(1,\pi)\left(\frac{M_i}{N_i} - \pi\right)\right].$$

Lemma SA-4. *Suppose 1,2,3 hold. Then condition on* $\mathbf{U}$ *such that* $A(\mathbf{U}) \in \mathscr{A} = \{A \in \mathbb{R}^{n\times n} : \min_{i\in[n]} \sum_{j\neq i} A_{ij} \geq 32 \log n\}$,

$$\begin{aligned}
&\Delta_{2,3,1} - \mathbb{E}[\Delta_{2,3,1}|\mathbf{E}, (f_i)_{i\in[n]}] \\
=&O_{\psi_{\mathtt{p}_{\beta,h}/2}}(n^{-\mathtt{r}_{\beta,h}}) + O_{\psi_{\beta,h},tc}(\max_i \mathbb{E}[N_i|\mathbf{U}]^{-1/2}) + O_{\psi_1,tc}(n^{-1/2}) \\
&\quad + O_{\psi_2,tc}(n^{\frac{1}{2}-\mathtt{a}_{\beta,h}} \max \mathbb{E}[N_i|\mathbf{U}]^{-1/2}).
\end{aligned}$$

Lemma SA-5. *Suppose 1,2,3 hold. If* $g_i(1,\cdot)$ *and* $g_i(-1,\cdot)$ *are 4-times continuously differentiable, then condition on* $\mathbf{U}$ *such that* $A(\mathbf{U}) \in \mathscr{A}$,

$$\begin{aligned}
&\Delta_{2,3,2} - \mathbb{E}[\Delta_{2,3,2}|\mathbf{E}, (f_i)_{i\in[n]}] \\
=&O_{\psi_{\mathtt{p}_{\beta,h}/2},tc}((\log n)^{-1/\mathtt{p}_{\beta,h}} n^{-2\mathtt{r}_{\beta,h}}) + O_{\psi_1,tc}((\log n)^{-1/\mathtt{p}_{\beta,h}}(\min_i \mathbb{E}[N_i|\mathbf{U}])^{-1}) \\
&+ O_{\psi_1,tc}\left(n^{1/2-\mathtt{a}_{\beta,h}}\left(\frac{\max_i \mathbb{E}[N_i|\mathbf{U}]^3}{\min_i \mathbb{E}[N_i|\mathbf{U}]^4}\right)^{1/2}\right) + O_{\psi_{2/(p+1)},tc}\left(n^{\mathtt{r}_{\beta,h}}(\min_i \mathbb{E}[N_i|\mathbf{U}]^{-(p+1)/2})\right).
\end{aligned}$$

SA-4.2 Hajek Estimator

Lemma SA-6. *Suppose Assumption 1, 2 and 3 hold. Then*

$$\widehat{\tau}_n - \widehat{\tau}_{n,UB} = -\Big(\frac{\mathbb{E}[g_i(1,\frac{M_i}{N_i})]}{\pi+1} + \frac{\mathbb{E}[g_i(-1,\frac{M_i}{N_i})]}{1-\pi}\Big)(1-\beta(1-\pi^2))(m-\pi) + O_{\psi_1}(n^{-2\mathtt{r}_{\beta,h}}).$$

SA-4.3 Stochastic Linearization

Lemma SA-7. *Suppose Assumptions 1, 2, and 3 hold. Define*

$$R_i = \frac{g_i(1,\frac{M_i}{N_i})}{1+\pi} + \frac{g_i(-1,\frac{M_i}{N_i})}{1-\pi}, \qquad Q_i = \mathbb{E}[\frac{G(U_i,U_j)}{\mathbb{E}[G(U_i,U_j)|U_j]}(g_j'(1,\pi) - g_j'(-1,\pi))|U_i].$$

Then,

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(\widehat{\tau}_n - \tau_n \leq t) - \mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}(R_i - \mathbb{E}[R_i] + Q_i)(W_i - \pi) \leq t)| = O\Big(\frac{\log n}{\sqrt{n\rho_n}} + \mathtt{r}_{n,\beta}\Big),$$

where $\mathtt{r}_{n,\beta} = \sqrt[4]{n}\sqrt{\log n}(n\rho_n)^{-\frac{p+1}{2}}$ *if* $\beta = 1, h = 0$*; and* $\sqrt{n \log n}(n\rho_n)^{-\frac{p+1}{2}}$ *if* $\beta < 1$ *or* $h \neq 0$.

Lemma SA-8. *Define Assumptions 1, 2, and 3 hold with* $h = 0$, $\beta \in [0,1]$*. Define*

$$R_i = \frac{g_i(1,\frac{M_i}{N_i})}{1+\pi} + \frac{g_i(-1,\frac{M_i}{N_i})}{1-\pi}, \qquad Q_i = \mathbb{E}[\frac{G(U_i,U_j)}{\mathbb{E}[G(U_i,U_j)|U_j]}(g_j'(1,\pi) - g_j'(-1,\pi))|U_i].$$

Then,

$$\sup_{\beta\in[0,1]} \sup_{t\in\mathbb{R}} |\mathbb{P}(\widehat{\tau}_n - \tau_n \leq t) - \mathbb{P}(\frac{1}{n}\sum_{i=1}^{n}(R_i - \mathbb{E}[R_i] + Q_i)(W_i - \pi) \leq t)| = o(1).$$

SA-5 Jacknife-Assisted Variance Estimation

Lemma SA-1. *Suppose Assumptions 1,2,3,4 hold, and $n\rho_n^3 \to \infty$ as $n \to \infty$. Suppose the non-parametric learner $\widehat{f}$ satisfies $\widehat{f}(\ell, \cdot) \in C_2([0,1])$, and $|\widehat{f}(\ell, \frac{1}{2}) - f(\ell, \frac{1}{2})| = o_{\mathbb{P}}(1)$, $|\partial_2 \widehat{f}(\ell, \frac{1}{2}) - \partial_2 f(\ell, \frac{1}{2})| = o_{\mathbb{P}}(1)$, for $\ell \in \{0,1\}$, where the rate in $o_{\mathbb{P}}(\cdot)$ does not depend on β. Suppose $\widehat{K}_n$ is the jacknife estimator from Algorithm 2. Then*

$$\widehat{K}_n = \mathbb{E}[(R_i - \mathbb{E}[R_i] + Q_i)^2] + o_{\mathbb{P}}(1),$$

where the rate in $o_{\mathbb{P}}(1)$ also does not depend on β.

Here we give a local-polynomial based learner $\widehat{f}$ that satisfies requirements of Lemma SA-1 (hence Theorem 4 in the main paper.)

Lemma SA-2. *Use a local polynomial estimator to fit the potential outcome functions: Take*

$$\begin{aligned}
&\widehat{f}(1, x) := \widehat{\gamma}_0 + \widehat{\gamma}_1 x, \\
&(\widehat{\gamma}_0, \widehat{\gamma}_1) := \underset{\gamma_0, \gamma_1}{\arg\min} \sum_{i=1}^{n} \Big(Y_i - \gamma_0 - \gamma_1 \frac{M_i}{N_i}\Big)^2 K_h\Big(\frac{M_i}{N_i}\Big) \mathbb{1}(T_i = 1),
\end{aligned}$$

where $K_h(\cdot) = h^{-1} K(\cdot/h)$ where K is a kernel function, h is the optimal bandwidth. Then $\widehat{f}(1,0) = f(1,0) + o_{\mathbb{P}}(1), \partial_2 \widehat{f}(1,0) = \partial_2 f(1,0) + o_{\mathbb{P}}(1)$, the same for control group. Moreover, the rate of convergence can be made not depending on β.

SA-6 Proof of Main Theorems

SA-6.1 Proof of Theorem 1

The conclusion follows from the stochastic linearization result in Lemma SA-6, and the Berry-Esseen result for Curie-Weiss magnetization with independent multipliers in Lemma SA-3.

SA-6.2 Proof of Theorem 2

The conclusion follows from the stochastic linearization result in Lemma SA-6, and the (uniform in β) Berry-Esseen result for Curie-Weiss magnetization with independent multipliers in Lemma SA-4.

SA-6.3 Proof of Theorem 3

The uniform approximation for $\sqrt{n}(\widehat{\beta}_n - 1)$ established in Lemma SA-3 implies

$$\inf_{\beta} \mathbb{P}_\beta(\beta \in \mathcal{I}(\alpha_1)) \geq \inf_{\beta} \mathbb{P}_\beta(\sqrt{n}(1-\beta) \geq q) \geq 1 - \alpha_1 + o_{\mathbb{P}}(1).$$

where q is the α_1 quantile of $\min\{\max\{\mathsf{T}_{c_{\beta,n},n}^{-2} - \mathsf{T}_{c_{\beta,n},n}^{2}/(3n), 0\}, 1\}$.

Then by a Bonferroni correction argument, the second step coverage can be lower bounded by

$$\inf_{\beta \in [0,1]} \mathbb{P}_\beta(\tau_n \in \widehat{\mathcal{C}}(\alpha_1, \alpha_2)) \geq \inf_{\beta \in [0,1]} \mathbb{P}_\beta(\tau_n \in \widehat{\mathcal{C}}(\alpha_1, \alpha_2), \beta \in \mathcal{I}(\alpha_1)) - \mathbb{P}_\beta(\beta \notin \mathcal{I}(\alpha_1)).$$

Observe that the event $\tau_n \in \widehat{\mathcal{C}}(\alpha_1, \alpha_2)$ conincides with the event $\widehat{\tau}_n - \tau_n \in [\inf_{c \in \mathcal{I}(\alpha_1)} law_{c_\beta}(1 - \frac{\alpha_2}{2}; \widehat{s}, n), \sup_{c \in \mathcal{I}(\alpha_1)} law_{c_\beta}(1 - \frac{\alpha_2}{2}; \widehat{s}, n)]$, where $\widehat{s} = (\widehat{K}_n, \widehat{K}_n^2)$. Hence

$$
\begin{aligned}
&\inf_{\beta \in [0,1]} \mathbb{P}_\beta(\tau_n \in \widehat{\mathcal{C}}(\alpha_1, \alpha_2), \beta \in \mathcal{I}(\alpha_1)) \\
\geq &\inf_{\beta \in [0,1]} \mathbb{P}_\beta(\widehat{\tau}_n - \tau_n \in [law_{c_\beta}(1 - \frac{\alpha_2}{2}; \widehat{s}, n), law_{c_\beta}(1 - \frac{\alpha_2}{2}; \widehat{s}, n)], \beta \in \mathcal{I}(\alpha_1)) \\
\geq &\inf_{\beta \in [0,1]} \mathbb{P}_\beta(\widehat{\tau}_n - \tau_n \in [law_{c_\beta}(1 - \frac{\alpha_2}{2}; \widehat{s}, n), law_{c_\beta}(1 - \frac{\alpha_2}{2}; \widehat{s}, n)]) - \mathbb{P}_\beta(\beta \in \mathcal{I}(\alpha_1)).
\end{aligned}
$$

Theorem 2 shows that the quantiles of the distributions of $\widehat{\tau}_n - \tau_n$ can be uniformly approximated by quantiles from $law_{c_{\beta,n}}$, if κ_1 and κ_2 are correctly specified, and the confidence interval is conservative, if we use upper boundds for κ_1 and κ_2. The conclusion then follows.

SA-6.4 Proof of Theorem 4

The conclusion follows from Theorem 3 and Lemma SA-1.

SA-7 Proofs

SA-7.1 Proofs for Section SA-2

SA-7.1.1 Proof of Lemma SA-2

Our proof is divided according to the different temperature regimes.

The High Temperature Regime.

We introduce the handy notation given by $F(v) := -\frac{1}{2}v^2 + \log\cosh(\sqrt{\beta}v + h)$. For the high temperature regime, we note that the term in the exponential can be expanded across its global minimum v^* (which satisfies the first order stationary point condition given by $v^* = \sqrt{\beta}\tanh(\sqrt{\beta}v^* + h)$) by

$$
\begin{aligned}
F(v) &= F(v^*) + F'(v^*)(v - v^*) + \frac{1}{2}F^{(2)}(v^*)(v - v^*)^2 + O((v - v^*)^3) \\
&= F(v^*) - \frac{1}{2}(1 - \beta\,\mathrm{sech}^2(\sqrt{\beta}v^* + h))(v - v^*)^2 + O((v - v^*)^3).
\end{aligned}
$$

Therefore, to obtain the limit of the expectation, we note that by the Laplace method given similar to the proof of Lemma SA-3 and the definition of $\mathsf{V}_n := n^{-1/2}\mathsf{U}_n$:

$$
\mathbb{E}[\mathsf{V}_n] = \frac{\int_{\mathbb{R}} v \exp\left(-nF(v)\right) dv}{\int_{\mathbb{R}} \exp(-nF(v)) dv} = v^*(1 + O(n^{-1})).
$$

Then, we note that for $\ell \in \mathbb{N}$, when $h = 0$ and $\beta < 1$ we use the Laplace method again to obtain that for all $\ell \in \mathbb{N}$,

$$
\begin{aligned}
\mathbb{E}\left[(\mathsf{V}_n - \mathbb{E}[\mathsf{V}_n])^{2\ell}\right] &= \frac{\int_{\mathbb{R}} (v - v^*)^{2\ell} \exp(-n(F(v) - F(v^*))) dv}{\int_{\mathbb{R}} \exp(-n(F(v) - F(v^*))) dv}(1 + O(n^{-1})) \\
&= \frac{1}{\sqrt{\pi}}\left(\frac{2}{n(1 - \beta\,\mathrm{sech}^2(\sqrt{\beta}v^* + h))}\right)^{\ell}\Gamma\left(\frac{2\ell + 1}{2}\right)(1 + O(n^{-1})).
\end{aligned}
$$

Then we can obtain that for all $t \in \mathbb{R}$, we have

$$\mathbb{E}[\exp(t(\mathsf{V}_n - \mathbb{E}[\mathsf{V}_n]))] = \sum_{\ell=0}^{\infty} \frac{t^\ell}{\ell!} \mathbb{E}[(\mathsf{V}_n - \mathbb{E}[\mathsf{V}_n])^\ell] = \sum_{\ell=0}^{\infty} \frac{t^{2\ell}}{(2\ell)!} \mathbb{E}[(\mathsf{V}_n - \mathbb{E}[\mathsf{V}_n])^{2\ell}]$$
$$\leq \exp\left(\frac{(1+o(1))t^2}{2n(1-\beta \operatorname{sech}^2(\sqrt{\beta}v^* + h))} \right),$$

which alternatively implies that

$$\|\mathsf{U}_n - \mathbb{E}[\mathsf{U}_n]\|_{\psi_2} = n^{1/2} \|\mathsf{V}_n - \mathbb{E}[\mathsf{V}_n]\|_{\psi_2} \leq (1+o(1))(1-\beta \operatorname{sech}^2(\sqrt{\beta}v^* + h))^{\frac{1}{2}}. \tag{SA-3}$$

The Critical Temperature Regime.

Then we study the critical temperature regime with $\beta = 1$. Note that one has $\mathbb{E}[\mathsf{U}_n] = 0$ and for all $\ell \in \mathbb{N}$ we have

$$F(v) = F(0) + F'(0)v + \frac{1}{2}F^{(2)}(0)v^2 + \frac{1}{6}F^{(3)}(0)v^3 + \frac{1}{24}F^{(4)}(0)v^4 + O(v^5)$$
$$= F(0) + \frac{1}{12}v^4 + O(v^5).$$

Then we can obtain that $\ell \in \mathbb{N}$,

$$\mathbb{E}\left[V_n^{2\ell}\right] = \frac{\int_{\mathbb{R}} v^{2\ell} \exp(-nF(v))dv}{\int_{\mathbb{R}} \exp(-nF(v))dv} = (1+o(1)) \cdot 2^{\ell - \frac{1}{2}} \cdot 3^{\frac{\ell}{2}+\frac{1}{4}} \frac{\Gamma\left(\frac{\ell}{2} + \frac{1}{4}\right)}{\Gamma(1/4)}$$
$$\leq (1+o(1)) \frac{1}{\sqrt{\pi}} \left(\frac{2^{3/2} \cdot 3^{3/4} \Gamma(3/4)}{n^{1/2}\Gamma(1/4)} \right)^\ell \Gamma\left(\frac{2\ell+1}{2} \right).$$

And we immediately obtain that

$$\mathbb{E}\left[\exp(t\mathsf{V}_n)\right] = \sum_{\ell=0}^{\infty} \frac{t^\ell \mathbb{E}[\mathsf{V}_n^{2\ell}]}{\Gamma(1+\ell)} \leq \sum_{\ell=0}^{\infty} \frac{1+o(1)}{\Gamma(1+2\ell)} \frac{1}{\sqrt{\pi}} \left(\frac{2^{1/2} \cdot 3^{3/4}\sqrt{2}\Gamma(3/4)}{n^{1/2}\Gamma(1/4)} \right)^\ell \Gamma\left(\frac{2\ell+1}{2} \right) t^\ell$$
$$\leq \exp\left(\frac{1+o(1)}{2} t^2 \left(\frac{2^{3/2} \cdot 3^{3/4}\Gamma(3/4)}{n^{1/2}\Gamma(1/4)} \right) \right),$$

which finally leads to

$$\|\mathsf{V}_n\|_{\psi_2} \leq (1+o(1)) \sqrt{\frac{2^{1/2} \cdot 3^{3/4}\Gamma(3/4)}{n^{1/2}\Gamma(1/4)}}. \tag{SA-4}$$

The Low Temperature Regime.

We shall note that at the low temperature regime the function $F(v)$ has two symmetric global minima $v_1 > 0 > v_2$, satisfying

$$F'(v_1) = F'(v_2) = 0 \quad \Rightarrow \quad v_\ell = \sqrt{\beta}\tanh(\sqrt{\beta}v_\ell + h) \quad \text{for } \ell \in \{1, 2\}.$$

Then we can check that by the Laplace method, for all $t > 0$ (following the path given by the high temperature regime) we have

$$\mathbb{E}[\exp(t(\mathsf{V}_n - \mathbb{E}[\mathsf{V}_n | \mathsf{V}_n > 0])) | \mathsf{V}_n > 0] = \frac{\int_{[0,\infty)} \exp\left(t(v - v_1) - nF(v)\right) dv}{\int_{[0,\infty)} \exp(-nF(v))dv}$$
$$= \exp\left(\frac{(1+o(1))t^2}{2n(1-\sqrt{\beta}\operatorname{sech}^2(\sqrt{\beta}v_1))} \right).$$

Then we similarly obtain that $\mathbb{E}[\exp(t(\mathsf{V}_n - \mathbb{E}[\mathsf{V}_n|\mathsf{V}_n < 0]))|\mathsf{V}_n < 0] = \exp\left(\frac{(1+o(1))t^2}{2n(1-\sqrt{\beta}\operatorname{sech}^2(\sqrt{\beta}v_1))}\right)$. Hence we obtain that

$$\begin{aligned}\|\mathsf{V}_n - \mathbb{E}[\mathsf{V}_n|\mathsf{V}_n < 0]|\mathsf{V}_n < 0\|_{\psi_2} &= \|\mathsf{V}_n - \mathbb{E}[\mathsf{V}_n|\mathsf{V}_n > 0]|\mathsf{V}_n > 0\|_{\psi_2} \\ &\leq (1+o(1))(1-\beta\operatorname{sech}^2(\sqrt{\beta}v_1))^{\frac{1}{2}}. \qquad \text{(SA-5)}\end{aligned}$$

The Drifting Sequence Case.

Then we consider the drifting case.

First consider $\beta = 1 - cn^{-\frac{1}{2}}$ with $c \in \mathbb{R}^+$ and $\beta \geq 0$. We will show that for any fixed n, $\|W_n\|_{\psi_2}$ is increasing in β when $\beta \in [0,1]$. This will imply that in the drifting case, $\|W_n\|_{\psi_2}$ will be no larger than its value at the critical regime.

For a comparison argument, denote $F_\beta(v) = -\frac{1}{2}v^2 + \log\cosh(\sqrt{\beta}v)$. Let $0 < \beta_1 < \beta_2 \leq 1$. Then

$$\frac{\exp(nF_{\beta_2}(v))}{\exp(nF_{\beta_1}(v))} = \exp(n\log\cosh(\sqrt{\beta_2}v) - n\log\cosh(\sqrt{\beta_1}v)),$$

where

$$\frac{d}{dv}\frac{\cosh(\sqrt{\beta_2}v)}{\cosh(\sqrt{\beta_1}v)} = \frac{(\sqrt{\beta_2}-\sqrt{\beta_1})\sinh((\sqrt{\beta_2}-\sqrt{\beta_1})v)}{\cosh^2(\sqrt{\beta_1}v)} > 0.$$

Hence for any $n \in \mathbb{N}$ and $t > 0$,

$$\mathbb{P}_\beta(|W_n| \geq t) = 2\frac{\int_t^\infty \exp(nF_\beta(v))dv}{\int_0^\infty \exp(nF_\beta(v))dv}$$

increases as $\beta \in [0,1]$ increases. This shows that $\|W_n\|_{\psi_2}$ increases as $\beta \in [0,1]$ increases. Together with Equation (SA-4), we have under $\beta_n = 1 - \frac{c}{\sqrt{n}}$, $0 \leq c \leq \sqrt{n}$,

$$\|\mathsf{V}_n\|_{\psi_2} \leq (1+o(1))\sqrt{\frac{2^{1/2}\cdot 3^{3/4}\Gamma(3/4)}{n^{1/2}\Gamma(1/4)}},$$

where $o(\cdot)$ is by an absolute constant.

Then we consider $\beta = 1 + cn^{-\frac{1}{2}}$. We shall note that under this situation it is not hard to check that

$$\begin{aligned}\mathbb{E}[\exp(t\mathsf{V}_n)] &= \frac{1}{2}\left(\mathbb{E}[\exp(t\mathsf{V}_n)|\mathsf{V}_n > 0] + \mathbb{E}[\exp(t\mathsf{V}_n)|\mathsf{V}_n < 0]\right) \\ &= \frac{1}{2}\left(\mathbb{E}[\exp(t(\mathsf{V}_n - v_+))|\mathsf{V}_n > 0]\exp(tv_+) + \mathbb{E}[\exp(t(\mathsf{V}_n - v_-))|\mathsf{V}_n < 0]\exp(tv_-)\right).\end{aligned}$$

Then, under this case we have by Taylor expanding F at 0 and the fact that $\sup_{v\in\mathbb{R}}|F^{(5)}(v)| < \infty$,

$$f_{\mathsf{V}_n}(v) \propto \sum_{l\in\{-,+\}} \mathbb{1}(v \in \mathcal{C}_l)\exp\Bigg(-cn^{\frac{1}{2}}(v-v_l)^2 - \frac{\sqrt{3c}}{3}n^{\frac{3}{4}}(v-v_l)^3 - \frac{1}{12}n(v-v_l)^4 - O(n(v-v_l)^5)\Bigg).$$

Before we start to upper bound the moments, we first use the fact that $v_+ = O(n^{-1/4})$ to obtain that

$$\int_{(-v_+,0)} v^{2\ell}\exp\left(-\sqrt{3c}v^3\right)dv \leq n^{-\frac{1}{4}}v_+^{2\ell}\exp(-\sqrt{3c}n^{-1/4}) = O\left(n^{-1/4-\ell/2}\right).$$

Then we obtain that

$$\begin{aligned}
\mathbb{E}[(\mathsf{V}_n - v_+)^{2\ell}|\mathsf{V}_n > 0] &= n^{-\frac{\ell}{2}} \frac{\int_{(-v_+,+\infty)} v^{2\ell} \exp\left(-cv^2 - \frac{\sqrt{3c}}{3}v^3 - \frac{1}{12}v^4\right) dv}{\int_{(-v_+,+\infty)} \exp\left(-cv^2 - \frac{\sqrt{3c}}{3}v^3 - \frac{1}{12}v^4\right) dv}(1+o(1)) \\
&\leq n^{-\frac{\ell}{2}}(1+o(1)) \frac{\int_{\mathbb{R}} v^{2\ell}\exp(-3cv^2)dv + \int_{(-v_+,+\infty)} v^{2\ell}\exp(-\sqrt{3c}v^3)dv + \int_{\mathbb{R}} v^{2\ell}\exp(-\frac{1}{4}v^4)dv}{\int_{(-v_+,+\infty)} \exp\left(-cv^2 - \frac{\sqrt{3c}}{3}v^3 - \frac{1}{12}v^4\right) dv} \\
&= n^{-\frac{\ell}{2}}(1+o(1)) \frac{\int_{\mathbb{R}} v^{2\ell}\exp(-3cv^2)dv + \int_{\mathbb{R}^+} v^{2\ell}\exp(-\sqrt{3c}v^3)dv + \int_{\mathbb{R}} v^{2\ell}\exp(-\frac{1}{4}v^4)dv}{\int_{(-v_+,+\infty)} \exp\left(-cv^2 - \frac{\sqrt{3c}}{3}v^3 - \frac{1}{12}v^4\right) dv} + O(n^{-1/4-\ell/2}) \\
&= n^{-\frac{\ell}{2}}(1+o(1))\left(\mathcal{C}_3\left(\frac{1}{3c}\right)^{\ell}\Gamma\left(\ell+\frac{1}{2}\right) + \mathcal{C}_4(3c)^{-\frac{\ell}{3}}\Gamma\left(\frac{2\ell}{3}+\frac{1}{3}\right) + \mathcal{C}_5 2^{\ell}\Gamma\left(\frac{\ell}{2}+\frac{1}{4}\right)\right),
\end{aligned}$$

with $\mathcal{C}_3 := \frac{(3c)^{-1/2}}{3\int_{(-v_+,+\infty)} \exp\left(-cv^2 - \frac{\sqrt{3c}}{3}v^3 - \frac{1}{12}v^4\right)dv}$, $\mathcal{C}_4 = \frac{1}{9\int_{(-v_+,+\infty)} \exp\left(-cv^2 - \frac{\sqrt{3c}}{3}v^3 - \frac{1}{12}v^4\right)dv}$, and $\mathcal{C}_5 = \frac{2^{-3/2}}{\int_{(-v_+,+\infty)} \exp\left(-cv^2 - \frac{\sqrt{3c}}{3}v^3 - \frac{1}{12}v^4\right)dv}$. Therefore, we can simply use the definition of the m.g.f. to obtain that

$$\begin{aligned}
\mathbb{E}[\exp(t^2(\mathsf{V}_n - v_+)^2)|\mathsf{V}_n > 0] &= \sum_{\ell=0}^{\infty} \frac{t^{2\ell}\mathbb{E}[(\mathsf{V}_n - v_+)^{2\ell}|\mathsf{V}_n > 0]}{\Gamma(2\ell+1)} \\
&\leq \sum_{\ell=0}^{\infty} \frac{(1+o(1))n^{-\ell/2}t^{2\ell}}{\Gamma(2\ell+1)}\left(\mathcal{C}_3\left(\frac{1}{3c}\right)^{\ell}\Gamma\left(\ell+\frac{1}{2}\right) + \mathcal{C}_4(3c)^{-\frac{\ell}{3}}\Gamma\left(\frac{2\ell}{3}+\frac{1}{3}\right) + \mathcal{C}_5 2^{\ell}\Gamma\left(\frac{\ell}{2}+\frac{1}{4}\right)\right) \\
&\leq \sum_{\ell=0}^{\infty} \frac{(1+o(1))n^{-\ell/2}t^{2\ell}}{\Gamma(2\ell+1)}\left(\mathcal{C}_3(3c)^{-1}\Gamma\left(\frac{3}{2}\right) + \mathcal{C}_4(3c)^{-1/3}\Gamma(1) + 2\mathcal{C}_5\Gamma\left(\frac{3}{4}\right)\right)^{\ell}\Gamma\left(\frac{2\ell+1}{2}\right) \\
&\leq (1 - 2t^2 n^{1/2}/\sigma^2)^{-\frac{1}{2}}, \qquad \sigma := \left(\mathcal{C}_3(3c)^{-1}\Gamma\left(\frac{3}{2}\right) + \mathcal{C}_4(3c)^{-1/3}\Gamma(1) + 2\mathcal{C}_5\Gamma\left(\frac{3}{4}\right)\right)^{\frac{1}{2}}.
\end{aligned}$$

Then we use the fact that $\mathbb{E}[\mathsf{V}_n|\mathsf{V}_n > 0] = v_+$ to obtain that (here we use proposition 2.5.2 in [7])

$$\mathbb{E}[\exp(t(\mathsf{V}_n - v_+))|\mathsf{V}_n > 0] \leq \exp\left(18e^2 n^{-1/2}\sigma^2 t^2\right).$$

Similarly one obtains that $\mathbb{E}[\exp(t(\mathsf{V}_n - v_-))|\mathsf{V}_n < 0] \leq \exp(18e^2 n^{-1/2}\sigma^2 t^2)$. And hence

$$\mathbb{E}[\exp(t\mathsf{V}_n)] \leq \frac{1}{2}\left(\exp(tv_+) + \exp(-tv_+)\right)\exp(18e^2 n^{-1/2}\sigma^2 t^2) \leq \exp\left(\frac{1}{2}t^2 v_+^2\right).$$

SA-7.1.2 Proof for Lemma SA-3 High Temperature

Throughout the proof, we denote by $\mathtt{C}$ an absolute constant, and $\mathtt{K}$ a constant that only depends on the distribution of X_i.

Take U_n to be a random variable with density

$$f_{\mathsf{U}_n}(u) = \frac{\exp\left(-\frac{1}{2}u^2 + n\log\cosh\left(\sqrt{\frac{\beta}{n}}u\right)\right)}{\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}v^2 + n\log\cosh\left(\sqrt{\frac{\beta}{n}}v\right)\right)dv}, \qquad u \in \mathbb{R}. \tag{SA-6}$$

By Lemma SA-3, condition on U_n, W_i are i.i.d Bernouli with

$$\mathbb{P}(W_i = 1|\mathsf{U}_n) = \frac{1}{2}(\tanh(\sqrt{\frac{\beta}{n}}\mathsf{U}_n) + 1).$$

We characterize the conditional mean and variance as

$$\begin{aligned} e(\mathsf{U}_n) &= \mathbb{E}\left[X_i W_i|\mathsf{U}_n\right] = \mathbb{E}\left[X_i\right]\tanh\left(\sqrt{\frac{\beta}{n}}\mathsf{U}_n\right), \\ v(\mathsf{U}_n) &= \mathbb{V}\left[X_i W_i|\mathsf{U}_n\right] = \mathbb{E}\left[X_i^2\right] - \mathbb{E}\left[X_i\right]^2\tanh^2\left(\sqrt{\frac{\beta}{n}}\mathsf{U}_n\right). \end{aligned} \tag{SA-7}$$

Moreover, we have $\mathbb{E}\left[\left|X_i^3(W_i-\pi)^3\right| |\mathsf{U}_n\right] \leq \mathbb{E}\left[|X_i|^3\right]$.

Step 1: Conditional Berry-Esseen.

Apply Berry-Esseen Theorem conditional on U_n,

$$\sup_{u\in\mathbb{R}}\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left(g_n \leq t|\mathsf{U}_n = u\right) - \Phi\left(\frac{t-\sqrt{n}\mathbb{E}\left[X_iW_i|\mathsf{U}_n=u\right]}{\mathbb{V}\left[X_iW_i|\mathsf{U}_n=u\right]^{1/2}}\right)\right| \leq \mathsf{C}\frac{\mathbb{E}\left[|X_i|^3\right]}{v(\mathsf{U}_n)}n^{-1/2}.$$

Since $v(\mathsf{U}_n) \geq \mathbb{V}[X_i] + \mathbb{E}[X_i]^2\,\mathrm{sech}^2(\sqrt{\beta/n}\mathsf{U}_n)$, and be Lemma SA-2, $\|\mathsf{U}_n\|_{\psi_1} \leq \mathsf{C}n^{1/4}$. Hence

$$\begin{aligned} & d_{\mathrm{KS}}\left(g_n, v(\mathsf{U}_n)^{1/2}\mathsf{Z} + \sqrt{n}e(\mathsf{U}_n)\right) \\ =& \sup_{t\in\mathbb{R}}\left|\int_{-\infty}^{\infty}(\mathbb{P}\left(g_n\leq t|\mathsf{U}_n=u\right) - \Phi\left(\frac{t-\sqrt{n}e(\mathsf{U}_n)}{v(\mathsf{U}_n)^{1/2}}\right))f_{\mathsf{U}_n}(u)du\right| \\ \leq& \mathsf{K}n^{-1/2}. \end{aligned}$$

Step 2: Approximation for U_n.

Take $\mathsf{U} \sim \mathsf{N}(0, (1-\beta)^{-1})$ independent to Z. Consider $\mathsf{V}_n = n^{-1/2}\mathsf{U}_n$. Then

$$f_{\mathsf{V}_n}(v) \propto \exp\left(-\frac{1}{2}nv^2 + n\log\cosh\left(\sqrt{\beta}v\right)\right) =: \exp\left(-n\phi(v)\right),$$

where $\phi(v) = -\frac{1}{2}v^2 + \log\cosh(\sqrt{\beta}v)$. And ϕ is maximized at 0 with $\phi''(0) = 1-\beta > 0$.

We will approximate the integral of f_{V_n} by Laplace method. By Equation (5.1.21) in [2],

$$\begin{aligned} \int_{-\infty}^{\infty}\exp\left(-n\phi(v)\right)dv &= \sqrt{\frac{2\pi}{n\phi''(0)}}\exp\left(-n\phi(0)\right) + O\left(\frac{\exp(-n\phi(0))}{n^{3/2}}\right) \\ &= \sqrt{\frac{2\pi}{n\phi''(0)}}\exp\left(-n\phi(0)\right)\left[1+O(n^{-1})\right], \end{aligned}$$

where the $O(n^{-1})$ term only depends on n and ϕ. It follows that

$$f_{\mathsf{V}_n}(v) = \sqrt{\frac{n\phi''(0)}{2\pi}}\exp\left(-n\phi(v)+n\phi(0)\right)\left[1+O(n^{-1})\right].$$

Then by a change of variable and the fact that $O(n^{-1})$ term does not depend on v,

$$f_{\mathsf{U}_n}(u) = \sqrt{\frac{\phi''(0)}{2\pi}} \exp\left(-n\phi(n^{-1/2}u) + n\phi(0)\right) [1 + O(n^{-1})]. \tag{SA-8}$$

Taylor expanding ϕ at 0, we get

$$-n\phi(n^{-1/2}u) + n\phi(0) = -\frac{\phi''(0)}{2}u^2 - \tanh(\sqrt{\beta}v_* + h)\operatorname{sech}^2(\sqrt{\beta}v_*)\frac{u^3}{3\sqrt{n}} \tag{SA-9}$$

$$= -\frac{1}{2}(1-\beta)(u^2 - \tanh(\sqrt{\beta}v_*)\operatorname{sech}^2(\sqrt{\beta}v_*)\frac{u^3}{3\sqrt{n}}, \tag{SA-10}$$

where v_* is some quantity between 0 and $n^{-1/2}u$. Then

$$\begin{aligned} d_{\mathrm{TV}}(\mathsf{U}_n, \mathsf{U}) &= \int_{-\infty}^{\infty} |f_{\mathsf{U}_n}(u) - f_U(u)|\, du \\ &\leq \int_{-\infty}^{\infty} \sqrt{\frac{\phi''(0)}{2\pi}} \exp\left(-\frac{1}{2}(1-\beta)u^2\right) \\ &\quad \cdot \left[\exp\left(-\tanh(\sqrt{\beta}v_*(u))\operatorname{sech}^2(\sqrt{\beta}v_*(u))\frac{u^3}{3\sqrt{n}}\right) - 1\right] du \left[1 + O(n^{-1})\right], \end{aligned}$$

where $v^*(u)$ is some random quantity between 0 and $n^{-1/2}u$. We will show that we can restrict the analysis to the region $[-\mathsf{c}_\beta\sqrt{\log n}, \mathsf{c}_\beta\sqrt{\log n}]$, which is where the bulk of mass lies, with $\mathsf{c}_\beta = (1-\beta)^{-1/2}$. Since $\mathsf{U} \sim N(0, (1-\beta)^{-1})$, $\mathbb{P}\left(|\mathsf{U}| \geq \mathsf{c}_\beta\sqrt{\log n}\right) \leq n^{-1}$. By Lemma SA-2, we also have $\mathbb{P}\left(|\mathsf{U}_n| \geq \mathsf{c}'_\beta\sqrt{\log n}\right) \leq n^{-1}$, where c'_β is a constant that only depends on β. Take $\mathsf{d}_\beta = \max\{\mathsf{c}_\beta, \mathsf{c}_{\beta'}\}$, and use the boundedness of tanh and sech and the Lipschitzness of exp when restricted to $[-1, 1]$, we have

$$\begin{aligned} d_{\mathrm{TV}}(\mathsf{U}_n, \mathsf{U}) &\leq \int_{-\mathsf{d}_\beta\sqrt{\log n}}^{\mathsf{d}_\beta\sqrt{\log n}} \sqrt{\frac{\phi''(0)}{2\pi}} \exp\left(-\frac{1}{2}(1-\beta)u^2\right) \\ &\quad \cdot \left[\exp\left(-\tanh(\sqrt{\beta}v_*(u))\operatorname{sech}^2(\sqrt{\beta}v_*(u))\frac{u^3}{3\sqrt{n}}\right) - 1\right] du \left[1 + O(n^{-1})\right] + O(n^{-1}) \\ &\leq \int_{-\mathsf{d}_\beta\sqrt{\log n}}^{\mathsf{d}_\beta\sqrt{\log n}} \sqrt{\frac{\phi''(0)}{2\pi}} \exp\left(-\frac{1}{2}(1-\beta)u^2\right) c_2\frac{|u|^3}{\sqrt{n}} du \left[1 + O(n^{-1})\right] + O(n^{-1}) \\ &= O(n^{-1/2}). \end{aligned}$$

Step 3: Data Processing Inequality.

We can use data processing inequality to get

$$d_{\mathrm{KS}}\left(v(\mathsf{U}_n)^{1/2}\mathsf{Z} + \sqrt{n}e(\mathsf{U}_n), v(\mathsf{U})^{1/2}\mathsf{Z} + \sqrt{n}e(\mathsf{U})\right) \leq d_{\mathrm{TV}}\left(\mathsf{U}_n, \mathsf{U}\right) = O(n^{-1/2}).$$

Step 4: Stabilization of Variance.

By independence between U and Z, we have

$$\begin{aligned}
&d_{\mathrm{KS}}\left(v(\mathsf{U})^{1/2}\mathsf{Z}+\sqrt{n}e(\mathsf{U})),\mathbb{E}[v(\mathsf{U})]^{1/2}Z+\sqrt{n}e(\mathsf{U})\right)\\
=&\sup_{t\in\mathbb{R}}\mathbb{E}\left[\Phi\left(\frac{t-\sqrt{n}e(\mathsf{U})}{v(\mathsf{U})^{1/2}}\right)-\Phi\left(\frac{t-\sqrt{n}e(\mathsf{U})}{\mathbb{E}[v(\mathsf{U})]^{1/2}}\right)\right]\\
\leq&\sup_{t\in\mathbb{R}}\mathbb{E}\left[\left|\phi\left(\frac{t-\sqrt{n}e(\mathsf{U})}{v^*(\mathsf{U})^{1/2}}\right)(t-\sqrt{n}e(\mathsf{U}))\left(v(\mathsf{U})^{-1/2}-\mathbb{E}[v(\mathsf{U})]^{-1/2}\right)\right|\right],
\end{aligned}$$

where $v^*(\mathsf{U})$ is some quantity between $\mathbb{E}[v(\mathsf{U})]$ and $v(\mathsf{U})$, and by Equation SA-7, $v^*(\mathsf{U})\geq \mathtt{C}^{-1}\mathbb{V}[X_i]$. It follows from boundedness of $v(\mathsf{U})$ and Lipshitzness of tanh in the expression of $v(\mathsf{U})$ that

$$\begin{aligned}
&d_{\mathrm{KS}}\left(v(\mathsf{U})^{1/2}Z+\sqrt{n}e(\mathsf{U})),\mathbb{E}[v(\mathsf{U})]^{1/2}Z+\sqrt{n}e(\mathsf{U})\right)\\
\leq&\sup_{t\in\mathbb{R}}\sup_{u\in\mathbb{R}}\left|\phi\left(\frac{t-\sqrt{n}e(u)}{\sqrt{2\mathbb{E}[X_i^2]}}\right)(t-\sqrt{n}e(u))\right|\frac{1}{2\sqrt{\mathtt{C}^{-1}\mathbb{V}[X_i]}}\mathbb{E}\left[|v(\mathsf{U})-\mathbb{E}[v(\mathsf{U})]|\right]\\
=&O(n^{-1/2}).
\end{aligned}$$

Step 5: Gaussian Approximation for $\sqrt{n}e(\mathsf{U})$.

In this step, we will show that $\sqrt{n}e(\mathsf{U})$ can be well-approximated by $\sqrt{\beta}\mathsf{U}$ and hence $\sqrt{n}g_n$ can be well-approximated by a Gaussian.

$$\begin{aligned}
&d_{\mathrm{KS}}\left(\mathbb{E}[v(\mathsf{U})]^{1/2}\mathsf{Z}+\sqrt{n}e(\mathsf{U}),\mathbb{E}[v(\mathsf{U})]^{1/2}\mathsf{Z}+\sqrt{\beta}\mathsf{U}\right)\\
\leq&\sup_{t\in\mathbb{R}}\mathbb{E}\left[\Phi\left(\frac{t-\sqrt{n}e(\mathsf{U})}{\mathbb{E}[v(\mathsf{U})]^{1/2}}\right)-\Phi\left(\frac{t-\sqrt{\beta}U}{\mathbb{E}[v(\mathsf{U})]^{1/2}}\right)\right]\\
\leq&\frac{\|\phi\|_\infty}{\mathbb{E}[v(\mathsf{U})^{1/2}]}\mathbb{E}\left[\left|\sqrt{n}e(\mathsf{U})-\sqrt{\beta}\mathsf{U}\right|\right].
\end{aligned}$$

Taylor expanding tanh at 0,

$$\begin{aligned}
\sqrt{n}e(\mathsf{U})&=\mathbb{E}[X_i]\sqrt{n}\tanh\left(\sqrt{\frac{\beta}{n}}\mathsf{U}\right)\\
&=\mathbb{E}[X_i]\sqrt{\beta}\mathsf{U}+O\left(\frac{\beta}{\sqrt{n}}\mathsf{U}^2\right)+O(n^{-1/2})\\
&=\mathbb{E}[X_i]\sqrt{\beta}\mathsf{U}+O\left(\frac{\beta}{\sqrt{n}}\mathsf{U}^2\right)+O(n^{-1/2}),
\end{aligned}$$

It follows that $\mathbb{E}\left[\left|\sqrt{n}e(\mathsf{U})-\sqrt{\beta}\mathsf{U}\right|\right]=O(n^{-1/2})$ and hence

$$d_{\mathrm{KS}}\left(\mathbb{E}[v(\mathsf{U})]^{1/2}\mathsf{Z}+\sqrt{n}e(\mathsf{U}),\mathbb{E}[v(\mathsf{U})]^{1/2}\mathsf{Z}+\mathbb{E}[X_i]\sqrt{\beta}\mathsf{U}\right)=O(n^{-1/2}).$$

Recall $\mathsf{U}\sim N(0,(1-\beta)^{-1})$, hence $\mathbb{E}[X_i]\sqrt{\beta}\mathsf{U}\sim N(0,\mathbb{E}[X_i]^2\frac{\beta}{1-\beta})$. Moreover,

$$\begin{aligned}
\mathbb{E}[v(\mathsf{U})]&=\mathbb{E}[\mathbb{E}[X_i^2]\mathbb{E}[W_i^2|\mathsf{U}]]-\mathbb{E}[\mathbb{E}[X_i]^2\mathbb{E}[W_i|\mathsf{U}]^2]\\
&=\mathbb{E}[X_i^2]-\mathbb{E}[X_i]^2\mathbb{E}[W_i|\mathsf{U}]^2\\
&=\mathbb{E}[X_i^2]+O(n^{-1/2}),
\end{aligned}$$

where the last line is because $\mathbb{E}[W_i|\mathsf{U}] = \tanh(\sqrt{\beta/n}\mathsf{U})$ and U is sub-Gaussian. Since $\mathsf{Z} \perp\!\!\!\perp \mathsf{U}$,

$$d_{\mathrm{KS}}\left(\mathbb{E}[v(\mathsf{U})]^{1/2}\mathsf{Z} + \mathbb{E}[X_i]\sqrt{\beta}U, N(0, \mathbb{E}[X_i^2] + \mathbb{E}[X_i]^2\frac{\beta}{1-\beta}\right) = O(n^{-1/2}).$$

Combining the previous five steps, we get

$$d_{\mathrm{KS}}\left(\sqrt{n}g_n, N\left(0, \mathbb{E}[X_i^2] + \mathbb{E}[X_i]^2\frac{\beta}{1-\beta}\right)\right) = O(n^{-1/2}).$$

SA-7.1.3 Proof for Lemma SA-3 Critical Temperature

Throughout the proof, we denote by $\mathtt{C}$ an absolute constant, and $\mathtt{K}$ a constant that only depends on the distribution of X_i. The proofs for the critical temperature case will have a similar structure as the proof for the high temperature case, based the same U_n defined in Equation (SA-6).

Step 1: Conditional Berry-Esseen.

The same argument as in the high-temperature case gives

$$d_{\mathrm{KS}}\left(g_n, v(\mathsf{U}_n)^{1/2}\mathsf{Z} + \sqrt{n}e(\mathsf{U}_n)\right) \leq \mathtt{K}n^{-1/2}.$$

Step 2: Approximation for U_n.

Take W to be a random variable with density function

$$f_{\mathsf{W}}(z) = \frac{\sqrt{2}}{3^{1/4}\Gamma(\frac{1}{4})}\exp\left(-\frac{1}{12}z^4\right), \quad z \in \mathbb{R},$$

independent to Z. Take $\mathsf{W}_n = n^{-1/4}\mathsf{U}_n$ and $\mathsf{V}_n = n^{-1/2}\mathsf{U}_n$. Again $f_{\mathsf{V}_n}(v) \propto \exp(-n\phi(v))$, where $\phi(v) := -\frac{1}{2}v^2 + \log\cosh(v)$. In particular, $\phi^{(v)}(0) = 0$ for all $0 \leq v \leq 3$, and $\phi^{(4)}(0) = -2 < 0$, $\phi^{(5)}(0) = 0$, $\phi^{(6)}(0) = 16 > 0$. Example 5.2.1 in [2] leads to

$$f_{\mathsf{V}_n}(v) = n^{\frac{1}{4}}\frac{\sqrt{2}}{3^{\frac{1}{4}}\Gamma(\frac{1}{4})}\exp(n\phi(v) - n\phi(0))(1 + o(1)),$$

which implies $f_{\mathsf{W}_n}(w) = f_W(w)(1 + o(1))$. Results in [2] do not give a rate, however. We will use a more cumbersome approach to obtain a slightly sub-optimal rate.

By a change of variable, $f_{\mathsf{W}_n}(w) = \frac{h_n(w)}{\int_{-\infty}^{\infty} h_n(u)du}$, where h_n can be written as

$$h_n(w) = \exp\left(-\frac{\sqrt{n}}{2}w^2 + n\log\cosh\left(n^{-\frac{1}{4}}w\right)\right) = \exp\left(-\frac{1}{12}w^4 + g(w)n^{-\frac{1}{2}}w^6\right).$$

The last equality follows from Taylor expanding the term in $\exp(\cdot)$ at $w = 0$, and g is some bounded function.

$$\int_{-10\sqrt{\log n}}^{10\sqrt{\log n}} h_n(w)dw = I_n(1 + O((\log n)^3 n^{-\frac{1}{2}})), \quad I_n := \int_{-10\sqrt{\log n}}^{10\sqrt{\log n}} \exp\left(-\frac{1}{12}w^4\right)dw$$

Moreover, $\int_{[-10\sqrt{\log n},10\sqrt{\log n}]^c} h_n(w)dw = O(n^{-1/2}) = I_n[1+O(n^{-\frac{1}{2}})]$. Hence for denominator, we have $\int_{-\infty}^{\infty} h_n(w)dw = I_n[1+O((\log n)^3 n^{-\frac{1}{2}})]$. It follows that

$$\begin{aligned}
&d_{\mathrm{TV}}(\mathsf{W}_n, \mathsf{W}) \\
\lesssim &\int_{-10\sqrt{\log n}}^{10\sqrt{\log n}} I_n^{-1} \exp\left(-\frac{1}{12}w^4\right) n^{-\frac{1}{2}} w^6 dw + \int_{-10\sqrt{\log n}}^{10\sqrt{\log n}} I_n^{-1} O((\log n)^3 n^{-\frac{1}{2}}) dw \\
&+ P(|\mathsf{W}_n| \geq 10\sqrt{\log n}) + \mathbb{P}(|\mathsf{W}| \geq 10\sqrt{\log n}) \\
=&O((\log n)^3 n^{-\frac{1}{2}}).
\end{aligned}$$

Step 3: Data Processing Inequality.

We can use data processing inequality to get

$$d_{\mathrm{KS}}\left(v(\mathsf{U}_n)^{1/2}\mathsf{Z} + \sqrt{n}e(\mathsf{U}_n), v(n^{1/4}\mathsf{W})^{1/2}\mathsf{Z} + \sqrt{n}e(n^{1/4}\mathsf{W})\right) \leq d_{\mathrm{TV}}\left(\mathsf{W}_n, \mathsf{W}\right) = O(n^{-1/2}).$$

Step 4: Non-Gaussian Approximation for $n^{\frac{1}{4}}e(n^{\frac{1}{4}}\mathsf{W})$

$$n^{1/4}e(n^{1/4}\mathsf{W})) = \mathbb{E}[X_i] n^{\frac{1}{4}} \tanh\left(n^{-\frac{1}{4}}\mathsf{W}\right) = \mathbb{E}[X_i]\left[\mathsf{W} - O\left(\frac{\mathsf{W}^2}{3\sqrt{n}}\right)\right],$$

where we have use the fact that $\tanh^{(2)}(0) = 0$. Hence there exists $C > 0$ such that for n large enough, for any $t > 0$,

$$\mathbb{P}\left(\mathbb{E}[X_i]\left[\mathsf{W} + C\frac{\mathsf{W}^2}{\sqrt{n}}\right] \leq t\right) \leq \mathbb{P}\left(n^{1/4}e(n^{1/4}\mathsf{W})) \leq t\right) \leq \mathbb{P}\left(\mathbb{E}[X_i]\left[\mathsf{W} - C\frac{\mathsf{W}^2}{\sqrt{n}}\right] \leq t\right). \quad (0)$$

We have showed that there exists $c > 0$ such that

$$\mathbb{P}(|\mathsf{W}| \geq c\sqrt{\log n}) \leq n^{-1/2}, \quad (1)$$

in which case $\mathsf{W}^2/\sqrt{n} \leq 1$ for large enough n. Hence for large enough n if $t/\mathbb{E}[X_i] > c\sqrt{\log n} + 1$, then

$$\mathbb{P}\left(\mathsf{W} + C\frac{\mathsf{W}^2}{\sqrt{n}} \leq \frac{t}{\mathbb{E}[X_i]}, |\mathsf{W}| \leq c\sqrt{\log n}\right) - \mathbb{P}\left(\mathsf{W} \leq \frac{t}{\mathbb{E}[X_i]}, |\mathsf{W}| \leq c\sqrt{\log n}\right) = 0. \quad (2)$$

If $0 < t/\mathbb{E}[X_i] < c\sqrt{\log n} + 1$, then

$$\begin{aligned}
&\left|\mathbb{P}\left(\mathsf{W} + \frac{\mathsf{W}^2}{\sqrt{n}} \leq \frac{t}{\mathbb{E}[X_i]}, |\mathsf{W}| \leq c\sqrt{\log n}\right) - \mathbb{P}\left(\mathsf{W} \leq \frac{t}{\mathbb{E}[X_i]}, |\mathsf{W}| \leq c\sqrt{\log n}\right)\right| \\
\leq &\mathbb{P}\left(\frac{t}{\mathbb{E}[X_i]} \leq \mathsf{W} \leq \frac{1 - \sqrt{1 - 4n^{-1/2}t/\mathbb{E}[X_i]}}{2n^{-1/2}}, |\mathsf{W}| \leq c\sqrt{\log n}\right).
\end{aligned}$$

Now we study $g(x;\alpha) = (1 - \sqrt{1 - 4x\alpha})/(2x), x > 0$. Then $\sup_{\alpha \leq \frac{1}{4}} \sup_{0 \leq x \leq \frac{1}{2}} |\theta'(x;\alpha)| \leq 2$ and $g(0;\alpha) = \alpha$. Since for large enough n, $0 < t/\mathbb{E}[X_i] < c\sqrt{\log n} + 1 \leq \frac{1}{4}$ and $0 \leq n^{-1/2} \leq \frac{1}{2}$, we have $\frac{1 - \sqrt{1 - 4n^{-1/2}t/\mathbb{E}[X_i]}}{2n^{-1/2}} \leq t/\mathbb{E}[X_i] + 2n^{-1/2}$. Hence if $0 < t/\mathbb{E}[X_i] < c\sqrt{\log n} + 1$,

$$\left|\mathbb{P}\left(\mathsf{W} + \frac{\mathsf{W}^2}{\sqrt{n}} \leq \frac{t}{\mathbb{E}[X_i]}, |\mathsf{W}| \leq c\sqrt{\log n}\right) - \mathbb{P}\left(\mathsf{W} \leq \frac{t}{\mathbb{E}[X_i]}, |\mathsf{W}| \leq c\sqrt{\log n}\right)\right| = O(n^{-1/2}). \quad (3)$$

Combining (1), (2), (3),

$$\sup_{t>0}\left|\mathbb{P}\left(\mathsf{W}+\frac{\mathsf{W}^2}{\sqrt{n}}\le\frac{t}{\mathbb{E}[X_i]}\right)-\mathbb{P}\left(\mathsf{W}\le\frac{t}{\mathbb{E}[X_i]}\right)\right|=O(n^{-1/2}).$$

By similar argument, we can show

$$\sup_{t>0}\left|\mathbb{P}\left(\mathsf{W}-\frac{\mathsf{W}^2}{\sqrt{n}}\le\frac{t}{\mathbb{E}[X_i]}\right)-\mathbb{P}\left(\mathsf{W}\le\frac{t}{\mathbb{E}[X_i]}\right)\right|=O(n^{-1/2}).$$

Noticing that W and $-W$ have the same distribution, the above two inequalities also hold for $t\le 0$. Hence it follows from (0) that

$$d_{\mathrm{KS}}\left(n^{1/4}e(n^{1/4}\mathsf{W})),\mathbb{E}[X_i]\mathsf{W}\right)=O(n^{-1/2}).$$

Step 5: Vanishing Variance Term. Denote by $f_{\mathsf{W}+n^{-1/4}\mathsf{Z}}$ the density of $\mathsf{W}+n^{-1/4}\mathsf{Z}$. Then

$$f_{\mathsf{W}+n^{-1/4}\mathsf{Z}}(y)=\int_{-\infty}^{\infty}\frac{\sqrt{2}}{3^{1/4}\Gamma(\frac{1}{4})}\exp\left(-\frac{1}{12}(y-x)^4\right)\frac{\exp(-\sqrt{n}x^2/2)}{\sqrt{2\pi n^{-1/2}}}dx.$$

We will use Laplace method to show $f_{\mathsf{W}+n^{-1/4}\mathsf{Z}}$ is close to f_{W}. However, to get uniformity over y, we need to work harder than in the high temperature case. Define $\varphi(x)=x^2/2$ and $g_y(t)=\exp(-(t-y)^4/12)$. Consider

$$I_{y,+}(\lambda)=\int_0^{\infty}g_y(t)\exp(-\lambda\varphi(t))dt,\qquad I_{y,-}(\lambda)=\int_{-\infty}^{0}g_y(t)\exp(-\lambda\varphi(t))dt.$$

Following Section 5.1 in [2], take $\tau>0$ such that $\varphi(t)=\tau$, by a change of variable,

$$I_{y,+}(\lambda)=\exp(-\lambda\varphi(0))\int_0^{\infty}\left[\frac{g_y(t)}{\varphi'(t)}\bigg|_{t=\varphi^{-1}(\tau)}\right]\exp(-\lambda\tau)d\tau=\int_0^{\infty}\frac{\exp(-(\sqrt{2\tau}-y)^4/12)}{\sqrt{2\tau}}\exp(-\lambda\tau)d\tau.$$

To get rate of convergence uniformly in y, we follow the proof of Watson's Lemma but consider only up to first order term. Taylor expanding $x\mapsto\exp(-x^4)/12$ up to first order at y, we have

$$\frac{\exp(-(\sqrt{2\tau}-y)^4/12)}{\sqrt{2\tau}}=\frac{\exp(-y^4/12)}{\sqrt{2\tau}}+\frac{1}{3}\exp(-y^4/12)y^3+\frac{h_y(\tau^*)}{2}\sqrt{2\tau},$$

where τ^* is some quantity between 0 and $\sqrt{2\tau}$ and

$$h_y(u)=-\exp(-(u-y)^4/12)(u-y)^2+\frac{1}{9}\exp(-(u-y)^4/12)(u-y)^6.$$

In particular, we have $\sup_{y\in\mathbb{R}}\sup_{u\in\mathbb{R}}|h_y(u)|<C$ for some absolute constant C. Then

$$\sup_{y\in\mathbb{R}}\left|\int_0^{\infty}\frac{h_y(\tau^*)}{2}\sqrt{2\tau}\exp(-\lambda\tau)d\tau\right|\le\frac{C}{\sqrt{2}}\Gamma\left(\frac{3}{2}\right)\lambda^{-3/2},\quad\forall\lambda>0.$$

Evaluating the first two terms, we get

$$\sup_{y\in\mathbb{R}}\left|I_{y,+}(\lambda)-\sqrt{\frac{\pi}{2\lambda}}\exp(-y^4/12)-\int_0^{\infty}\frac{1}{3}\exp(-y^4/12)y^3\exp(-\lambda\tau)d\tau\right|\le\frac{C}{\sqrt{2}}\Gamma\left(\frac{3}{2}\right)\lambda^{-3/2},\forall\lambda>0.$$

Similarly, for $I_{y,-}$, change of variable by taking $\tau < 0$ such that $\varphi(t) = \tau$, we have

$$\sup_{y \in \mathbb{R}} \left| I_{y,-}(\lambda) - \sqrt{\frac{\pi}{2\lambda}} \exp(-y^4/12) + \int_0^\infty \frac{1}{3} \exp(-y^4/12) y^3 \exp(-\lambda\tau) d\tau \right| \leq \frac{C}{\sqrt{2}} \Gamma\left(\frac{3}{2}\right) \lambda^{-3/2}, \forall \lambda > 0.$$

Combining the two parts, we get

$$\sup_{y \in \mathbb{R}} \left| \int_{-\infty}^{\infty} g_y(t) \exp(-\lambda\varphi(t)) dt - \sqrt{\frac{2\pi}{\lambda}} \exp(-y^4/12) \right| \leq C\sqrt{2}\Gamma\left(\frac{3}{2}\right) \lambda^{-3/2}, \quad \forall \lambda > 0.$$

Now take $\lambda = \sqrt{n}$ and multiply both sides by $\frac{n^{1/4}}{3^{1/4}\Gamma(\frac{1}{4})\sqrt{\pi}}$, we get

$$\sup_{y \in \mathbb{R}} \left| f_{\mathsf{W}+n^{-1/4}\mathsf{Z}}(y) - \frac{\sqrt{2}}{3^{1/4}\Gamma(\frac{1}{4})} \exp(-y^4/12) \right| \leq C \frac{\sqrt{2}\Gamma(\frac{3}{2})}{3^{1/4}\Gamma(\frac{1}{4})\sqrt{\pi}} n^{-1/2}.$$

By a truncation argument, we have

$$\begin{aligned} d_{\mathrm{KS}}(\mathsf{W} + n^{-1/4}\mathsf{Z}, \mathsf{W}) &\leq d_{\mathrm{TV}}(\mathsf{W} + n^{-1/4}\mathsf{Z}, \mathsf{W}) \\ &= \int_{-\sqrt{\log n}}^{\sqrt{\log n}} |f_{\mathsf{W}+n^{-1/4}\mathsf{Z}}(y) - f_{\mathsf{W}}(y)| dy + \mathbb{P}(|\mathsf{W} + n^{-1/4}\mathsf{Z}| \geq \sqrt{\log n}) \\ &\quad + \mathbb{P}(|\mathsf{W}| \geq \sqrt{\log n}) \\ &\leq C\sqrt{n^{-1}\log n}. \end{aligned}$$

Together with the fact that

$$\begin{aligned} n^{-1/4} v(n^{1/4}\mathsf{W}) &= n^{-1/4}(\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2 \tanh^2(\sqrt{\beta} n^{-1/4}\mathsf{W}))^{1/2} \\ &= n^{-1/4}\mathbb{E}[X_i^2]^{1/2}(1 + O_{\psi_2}(n^{-1/4})), \end{aligned}$$

we know

$$d_{\mathrm{KS}}(n^{-1/4} v(n^{1/4}\mathsf{W})^{1/2}\mathsf{Z} + n^{1/4} e(n^{1/4}\mathsf{W}), \mathsf{W}) = O(\sqrt{\log n}\, n^{-1/2}).$$

Putting together all previous steps, we have

$$d_{\mathrm{KS}}(n^{1/4} g_n, \mathbb{E}[X_i]\mathsf{W}) = O((\log n)^3 n^{-1/2}).$$

SA-7.1.4 Proof for Lemma SA-3 Low Temperature

Throughout the proof, we denote by $\mathtt{C}$ an absolute constant, and $\mathtt{K}$ a constant that only depends on the distribution of X_i. The proofs are based on essentially the same argument as in the high temperature case.

Instead of using sub-Gaussianity of U_n, here we use U_n is sub-Gaussian condition on $\mathsf{U}_n \in \mathcal{I}_\ell$, $\ell \in \{-, +\}$. In particular, the previous step 2 by:

Step 2: Approximation for U_n.

In case $\beta > 1$, $\phi(v) = \frac{1}{2}v^2 - \log(\cosh(\sqrt{\beta}v))$ has two global minimum v_+ and v_-, which are the two solutions of $v - \sqrt{\beta}\tanh(\sqrt{\beta}v) = 0$. We want to show $\phi^{(2)}(v_+) = \phi^{(2)}(v_-) = 1 - \beta + v_+^2 > 0$.

It sufffices to show $v_+ > \sqrt{\beta - 1}$. Since $\phi'(v) < 0$ for $v \in (0, v_+)$ and $\phi'(v) > 0$ for $v \in (v_+, \infty)$, it suffices to show $\phi'(\sqrt{\beta - 1}) < 0$. But

$$\phi'(\sqrt{\beta-1}) < 0 \Leftrightarrow \sqrt{\beta-1} - \sqrt{\beta}\tanh(\sqrt{\beta(\beta-1)}) < 0 \Leftrightarrow \beta > 1.$$

Hence $\phi^{(2)}(v_+) = \phi^{(2)}(v_-) > 0$. Observe that on $\mathcal{I}_- = (-\infty, 0)$ and $\mathcal{I}_+ = (0, \infty)$ respectively, the absolute minimum of ϕ occurs at v_- and v_+, and ϕ' is non-zero on $\mathcal{I}_-$ and $\mathcal{I}_+$ except at v_- and v_+. Hence we can apply Laplace method (Equation 5.1.21 in [2]) sperarately on $\mathcal{I}_-$ and $\mathcal{I}_+$ to get

$$\int_{-\infty}^{0} \exp(-n\phi(v))dv = \sqrt{\frac{2\pi}{n\phi^{(2)}(v_-)}} \exp(-n\phi(v_-))(1 + O(n^{-1})),$$
$$\int_{0}^{\infty} \exp(-n\phi(v))dv = \sqrt{\frac{2\pi}{n\phi^{(2)}(v_+)}} \exp(-n\phi(v_+))(1 + O(n^{-1})).$$

It follows from the definition of f_{V_n} and a change of variable that the density of $\mathsf{U}_n = \sqrt{n}\mathsf{V}_n$ can be approximated by

$$f_{\mathsf{U}_n}(u) = \sum_{l=+,-} \mathbb{1}(u \in \mathcal{C}_l)\sqrt{\frac{\phi^{(2)}(v_-)}{8\pi}} \exp(-n\phi(n^{-1/2}u) + n\phi(n^{-1/2}u_l))(1 + O(n^{-1})),$$

where $u_l = \sqrt{n}v_l, l \in \{+,-\}$. Since $\mathbb{P}(\mathsf{U}_n \in \mathcal{I}_+) = \mathbb{P}(\mathsf{U}_n \in \mathcal{I}_-) = \frac{1}{2}$, condition on $\mathsf{U}_n \in \mathcal{I}_+$,

$$f_{\mathsf{U}_n|\mathsf{U}_n\in\mathcal{I}_+}(u) = \sqrt{\frac{\phi^{(2)}(v_+)}{2\pi}} \exp(-n\phi(n^{-1/2}u) + n\phi(n^{-1/2}u_+))(1 + O(n^{-1})).$$

It then follows from Equation SA-9 that if we define U_+ to be a random variable with density

$$f_{\mathsf{U}_+}(u) = \sqrt{\frac{1-\beta+v_+^2}{2\pi}} \exp(-(1-\beta+v_+^2)(u-u_+)^2/2),$$

then by Taylor expanding ϕ at $v_+ = n^{-1/2}u_+$ and a similar argument as in the proof for high temperature case,

$$d_{\mathrm{TV}}(\mathsf{U}_n|\mathsf{U}_n \in \mathcal{I}_+, \mathsf{U}_+) = O(n^{-1/2}).$$

The rest follows from the same argument as in the proof for high temperature case and is sub-Gaussianity of U_n condition on $\mathsf{U}_n \in \mathcal{I}_\ell$, $\ell \in \{-,+\}$.

SA-7.1.5 Proof for Lemma SA-4 Drifting from High Temperature

Throughout the proof, we denote by C an absolute constant, and K a constant that only depends on the distribution of X_i.

Let $\mathsf{U}_n(c)$, $e(\mathsf{U}_n(c))$, $v(\mathsf{U}_n(c))$ be the latent variable, conditional mean, and conditional variance as previously defined when $\beta_n = 1 + cn^{-\frac{1}{2}}$, $c < 0$. For notational simplicity, we abbreviate the c, and call them $\mathsf{U}_n, e(\mathsf{U}_n), v(\mathsf{U}_n)$ respectively. By Lemma SA-2, $\|\mathsf{U}_n\|_{\psi_2} \leq \mathsf{C}n^{1/4}$.

Step 1: Conditional Berry-Esseen.

Apply Berry-Esseen Theorem conditional on U_n in the same way as in the high temperature case, we get

$$d_{\mathrm{KS}}\left(g_n, v(\mathsf{U}_n)^{1/2}\mathsf{Z} + \sqrt{n}e(\mathsf{U}_n)\right) \leq \mathtt{K}n^{-1/2}.$$

Step 2: Non-Normal Approximation for $n^{-\frac{1}{4}}\mathsf{U}_n$.

Consider $\mathsf{W}_n = n^{-1/4}\mathsf{U}_n$. Then $f_{\mathsf{W}_n}(w) = I_n(c)^{-1}h_n(w)$, with $I_n(c) = \int_{-\infty}^{\infty} h_n(w)dw$, and

$$h_n(w) = \exp\left(-\frac{\sqrt{n}}{2}w^2 + n\log\cosh\left(n^{-\frac{1}{4}}\sqrt{\beta_n}w\right)\right) = \exp\left(-\frac{c}{2}w^2 - \frac{\beta_n^2}{12}w^4 + g(w)\beta_n^3 n^{-\frac{1}{2}}w^6\right),$$

where by smoothness of $\log(\cosh(\cdot))$, $\|\theta\|_\infty \leq \mathtt{K}$. Then

$$\int_{-\mathtt{C}\sqrt{\log n}}^{\mathtt{C}\sqrt{\log n}} h_n(w)dw = \int_{-\mathtt{C}\sqrt{\log n}}^{\mathtt{C}\sqrt{\log n}} \exp(-\frac{c}{2}w^2 - \frac{\beta_n^2}{12}w^4)dw[1 + O(\mathtt{C}^6(\log n)^3 n^{-\frac{1}{2}})] \tag{SA-11}$$

$$= I(c)[1 + O(\mathtt{C}^6(\log n)^3 n^{-\frac{1}{2}})]. \tag{SA-12}$$

Moreover, by a change of variable and the fact that $\beta_n \leq 1$,

$$I_n(c) := \int_{-\infty}^{\infty} h_n(w)dw = n^{-\frac{1}{4}}\int_{-\infty}^{\infty} \exp\left(-n\left(\frac{v^2}{2} - \log\cosh(\sqrt{\beta_n v})\right)\right)dv$$
$$\leq n^{-\frac{1}{4}}\int_{-\infty}^{\infty} \exp\left(-n\left(\frac{v^2}{2} - \log\cosh(\sqrt{v})\right)\right)dv \leq \mathtt{C}.$$

Since $\|\mathsf{W}_n(c)\|_{\psi_2} \leq \mathtt{C}$, $I_n(c)^{-1}\int_{(-\mathtt{C}\sqrt{\log n}, \mathtt{C}\sqrt{\log n})^c} h_n(w)dw \leq \mathtt{C}n^{-1/2}$. It follows that

$$\int_{(-\mathtt{C}\sqrt{\log n}, \mathtt{C}\sqrt{\log n})^c} h_n(w)dw \leq \mathtt{C}n^{-1/2}. \tag{SA-13}$$

Combining Equation SA-11 and SA-13, we have $I_n(c) = I(c)[1 + O(\mathtt{C}^6(\log n)^3 n^{-1/2})]$. It follows that

$$\begin{aligned}
&d_{\mathrm{TV}}(\mathsf{W}_n, \mathsf{W}) \\
\leq &\int_{-\mathtt{C}\sqrt{\log n}}^{\mathtt{C}\sqrt{\log n}} \Big|\frac{h_n(w)}{I_n(c)} - \frac{h(w)}{I(c)}\Big|dw + \mathbb{P}(|\mathsf{W}_n| \geq \mathtt{C}\sqrt{\log n}) + \mathbb{P}(|\mathsf{W}| \geq \mathtt{C}\sqrt{\log n}) \\
\leq &\int_{-\mathtt{C}\sqrt{\log n}}^{\mathtt{C}\sqrt{\log n}} \Big|\frac{h_n(w) - h(w)}{I(c)}\Big| + h_n(w)\Big|\frac{1}{I(c)} - \frac{1}{I_n(c)}\Big|dw + O(n^{-\frac{1}{2}}) \\
\leq &\int_{-\mathtt{C}\sqrt{\log n}}^{\mathtt{C}\sqrt{\log n}} \exp\left(-\frac{c}{2}w^2 - \frac{\beta_n^2}{12}w^4\right)\frac{w^6}{\sqrt{n}I(c)}dw + \int_{-\mathtt{C}\sqrt{\log n}}^{\mathtt{C}\sqrt{\log n}} \frac{1}{I(c)}O(\mathtt{C}^6(\log n)^3 n^{-\frac{1}{2}})dw + O(n^{-\frac{1}{2}}) \\
\leq &\mathtt{C}(\log n)^3 n^{-1/2}.
\end{aligned}$$

Step 3: A Reduction through TV-distance Inequality.

Since $\mathsf{Z} \perp\!\!\!\perp (\mathsf{U}_n, \mathsf{W}_n)$, we can use data processing inequality to get

$$\begin{aligned}
d_{\mathrm{KS}}\left(n^{-\frac{1}{4}}v(\mathsf{U}_n)^{\frac{1}{2}}\mathsf{Z} + n^{\frac{1}{4}}e(\mathsf{U}_n), n^{-\frac{1}{4}}v(n^{\frac{1}{4}}\mathsf{W})^{\frac{1}{2}}\mathsf{Z} + n^{\frac{1}{4}}e(n^{\frac{1}{4}}\mathsf{W})\right) &\leq d_{\mathrm{TV}}(\mathsf{W}_n, \mathsf{W}) \\
&\leq \mathtt{C}(\log n)^3 n^{-1/2}.
\end{aligned}$$

Step 4: Non-Gaussian Approximation for $n^{\frac{1}{4}} e(n^{\frac{1}{4}} \mathsf{W})$.

This is essentially the same as the proof for step 4 from the critical temperature case in Lemma SA-3.

$$d_{\mathrm{KS}}\Big(n^{1/4} e(n^{1/4}\mathsf{W}), \mathbb{E}[X_i]\mathsf{W}\Big) \leq \mathtt{K}\frac{\log n}{\sqrt{n}}.$$

Step 5: Stabilization of Variance.

Using the same argument as Step 4 in the high temperature case for Lemma SA-3, and $\|\mathsf{W}\| \leq \mathtt{K}$,

$$d_{\mathrm{KS}}(n^{-\frac{1}{4}} v(n^{\frac{1}{4}}\mathsf{W})^{\frac{1}{2}}\mathsf{Z} + n^{\frac{1}{4}} e(n^{\frac{1}{4}}\mathsf{W}), n^{-\frac{1}{4}}\mathbb{E}[X_i^2]^{\frac{1}{2}}\mathsf{Z} + \mathbb{E}[X_i]\mathsf{W})) \leq \mathtt{K}\frac{\log n}{\sqrt{n}}.$$

The conclusion then follows from putting together the previous five steps.

SA-7.1.6 Proof for Lemma SA-4 Drifting from Low Temperature

Consider the same U_n defined in Equation (SA-6). Recall $\phi(v) = \frac{v^2}{2} - \log\cosh(\sqrt{\beta_n}v)$, $\phi'(v) = v - \sqrt{\beta_n}\tanh(\sqrt{\beta_n}v)$, $\phi^{(2)}(v) = 1 - \beta_n \operatorname{sech}^2(\sqrt{\beta_n}v)$. And we take $v_+ > 0$, $v_- < 0$ to be the two solutions of $v - \sqrt{\beta_n}\tanh(\sqrt{\beta_n}v) = 0$.

Step 2': Non-Normal Approximation for $n^{-\frac{1}{4}}\mathsf{U}_n$.

Take $\mathsf{V}_n = n^{-1/2}\mathsf{U}_n$. Then $f_{\mathsf{V}_n}(v) \propto \exp(-n\phi(v))$. Taylor expanding ϕ' at 0, we know there exists some function g that is uniformly bounded such that $\phi'(v) = (1-\beta_n)v + \frac{1}{3}\beta_n^2 v^3 + \beta_n^3 g(v) v^5$. Hence

$$v_+ = \sqrt{\frac{3(\beta_n - 1)}{\beta_n^2}} + O(\beta_n - 1) = \sqrt{3c}n^{-1/4} + O(n^{-1/2}).$$

Taylor expand tanh and sech at 0,

$$\begin{aligned}
\phi^{(2)}(v_+) &= 1 - \beta_n + v_+^2 \\
&= -cn^{-1/2} + 3cn^{-1/2}(1 + O(cn^{-1/2}))^{-2} + O((cn^{-1/2})^{5/2}) \\
&= 2cn^{-1/2}(1 + O(cn^{-1/2})), \\
\phi^{(3)}(v_+) &= 2(\beta_n - v_+^2)v_+^2 \\
&= 2\beta_n^{3/2}\operatorname{sech}^2(\sqrt{\beta_n}v_+)\tanh(\sqrt{\beta_n}v_+) \\
&= 2(1 + O(cn^{-1/2}))(1 + O(v_+^2))(\sqrt{\beta_n}v_+ + O(v_+^3)) \\
&= 2\sqrt{3c}n^{-1/4}(1 + O(cn^{-1/2})), \\
\phi^{(4)}(v_+) &= 2(\beta - v_+^2)(\beta - 3v_+^2) \\
&= 2\beta_n^2\operatorname{sech}^4(\sqrt{\beta_n}v_+) - 4\beta_n^2\operatorname{sech}^2(\sqrt{\beta_n}v_+)\tanh^2(\sqrt{\beta_n}v_+) \\
&= 2(1 + O(cn^{-1/2})).
\end{aligned}$$

Take $\mathsf{W}_n = n^{1/4}\mathsf{V}_n = n^{-1/4}\mathsf{U}_n$, $\mathfrak{w}_+ = n^{1/4}v_+ = \sqrt{3c} + O(n^{-1/4})$, and $\mathfrak{w}_- = n^{1/4}v_-$. Define

$$\begin{aligned}
&h_{c,n}(w) \\
&= -\frac{\sqrt{n}\phi^{(2)}(v_+)}{2}(w - \mathfrak{w}_{\mathrm{sgn}(w)})^2 - \frac{n^{1/4}\phi^{(3)}(v_+)}{6}(w - \mathfrak{w}_{\mathrm{sgn}(w)})^3 - \frac{\phi^{(4)}(v_+)}{24}(w - \mathfrak{w}_{\mathrm{sgn}(w)})^4.
\end{aligned}$$

By a change of variable and Taylor expansion, the density for W_n satisfies

$$f_{\mathsf{W}_n}(w) \propto g_{c,\gamma}(w) = \exp\Big(h_{c,n}(w) + O(\|\phi^{(6)}\|_\infty/6!\,)\frac{(w - \mathscr{w}_{\mathrm{sgn}(w)})^6}{\sqrt{n}}\Big). \qquad \text{(SA-14)}$$

By Lemma SA-2, for $\ell \in \{-,+\}$, condition on $\mathsf{W}_n \in \mathcal{I}_{c,n,\ell}$, $\mathsf{W}_n - \mathscr{w}_\ell$ is sub-Gaussian with ψ_2-norm bounded by $\mathtt{C}$. Let $\mathsf{W}_{c,n}$ be a random variable with density at w proportional to $\exp(h_{c,n}(w))$. By similar argument as Equations SA-11 and SA-13,

$$d_{\mathrm{KS}}(\mathsf{W}_n|\mathsf{W}_n \in \mathcal{I}_{c,n,\ell}, \mathsf{W}_{c,n}|\mathsf{W}_{c,n} \in \mathcal{C}_l) \leq \mathtt{C}(\log n)^3 n^{-1/2}).$$

The other steps, *conditional Berry-Esseen*, *reduction through TV-distance inequality*, and *non-Gaussian approximation for* $n^{\frac{1}{4}}e(n^{\frac{1}{4}}\mathsf{W}_{c,n})$ can be proceeded in the same way as in the proof for Lemma SA-3, with $\mathsf{W}_n - \mathscr{w}_\ell$ sub-Gaussian condition on $\mathsf{W}_n \in \mathcal{I}_{c,n,\ell}$ with ψ_2-norm bounded by $\mathtt{C}$, and respectively for $\mathsf{W}_{c,n}$.

SA-7.1.7 Proof for Lemma SA-5 Knife-Edge Representation

Again we take U_n to be the latent variable from Lemma SA-1, and $\mathsf{W}_n = n^{-1/4}\mathsf{U}_n$. From Step 2 in the proof of Lemma SA-4, $f_{\mathsf{W}_n}(w) = I_n(c)^{-1}h_n(w)$, with $I_n(c) = \int_{-\infty}^{\infty} h_n(w)dw$, and

$$h_n(w) = \exp(-\frac{\sqrt{n}}{2}w^2 + n\log\cosh(n^{-\frac{1}{4}}\sqrt{\beta_n}w)) = \exp(-\frac{c_n}{2}w^2 - \frac{\beta_n^2}{12}w^4 + g(w)\beta_n^3 n^{-\frac{1}{2}}w^6),$$

where by smoothness of $\log(\cosh(\cdot))$, $\|\theta\|_\infty \leq \mathtt{K}$.

Case 1: When $\sqrt{n}(\beta_n - 1) = o(1)$. We can apply Berry-Esseen conditional on U_n the same way as in the proof of Lemma SA-4, and its Step 2 can also be applied here to show that if we take $\widetilde{\mathsf{W}}_c$ to be a random variable with density proportional to $\exp(-c_n^2/2w^2 - \beta_n^2/12w^4)$, then $d_{\mathrm{KS}}(\mathsf{W}_n, \widetilde{\mathsf{W}}_c) = O((\log n)^3 n^{-1/2})$. Moreover, $c_n = o(1)$ and $\beta_n = 1 - o(1)$. Hence $d_{\mathrm{KS}}(\mathsf{W}_n, \mathsf{W}_0) = o(1)$. The rest of the proof then follows from Step 3 to Step 5 in the proof for the critical regime case in Lemma SA-3.

Case 2: When $\sqrt{n}(1 - \beta_n) \gg 1$. Again we still have $\|\mathsf{U}_n\|_{\psi_2} = O(n^{1/4})$. Similarly as in the previous case, the first two steps in the proof of Lemma SA-4 implies $d_{\mathrm{KS}}(\mathsf{W}_n, \widetilde{\mathsf{W}}_c) = o(1)$, where the density of W_c is proportional to $\exp(-c_n^2/2w^2 - \beta_n^2/12w^4)$. Since $c_n \gg 1$, the first term in the exponent dominates, and we can show $d_{\mathrm{KS}}(\mathsf{W}_n, \mathsf{W}_c^\dagger) = o(1)$, where $\mathsf{W}_c^\dagger$ has density proportional to $\exp(-c_n^2/2w^2)$. Again, we can Taylor expand to get $n^{1/4}e(n^{1/4}\mathsf{W})) = \mathbb{E}[X_i]n^{\frac{1}{4}}\tanh\Big(n^{-\frac{1}{4}}\mathsf{W}\Big) = \mathbb{E}[X_i][\mathsf{W} - O(\frac{\mathsf{W}^2}{3\sqrt{n}})]$, and show $d_{\mathrm{KS}}(n^{1/4}e(n^{1/4}\mathsf{W}_c^\dagger), \mathbb{E}[X_i]\mathsf{W}_c^\dagger) = o(1)$. Combining with stablization of variance as in the proof of Lemma SA-2 (high temperature case), we can show

$$d_{\mathrm{KS}}(g_n, n^{-1/4}\mathbb{E}[X_i^2]^{1/2}\mathsf{Z} + \mathbb{E}[X_i]\mathsf{W}_c^\dagger) = o(1).$$

Since Z and $\mathsf{W}_c^\dagger$ are independent Gaussian random variables, we also have $d_{\mathrm{KS}}(g_n/\sqrt{\mathbb{V}[g_n]}, \mathsf{Z}) = o(1)$.

Case 3: When $\sqrt{n}(\beta_n - 1) \gg 1$. By Lemma SA-4 (2),

$$\sup_{t\in\mathbb{R}}\Big|\mathbb{P}(n^{\frac{1}{4}}g_n \leq t|m \in \mathcal{I}_{c,\ell}) - \mathbb{P}(n^{-\frac{1}{4}}\mathbb{E}[X_i^2]^{\frac{1}{2}}\mathsf{Z} + \beta_n^{\frac{1}{2}}\mathbb{E}[X_i]\mathsf{W}_{c_n,n} \leq t|\mathsf{W}_{c_n,n} \in \mathcal{I}_{c,\ell})\Big| = o(1), \qquad \text{(SA-15)}$$

where $\mathsf{W}_{c,n}$ has density proportional to $\exp(h_{c,n}(w))$, with

$$h_{c,n}(w) = -\frac{\sqrt{n}\phi^{(2)}(v_+)}{2}(w - \mathfrak{w}_{\mathrm{sgn}(w)})^2 - \frac{n^{1/4}\phi^{(3)}(v_+)}{6}(w - \mathfrak{w}_{\mathrm{sgn}(w)})^3 - \frac{\phi^{(4)}(v_+)}{24}(w - \mathfrak{w}_{\mathrm{sgn}(w)})^4,$$

and $\mathcal{I}_{c,n,-} = (-\infty, K_{c,n,-})$ and $\mathcal{I}_{c,n,+} = (K_{c,n,+}, \infty)$ such that $\mathbb{E}[\mathsf{W}_{c,n}|\mathsf{W}_{c,n} \in \mathcal{I}_{c,n,\ell}] = w_{c,n,\ell}$ for $\ell \in \{-,+\}$. Now we calculate the order of the coefficients under $\sqrt{n}(\beta_n - 1) \gg 1$. First, suppose $\beta_n = 1 + cn^{\gamma}$ for some $\gamma \in (0,\infty)$ and c not depending on n. Then $v_+ = \sqrt{\frac{3(\beta_n-1)}{\beta_n^2}} + O(\beta_n - 1) = \sqrt{3c}n^{-\gamma/2} + O(n^{-\gamma})$. Taylor expand tanh and sech at 0,

$$\begin{aligned}
\phi^{(2)}(v_+) &= 1 - \beta_n + v_+^2 = -cn^{-\gamma} + cn^{-\gamma}3(1 + cn^{-\gamma})^{-2} + O((cn^{-\gamma})^{5/2}) \\
&= 2cn^{-\gamma}(1 + O(cn^{-\gamma})), \\
\phi^{(3)}(v_+) &= 2\beta_n^{3/2}\operatorname{sech}^2(\sqrt{\beta_n}v_+)\tanh(\sqrt{\beta_n}v_+) \\
&= 2(1 + O(cn^{-\gamma}))(1 + O(v_+^2))(\sqrt{\beta_n}v_+ + O(v_+^3)) \\
&= 2\sqrt{3c}n^{-\gamma/2}(1 + O(cn^{-\gamma})), \\
\phi^{(4)}(v_+) &= -2\beta_n^4\operatorname{sech}^4(\sqrt{\beta_n}v) + 4\operatorname{sech}^2(\sqrt{\beta_n}v)\tanh^2(\sqrt{\beta_n}v) \\
&= -2(1 + O(cn^{-\gamma})).
\end{aligned}$$

We see when $\gamma = 1/2$, all of $\sqrt{n}\phi^{(2)}(v_+)$, $n^{1/4}\phi^{(3)}(v_+)$ and $\phi^{(4)}(v_+)$ are of order 1. And when $c_n = \sqrt{n}(\beta_n - 1) \gg 1$, we have $\sqrt{n}\phi^{(2)}(v_+) \gg n^{1/4}\phi^{(3)}(v_+) \gg \phi^{(4)}(v_+)$. Since $w_+ = n^{1/4}v_+ = \sqrt{3c_n} \gg 1$, and similarly, $|w_-| \gg 1$, condition on $\mathsf{W}_{c,n} \in [n]$, $\mathsf{W}_{c,n} - \mathbb{E}[\mathsf{W}_{c,n}|\mathsf{W}_{c,n} \in [n]]$ is C-sub-Gaussian, $\ell \in \{-,+\}$. By similar concentration arguments as in the proof for Step 2 in Lemma SA-4 (1), we can show the second order term in $h_{c,n}$ dominates, and for $\ell \in \{-,+\}$,

$$\sup_{t \in \mathbb{R}} |\mathbb{P}(\mathsf{W}_{c,n} - \mathbb{E}[\mathsf{W}_{c,n}|\mathsf{W}_{c,n} \in \mathcal{I}_\ell] \le t | \mathsf{W}_{c,n} \in \mathcal{I}_\ell) - \Phi(\sqrt{n(1 - \beta_n + v_\ell^2)}t)| = o(1).$$

The conclusion then follows from pluggin the (conditional) Gaussian approximation for $\mathsf{W}_{c_n,n}$ back into Equation (SA-15), and the fact that Z is independent to $\mathsf{W}_{c,n}$ and also Gaussian.

SA-7.2 Proof of Section SA-3

SA-7.2.1 Proof of Lemma SA-1

Our proof is constructive. We show that consistent estimate of $n\mathbb{V}[\widehat{\tau}_n]$ would imply that one can distinguish between two constructed hypotheses easily. Let $\mathscr{P}_n$ be the class of distributions of random vectors $(\mathbf{W} = (W_1, \cdots, W_n), \mathbf{Y} = (Y_1, \cdots, Y_n))$ taking values in $\mathbb{R}^{2n}$ that satisfies Assumptions 1,2,3. Consider the following two data generating processes:

$$\begin{aligned}
&\mathrm{DGP}_0: \quad \beta = 0, \quad G(\cdot,\cdot) \equiv 1, \quad \rho_n = 1, \quad Y_i(\cdot,\cdot) = f_i(\cdot,\cdot) + \varepsilon_i, \quad f_i(\cdot,\cdot) \equiv 1, \\
&\mathrm{DGP}_1: \quad \beta = u, \quad G(\cdot,\cdot) \equiv 1, \quad \rho_n = 1, \quad Y_i(\cdot,\cdot) = f_i(\cdot,\cdot) + \varepsilon_i, \quad f_i(\cdot,\cdot) \equiv 1,
\end{aligned}$$

where $0 < u < 1$, and in both cases $(\varepsilon_i : 1 \le i \le n)$ are i.i.d $\mathsf{N}(0,1)$ random variables, independent to $\mathbf{W}$. Denote by $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ the laws of $(\mathbf{W},\mathbf{Y})$ under DGP_0 and DGP_1. Then

$$\begin{aligned}
d_{\mathrm{KL}}(\mathbb{P}_{0,n}(\mathbf{W},\mathbf{Y}), \mathbb{P}_{1,n}(\mathbf{W},\mathbf{Y})) &= d_{\mathrm{KL}}(\mathbb{P}_{0,n}(\mathbf{W}), \mathbb{P}_{1,n}(\mathbf{W})) + d_{\mathrm{KL}}(\mathbb{P}_{0,n}(\mathbf{Y}|\mathbf{W}), \mathbb{P}_{1,n}(\mathbf{Y}|\mathbf{W})) \\
&= d_{\mathrm{KL}}(\mathbb{P}_{0,n}(\mathbf{W}), \mathbb{P}_{1,n}(\mathbf{W})),
\end{aligned}$$

the first line uses chain rule of d_{KL}, the second line uses

$$d_{\mathrm{KL}}(\mathbb{P}_{0,n}(\mathbf{Y}|\mathbf{W}),\mathbb{P}_{1,n}(\mathbf{Y}|\mathbf{W})) = d_{\mathrm{KL}}(\mathbb{P}_{0,n}(\mathbf{Y}),\mathbb{P}_{1,n}(\mathbf{Y})) = 0.$$

From Theorem 2.3 (and its proof) in [1],

$$M := \lim_{n\to\infty} d_{\mathrm{KL}}(\mathbb{P}_{0,n}(\mathbf{W}),\mathbb{P}_{1,n}(\mathbf{W})) < \infty.$$

Hence for large enough n,

$$\begin{aligned} d_{\mathrm{TV}}(\mathbb{P}_{0,n}(\mathbf{W},\mathbf{Y}),\mathbb{P}_{1,n}(\mathbf{W},\mathbf{Y})) &\leq 1-\frac{1}{2}\exp(-d_{\mathrm{KL}}(\mathbb{P}_{0,n}(\mathbf{W},\mathbf{Y}),\mathbb{P}_{1,n}(\mathbf{W},\mathbf{Y})))\\ &\leq 1-\frac{1}{2}\exp(-M). \end{aligned}$$

Le Cam's method (Section 15.2.1 in [8]) gives for large enough n,

$$\begin{aligned} &\inf_{\widehat{\mathbb{V}}}\sup_{\mathbb{P}_n\in\mathscr{P}_n}\mathbb{E}_{\mathbb{P}_n}[n(\widehat{\mathbb{V}}[\widehat{\tau}-\tau]-\mathbb{V}[\widehat{\tau}-\tau])]\\ &\qquad\geq n|\mathbb{V}_{\mathbb{P}_{n,0}}[\widehat{\tau}-\tau]-\mathbb{V}_{\mathbb{P}_{n,1}}[\widehat{\tau}-\tau]|(1-d_{\mathrm{TV}}(\mathbb{P}_{0,n}(\mathbf{W},\mathbf{Y}),\mathbb{P}_{1,n}(\mathbf{W},\mathbf{Y})))\\ &\qquad\geq\varepsilon\exp(-M)/2, \end{aligned}$$

in the last line we used Theorem 2 (1) to get $n\mathbb{V}_{\mathbb{P}_{n,0}}[\widehat{\tau}-\tau]-n\mathbb{V}_{\mathbb{P}_{n,1}}[\widehat{\tau}-\tau]=\varepsilon(1+o(1))$.

SA-7.2.2 Proof of Lemma SA-2

The following discussions will be organized according to the three different cases: (1) When $\beta<1$. (2) When $\beta\geq 1$, m concentrates around 0. (3) When $\beta\geq 1$ and m concentrates around two symmetric locations $w_+>0$ and $w_-<0$ with $|w_+|=|w_-|$.

We have required $\widehat{\beta}\in[0,1]$. For analysis, consider an unrestricted pseud-likelihood estimator,

$$\widehat{\beta}_{\mathrm{UR}} = \arg\max_{\beta\in\mathbb{R}} l(\beta;\mathbf{W}),$$

where $l(\beta;\mathbf{W})$ is the pseudo log-likelihood given by

$$l(\beta;\mathbf{W}) = \sum_{i\in[n]}\log\mathbb{P}_\beta\left(W_i \mid \mathbf{W}_{-i}\right) = \sum_{i\in[n]} -\log\left(\frac{1}{2}W_i\tanh(\beta m_i)+\frac{1}{2}\right).$$

We show that $l(\beta;\mathbf{W})$ is concave.

$$\begin{aligned} \frac{\partial}{\partial\beta}l(\beta;\mathbf{W}) &= -\frac{1}{n}\sum_{i=1}^{n}\frac{(n^{-1}\sum_{j\neq i}W_j)W_i\,\mathrm{sech}^2(\beta n^{-1}\sum_{j\neq i}W_j)}{W_i\tanh(\beta n^{-1}\sum_{j\neq i}W_j)+1}\\ &= -\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j\neq i}W_j\right)(W_i-\tanh(\beta n^{-1}\sum_{j\neq i}W_j)), \end{aligned}$$

and

$$l^{(2)}(\beta;\mathbf{W}) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j\neq i}W_j\right)^2\mathrm{sech}^2\left(\frac{\beta}{n}\sum_{j\neq i}W_j\right) > 0.$$

Hence $l(\cdot;\mathbf{W})$ is concave everywhere in $\mathbb{R}$. This shows $\widehat{\beta}=\min\{\max\{\widehat{\beta}_{\mathrm{UR}},0\},1\}$. Now we study limiting distribution of $\widehat{\beta}_{\mathrm{UR}}$

1. High and critical temperature regime.

To obtain a more precise distribution for $\widehat{\beta}_{\mathrm{UR}}$, we use Fermat's condition to obtain that

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j\neq i}W_j\right)\left(W_i - \tanh\left(\widehat{\beta}_{\mathrm{UR}} n^{-1}\sum_{j\neq i}W_j\right)\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(m-\frac{W_i}{n}\right)\left(W_i - \tanh(\widehat{\beta}_{\mathrm{UR}} m) + \operatorname{sech}^2(\widehat{\beta}_{\mathrm{UR}} m)\frac{\widehat{\beta}_{\mathrm{UR}} W_i}{n} + O(n^{-2})\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\left(m-\frac{W_i}{n}\right)\left(\left(1+\operatorname{sech}^2(\widehat{\beta}_{\mathrm{UR}} m)\frac{\widehat{\beta}_{\mathrm{UR}}}{n}\right)W_i - \tanh(\widehat{\beta}_{\mathrm{UR}} m) + O(n^{-2})\right)\\
&= \left(1+\frac{\widehat{\beta}_{\mathrm{UR}}}{n}\operatorname{sech}^2(\widehat{\beta}_{\mathrm{UR}} m)\right)\left(m^2-\frac{1}{n}\right) - \frac{n-1}{n} m\tanh(\widehat{\beta}_{\mathrm{UR}} m) + O(n^{-2})m,
\end{aligned}
$$

here $O(\cdot)$'s are all up to an absolute constant. By Lemma SA-4 with $X_i = 1$, we can show $\mathbb{E}[|(nm)^{-1}|] \leq \mathtt{C}n^{-1/2}$. By Markov inequality, $(nm)^{-1} = O_{\mathbb{P}}(n^{-1/2})$. Taylor expanding tanh, we have

$$
\begin{aligned}
\widehat{\beta}_{\mathrm{UR}} &= \frac{n}{(n-1)m}\tanh^{-1}\left(m-\frac{1}{nm}\right)\\
&= \frac{n}{(n-1)m}\left(m-\frac{1}{nm}+\frac{1}{3}\left(m-\frac{1}{nm}\right)^3 + O\left(\left(m-\frac{1}{nm}\right)^5\right)\right)\\
&= 1-\frac{1}{nm^2}+\frac{m^2}{3}+O_{\mathbb{P}}(n^{-1}),
\end{aligned}
\tag{SA-16}
$$

where in the above equation, both $O(\cdot)$ and $O_{\mathbb{P}}(\cdot)$ are up to absolute constants. The rest of the results are given according to the different temperature regimes.

(1) The High Temperature Regime. Using Lemma SA-2 with $X_i = 1$, our result for the high temperature regime with $\beta < 1$ implies that $n^{\frac{1}{2}}m \xrightarrow{d} N(0, \frac{1}{1-\beta}) \Rightarrow (1-\beta)nm^2 \xrightarrow{d} \chi^2(1)$. Therefore we conclude that $\frac{1-\beta}{1-\widehat{\beta}_{\mathrm{UR}}} \xrightarrow{d} \chi^2(1)$. The conclusion then follows from $\widehat{\beta} = \min\{\max\{\widehat{\beta}_{\mathrm{UR}}, 0\}, 1\}$.

(2) The Critical Temperature Regime. Using Lemma SA-2 with $X_i = 1$, we have $d_{\mathrm{KS}}(n^{\frac{1}{4}}m, \mathsf{W}_0) = o(1)$. This implies $n^{\frac{1}{2}}(\widehat{\beta}_{\mathrm{UR}} - 1) \xrightarrow{d} \mathit{Law}(\frac{\mathsf{W}_0^2}{3} - \frac{1}{\mathsf{W}_0^2})$. Since $\mathsf{W}_0 = O_{\mathbb{P}}(1)$, $\mathbb{P}(\widehat{\beta}_{\mathrm{UR}} < 0) = o(1)$. The conclusion then follows from $\widehat{\beta} = \min\{\max\{\widehat{\beta}_{\mathrm{UR}}, 0\}, 1\}$.

2. The low temperature regime.

When m concentrates around π_+ and π_- we have when $m > 0$, use the fact that $\pi_\ell = \tanh(\beta\pi_\ell)$ for $\ell \in \{+,-\}$,

$$
\begin{aligned}
\widehat{\beta}_{\mathrm{UR}} - \beta &= \frac{(1-O(n^{-1}))(m-\tanh(\beta m))}{m\operatorname{sech}^2(\beta m)} + mO(\delta^2) + O(n^{-1})\\
&= \frac{(1-O(n^{-1}))\left((m-\pi_\ell) - (\tanh(\beta m)-\tanh(\beta\pi_\ell))\right)}{\pi_\ell\left(\operatorname{sech}^2(\beta\pi_\ell) - 2(m-\pi_\ell)\tanh(\beta\pi_\ell)\operatorname{sech}^2(\beta\pi_\ell) + O(m-\pi_\ell)^2\right)\left(1+\frac{m-\pi_\ell}{\pi_\ell}\right)}\\
&\quad + mO(\delta^2) + O(n^{-1})\\
&= (1-O(n^{-1}))\frac{(1-\beta\operatorname{sech}^2(\beta\pi_\ell))(m-\pi_\ell)}{\pi_\ell\operatorname{sech}^2(\beta\pi_\ell)}(1+O(m-\pi_\ell)) + mO(\delta^2) + O(n^{-1}).
\end{aligned}
$$

and the similar argument gives

$$m(\widehat{\beta}_{\mathrm{UR}} - \beta^*) = \frac{1 - \beta^* \operatorname{sech}^2(\beta^* \pi_\ell)}{\operatorname{sech}^2(\beta^* \pi_\ell)}(m - \pi_\ell) + O_{\psi_1}(n^{-1}).$$

The conclusion then Lemma SA-3 (3) and the convergence of m to π_+ or π_-.

SA-7.2.3 Proof of Lemma SA-3

Again we consider the unrestricted PMLE given by

$$\widehat{\beta}_{\mathrm{UR}} = \arg\max_{\beta \in \mathbb{R}} l(\beta; \mathbf{W}),$$

where $l(\beta; \mathbf{W})$ is the pseudo log-likelihood given by

$$l(\beta; \mathbf{W}) = \sum_{i \in [n]} \log \mathbb{P}_\beta \left(W_i \mid \mathbf{W}_{-i} \right) = \sum_{i \in [n]} - \log \left(\frac{1}{2} W_i \tanh(\beta m_i) + \frac{1}{2} \right).$$

For $\beta \in [0, 1]$, that is $c_\beta = \sqrt{n}(\beta - 1) \leq 0$, Equation (SA-16) and the approximation of m by $n^{-1/2}\mathsf{Z} + n^{-1/4}\mathsf{W}_c$ from Lemma SA-4 gives

$$\sup_{\beta \in [0,1]} \sup_{t \in \mathbb{R}} |\mathbb{P}(1 - \widehat{\beta} \leq t) - \mathbb{P}(z_{\beta,n}^{-2} - \frac{3}{n} z_{\beta,n}^2 \leq t)| = o(1).$$

The conclusion follows from the fact that $x \mapsto \max\{\min\{x, 0\}, 1\}$ is 1-Lipschitz.

SA-7.3 Proofs for Section SA-4

SA-7.3.1 Preliminary Lemmas

Lemma SA-1. *Suppose $\pi = \mathbb{E}[W_i]$ where $\mathbf{W} = (W_i)_{1 \leq i \leq n}$ takes value in $\{-1, 1\}^n$ and*

$$\mathbb{P}\left(\mathbf{W} = \mathbf{w} \right) = \frac{1}{Z} \exp \left(\frac{\beta}{n} \sum_{i<j} W_i W_j + h \sum_{i=1}^n W_i \right), \quad \beta = 1, h = 0.$$

Suppose either $h \neq 0$ or $h = 0, 0 \leq \beta \leq 1$ holds. Then $\pi = \tanh(\beta \pi + h) + O(n^{-1})$.

Proof. First, if $h = 0$, then $\pi = \tanh(\beta \pi + h) = 0$. Now, consider $\pi \neq 0$. Using concentration of $m := \frac{1}{n} \sum_{i=1}^n W_i$ towards π from Lemma SA-3,

$$\begin{aligned}
\pi &= \mathbb{E}[\mathbb{E}[W_i | W_{-i}]] = \mathbb{E}[\tanh(\beta m_i + h)] \\
&= \mathbb{E}[\tanh(\beta \pi + h) + \operatorname{sech}^2(\beta \pi + h)(m_i - \pi) - \operatorname{sech}^2(\beta m^* + h) \tanh(\beta m^* + h)(m_i - \pi)^2] \\
&= \tanh(\beta \pi + h) + O(n^{-1}).
\end{aligned}$$

□

Lemma SA-2. *Suppose Assumption 1, and Assumption 2, 3 hold. Then (1)*

$$\max_{i \in [n]} \left| \frac{M_i}{N_i} - \pi \right| = O_{\psi_{\beta,\gamma}}(n^{-\mathtt{r}_{\beta,h}}) + O_{\psi_2}(N_i^{-1/2}).$$

(2) Define $A(\mathbf{U}) = (G(U_i, U_j))_{1\le i,j\le n}$. Condition on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathscr{A} = \{A \in \mathbb{R}^{n\times n} : \min_{i\in[n]} \sum_{j\neq i} A_{ij} \ge 32\log n\}$, for large enough n, for each $i \in [n]$ and $t > 0$,

$$\mathbb{P}\left(\left|\frac{M_i}{N_i} - \pi\right| \ge 4\mathbb{E}[N_i|\mathbf{U}]^{-1/2}t^{1/2} + C_{\beta,h}n^{-\mathtt{r}_{\beta,h}}t^{\mathtt{p}_{\beta,h}}\middle|\mathbf{U}\right) \le 2\exp(-t) + n^{-98},$$

where $C_{\beta,h}$ is some constant that only depends on β, h.

(3) When $h = 0$, and $\beta \in [0,1]$, then there exists a constant $\mathtt{K}$ that does not depend on β, such that for large enough n, for each $i \in [n]$ and $t > 0$,

$$\mathbb{P}\left(\left|\frac{M_i}{N_i} - \pi\right| \ge 4\mathbb{E}[N_i|\mathbf{U}]^{-1/2}t^{1/2} + \mathtt{K}n^{-\mathtt{r}_{\beta,h}}t\middle|\mathbf{U}\right) \le 2\exp(-t) + n^{-98}.$$

Proof. Take U_n to be a random variable with density

$$f_{\mathsf{U}_n}(u) = \frac{\exp\left(-\frac{1}{2}u^2 + n\log\cosh\left(\sqrt{\frac{\beta}{n}}u + h\right)\right)}{\int_{-\infty}^{\infty}\exp\left(-\frac{1}{2}v^2 + n\log\cosh\left(\sqrt{\frac{\beta}{n}}v + h\right)\right)dv}.$$

Condition on U_n, W_i's are i.i.d. Decompose by

$$\frac{M_i}{N_i} - \pi = \sum_{j\neq i}\frac{E_{ij}}{N_i}\left(W_j - \mathbb{E}[W_j|\mathsf{U}_n]\right) + \mathbb{E}[W_j|\mathsf{U}_n] - \pi.$$

Condition on U_n, W_i's are i.i.d. Berry-Esseen theorem condition on U_n and $\mathbf{E}$ gives,

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\Big(\frac{M_i}{N_i} - \pi \le t\Big|\mathbf{E}\Big) - \mathbb{P}\Big(\sqrt{\frac{v(\mathsf{U}_n)}{N_i}}Z + e(\mathsf{U}_n) \le t\Big|\mathbf{E}\Big)\right| = O(n^{-\frac{1}{2}}), \tag{SA-17}$$

where $e(\mathsf{U}_n) := \mathbb{E}[W_i|\mathsf{U}_n] - \pi = \tanh(\sqrt{\beta/n}\mathsf{U}_n + h) - \pi$, and $v(\mathsf{U}_n) := \mathbb{V}[W_i - \pi|\mathsf{U}_n]$. By McDiarmid's inequality,

$$\mathbb{P}\Big(|\sum_{j\neq i}\frac{E_{ij}}{N_i}\left(W_j - \mathbb{E}[W_j|\mathsf{U}_n]\right)| \ge 2N_i^{-1/2}t\Big|\mathbf{E}\Big) \le 2\exp(-t^2).$$

Plugging into Equation (SA-17), we can show (1) holds.

Next, we want to show condition on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathscr{A}$, $\mathbb{P}(N_i \le \mathbb{E}[N_i|\mathbf{U}]/3|\mathbf{U}) \le n^{-100}$:

Notice that for any $\mathbf{U}$ such that $\rho_n \min_{i\in[n]}\sum_{j\neq i}A_{ij}(\mathbf{U}) \to \infty$, Condition on A such that $A \in \mathscr{A}$, $E_{ij} = \rho A_{ij}\iota_{ij}$, $1 \le i \le j \le n$ are i.i.d Bernouli random variables, and for each i, j, $\sum_{k\neq i,j}A_{ki} \ge 32\log n - 1 \ge 31\log n$ for $n \ge 3$. By bounded difference inequality, for all $t > 0$,

$$\mathbb{P}\Big(\Big|\sum_{k\neq i,j}E_{ki} - \sum_{k\neq i,j}\rho_n A_{ki}\Big| \ge \rho_n\sqrt{\sum_{k\neq i,j}A_{i,j}^2 t}\Big) \le 2\exp(-2t^2).$$

Hence condition on A, with probability at least $1-n^{-100}$,

$$\begin{aligned}\sum_{k\neq i,j} E_{ki} &\geq \sum_{k\neq i,j}\rho_n A_{ki} - 8\sqrt{\log n}\rho_n\sqrt{\sum_{k\neq i,j} A_{ij}^2} \geq \rho_n \sum_{k\neq i,j} A_{ki} - 8\sqrt{\log n}\rho_n\sqrt{\sum_{k\neq i,j} A_{ki}} \\ &\geq \rho_n \sqrt{\sum_{k\neq i,j} A_{ki}}\left(\sqrt{\sum_{k\neq i,j} A_{ki}} - 8\sqrt{\log n}\right) \\ &\geq \rho_n \sqrt{\sum_{k\neq i,j} A_{ki}}\left(\sqrt{\sum_{k\neq i,j} A_{ki}} - 8\sqrt{31^{-1}\sum_{k\neq i,j} A_{ij}}\right) \\ &\geq \rho_n \sum_{k\neq i,j} A_{ij}/3 \geq \frac{31}{3}\log n, \end{aligned} \tag{SA-18}$$

and since $\rho_n A_{i,j} = \mathbb{E}[E_{ij}|\mathbf{U}] \in [0,1]$, $\sum_{k\neq i,j} E_{ki} + 1 \geq \mathbb{E}[N_j|\mathbf{A}]/3$. By Equation SA-18, condition on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathscr{A}$, $\mathbb{P}(N_i \leq \mathbb{E}[N_i|\mathbf{U}]/3|\mathbf{U}) \leq n^{-100}$.

Hence we can disintegrate over the distribution of $\mathbf{E}$ to get

$$\mathbb{P}\left(|\sum_{j\neq i}\frac{E_{ij}}{N_i}\left(W_j - \mathbb{E}[W_j|\mathsf{U}_n]\right)| \geq 4\mathbb{E}[N_i|\mathbf{U}]^{-1/2}t \middle| \mathbf{U}\right) \leq 2\exp(-t^2) + n^{-100}.$$

By Equation SA-7 and Lemma SA-2, and the Lipschitzness of tanh that

$$\mathbb{E}[W_i|\mathsf{U}_n] - \pi = O_{\psi_{\beta,h}}\left(n^{-\mathtt{r}_{\beta,h}}\right).$$

Plugging into Equation (SA-17), we can show (2) holds.

Under the setting of (3), the only part that depends on β in our proof is U_n. Since we show in Lemma SA-2 $\|\mathsf{U}_n\|_{\psi_1} \leq \mathtt{K} n^{1/4}$ for some absolute constant $\mathtt{K}$, which is essentially the $\beta = 1$ rate, the conclusion of (3) then follows. □

SA-7.3.2 Proof of Lemma SA-1

Since we use the conditional probability p_i in the inverse probability weight, we have

$$\begin{aligned}\mathbb{E}[\widehat{\tau}_{n,\mathrm{UB}}|(f_i)_{i\in[n]}, \mathbf{E}] &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\frac{T_iY_i}{p_i} - \frac{(1-T_i)Y_i}{1-p_i}\middle|(f_i)_{i\in[n]}, \mathbf{E}\right] \\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[\mathbb{E}\left[\frac{T_iY_i}{p_i} - \frac{(1-T_i)Y_i}{1-p_i}\middle|\mathbf{T}_{-i}, (f_i)_{i\in[n]}, \mathbf{E}\right]\middle|(f_i)_{i\in[n]}, \mathbf{E}\right], \end{aligned}$$

and the conclusion follows from $\mathbb{E}[T_i|\mathbf{T}_{-i}, (f_i)_{i\in[n]}, \mathbf{E}] = p_i$.

SA-7.3.3 Proof of Lemma SA-2

First consider the treatment part.

$$n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n \frac{T_i}{p_i} g_i(1,\pi) = n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n g_i(1,\pi) + n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n \frac{T_i - p_i}{p_i} g_i(1,\pi).$$

For the second term, taylor expand p_i^{-1}, p_i as follows:

$$
\begin{aligned}
p_i^{-1} =& 1+\exp\left(-2\beta m_i-2h\right) = 1+\exp\left(-2\beta\frac{n-1}{n}\pi-2h\right)\\
&-\exp\left(-2\beta\frac{n-1}{n}\pi-2h\right)2\beta\left(m_i-\frac{n-1}{n}\pi\right)+\frac{1}{2}\exp(-\xi_i^*)4\beta^2\left(m_i-\frac{n-1}{n}\pi\right)^2,
\end{aligned}
\tag{SA-19}
$$

where ξ_i^* is some random quantity that lies between $4\frac{\beta}{n}\sum_{j\neq i}W_j$ and $4\frac{\beta}{n}\sum_{j\neq i}\pi$. Taking the parameters $c_i^+ = g_i\left(1,\pi\right)\left(1+\exp(-2\beta\pi-2h)\right)$, $d^+ = \beta(1-\tanh(\beta\pi+h))\mathbb{E}[g_i(1,\pi)]$. Then

$$
\begin{aligned}
& n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n \frac{T_i-p_i}{p_i}g_i\left(1,\pi\right)\\
\stackrel{(1)}{=}& n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n (T_i-p_i)g_i\left(1,\pi\right)\left(1+\exp\left(-2\beta\pi-2h\right)-\exp\left(-2\beta\pi-2h\right)2\beta(m_i-\pi)\right)\\
&+O_{\psi_{\beta,h},tc}(n^{-\mathtt{r}_{\beta,h}})\\
\stackrel{(2)}{=}& n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n c_i(T_i-p_i)+O_{\psi_{\beta,h},tc}((\log n)^{1/2}n^{-\mathtt{r}_{\beta,h}})\\
\stackrel{(3)}{=}& n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n c_i^+\left[T_i-\frac{1}{1+\exp(-2\beta\pi-2h)}-\frac{2\beta\exp(2\beta\pi+2h)}{(1+\exp(2\beta\pi+2h))^2}(m_i-\pi)\right]\\
&+O_{\psi_{\beta,h},tc}((\log n)^{1/2}n^{-\mathtt{r}_{\beta,h}})\\
=& n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n \frac{c_i^+}{2}\left(W_i-\tanh(\beta\pi+h)\right)\\
&-n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n\frac{2\beta\exp(2\beta\pi+2h)}{(1+\exp(2\beta\pi+2h))^2}\Big(\frac{1}{n}\sum_{j\neq i}c_j^+\Big)\left(W_i-\pi\right)+O_{\psi_{\beta,h},tc}((\log n)^{1/2}n^{-\mathtt{r}_{\beta,h}})\\
\stackrel{(4)}{=}& n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n\left[g_i\left(1,\pi\right)+\left(c_i^+/2-d^+\right)\left(W_i-\pi\right)\right]+O_{\psi_{\beta,h},tc}((\log n)^{1/2}n^{-\mathtt{r}_{\beta,h}}).
\end{aligned}
$$

Proof of (1): By Lemma SA-3, $m-\pi = O_{\psi_{\beta,h}}(n^{-\mathtt{r}_{\beta,h}})$. The claim follows from Equation SA-19 and a union bound argument.

Proof of (2):

$$
\begin{aligned}
n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n (T_i-p_i)g_i(1,\pi)(m_i-\pi) =& \frac{1}{2}(m-\pi)n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n (W_i-\tanh(\beta m+h))g_i(1,\pi)\\
&+O(n^{-\mathtt{a}_{\beta,h}}).
\end{aligned}
$$

By Lemma SA-3,

$$
m-\pi = O_{\psi_{\beta,h},tc}(n^{-\mathtt{r}_{\beta,h}}).
$$

Taylor expand $\tanh(x)$ at $x = \beta\pi + h$, we have

$$
\begin{aligned}
&n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} g_i(1,\pi)(W_i - \tanh(\beta m + h)) \\
=&n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} g_i(1,\pi)(W_i - \tanh(\beta\pi + h) - \beta \operatorname{sech}^2(\beta\pi + h)(m - \pi) + \tanh(\beta\pi + h) \operatorname{sech}^2(\beta\pi + h)(m - \pi)^2 \\
&\quad + O((m - \pi)^3)) \\
=&O_{\psi_{\beta,h,tc}}(1).
\end{aligned}
$$

hence

$$
n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} (T_i - p_i) g_i(1,\pi)(m_i - \pi) = O_{\psi_{\beta,h},tc}((\log n)^{1/2} n^{-\mathtt{r}_{\beta,h}}).
$$

Proof of (3): The first line follows from a Taylor expansion of $p_i = (1 + \exp(2\beta m_i + 2h))^{-1}$ at π, and $m_i - \pi = O_{\psi_{\beta,h}}(n^{-\mathtt{r}_{\beta,h}})$, noticing that c_i, $\|\psi''\|_\infty$ are bounded. The second line follows by reordering the terms.

Proof of (4): By Lemma SA-1, $\tanh(\beta\pi + h) = \pi + O(n^{-1})$. By boundedness and i.i.d of $g_i(1,\pi)$, $\frac{1}{n}\sum_{j \neq i} c_j = \bar{c} + O(n^{-1}) = \mathbb{E}[c_i] + O_{\mathbb{P}}(n^{-1/2}) + O(n^{-1})$. Similarly, for the control part, taking the parameters $c_i^- = g_i(-1,\pi)(1 + \exp(2\beta\pi + 2h))$, $d^- = \beta(1 - \tanh(-\beta\pi - h))\mathbb{E}[g_i(-1,\pi)]$.

$$
\begin{aligned}
&- n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \frac{1 - T_i}{1 - p_i} g_i(-1,\pi) \\
=&- n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} g_i(-1,\pi) + n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} (c_i^-/2 - d^-)(W_i - \pi) + O_{\psi_{\beta,h},tc}((\log n)^{1/2} n^{-\mathtt{r}_{\beta,h}}).
\end{aligned}
$$

Using Lemma SA-1 again, we can show $(1+\exp(-2\beta\pi - 2h))/2 = 1/\pi + O(n^{-1})$ and $(1+\exp(2\beta\pi + 2h))/2 = 1/(1-\pi) + O(n^{-1})$, $\tanh(-\beta\pi - h) = -\pi + O(n^{-1})$. The result then follows from replacing these quantities in c_i^+, c_i^-, d^+, d^- by corresponding ones using π.

SA-7.3.4 Proof of Lemma SA-3

We decompose by $\Delta_{2,2} = \Delta_{2,2,1} + \Delta_{2,2,2}$, where

$$
\begin{aligned}
\Delta_{2,2,1} &= n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \frac{T_i - \mathbb{E}[p_i]}{\mathbb{E}[p_i]} g_i'(1,\pi) \left(\frac{M_i}{N_i} - \pi \right), \\
\Delta_{2,2,2} &= n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} T_i \left(p_i^{-1} - \mathbb{E}[p_i]^{-1} \right) g_i'(1,\pi) \left(\frac{M_i}{N_i} - \pi \right).
\end{aligned}
$$

Notice that the first term is a quadractic form. Define $\mathbf{H}$ such that $H_{ij} = \frac{g_i'(1,\pi)E_{ij}}{2\mathbb{E}[p_i]N_i}$. Then $\Delta_{2,2,1} = n^{-\mathtt{a}_{\beta,h}}(\mathbf{W} - \pi)^{\mathrm{T}}\mathbf{H}(\mathbf{W} - \pi)$. Take U_n to be the latent variable from Lemma SA-1. Then we decompose

$$
\Delta_{2,2,1} = \Delta_{2,2,1,a} + \Delta_{2,2,1,b} + \Delta_{2,2,1,c} + \Delta_{2,2,1,d},
$$

where

$$\Delta_{2,2,1,a} = n^{-\mathtt{a}_{\beta,h}}(\mathbf{W} - \mathbb{E}[\mathbf{W}|\mathsf{U}_n])^{\mathrm{T}}\mathbf{H}(\mathbf{W} - \mathbb{E}[\mathbf{W}|\mathsf{U}_n]),$$
$$\Delta_{2,2,1,b} = n^{-\mathtt{a}_{\beta,h}}(\mathbb{E}[\mathbf{W}|\mathsf{U}_n] - \pi)^{\mathrm{T}}\mathbf{H}(\mathbf{W} - \mathbb{E}[\mathbf{W}|\mathsf{U}_n]),$$
$$\Delta_{2,2,1,c} = n^{-\mathtt{a}_{\beta,h}}(\mathbf{W} - \mathbb{E}[\mathbf{W}|\mathsf{U}_n])^{\mathrm{T}}\mathbf{H}(\mathbb{E}[\mathbf{W}|\mathsf{U}_n] - \pi),$$
$$\Delta_{2,2,1,d} = n^{-\mathtt{a}_{\beta,h}}(\mathbb{E}[\mathbf{W}|\mathsf{U}_n] - \pi)^{\mathrm{T}}\mathbf{H}(\mathbb{E}[\mathbf{W}|\mathsf{U}_n] - \pi).$$

Since $\|\mathbf{H}\|_2 \leq \|\mathbf{H}\|_F \leq \frac{B}{2\pi}\sqrt{n}(\min_i N_i)^{-1/2}$, we can apply Hanson-Wright inequality conditional on $\mathsf{U}_n, \mathbf{E}$,

$$\Delta_{2,2,1,a} = O_{\psi_1}(n^{\frac{1}{2}-\mathtt{a}_{\beta,h}}(\min_i N_i)^{-1/2}).$$

Since $g_i'(1,\pi)$'s are independent to W_i, by Lemma SA-3,

$$n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}(W_i - \pi)g_i'(1,\pi) = O_{\psi_{\beta,h},tc}(1).$$

By Equation SA-7, Lipschitzness of tanh and Lemma SA-2, $\mathbb{E}[W_i|\mathsf{U}_n] - \pi = O_{\psi_{\beta,h}}(n^{-\mathtt{r}_{\beta,h}})$, hence

$$\Delta_{2,2,1,b} = (\mathbb{E}[W_i|\mathsf{U}_n] - \pi)\, n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\frac{T_i - \mathbb{E}[p_i]}{\mathbb{E}[p_i]}g_i'(1,\pi) = O_{\psi_{\beta,h},tc}\left((\log n)^{-1/2}n^{-\mathtt{r}_{\beta,h}}\right).$$

Then by concentration of $\frac{M_i}{N_i}$ from Lemma SA-2, we have

$$\begin{aligned}
|\Delta_{2,2,1,c}| &= \left|\frac{\mathbb{E}[W_i|\mathsf{U}_n] - \pi}{2\mathbb{E}[p_i]}n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}g_i'(1,\pi)\left(\frac{M_i}{N_i} - \pi\right)\right| \\
&\leq n^{\mathtt{r}_{\beta,h}}\left|\frac{\mathbb{E}[W_i|\mathsf{U}_n] - \pi}{2\mathbb{E}[p_i]}\right| \cdot \max_{i\in[n]}\left|\frac{M_i}{N_i} - \pi\right| \\
&= O_{\psi_2,tc}\left(\log n \max_{i\in[n]}\mathbb{E}[N_i|\mathbf{U}]^{-1/2}\right) + O_{\psi_{\beta,\gamma},tc}(n^{-\mathtt{r}_{\beta,h}}).
\end{aligned}$$

The bound for $\Delta_{2,2,1,d}$ follows from the definition of $\mathbf{H}$ and U_n,

$$\Delta_{2,2,1,d} = n^{\mathtt{r}_{\beta,h}}\left(\tanh\left(\sqrt{\frac{\beta}{n}}\mathsf{U}_n + h\right) - \mathbb{E}\left[\tanh\left(\sqrt{\frac{\beta}{n}}\mathsf{U}_n + h\right)\right]\right)^2 = O_{\psi_{\beta,\gamma}}(n^{-\mathtt{r}_{\beta,h}}).$$

SA-7.3.5 Proof of Lemma SA-4

Take U_n to be the latent variable given in Lemma SA-1. We further decompose by

$$\Delta_{2,3,1} = n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\frac{1}{2}g_i^{(2)}(1,\eta_i^*)\left(\sum_{j\neq i}\frac{E_{ij}}{N_i}(W_i - \pi)\right)^2 = \Delta_{2,3,1,a} + \Delta_{2,3,1,b} + \Delta_{2,3,1,c},$$

where η_i^* is some value between π and M_i/N_i, and

$$\Delta_{2,3,1,a} = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \frac{1}{2} g_i^{(2)}(1, \eta_i^*) \left(\sum_{j \neq i} \frac{E_{ij}}{N_i} (W_j - \mathbb{E}[W_j|\mathsf{U}_n]) \right)^2,$$

$$\Delta_{2,3,1,b} = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \frac{1}{2} g_i^{(2)}(1, \eta_i^*) \left(\sum_{j \neq i} \frac{E_{ij}}{N_i} (W_j - \mathbb{E}[W_j|\mathsf{U}_n]) \right) (\mathbb{E}[W_j|\mathsf{U}_n] - \pi),$$

$$\Delta_{2,3,1,c} = n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \frac{1}{2} g_i^{(2)}(1, \eta_i^*) (\mathbb{E}[W_j|\mathsf{U}_n] - \pi)^2.$$

Part I: $\Delta_{2,3,1,c}$.

$\mathbb{E}[W_i|\mathsf{U}_n, \mathbf{U}] = \tanh\left(\sqrt{\frac{\beta}{n}}\mathsf{U}_n + h\right)$, hence $\mathbb{E}[W_i|\mathsf{U}_n] - \pi = O_{\psi_{\beta,h}}(n^{-\mathtt{r}_{\beta,h}})$ and $(\mathbb{E}[W_i|\mathsf{U}_n] - \pi)^2 = O_{\psi_{\mathtt{p}_{\beta,h}/2}}(n^{-2\mathtt{r}_{\beta,h}})$. It then follows from boundness of $g_i^{(2)}(1, \eta_i^*)$ that

$$\Delta_{2,3,1,c} = O_{\psi_{\mathtt{p}_{\beta,h}/2}}(n^{-\mathtt{r}_{\beta,h}}).$$

Part II: $\Delta_{2,3,1,b}$.

Condition on U_n, W_i's are i.i.d. By Mc-Diarmid inequality conditional on U_n for each $\sum_{j \neq i} \frac{E_{ij}}{N_i}(W_j - \mathbb{E}[W_j|\mathsf{U}_n])$ and using a union bound over $i \in [n]$, for all $i \in [n]$, for all $t > 0$,

$$\mathbb{P}\left(|\Delta_{2,3,1,b}| \geq 2 \max_i N_i^{-1/2} n^{\mathtt{r}_{\beta,h}} |\mathbb{E}[W_j|\mathsf{U}_n] - \pi| \sqrt{t} \middle| \mathsf{U}_n, \mathbf{E} \right) \leq 2n \exp(-t).$$

The tails for $n^{\mathtt{r}_{\beta,h}}(\mathbb{E}[W_j|\mathsf{U}_n] - \pi)$ are also controlled,

$$\mathbb{P}\left(n^{\mathtt{r}_{\beta,h}} |\mathbb{E}[W_j|\mathsf{U}_n] - \pi| \geq C_{\beta,h} (\log n)^{1/\mathtt{p}_{\beta,h}} \right) \leq n^{-1/2}.$$

Integrate over the distribution of U_n and using a union bound, for large n, for all $t > 0$,

$$\mathbb{P}\left(|\Delta_{2,3,1,b}| \geq 2 C_{\beta,h} \max_i N_i^{-1/2} t^{1/\mathtt{p}_{\beta,h}} \middle| \mathbf{E} \right) \leq 2n \exp(-t) + C_{\beta,h} n^{-1/2}.$$

By Equation SA-18, condition on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathscr{A}$, $\mathbb{P}(N_i \leq \mathbb{E}[N_i|\mathbf{U}]/3|\mathbf{U}) \leq n^{-100}$. Hence for such $\mathbf{U}$,

$$\mathbb{P}\left(|\Delta_{2,3,1,b}| \geq 4 C_{\beta,h} \max_i \mathbb{E}[N_i|\mathbf{U}]^{-1/2} t^{1/\mathtt{p}_{\beta,h}} \middle| \mathbf{U} \right) \leq 2n \exp(-t) + C_{\beta,h} n^{-1/2}.$$

In other words, conditional on $\mathbf{U}$ s.t. $A(\mathbf{U}) \in \mathscr{A}$,

$$\Delta_{2,3,1,b} = O_{\psi_{\beta,h},tc}(\max_i \mathbb{E}[N_i|\mathbf{U}]^{-1/2}).$$

Part III: $\Delta_{2,3,1,a}$.

For notational simplicity, we will denote

$$A_i = \frac{1}{2} g_i^{(2)}(1, \eta_i^*) \left(\sum_{j \neq i} \frac{E_{ij}}{N_i} (W_j - \mathbb{E}[W_j | \mathsf{U}_n]) \right)^2$$
$$= \frac{1}{2} \theta \left(\frac{M_i}{N_i} \right) \left(\sum_{j \neq i} \frac{E_{ij}}{N_i} (W_j - \mathbb{E}[W_j | \mathsf{U}_n]) \right)^2 =: F(\mathbf{W}, \mathsf{U}_n),$$

and since we assume $g_i(\ell, \cdot)$ is C^4 for $\ell \in \{-1, 1\}$, we know $\theta(\ell, \cdot)$ is C^2 for $\ell \in \{-1, 1\}$. Then we can decompose $\Delta_{2,3,1,a} - \mathbb{E}[\Delta_{2,3,1,a} | \mathbf{E}]$ as

$$\Delta_{2,3,1,a} - \mathbb{E}[\Delta_{2,3,1,a} | \mathbf{E}] = n^{-\mathsf{a}_{\beta,h}} \sum_{i=1}^n (A_i - \mathbb{E}[A_i | \mathsf{U}_n, \mathbf{E}]) + n^{-\mathsf{a}_{\beta,h}} \sum_{i=1}^n (\mathbb{E}[A_i | \mathsf{U}_n, \mathbf{E}] - \mathbb{E}[A_i | \mathbf{E}]) .$$

where F is a function that possibly depends on $\beta(\mathbf{U})$ and $\mathbf{E}$.

First part of $\Delta_{2,3,1,a}$: The first two terms have a quadratic form in $W_j - \mathbb{E}[W_j | \mathsf{U}_n]$, except for the term $\theta(M_i/N_i)$. We will handle it via a generalized version of Hanson-Wright inequality. Fix U_n and $\mathbf{E}$, consider

$$H(\mathbf{W}) = n^{-1/2} \sum_{i=1}^n \frac{1}{2} \theta \left(\frac{M_i}{N_i} \right) \left(\sum_{j \neq i} \frac{E_{ij}}{N_i} (W_j - \mathbb{E}[W_j | \mathsf{U}_n]) \right)^2 .$$

Denoting by $D_k H$ the partial derivative of H w.r.p to W_k and $D_{k,l}$ the mixed partials, then

$$D_k H(\mathbf{W}) = n^{-1/2} \sum_{i \neq k}^n \frac{1}{2} \theta' \left(\frac{M_i}{N_i} \right) \frac{E_{ik}}{N_i} \left(\sum_{j \neq i} \frac{E_{ij}}{N_i} (W_j - \mathbb{E}[W_j | \mathsf{U}_n]) \right)^2$$
$$+ n^{-1/2} \sum_{i \neq k}^n \theta \left(\frac{M_i}{N_i} \right) 2 \left(\sum_{j \neq i} \frac{E_{ij}}{N_i} (W_j - \mathbb{E}[W_j | \mathsf{U}_n]) \right) \frac{E_{ik}}{N_i}.$$

Since we have assumed f is at least 4-times continuously differentiable, we can apply standard concentration inequalities for $\sum_{j \neq i} \frac{E_{ij}}{N_i} (W_j - \mathbb{E}[W_j | \mathsf{U}_n])$ to get

$$|\mathbb{E}[D_k H(\mathbf{W}) | \mathsf{U}_n, \mathbf{E}]| \lesssim n^{-1/2} \sum_{i=1}^n E_{ik} N_i^{-3/2}.$$

Hence the gradient of H is bounded by

$$\|\mathbb{E}[DH(\mathbf{W}) | \mathsf{U}_n, \mathbf{E}]\|_2^2 \lesssim \sum_{k=1}^n n^{-1} \left(\sum_{i=1}^n E_{ik} N_i^{-3/2} \right)^2$$
$$\lesssim \sum_{k=1}^n n^{-1} \left(\sum_{i=1}^n E_{ik} N_i^{-3} + \sum_{j_1=1} \sum_{j_2 \neq j_1} \frac{E_{j_1 k} E_{j_2 k}}{N_{j_1}^{3/2} N_{j_2}^{3/2}} \right)$$
$$\lesssim \frac{\max_i N_i^2}{\min_i N_i^3}.$$

Moreover, the mix partials are

$$D_{k,l}H(\mathbf{W}) = n^{-1/2}\sum_{i\neq k,l}^{n}\theta''\left(\frac{M_i}{N_i}\right)\left(\sum_{j\neq i}\frac{E_{ij}}{N_i}(W_j-\mathbb{E}[W_j|\mathsf{U}_n])\right)^2\frac{E_{ik}E_{il}}{N_i^2}$$
$$+2n^{-1/2}\sum_{i=1}^{n}\theta'\left(\frac{M_i}{N_i}\right)2\left(\sum_{j\neq i}\frac{E_{ij}}{N_i}(W_j-\mathbb{E}[W_j|\mathsf{U}_n])\right)\frac{E_{ik}E_{il}}{N_i^2}$$
$$+n^{-1/2}\sum_{i=1}^{n}\theta\left(\frac{M_i}{N_i}\right)\frac{E_{ik}E_{il}}{N_i^2}.$$

Hence $\|D_{k,l}H(\mathbf{W})\|_\infty \lesssim n^{-1/2}\sum_{i=1}^n \frac{E_{ik}E_{il}}{N_i^2}$. Hence

$$\|\|HF\|_F^2\|_\infty \lesssim \sum_{k=1}^n\sum_{l=1}^n\left(n^{-1/2}\sum_{i=1}^n\frac{E_{ik}E_{il}}{N_i^2}\right)^2 \lesssim n^{-1}\sum_{i_1=1}^n\sum_{l=1}^n\frac{E_{i_1l}}{N_{i_1}}\sum_{k=1}^n\frac{E_{i_1k}}{N_{i_1}}\sum_{i_2=1}^n\frac{E_{i_2k}}{N_{i_2}}\frac{1}{N_{i_2}}\lesssim\frac{\max_i N_i}{\min_i N_i^2}.$$

Moreover, since HF is symmetric,

$$\|\|HF\|_2\|_\infty \le \|\|HF\|_1\|_\infty \lesssim \max_k\sum_{l=1}^n n^{-1/2}\sum_{i=1}^n\frac{E_{ik}E_{il}}{N_i^2}\lesssim n^{-1/2}\frac{\max_i N_i}{\min_i N_i}.$$

Hence by Theorem 3 from [4], for all $t>0$,

$$\mathbb{P}\left(\left|n^{-1/2}\sum_{i=1}^n(A_i-\mathbb{E}[A_i|\mathsf{U}_n,\mathbf{E}])\right|\ge t\middle|\mathsf{U}_n,\mathbf{E}\right)$$
$$\le\exp\left(-c\min\left(\frac{t^2}{\frac{\max_i N_i^2}{\min_i N_i^3}+\frac{\max_i N_i}{\min_i N_i^2}},\frac{t}{n^{-1/2}\frac{\max_i N_i}{\min_i N_i}}\right)\right).$$

By Equation SA-18 and a similar argument for upper bound, for each $i\in[n]$, conditional on $\mathbf{U}$ such that $A(\mathbf{U})\in\mathscr{A}$, with probability at least $1-n^{-100}$, $\mathbb{E}[N_i|\mathbf{U}]/2\le N_i\le 2\mathbb{E}[N_i]$. Hence for each $t>0$,

$$\mathbb{P}\left(\left|n^{-1/2}\sum_{i=1}^n(A_i-\mathbb{E}[A_i|\mathsf{U}_n,\mathbf{E}])\right|\ge 8\max_i\mathbb{E}[N_i|\mathbf{U}]^{-1/2}\sqrt{t}+8C_{\beta,h}n^{-1/2}t\middle|\mathbf{U}\right)\le\exp(-t)$$
$$+n^{-99},$$

that is

$$n^{-\mathsf{a}_{\beta,h}}\sum_{i=1}^n(A_i-\mathbb{E}[A_i|\mathsf{U}_n,\mathbf{E}])=O_{\psi_2,tc}\left(n^{\frac{1}{2}-\mathsf{a}_{\beta,h}}\max\mathbb{E}[N_i|\mathbf{U}]^{-1/2}\right)+O_{\psi_1,tc}\left(n^{-1/2}\right).\qquad\text{(SA-20)}$$

Second part of $\Delta_{2,3,1,a}$: Next, we will show $n^{1-\mathsf{a}_{\beta,h}}(\mathbb{E}[A_i|U,\mathbf{U},\mathbf{E}]-\mathbb{E}[A_i|\mathbf{E}])$, is small. There exists a function F that possibly depends on β and $\mathbf{E}$ such that

$$F(\mathbf{W},\mathsf{U}_n)=\frac{1}{2}\theta\left(\frac{M_i}{N_i}\right)\left(\sum_{j\neq i}\frac{E_{ij}}{N_i}(W_j-\mathbb{E}[W_j|\mathsf{U}_n])\right)^2.$$

Define $p(u) = \mathbb{P}(W_j = 1|U, \mathbf{U})$. Then

$$\mathbb{E}[A_i|U = u, \mathbf{U}, \mathbf{E}] = \mathbb{E}[F(\mathbf{W}, U)|U = u] = \sum_{\mathbf{w}\in\{-1,1\}^n} \prod_{l=1}^{n} p(u)^{w_l}(1-p(u))^{1-w_l} F(\mathbf{w}, u).$$

Using chain rule and product rule for derivatives,

$$\begin{aligned}
&\partial_u \mathbb{E}\left[A_i|U = u, \mathbf{U}\right] \\
&= \sum_{\mathbf{w}\in\{-1,1\}^n} \Bigg[\sum_{l=1}^{n} \prod_{s\neq l} p(u)^{w_s}(1-p(u))^{1-w_s} \left(F((\mathbf{w}_{-l}, w_l = 1), u) - F((\mathbf{w}_{-l}, w_l = -1), u)\right) \\
&\qquad + \prod_{i=1}^{n} p(u)^{w_i}(1-p(u))^{1-w_i} \partial_u F(\mathbf{w}, u)\Bigg] p'(u) \\
&= \sum_{l=1}^{n} \mathbb{E}_{\mathbf{W}_{-l}}\left[F((\mathbf{W}_{-l}, W_l = 1), u) - F((\mathbf{W}_{-l}, W_l = -1), u)\right] p'(u) + \mathbb{E}_{\mathbf{W}}\left[\partial_u F(\mathbf{W}, u)\right] p'(u) \\
&= \sum_{l=1}^{n} O_{\mathbb{P}}\left(\frac{1}{\sqrt{N_i}} \frac{E_{il}}{N_i}\right) \|p'\|_\infty + O_{\mathbb{P}}\left(\frac{1}{\sqrt{N_i}} \|p'\|_\infty\right) \|p'\|_\infty = O_{\mathbb{P}}((nN_i)^{-0.5}),
\end{aligned}$$

where in the last line, we have used

$$\begin{aligned}
&|D_{W_l} F(\mathbf{w}, u)| \lesssim \|\theta'\|_\infty \frac{E_{il}}{N_i} \left(\sum_{j\neq i} \frac{E_{ij}}{N_i}(W_j - \mathbb{E}[W_j|U, \mathbf{U}])\right)^2 + \|\theta\|_\infty \left|\sum_{j\neq i} \frac{E_{ij}}{N_i}(W_j - \mathbb{E}[W_j|U, \mathbf{U}])\right| \frac{E_{il}}{N_i}, \\
&|\partial_u F(\mathbf{w}, u)| \lesssim \|\theta\|_\infty \|p'\|_\infty \left|\sum_{j\neq i} \frac{E_{ij}}{N_i}(W_j - \mathbb{E}[W_j|U, \mathbf{U}])\right|,
\end{aligned}$$

and that fact that $\|p'\|_\infty = O((2\beta/n)^{0.5})$ and Hoeffding's inequality for $\sum_{j\neq i} \frac{E_{ij}}{N_i}(W_j - \mathbb{E}[W_j|\mathsf{U}_n])$,

$$|\partial_u \mathbb{E}\left[F(\mathbf{w}, \mathsf{U}_n)|\mathsf{U}_n = u, \mathbf{E}\right]| \leq \mathbb{E}\left[|\partial_u F(\mathbf{w}, \mathsf{U}_n)||\mathsf{U}_n = u\right] = O\left(n^{-1/2} \min_i N_i^{-1/2}\right).$$

Since $\mathsf{U}_n = O_{\psi_{\beta,h}}(n^{\mathsf{a}_{\beta,h}-1/2})$, we have

$$\begin{aligned}
n^{-\mathsf{a}_{\beta,h}} \sum_{i=1}^{n} \left(\mathbb{E}[A_i|\mathsf{U}_n, \mathbf{U}] - \mathbb{E}[A_i|\mathbf{U}]\right) &= O_{\psi_{\beta,h}}\left(n^{1-\mathsf{a}_{\beta,h}} n^{-1/2} \min_i N_i^{-1/2} n^{\mathsf{a}_{\beta,h}-1/2}\right) \\
&= O_{\psi_{\beta,h}}\left(\min_i N_i^{-1/2}\right). \qquad \text{(SA-21)}
\end{aligned}$$

Combining Equations SA-20 and SA-21, conditional on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathscr{A}$,

$$\begin{aligned}
n^{-\mathsf{a}_{\beta,h}} \sum_{i=1}^{n} \left(A_i - \mathbb{E}[A_i|\mathbf{E}]\right) &= O_{\psi_2,tc}\left(n^{\frac{1}{2}-\mathsf{a}_{\beta,h}} \max \mathbb{E}[N_i|\mathbf{U}]^{-1/2}\right) + O_{\psi_1,tc}\left(n^{-1/2}\right) \\
&\quad + O_{\psi_{\beta,h},tc}\left(\max_i \mathbb{E}[N_i]^{-1/2}\right).
\end{aligned}$$

Combining the bounds for $\Delta_{2,3,1,a}, \Delta_{2,3,1,b}, \Delta_{2,3,1,c}$, we get the desired result.

SA-7.3.6 Proof of Lemma SA-5

Recall

$$\Delta_{2,3,2}=n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\frac{1}{2}\frac{W_i-\mathbb{E}[W_i|\mathbf{W}_{-i}]}{p_i}\left[g_i\left(1,\frac{M_i}{N_i}\right)-g_i\left(1,\pi\right)-g_i'\left(1,\pi\right)\left(\frac{M_i}{N_i}-\pi\right)\right].$$

First, we will consider the effect of fluctuation of p_i and $\mathbb{E}[W_i|\mathbf{W}_{-i}]$. Recall

$$\mathbb{E}[W_i|\mathbf{W}_{-i}]=\tanh\left(\beta m_i+h\right),\quad p_i=\left(1+\exp\left(-2\beta m_i-2h\right)\right)^{-1}.$$

It follows from the boundeness of βm_i+h, $m_i-\pi=O_{\psi_{\beta,h}}(n^{-\mathtt{r}_{\beta,h}})$ that for each $i\in[n]$,

$$\frac{W_i-\mathbb{E}[W_i|\mathbf{W}_{-i}]}{p_i}=2\frac{W_i-\pi}{\pi+1}+O_{\psi_{\beta,h}}(n^{-\mathtt{r}_{\beta,h}}).$$

Moreover for some η_i^* between M_i/N_i and π, using Lemma SA-2 we have

$$\begin{aligned}&g_i\left(1,\frac{M_i}{N_i}\right)-g_i\left(1,\pi\right)-g_i'\left(1,\pi\right)\left(\frac{M_i}{N_i}-\pi\right)\\=&\frac{1}{2}g_i''(1,\eta_i^*)\left(\frac{M_i}{N_i}-\pi\right)^2=O_{\psi_{\mathtt{p}_{\beta,h}/2},tc}(n^{-2\mathtt{r}_{\beta,h}})+O_{\psi_1,tc}(N_i^{-1}).\end{aligned}$$

Using a union bound over i and an argument for the product of two terms with bounded Orlicz norm with tail control, we have

$$\begin{aligned}\Delta_{2,3,2}=&n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\frac{W_i-\pi}{\pi+1}\left[g_i\left(1,\frac{M_i}{N_i}\right)-g_i(1,\pi)-g_i'(1,\pi)\left(\frac{M_i}{N_i}-\pi\right)\right]\\&+O_{\psi_{\mathtt{p}_{\beta,h}/2},tc}((\log n)^{-1/\mathtt{p}_{\beta,h}}n^{-2\mathtt{r}_{\beta,h}})+O_{\psi_1,tc}((\log n)^{-1/\mathtt{p}_{\beta,h}}N_i^{-1}).\end{aligned}$$

Next, we will show $n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\frac{W_i-\pi}{\pi+1}\left[g_i\left(1,\frac{M_i}{N_i}\right)-g_i(1,\pi)-g_i'(1,\pi)\left(\frac{M_i}{N_i}-\pi\right)\right]$ is small. Suppose $g_i(1,\cdot)$ is p-times continuously differentiable. Define

$$\delta_p=n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\frac{W_i-\pi}{\pi+1}g_i^{(p)}\left(1,\pi\right)\left(\frac{M_i}{N_i}-\pi\right)^p.$$

We will use the conditioning strategy to analyse δ_p: Decompse by

$$\delta_p=\delta_{p,1}+\delta_{p,2}+\delta_{p,3},$$

with

$$\begin{aligned}\delta_{p,1}&=n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\frac{W_i-\mathbb{E}[W_i|\mathsf{U}_n]}{\pi+1}g_i^{(p)}\left(1,\pi\right)\left(\frac{M_i}{N_i}-\mathbb{E}[W_i|\mathsf{U}_n]\right)^p,\\\delta_{p,2}&=n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\frac{\mathbb{E}[W_i|\mathsf{U}_n]-\pi}{\pi+1}g_i^{(p)}\left(1,\pi\right)\left(\frac{M_i}{N_i}-\mathbb{E}[W_i|\mathsf{U}_n]\right)^p,\\\delta_{p,3}&=n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^{n}\frac{W_i-\pi}{\pi+1}g_i^{(p)}(1,\pi)\left[\left(\frac{M_i}{N_i}-\mathbb{E}[W_i|\mathsf{U}_n]\right)^p-\left(\frac{M_i}{N_i}-\pi\right)^p\right].\end{aligned}$$

First, we will show $\delta_{p,2}$ and $\delta_{p,3}$ are small. By Hoeffding inequality, $M_i/N_i - \mathbb{E}[W_i|\mathsf{U}_n] = O_{\psi_2}(N_i^{-1/2})$. Moreover, $\mathbb{E}[W_i|\mathsf{U}_n] - \pi = O_{\psi_{\beta,h}}(n^{-\mathtt{r}_{\beta,h}})$. Hence

$$\delta_{p,2} = O_{\psi_{\beta,h},tc}(\max_i N_i^{-1/2}).$$

For $\delta_{p,3}$, we have

$$\left(\frac{M_i}{N_i} - \mathbb{E}[W_i|\mathsf{U}_n]\right)^p - \left(\frac{M_i}{N_i} - \pi\right)^p = p\left(\frac{M_i}{N_i} - \xi^*\right)^{p-1}\left(\mathbb{E}[W_i|\mathsf{U}_n] - \pi\right),$$

where ξ^* is some quantity between $\mathbb{E}[W_i|\mathsf{U}_n]$ and π. Since $x \mapsto x^{p-1}$ is either monotone or convex and none-negative, condition on $\mathbf{E}$,

$$\begin{aligned}\left|\frac{M_i}{N_i} - \xi^*\right|^{p-1} &\le \max\left\{\left|\frac{M_i}{N_i} - \mathbb{E}[W_i|\mathsf{U}_n]\right|^{p-1}, \left|\frac{M_i}{N_i} - \pi\right|^{p-1}\right\} \\ &= O_{\psi_{\frac{\mathtt{p}_{\beta,h}}{p-1}}}(n^{-(p-1)\mathtt{r}_{\beta,h}}) + O_{\psi_{\frac{2}{p-1}}}(N_i^{-\frac{p-1}{2}}).\end{aligned}$$

Combining with boundedness of $g_i^{(p)}(1,\pi)$ and tail control of $\mathbb{E}[W_i|\mathsf{U}_n]$, we have

$$\delta_{p,3} = O_{\psi_{\frac{\mathtt{p}_{\beta,h}}{p-1}}}\left((\log n)^{\frac{1}{\mathtt{p}_{\beta,h}}} n^{-(p-1)\mathtt{r}_{\beta,h}}\right) + O_{\psi_{\frac{2}{p-1}}}\left((\log n)^{\frac{1}{\mathtt{p}_{\beta,h}}} N_i^{-\frac{p-1}{2}}\right).$$

For $\delta_{p,1}$, we will again use the generalized version of Hanson-Wright inequality. For each $k \in [n]$,

$$\begin{aligned}\partial_k \delta_{p,1} =& n^{-\mathtt{a}_{\beta,h}} \sum_{i \neq k} \frac{W_i - \mathbb{E}[W_i|\mathsf{U}_n]}{\pi + 1} g_i^{(p)}(1,\pi) p \left(\frac{M_i}{N_i} - \mathbb{E}[W_i|\mathsf{U}_n]\right)^{p-1} \frac{E_{ik}}{N_i} \\ &+ n^{-\mathtt{a}_{\beta,h}} g_k^{(p)}(1,\pi)\left(\frac{M_k}{N_k} - \mathbb{E}[W_i|\mathsf{U}_n]\right)^p.\end{aligned}$$

Hence condition on $\mathbf{E}$,

$$\|\mathbb{E}\left[\nabla \delta_{p,1}\right]\| = O\left(n^{1/2-\mathtt{a}_{\beta,h}} N_i^{-(p-1)/2}\right).$$

Taking mixed partials w.r.p $\delta_{p,1}$ and using boundedness of $g_i^{(p)}$, we have

$$\|\partial_k \partial_l \delta_{p,1}\|_\infty \lesssim n^{-\mathtt{a}_{\beta,h}} \sum_{i \neq k,l} \frac{E_{ik}E_{il}}{N_i^2} + n^{-\mathtt{a}_{\beta,h}} \frac{E_{lk}}{N_l} + n^{-\mathtt{a}_{\beta,h}} \frac{E_{kl}}{N_k}.$$

It follows that

$$\|\|\mathrm{Hess}(\delta_{p,1})\|_2\|_\infty \lesssim \|\|\mathrm{Hess}(\delta_{p,1})\|_F\|_\infty \lesssim n^{1/2-\mathtt{a}_{\beta,h}}\left(\frac{\max_i N_i^3}{\min_i N_i^4}\right)^{1/2}.$$

It then follows from Equation SA-18 and Theorem 3 in [4] that conditional on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathcal{A}$,

$$\delta_{p,1} - \mathbb{E}[\delta_{p,1}|\mathbf{E}] = O_{\psi_1,tc}\left(n^{1/2-\mathtt{a}_{\beta,h}}\left(\frac{\max_i \mathbb{E}[N_i|\mathbf{U}]^3}{\min_i \mathbb{E}[N_i|\mathbf{U}]^4}\right)^{1/2}\right).$$

Trade-off Between Smoothness of $g_i(1,\cdot)$ and Sparsity of Graph Assume $g_i(1,\cdot)$ is $p+1$-times continuously differentiable. Then by the decomposition of $\Delta_{2,3,2}$, condition on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathscr{A}$,

$$\begin{aligned}
&\Delta_{2,3,2} - \mathbb{E}[\Delta_{2,3,2}|\mathbf{E}] \\
=& \sum_{l=2}^{p} \delta_l - \mathbb{E}[\delta_l|\mathbf{E}] + n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} \left[\frac{Y_i^{(p+1)}(1,\xi_i^*)}{(p+1)!} \left(\frac{M_i}{N_i} - \pi\right)^{p+1} - \mathbb{E}\left[\frac{Y_i^{(p+1)}(1,\xi_i^*)}{(p+1)!} \left(\frac{M_i}{N_i} - \pi\right)^{p+1} \middle| \mathbf{E} \right] \right] \\
&+ O_{\psi_{\mathtt{p}_{\beta,h}/2},tc}((\log n)^{-1/\mathtt{p}_{\beta,h}} n^{-2\mathtt{r}_{\beta,h}}) + O_{\psi_1,tc}((\log n)^{-1/\mathtt{p}_{\beta,h}} (\min_i \mathbb{E}[N_i|\mathbf{U}])^{-1}).
\end{aligned}$$

Then by the concentration of $M_i/N_i - \pi$ given in Lemma SA-2, we have

$$\begin{aligned}
&\Delta_{2,3,2} - \mathbb{E}[\Delta_{2,3,2}|\mathbf{E}] \\
=& O_{\psi_{\mathtt{p}_{\beta,h}/2},tc}((\log n)^{-1/\mathtt{p}_{\beta,h}} n^{-2\mathtt{r}_{\beta,h}}) + O_{\psi_1,tc}((\log n)^{-1/\mathtt{p}_{\beta,h}} (\min_i \mathbb{E}[N_i|\mathbf{U}])^{-1}) \\
&+ O_{\psi_1,tc}\left(n^{1/2-\mathtt{a}_{\beta,h}} \left(\frac{\max_i \mathbb{E}[N_i|\mathbf{U}]^3}{\min_i \mathbb{E}[N_i|\mathbf{U}]^4} \right)^{1/2} \right) + O_{\psi_{2/(p+1)},tc}\left(n^{\mathtt{r}_{\beta,h}} (\min_i \mathbb{E}[N_i|\mathbf{U}]^{-(p+1)/2}) \right).
\end{aligned}$$

SA-7.3.7 Proof of Lemma SA-6

For notational simplicity, denote $\widehat{p} = \frac{1}{n}\sum_{i=1}^n T_i$ and $p = \frac{1}{2}\tanh(\beta\pi + h) + \frac{1}{2} = \frac{1}{2}\pi + \frac{1}{2}$. Then

$$\frac{1}{n}\sum_{i=1}^{n} \frac{T_i Y_i}{\widehat{p}} - \frac{1}{n}\sum_{i=1}^{n} \frac{T_i Y_i}{p} = \frac{1}{n}\sum_{i=1}^{n} \frac{T_i Y_i}{\widehat{p}} \frac{p - \widehat{p}}{p}.$$

Taylor expand $x \mapsto \tanh(\beta x + h)$ at $x = \pi$, we have

$$\begin{aligned}
2(\widehat{p} - p) =& m - \tanh(\beta m + h) \\
=& \pi + m - \pi - \tanh(\beta\pi + h) - \beta \operatorname{sech}^2(\beta\pi + h)(m - \pi) + O((m-\pi)^2) \\
=& (1 - \beta \operatorname{sech}^2(\beta\pi + h))(m - \pi) + O((m-\pi)^2),
\end{aligned}$$

where $O(\cdot)$ is up to a universal constant. Together with concentration of $\frac{1}{n}\sum_{i=1}^n T_i Y_i$ towards $p\mathbb{E}[Y_i]$, we have

$$\frac{1}{n}\sum_{i=1}^{n} \frac{T_i Y_i}{\widehat{p}} - \frac{1}{n}\sum_{i=1}^{n} \frac{T_i Y_i}{p} = -\frac{1 - \beta(1-\pi^2)}{1+\pi} \mathbb{E}[Y_i(1, \frac{M_i}{N_i})] + O_{\psi_1}(n^{-2\mathtt{r}_{\beta,h}}).$$

SA-7.3.8 Proof of Lemma SA-7

By Lemma SA-2 to Lemma SA-6, we show

$$n^{\mathtt{r}_{\beta,h}}(\widehat{\tau}_n - \tau_n) \tag{SA-22}$$

$$= n^{-\mathtt{a}_{\beta,h}} \sum_{i=1}^{n} (R_i - \mathbb{E}[R_i] + b_i)(W_i - \pi) + \varepsilon, \tag{SA-23}$$

where $R_i = \frac{g_i(1,\frac{M_i}{N_i})}{1+\pi} + \frac{g_i(-1,\frac{M_i}{N_i})}{1-\pi}$, and $b_i = \sum_{j\neq i}\frac{E_{ij}}{N_j}g_j'(1,\pi)$, and ε is such that condition on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathscr{A} = \{A \in \mathbb{R}^{n\times n} : \min_{i\in[n]}\sum_{j\neq i}A_{ij} \geq 32\log n\}$,

$$\begin{aligned}\varepsilon =& O_{\psi_1,tc}\left(\log n \max_{i\in[n]}\mathbb{E}[N_i|\mathbf{U}]^{-1/2}\right) + O_{\psi_1,tc}(\sqrt{\log n}n^{-\mathtt{r}_{\beta,h}})\\ &+ O_{\psi_1,tc}\left(n^{1/2-\mathtt{a}_{\beta,h}}\left(\frac{\max_i \mathbb{E}[N_i|\mathbf{U}]^3}{\min_i \mathbb{E}[N_i|\mathbf{U}]^4}\right)^{1/2}\right) + O_{\psi_{2/(p+1)},tc}\left(n^{\mathtt{r}_{\beta,h}}(\min_i \mathbb{E}[N_i|\mathbf{U}]^{-(p+1)/2})\right). \quad \text{(SA-24)}\end{aligned}$$

Following the strategy as in the proof of Theorem 4 in [6], we will show b_i is close to R_i: First, decompose by

$$\begin{aligned}&\sum_{j\neq i}\frac{E_{ij}}{N_j}g_j'(1,\pi) - R_i\\ =&\sum_{j\neq i}\frac{E_{ij}}{N_j}g_j'(1,\pi) - \sum_{j\neq i}\frac{E_{ij}}{n\mathbb{E}[G(U_i,U_j)|U_j]}g_j'(1,\pi) + \sum_{j\neq i}\frac{E_{ij}}{n\mathbb{E}[G(U_i,U_j)|U_j]}g_j'(1,\pi) - R_i.\end{aligned}$$

By Equation SA-18, condition on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathscr{A}$,

$$|\sum_{j\neq i}\frac{E_{ij}}{N_j}g_j'(1,\pi) - \sum_{j\neq i}\frac{E_{ij}}{n\mathbb{E}[G(U_i,U_j)|U_j]}g_j'(1,\pi)| \leq Cn^{-1/2}$$

with probability at least $1 - n^{-99}$. Moreover, $\frac{E_{ij}}{\mathbb{E}[G(U_i,U_j)|U_j]}g_j'(1,\pi), j \neq i$ are i.i.d condition on U_i, hence $\sum_{j\neq i}\frac{E_{ij}}{n\mathbb{E}[G(U_i,U_j)|U_j]}g_j'(1,\pi) - R_i = O_{\psi_2}((n\mathbb{E}[G(U_i,U_j)|U_j]^{-1/2}) = O_{\psi_2}(\mathbb{E}[N_j|X]^{-1/2})$. It follows that conditional on $\mathbf{U}$ such that $A(\mathbf{U}) \in \mathscr{A}$,

$$\max_i|\sum_{j\neq i}\frac{E_{ij}}{N_j}g_j'(1,\pi) - R_i| = O_{\psi_2,tc}(\max_i \mathbb{E}[N_i|\mathbf{U}]^{-1/2}). \quad \text{(SA-25)}$$

Again using the conditional i.i.d decomposition, Hoeffiding inequality and U_n's concentration for the two terms respectively,

$$\begin{aligned}&|n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n[\sum_{j\neq i}\frac{E_{ij}}{N_j}g_j'(1,\pi) - R_i](W_i - \pi)|\\ \leq&|n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n[\sum_{j\neq i}\frac{E_{ij}}{N_j}g_j'(1,\pi) - R_i](W_i - \mathbb{E}[W_i|\mathsf{U}_n])|\\ &+ n^{\mathtt{r}_{\beta,h}}|\mathbb{E}[W_i|\mathsf{U}_n] - \pi|\max_i|\sum_{j\neq i}\frac{E_{ij}}{N_j}g_j'(1,\pi) - R_i|\\ =&O_{\psi_2}(n^{\frac{1}{2}-\mathtt{a}_{\beta,h}}\max_i \mathbb{E}[N_i|\mathbf{U}]^{-1/2}) + O_{\psi_{\beta,h},tc}((\log n)^{1/\mathtt{p}_{\beta,h}}\max_i \mathbb{E}[N_i|\mathbf{U}]^{-1/2}) = \varepsilon'.\end{aligned}$$

Hence denote the term of stochastic linearization by G_n, i.e.

$$G_n = n^{-\mathtt{a}_{\beta,h}}\sum_{i=1}^n(R_i - \mathbb{E}[R_i] + Q_i)(W_i - \pi).$$

Since $R_i - \mathbb{E}[R_i] + Q_i$'s are i.i.d independent to W_i's with bounded third moment, we know from Lemma SA-3 that G_n can be approximated by either a Gaussian or non-Gaussian law, that is order 1, this gives

$$
\begin{aligned}
&\sup_{t\in\mathbb{R}} \mathbb{P}\left(\widehat{\tau}_n - \tau_n|\mathbf{U}) \le t\right) - \mathbb{P}\left(G_n \le t|\mathbf{U}\right) \\
\le& \sup_{t\in\mathbb{R}} \min_{u>0} \mathbb{P}\left(G_n \le t+u\right) - \mathbb{P}\left(G_n \le t\right) + \mathbb{P}(\varepsilon+\varepsilon' \ge u) \\
\le& \sup_{t\in\mathbb{R}} \min_{u>0} \mathbb{P}\left(G_n \le t+u\right) - \mathbb{P}\left(G_n \le t+u\right) + \mathbb{P}(\varepsilon+\varepsilon' \ge u) + \mathbb{P}\left(t \le G_n \le t+u\right) \\
\le& O(n^{-1/2}) + \min_{u>0} \exp(-(u/\mathtt{r})^{\mathtt{a}}) + \mathtt{c}u \\
=& O((\log n)^{\mathtt{a}} \mathtt{r}(\mathbf{U})),
\end{aligned}
$$

where $O(\cdot)$ does not depend on the value of $\mathbf{U}$ and

$$
\begin{aligned}
\mathtt{r}(\mathbf{U}) =& n^{-\mathtt{r}_{\beta,h}} + \max_i \mathbb{E}[N_i|\mathbf{U}]^{-1/2} + n^{1/2-\mathtt{a}_{\beta,h}} \left(\frac{\max_i \mathbb{E}[N_i|\mathbf{U}]^3}{\min \mathbb{E}[N_i|\mathbf{U}]^4}\right)^{1/2} \\
&+ n^{\mathtt{r}_{\beta,h}} \max_i \mathbb{E}[N_i|\mathbf{U}]^{-(p+1)/2}.
\end{aligned}
$$

To analyse the second term, recall $\mathbb{E}[N_i|\mathbf{U}] = \rho_n \sum_{j\neq i} G(U_i, U_j)$. Hence

$$
\begin{aligned}
&\mathbb{E}\left[\max_i \left(\mathbb{E}[N_i|\mathbf{U}]\right)^{-1/2} \mathbb{1}(A(\mathbf{U}) \in \mathcal{A})\right] \\
=& (n\rho_n)^{-1/2} \mathbb{E}\left[\max_i \left(\frac{1}{n}\sum_{j\neq i} G(U_i,U_j)\right)^{-1/2} \mathbb{1}(A(\mathbf{U}) \in \mathcal{A})\right] \\
=& O(\sqrt{\log n}(n\rho_n)^{-1/2}),
\end{aligned}
$$

the last line is because with probability at least $1-n^{-98}$, $E = \{\frac{1}{2}g(U_i) \le \frac{1}{n}\sum_{j\neq i} G(U_i,U_j) \le 2g(U_i), \forall 1 \le i \le n\}$ happens, and by maximal inequality, $\max_i |g(U_i)|^{-1/2} = O_{\psi_2}(\sqrt{\log n})$. And on $\{A(\mathbf{U}) \in \mathcal{A}\} \cap E$, $\max_i (\frac{1}{n}\sum_{j\neq i} G(U_i,U_j))^{-1/2} \le (32\log n/n)^{-1/2}$, since we assume G is positive. By similar argument for the last two terms in $\mathtt{r}(\mathbf{U})$, we have

$$
\mathbb{E}\left[r(\mathbf{U})\mathbb{1}(A(\mathbf{U}) \in \mathcal{A})\right] \le n^{-\mathtt{r}_{\beta,h}} + \sqrt{\log n}(n\rho_n)^{-1/2} + \sqrt{\log n} n^{\mathtt{r}_{\beta,h}} (n\rho_n)^{-(p+1)/2}.
$$

Recall that $\mathcal{A} = \{A(\mathbf{U}) : \min_i \sum_{j\neq i} A_{ij}(\mathbf{U}) \ge 32\log n\}$. Since $\sum_{j\neq i} A_{ij}(\mathbf{U}) \sim \mathrm{Bin}(n-1, \mathbb{E}[G(X_1,X_2)])$, we know from Chernoff bound for Binomials and union bound over i that $\mathbb{P}(A(\mathbf{U}) \notin \mathcal{A}) \le n^{-99}$. The conclusion then follows.

SA-7.3.9 Proof of Lemma SA-8

Our proof for Lemma SA-2 to Lemma SA-6 relies on the following devices:

(1) Taylor expansion of $\tanh(\cdot)$ in the inverse probability weighting for unbiased estimator, and taylor expansion of $Y_i(\ell,\cdot)$ at $\mathbb{E}[T_i]$ for $\ell \in \{0,1\}$. Then the higher order terms are in terms of $m-\pi$ and $\frac{M_i}{N_i} - \pi$. In Lemma SA-4 (taking $X_i \equiv 1$), we show

$$
\|m\|_{\psi_1} \le \mathtt{K} n^{-1/4},
$$

and in Lemma SA-2, we show

$$\|\frac{M_i}{N_i}\|_{\psi_1} \leq \mathtt{K}n^{-1/4} + \mathtt{K}(n\rho_n)^{-1/2},$$

where $\mathtt{K}$ is some constant that does not depend on β. This shows for the higher order terms, we always have

$$m^2 = m(1 + o_{\mathbb{P}}(1)), \qquad (M_i/N_i)^2 = (M_i/N_i)(1 + o_{\mathbb{P}}(1)),$$

where the $o_{\mathbb{P}}(\cdot)$ terms does not depend on β.

(2) Condition i.i.d decomposition based on the de-Finetti's lemma (Lemma SA-1). Suppose U_n is the latent variable from Lemma SA-1, we use decompositions based on U_n: For Lemma SA-3 to Lemma SA-5, we break down higher order terms in the form

$$\begin{aligned}&F(\mathbf{W},\mathbf{E}) - \mathbb{E}[F(\mathbf{W},\mathbf{E})|\mathbf{E}]\\=&F(\mathbf{W},\mathbf{E}) - \mathbb{E}[F(\mathbf{W},\mathbf{E})|\mathbf{E},\mathsf{U}_n] + \mathbb{E}[F(\mathbf{W},\mathbf{E})|\mathbf{E},\mathsf{U}_n] - \mathbb{E}[F(\mathbf{W},\mathbf{E})|\mathbf{E}].\end{aligned}$$

For the first part $F(\mathbf{W},\mathbf{E}) - \mathbb{E}[F(\mathbf{W},\mathbf{E})|\mathbf{E},\mathsf{U}_n]$, we use the conditional i.i.d of W_i's given U_n. For the second part, we use concentration from Lemma SA-2 that there exists a constant $\mathtt{K}$ not depending on β or n, such that $\|\mathsf{U}_n\|_{\psi_1} \leq \mathtt{K}n^{1/4}$ and the effective term $\|\tanh(\sqrt{\frac{\beta}{n}}\mathsf{U}_n)\|_{\psi_1} \leq \mathtt{K}n^{-1/4}$.

In particular, the rate of concentration for conditional i.i.d Berry-Esseen and concentration of $\tanh(\sqrt{\frac{\beta}{n}}\mathsf{U}_n)$ does not depend on β.

By the same proof from Lemma SA-2 to Lemma SA-6, we can show in $\widehat{\tau}_n - \tau_n$, the second and higher order terms in terms of $W_i - \pi$ can always be dominated by the first order terms, with a rate that does not depend on β.

The conclusion then follows from the two devices and the same proof logic of Lemma SA-2 to Lemma SA-6.

SA-7.4 Proof for Section SA-5

SA-7.4.1 Proof of Lemma SA-1

Define $g(U_j) = \mathbb{E}[G(U_i, U_j)|U_j]$, for $i \neq j$. Reordering the terms,

$$\overline{\tau}^a = \frac{n-1}{n^2}\sum_{j\in[n]}\frac{T_j}{1/2}h_j(1, M_j/N_j) - \frac{1-T_j}{1-1/2}h_j(-1, M_j/N_j).$$

Hence $\tau^a_{(i)} - \overline{\tau}^a$ has the representation given by

$$
\begin{aligned}
&\tau^a_{(i)} - \overline{\tau}^a \\
=& -\frac{1}{n}\frac{T_i}{1/2}h_i\Big(1, \frac{M_i}{N_i}\Big) + \frac{1}{n^2}\sum_{j\in[n]}\frac{T_j}{1/2}h_j\Big(1, \frac{M_j}{N_j}\Big) + \frac{1}{n}\frac{1-T_i}{1-1/2}h_i\Big(1, \frac{M_i}{N_i}\Big) \\
&\quad - \frac{1}{n^2}\sum_{j\in[n]}\frac{1-T_j}{1-1/2}h_j\Big(1, \frac{M_j}{N_j}\Big) &\text{(SA-26)}\\
=& -\frac{1}{n}\Big(\frac{T_i}{1/2}h_i(1,0) - 1/2\mathbb{E}[h_i(1,0)]\Big) + \frac{1}{n}\Big(\frac{1-T_i}{1-1/2}h_i(-1,0) - (1-1/2)\mathbb{E}[h_i(-1,0)]\Big) \\
&\quad + O_{\psi_{2,tc}}(n^{-1}(n\rho_n)^{-\frac{1}{2}}) &\text{(SA-27)}\\
=& -\frac{1}{n}\Big(\frac{h_i(1,0)}{1/2} + \frac{h_i(-1,0)}{1-1/2}\Big)(T_i - 1/2) + O_{\psi_{2,tc}}(n^{-1}(n\rho_n)^{-\frac{1}{2}}) &\text{(SA-28)}\\
=& -\frac{1}{n}\Big(\frac{f_i(1,0)}{1/2} + \frac{f_i(-1,0)}{1-1/2}\Big)(T_i - 1/2) + O_{\psi_{2,tc}}(n^{-1}(n\rho_n)^{-\frac{1}{2}}) + o_{\mathbb{P}}(n^{-1}), &\text{(SA-29)}
\end{aligned}
$$

where the second to last line is due to $-\frac{1}{n}\frac{1}{1/2}1/2(h_i(1,0)-\mathbb{E}[h_i(1,0)]) + \frac{1}{n}\frac{1}{1-1/2}(1-1/2)(h_i(-1,0) - \mathbb{E}[h_i(-1,0)]) = -\frac{2}{n}\varepsilon_i + \frac{2}{n}\varepsilon_i = 0$.

Now we look at b-part. For representation purpose, we look at only the treatment part. The control part can be analyzed by in the same way. Reordering the terms,

$$
\begin{aligned}
\overline{\tau}^b = \frac{1}{n}\sum_{i\in[n]}\tau^b_{(i)} &= \frac{1}{n}\sum_{i\in[n]}\frac{1}{n}\sum_{j\in[n]}\frac{T_j}{1/2}\Big[h_j\Big(1, \frac{M_j}{N_j}_{(i)}\Big) - h_j\Big(1, \frac{M_j}{N_j}\Big)\Big] \\
&= \frac{1}{n}\sum_{j\in[n]}\frac{T_j}{1/2}\frac{1}{n}\sum_{i\in[n]}\Big[h_j\Big(1, \frac{M_j}{N_j}_{(i)}\Big) - h_j\Big(1, \frac{M_j}{N_j}\Big)\Big].
\end{aligned}
$$

Hence $\tau^b_{(i)} - \overline{\tau}^b$ has the representation given by

$$
\tau^b_{(i)} - \overline{\tau}^b = \frac{1}{n}\sum_{j\in[n]}\frac{T_j}{1/2}\Big[h_j\Big(1, \frac{M_j}{N_j}_{(i)}\Big) - \frac{1}{n}\sum_{\iota\in[n]}h_j\Big(1, \frac{M_j}{N_j}_{(\iota)}\Big)\Big]. \tag{SA-30}
$$

The analysis follows from a Taylor expansion of $h_j(1,\cdot)$. For some $\xi^*_{j,i}$ between $\frac{M_j}{N_j}_{(i)}$ and 0 for each j, i,

$$
\begin{aligned}
h_j\Big(1, \frac{M_j}{N_j}_{(i)}\Big) =& h_j(1,0) + \partial_2 h(1,0)\Big(\frac{M_j}{N_j}_{(i)} - 0\Big) + \frac{1}{2}\partial_{2,2}h(1,0)\Big(\frac{M_j}{N_j}_{(i)} - 0\Big)^2 &\text{(SA-31)}\\
&+ \frac{1}{6}\partial_{2,2,2}h(1,\xi^*_{j,i})\Big(\frac{M_j}{N_j}_{(i)} - 0\Big)^3, &\text{(SA-32)}
\end{aligned}
$$

where we have used $\partial_2 h_j(1,\cdot) = \partial_2[h(1,\cdot) + \varepsilon_j] = \partial_2 h(1,\cdot)$.

Part 1: Linear Terms

$$
\begin{aligned}
\frac{M_j}{N_j}_{(i)} - \frac{1}{n}\sum_{\iota\in[n]}\frac{M_j}{N_j}_{(\iota)} &= \sum_{l\neq i}\frac{E_{lj}}{N_j^{(i)}}W_l - \frac{1}{n}\sum_{\iota\in[n]}\sum_{l\neq\iota}\frac{E_{lj}}{N_j^{(\iota)}}W_l \\
&= \sum_{l=1}^{n}E_{lj}W_l\Big(\frac{1}{N_j^{(i)}} - \frac{1}{n}\sum_{\iota\in[n],\iota\neq l}\frac{1}{N_j^{(\iota)}}\Big) - \frac{E_{ij}}{N_j^{(i)}}W_i.
\end{aligned}
\tag{SA-33}
$$

By a decomposition argument,

$$\begin{aligned}\frac{1}{N_j^{(i)}} - \frac{1}{n}\sum_{\iota\in[n],\iota\neq l}\frac{1}{N_j^{(\iota)}} &= \frac{1}{N_j^{(i)}} - \frac{1}{n-1}\sum_{\iota\in[n],\iota\neq l}\frac{1}{N_j^{(\iota)}} + \frac{1}{(n-1)n}\sum_{\iota\in[n],\iota\neq l}\frac{1}{N_j^{(\iota)}}\\ &= \frac{1}{n-1}\sum_{\iota\in[n],\iota\neq l}\frac{E_{ji}-E_{j\iota}}{N_j^{(i)}N_j^{(\iota)}} + \frac{1}{(n-1)n}\sum_{\iota\in[n],\iota\neq l}\frac{1}{N_j^{(\iota)}}\\ &= n^{-1}(n\rho_n)^{-1}\frac{E_{ij}-\rho_n g(U_j)}{\rho_n g(U_j)^2} + \frac{1}{(n-1)n}\sum_{\iota\in[n],\iota\neq l}\frac{1}{N_j^{(\iota)}}.\end{aligned}$$

Hence

$$\begin{aligned}&\sum_{l=1}^{n}E_{lj}W_l\Big(\frac{1}{N_j^{(i)}} - \frac{1}{n}\sum_{\iota\in[n],\iota\neq l}\frac{1}{N_j^{(\iota)}}\Big)\\ =&(n\rho_n)^{-1}\frac{E_{ij}-\rho_n g(U_j)}{\rho_n g(U_j)^2}\frac{1}{n}\sum_{l=1}^{n}E_{lj}W_l + \frac{\sum_{l=1}^{n}E_{lj}W_l}{N_j^{(i)}}O_{\psi_{2,tc}}((n\rho_n)^{-\frac{3}{2}})\\ &+ \frac{1}{n-1}\sum_{\iota\in[n],\iota\neq l}\frac{\sum_{l=1}^{n}E_{lj}W_l}{nN_j^{(\iota)}}.\end{aligned}$$

Condition on U_j, $(E_{lj}W_l : l \neq j)$ are i.i.d mean-zero, hence Bernstein inequality gives $\frac{1}{n}\sum_{l=1}^{n}E_{lj}W_l = O_{\psi_2}(\sqrt{n^{-1}\rho_n}) + O_{\psi_1}(n^{-1})$, which implies

$$\begin{aligned}(n\rho_n)^{-1}\frac{E_{ij}-\rho_n g(U_j)}{\rho_n g(U_j)^2}\frac{1}{n}\sum_{l=1}^{n}E_{lj}W_l &= O_{\psi_2}((n\rho_n)^{-\frac{3}{2}}) + O_{\psi_1}((n\rho_n)^{-2}),\\ \frac{1}{n-1}\sum_{\iota\in[n],\iota\neq l}\frac{\sum_{l=1}^{n}E_{lj}W_l}{nN_j^{(\iota)}} &= O_{\psi_2}(n^{-\frac{3}{2}}\rho_n^{-\frac{1}{2}}) + O_{\psi_1}(n^{-2}).\end{aligned}$$

Putting back into Equation (SA-33),

$$\frac{M_j}{N_j}_{(i)} - \frac{1}{n}\sum_{\iota\in[n]}\frac{M_j}{N_j}_{(\iota)} = -\frac{E_{ij}}{N_j^{(i)}}W_i + O_{\psi_1}((n\rho_n)^{-\frac{3}{2}}).$$

Looking at contribution from the first order term in Taylor expanding $h_j(1,\cdot)$ to $\tau_{(i)}^b - \overline{\tau}^b$ in Equation (SA-30),

$$\begin{aligned}&\frac{1}{n}\sum_{j\in[n]}\partial_2 h(1,0)\frac{T_j}{1/2}\Big[\frac{M_j}{N_j}_{(i)} - \frac{1}{n}\sum_{\iota\in[n]}\frac{M_j}{N_j}_{(\iota)}\Big]\\ =&-\sum_{j\in[n]}\partial_2 h(1,0)W_i\frac{1}{n}\frac{E_{ij}}{N_j^{(i)}}\frac{T_j}{1/2} + O_{\psi_{1,tc}}((n\rho_n)^{-\frac{3}{2}})\\ =&-W_i\frac{1}{n}\sum_{j\in[n]}\partial_2 h(1,0)\frac{E_{ij}}{n\rho_n g(U_j)}\frac{T_j}{1/2} - W_i\frac{1}{n}\sum_{j\in[n]}\partial_2 h(1,0)\frac{E_{ij}}{N_j^{(i)}}\frac{n\rho_n g(U_j)-N_j}{n\rho_n g(U_j)}\frac{T_j}{1/2}\\ &+ O_{\psi_{1,tc}}((n\rho_n)^{-\frac{3}{2}})\\ =&-W_i\frac{1}{n}\sum_{j\in[n]}\partial_2 h(1,0)\frac{E_{ij}}{n\rho_n g(U_j)}\frac{T_j}{1/2} + O_{\psi_{1,tc}}((n\rho_n)^{-\frac{3}{2}}).\end{aligned}$$

Since $(E_{ij}T_j/g(U_j) : j \in [n])$ are independent condition on U_i, standard concentration inequality gives

$$\begin{aligned}
&\frac{1}{n}\sum_{j\in[n]}\partial_2 h(1,0)\frac{T_j}{1/2}\left[\frac{M_j}{N_j}_{(i)} - \frac{1}{n}\sum_{\iota\in[n]}\frac{M_j}{N_j}_{(\iota)}\right] \\
=& -W_i\frac{1}{n}\sum_{j\in[n]}\partial_2 h(1,0)\frac{E_{ij}}{n\rho_n g(U_j)}\frac{T_j}{1/2} + O_{\psi_{1,tc}}((n\rho_n)^{-\frac{3}{2}}) \\
=& -W_i\partial_2 h(1,0)\frac{1}{n}\sum_{j\in[n]}\frac{E_{ij}}{n\rho_n g(U_j)}\frac{T_j}{1/2} + O_{\psi_{1,tc}}((n\rho_n)^{-\frac{3}{2}}) \\
=& -\partial_2 h(1,0)\frac{W_i}{n}\mathbb{E}\left[\frac{E_{ij}}{\rho_n g(U_j)}\middle|U_i\right] + O_{\psi_{1,tc}}((n\rho_n)^{-\frac{3}{2}}).
\end{aligned}$$

Since we assumed $\partial_2 h(1,0) = \partial_2 f(1,0) + o_{\mathbb{P}}(1) = \partial_2 f_j(1,0) + o_{\mathbb{P}}(1)$ where

$$\begin{aligned}
&\frac{1}{n}\sum_{j\in[n]}\partial_2 h(1,0)\frac{T_j}{1/2}\left[\frac{M_j}{N_j}_{(i)} - \frac{1}{n}\sum_{\iota\in[n]}\frac{M_j}{N_j}_{(\iota)}\right] \\
=& -\frac{W_i}{n}\mathbb{E}\left[\frac{E_{ij}\partial_2 f_j(1,0)}{\rho_n g(U_j)}\middle|U_i\right] + O_{\psi_{1,tc}}((n\rho_n)^{-\frac{3}{2}}) + o_{\mathbb{P}}(n^{-1}).
\end{aligned}$$

Together with the leading term in Equation (SA-30), we have

$$\begin{aligned}
&n\sum_{i\in[n]}\left(\frac{1}{n}\sum_{j\in[n]}\partial_2 h_j(1,0)\frac{T_j}{1/2}\left[\frac{M_j}{N_j}_{(i)} - \frac{1}{n}\sum_{\iota\in[n]}\frac{M_j}{N_j}_{(\iota)}\right] + \tau^a_{(i)} - \overline{\tau}^a\right)\cdot \\
&\qquad\left(\frac{2}{n_q}\sum_{j\in\mathcal{I}_q}\partial_2 h_j(1,0)\frac{T_j}{\theta_q}\left[\frac{M_j}{N_j}_{(i)} - \frac{1}{n}\sum_{\iota\in[n]}\frac{M_j}{N_j}_{(\iota)}\right] + \tau^a_{(i)} - \overline{\tau}^a\right) \\
=&\frac{n}{n^2}\sum_{i\in[n]}\left(\mathbb{E}\left[\frac{E_{ij}\partial_2 f_j(1,0)}{\rho_n g(U_j)}\middle|U_i\right] + \frac{f_i(1,0)}{1/2}(T_i - 1/2)\right)\cdot \\
&\qquad\left(\mathbb{E}\left[\frac{E_{ij}\partial_2 f_j(1,0)}{\rho_n g(U_j)}\middle|U_i\right] + \frac{f_i(1,0)}{1/2}(T_i - 1/2)\right) + O_{\psi_{1,tc}}((n\rho_n^3)^{-1}) + o_{\mathbb{P}}(1) \\
=&\frac{n_l^2}{n^2}\mathbb{E}\left[\left(\mathbb{E}\left[\frac{E_{ij}\partial_2 f_j(1,0)}{\rho_n g(U_j)}\middle|U_i\right] + \frac{f_i(1,0)}{1/2}(T_i - 1/2)\right)\cdot\right. \\
&\qquad\left.\left(\mathbb{E}\left[\frac{E_{ij}\partial_2 f_j(1,0)}{\rho_n g(U_j)}\middle|U_i\right] + \frac{f_i(1,0)}{1/2}(T_i - 1/2)\right)\right] + O_{\psi_{1,tc}}((n\rho_n^3)^{-1}) + o_{\mathbb{P}}(1) \\
=&\mathbf{e}_s^\top\mathbb{E}[\mathbf{S}_\ell\mathbf{S}_\ell^\top]\mathbf{e}_q + O_{\psi_{1,tc}}((n\rho_n^3)^{-1}) + o_{\mathbb{P}}(1).
\end{aligned}$$

Part 2: Higher Order Terms For the second order terms, first notice that if $l \notin [n]$, then

$$\begin{aligned}
&\left(\frac{M_j}{N_j}_{(i)}\right)^2 - \frac{1}{n}\sum_{\iota\in[n],\iota\neq l}\left(\frac{M_j}{N_j}_{(\iota)}\right)^2 \\
=&\frac{1}{n}\sum_{\iota\in[n],\iota\neq l}\left(\frac{M_j}{N_j}_{(i)} + \frac{M_j}{N_j}_{(\iota)}\right)\frac{M_j(E_{ij} - E_{\iota j}) - (E_{ij}W_i - E_{\iota j}W_\iota)N_j + E_{ij}E_{\iota j}(W_i - W_\iota)}{N_j^{(i)}N_j^{(\iota)}} \\
=&O_{\psi_{2,tc}}((n\rho_n)^{-\frac{3}{2}}),
\end{aligned}$$

where we have used $(M_j/N_j)_\iota = O_{\psi_2}((n\rho_n)^{-\frac{1}{2}})$ and $N_j^{-1} = O_{\psi_2}((n\rho_n)^{-1})$. If $l \in [n]$, then again

$$\begin{aligned}
&\Big(\frac{M_j}{N_j}_{(i)}\Big)^2 - \frac{1}{n}\sum_{\iota\in[n],\iota\neq l}\Big(\frac{M_j}{N_j}_{(\iota)}\Big)^2 \\
=&\Big(\frac{M_j}{N_j}_{(i)}\Big)^2 - \frac{1}{n-1}\sum_{\iota\in[n],\iota\neq l}\Big(\frac{M_j}{N_j}_{(\iota)}\Big)^2 + \frac{1}{(n-1)n}\sum_{\iota\in[n],\iota\neq l}\Big(\frac{M_j}{N_j}_{(\iota)}\Big)^2 \\
=&O_{\psi_{2,tc}}((n\rho_n)^{-\frac{3}{2}}).
\end{aligned}$$

Hence

$$\begin{aligned}
&n\sum_{i\in[n]}\Big(\partial_{2,2}h(1,0)\frac{2}{n}\sum_{j\in[n]}T_j\Big[\Big(\frac{M_j}{N_j}_{(i)}\Big)^2 - \frac{1}{n}\sum_{\iota\in[n]}\Big(\frac{M_j}{N_j}_{(\iota)}\Big)^2\Big]\Big)\cdot \\
&\qquad\Big(\partial_{2,2}h(1,0)\frac{2}{n_q}\sum_{j\in\mathcal{I}_q}T_j\Big[\Big(\frac{M_j}{N_j}_{(i)}\Big)^2 - \frac{1}{n}\sum_{\iota\in[n]}\Big(\frac{M_j}{N_j}_{(\iota)}\Big)^2\Big]\Big) = O_{\psi_{2,tc}}((n\rho_n^3)^{-1}).
\end{aligned}$$

For the third order residual, observe that $(\frac{M_j}{N_j}_{(\iota)})^3 = O_{\psi_2}((n\rho_n)^{-3/2})$. Then

$$\begin{aligned}
&n\sum_{i\in[n]}\Big(\frac{2}{n}\sum_{j\in[n]}T_j\Big[\partial_{2,2,2}h\Big(1,\xi^*_{j,i}\Big)\Big(\frac{M_j}{N_j}_{(i)}\Big)^3 - \frac{1}{n}\sum_{\iota\in[n]}\partial_{2,2,2}h\Big(1,\xi^*_{j,\iota}\Big)\Big(\frac{M_j}{N_j}_{(\iota)}\Big)^3\Big]\Big)\cdot \\
&\quad\Big(\frac{2}{n_q}\sum_{j\in\mathcal{I}_q}T_j\Big[\partial_{2,2,2}h\Big(1,\xi^*_{j,i}\Big)\Big(\frac{M_j}{N_j}_{(i)}\Big)^3 - \frac{1}{n}\sum_{\iota\in[n]}\partial_{2,2,2}h\Big(1,\xi^*_{j,\iota}\Big)\Big(\frac{M_j}{N_j}_{(\iota)}\Big)^3\Big]\Big) \\
&\quad= O_{\psi_{2,tc}}((n\rho_n^3)^{-1}).
\end{aligned}$$

The conclusion then follows from Equations (SA-26), (SA-30) and (SA-31).

SA-7.5 Proof of Lemma SA-2

Define $\mathtt{r}(x) = (1,x)^\top$. Denote $\pi = \mathbb{E}[W_i] = 2\mathbb{E}[T_i] - 1$. Then

Case 1: $\beta < 1$

First, consider the gram-matrix. Take $\zeta_i := \sqrt{n\rho_n}(\frac{M_i}{N_i} - \pi)$. Then $1 \lesssim \mathbb{V}[\zeta_i] \lesssim 1$. Take $b_n = \sqrt{n\rho_n}h_n$. Take

$$\mathbf{B}_n := \frac{1}{nb_n}\sum_{i=1}^n \mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)\mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)^\top K\Big(\frac{\zeta_i}{b_n}\Big),$$

where $\mathtt{r}: \mathbb{R} \to \mathbb{R}^2$ is given by $\mathtt{r}(u) = (1,u)^\top$. Take Q to be the probability measure of ζ_i given $\mathbf{E}$. Then

$$\mathbf{B} := \mathbb{E}[\mathbf{B}_n|\mathbf{E}] = \begin{bmatrix} \int_{-\infty}^{\infty}\frac{1}{b_n}K(\frac{x}{b_n})dQ(x) & \int_{-\infty}^{\infty}\frac{x}{b_n}\frac{1}{b_n}K(\frac{x}{b_n})dQ(x) \\ \int_{-\infty}^{\infty}\frac{x}{b_n}\frac{1}{b_n}K(\frac{x}{b_n})dQ(x) & \int_{-\infty}^{\infty}(\frac{x}{b_n})^2\frac{1}{b_n}K(\frac{x}{b_n})dQ(x)\end{bmatrix}.$$

In particular, $\lambda_{\min}(\mathbf{B}) \gtrsim 1$. Now we want to show each entry of $\mathbf{B}_n$ converge to those of $\mathbf{B}$. Take

$$F_{p,q}(\mathbf{W}) := \mathbf{e}_p^\top\mathbf{B}_n\mathbf{e}_q = \frac{1}{nb_n}\sum_{i=1}^n\Big(\frac{\zeta_i}{b_n}\Big)^{p+q}K\Big(\frac{\zeta_i}{b_n}\Big), \qquad p,q\in\{0,1\}.$$

Denote ∂_j to be the partial derivative w.r.p to W_j. Since K is Lipschitz with bounded support,

$$|\partial_j F_{p,q}(\mathbf{W})| \lesssim \frac{1}{b_n^2}\frac{1}{n}\sum_{i=1}^{n}\Big|\partial_j\Big(\frac{M_i}{N_i}-\pi\Big)\Big| \lesssim \frac{1}{b_n^2}\frac{1}{n}\sum_{i=1}^{n}\frac{E_{ij}}{N_i}. \tag{SA-34}$$

Condition on $\mathbf{E}$,

$$F_{p,q}(\mathbf{W}) = \mathbb{E}[F_{p,q}(\mathbf{W})|\mathbf{E}] + O_{\psi_2}\Big(\sum_{j=1}^{n}|\partial_j F_{p,q}(\mathbf{W})|^2\Big) = \mathbf{e}_p^\top \mathbf{B}\mathbf{e}_q + O_{\psi_2}\Big(\frac{1}{nb_n^4}\frac{1}{n}\sum_{j=1}^{n}\Big(\sum_{i=1}^{n}\frac{E_{ij}}{N_i}\Big)^2\Big).$$

Hence for all $p, q \in \{0, 1\}$,

$$\mathbf{e}_p^\top \mathbf{B}_n \mathbf{e}_q = \mathbf{e}_p^\top \mathbf{B}\mathbf{e}_q + O_{\psi_2}((nb_n^4)^{-1}).$$

Since both $\mathbf{B}_n$ and $\mathbf{B}$ are two by two matrices, $\|\mathbf{B}_n - \mathbf{B}\|_{\mathrm{op}} \lesssim O_{\psi_2}((nb_n^4)^{-1})$. By Weyl's Theorem,

$$|\lambda_{\min}(\mathbf{B}_n) - \lambda_{\min}(\mathbf{B})| \leq \|\mathbf{B}_n - \mathbf{B}\|_{\mathrm{op}} \lesssim (nb_n^4)^{-1}, \tag{SA-35}$$

and together with $\lambda_{\min}(\mathbf{B}) \gtrsim 1$, implies $\lambda_{\min}(\mathbf{B}_n) \gtrsim 1$. Take

$$\mathbf{\Sigma}_n := \frac{1}{nb_n^2}\sum_{i=1}^{n}\mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)\mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)^\top K^2\Big(\frac{\zeta_i}{b_n}\Big)\mathbb{V}[Y_i|\zeta_i].$$

Hence variance can be bounded by

$$\mathbb{V}[\widehat{\gamma}_0|\mathbf{E},\mathbf{W}] = \mathbf{e}_0^{\mathbf{T}}\mathbf{B}_n^{-1}\mathbf{\Sigma}_n\mathbf{B}_n^{-1}\mathbf{e}_0 \lesssim (nb_n)^{-1}, \tag{SA-36}$$

$$\mathbb{V}[\widehat{\gamma}_1|\mathbf{E},\mathbf{W}] = n\rho_n\mathbf{e}_1^{\mathbf{T}}\mathbf{B}_n^{-1}\mathbf{\Sigma}_n\mathbf{B}_n^{-1}\mathbf{e}_1 \lesssim (n\rho_n)(nb_n^3)^{-1} = \rho_n b_n^{-3}. \tag{SA-37}$$

Next, consider the bias term. Since $f(1,\cdot) \in C^2$, whenever $|\frac{M_i}{N_i} - \pi| \leq h_n = (n\rho_n)^{-1/2}b_n$,

$$\begin{aligned} f(1, M_i/N_i) &= f(1,\pi) + \partial_2 f(1,\pi)\Big(\frac{M_i}{N_i}-\pi\Big) + O\Big(\Big(\frac{M_i}{N_i}-\pi\Big)^2\Big) \\ &= f(1,\pi) + \partial_2 f(1,\pi)\Big(\frac{M_i}{N_i}-\pi\Big) + O((n\rho_n)^{-1}b_n^2). \end{aligned}$$

Hence using the fourth and third lines above respectively,

$$\begin{aligned} \mathbb{E}[\widehat{\gamma}_0|\mathbf{E},\mathbf{W}] &= \mathbf{e}_0^{\mathbf{T}}\mathbf{B}_n^{-1}\Big[\frac{1}{nb_n}\sum_{i=1}^{n}\mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)K\Big(\frac{\zeta_i}{b_n}\Big)f\Big(1,\frac{M_i}{N_i}\Big)\Big] \\ &= \mathbf{e}_0^{\mathbf{T}}\mathbf{B}_n^{-1}\Big[\frac{1}{nb_n}\sum_{i=1}^{n}\mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)K\Big(\frac{\zeta_i}{b_n}\Big)\Big(\mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)^\top(f(1,\pi),\frac{1}{\sqrt{n\rho_n}}\partial_2 f(1,\pi))^\top + O_{\psi_2}((n\rho_n)^{-\frac{1}{2}})\Big)\Big] \\ &= f(1,\pi) + O_{\psi_2}((n\rho_n)^{-\frac{1}{2}}), \\ \mathbb{E}[\widehat{\gamma}_1|\mathbf{E},\mathbf{W}] &= \sqrt{n\rho_n}\mathbf{e}_1^{\mathbf{T}}\mathbf{B}_n^{-1}\Big[\frac{1}{nb_n}\sum_{i=1}^{n}\mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)K\Big(\frac{\zeta_i}{b_n}\Big)f\Big(1,\frac{M_i}{N_i}\Big)\Big] \\ &= \sqrt{n\rho_n}\mathbf{e}_1^{\mathbf{T}}\mathbf{B}_n^{-1}\Big[\frac{1}{nb_n}\sum_{i=1}^{n}\mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)K\Big(\frac{\zeta_i}{b_n}\Big)\Big(\mathtt{r}\Big(\frac{\zeta_i}{b_n}\Big)^\top(f(1,\pi),\frac{1}{\sqrt{n\rho_n}}\partial_2 f(1,\pi))^\top + O_{\psi_2}((n\rho_n)^{-1})\Big)\Big] \\ &= \partial_2 f(1,\pi) + O_{\psi_2}((n\rho_n)^{-\frac{1}{2}}), \end{aligned} \tag{SA-38}$$

Putting together Equations (SA-36) and (SA-38),

$$\widehat{\gamma}_0 - \gamma_0 = O_{\mathbb{P}}((n\rho_n)^{-\frac{1}{2}} + (nb_n)^{-\frac{1}{2}}), \quad \widehat{\gamma}_1 - \gamma_1 = O_{\mathbb{P}}((n\rho_n)^{-\frac{1}{2}} + \rho_n b_n^{-3}).$$

Hence any b_n such that $b_n = \Omega(n^{-1/4} + \rho_n^{1/3})$ will make $(\widehat{\gamma}_0, \widehat{\gamma}_1)$ a consistent estimator for (γ_0, γ_1). For any $0 \leq \rho_n \leq 1$ such that $n\rho_n \to \infty$, such a sequence b_n exists.

Case 2: $\beta = 1$

The order $\frac{M_i}{N_i}$ is $n^{-1/4}$ if $\liminf_{n\to\infty} n\rho_n^2 > c$ for some $c > 0$; and is $(n\rho_n)^{-1/2}$ if $n\rho_n^2 = o(1)$. We consider these two cases separately.

Case 2.1: $\liminf_{n\to\infty} n\rho_n^2 > c$ **for some** $c > 0$ Take $\eta_i = n^{\frac{1}{4}}(\frac{M_i}{N_i} - \pi)$. Take $d_n = n^{1/4}h_n$. And with the same $\mathtt{r}$ defined in Case 1,

$$\mathbf{D}_n := \frac{1}{nd_n}\sum_{i=1}^n \mathtt{r}\Big(\frac{\eta_i}{d_n}\Big)\mathtt{r}\Big(\frac{\eta_i}{d_n}\Big)^\top K\Big(\frac{\eta_i}{d_n}\Big), \qquad \mathbf{D} = \mathbb{E}[\mathbf{D}_n].$$

Under the assumption $\liminf_{n\to\infty} n\rho_n^2 \leq c$ for some $c > 0$, we have $1 \lesssim \mathbb{V}[\eta_i] \lesssim 1$. Hence $\lambda_{\min}(\mathbf{D}) \gtrsim 1$. To study the convergence between $\mathbf{D}_n$ and $\mathbf{D}$, again consider for $p, q \in \{0, 1\}$,

$$G_{p,q}(\mathbf{W}) := \mathbf{e}_p^\top \mathbf{D}_n \mathbf{e}_q = \frac{1}{nd_n}\sum_{i=1}^n \Big(\frac{\eta_i}{d_n}\Big)^{p+q} K\Big(\frac{\eta_i}{d_n}\Big) = \frac{1}{n^{5/4}h_n}\sum_{i=1}^n \Big(h_n^{-1}(\frac{M_i}{N_i} - \pi)\Big)^{p+q} K\Big(h_n^{-1}(\frac{M_i}{N_i} - \pi)\Big).$$

Still let U_n be the latent variable from Lemma SA-1, W_i's are independent conditional on U_n. Hence by similar argument as Equation (SA-34), we can show

$$G_{p,q}(\mathbf{W}) = \mathbb{E}[G_{p,q}(\mathbf{W})|\mathsf{U}_n, \mathbf{E}] + O_{\psi_2}((nd_n^4)^{-1}).$$

Moreover, recall we denote by $\omega_i \in [k]$ the block unit i belongs to, then

$$\mathbb{E}[G_{p,q}(\mathbf{W})|\mathsf{U}_n, \mathbf{E}] = \sum_{\mathbf{W}\in\{-1,1\}^n}\prod_{i=1}^n p(U_{\omega_i})^{W_s}(1 - p(U_{\omega_i}))^{1-W_s} G_{p,q}(\mathbf{W}),$$

$p(U_l) = \mathbb{P}(W_i = 1|U_\ell) = \frac{1}{2}(\tanh(\sqrt{\beta_\ell/n}U_n + h_\ell) + 1)$, $i \in \mathcal{I}_\ell$. Take the derivative term by term,

$$\partial_{U_\ell}\mathbb{E}[G_{p,q}(\mathbf{W})|\mathsf{U}_n, \mathbf{E}] = \sum_{j\in\mathcal{I}_\ell}\mathbb{E}_{\mathbf{W}_{-j}}[G_{p,q}(W_j = 1, W_{-j}) - G_{p,q}(W_j = -1, W_{-j})]p'(U_\ell).$$

Using Lipschitz property of $x \mapsto (x/h_n)^{p+q}K(x/h_n)$,

$$|G_{p,q}(W_j = 1, W_{-j}) - G_{p,q}(W_j = -1, W_{-j})| \lesssim \frac{1}{n^{5/4}h_n}\sum_{i=1}^n \frac{1}{h_n}\frac{E_{ij}}{N_i}.$$

Hence for all $\ell \in \mathcal{C}$,

$$|\partial_{U_\ell}\mathbb{E}[G_{p,q}(\mathbf{W})|\mathsf{U}_n, \mathbf{E}]| \lesssim \sum_{j\in\mathcal{I}_\ell}\frac{1}{n^{5/4}h_n}\sum_{i=1}^n \frac{1}{h_n}\frac{E_{ij}}{N_i}\|p'\|_\infty \lesssim \frac{1}{n^{3/4}h_n^2}.$$

Moreover, for all $\ell \in \mathcal{C}$, $\|U_\ell\|_{\varphi_2} \lesssim n^{1/4}$. Together, this gives

$$\mathbb{E}[G_{p,q}(\mathbf{W})|\mathsf{U}_n, \mathbf{E}] - \mathbb{E}[G_{p,q}(\mathbf{W})|\mathbf{E}] = O_{\mathbb{P}}((n^{1/2}h_n^2)^{-1}) = O_{\mathbb{P}}(d_n^{-2}).$$

Hence if we take $d_n \gg 1$ (which implies $nd_n^4 \gg 1$), then $G_{p,q}(\mathbf{W}) = \mathbb{E}[G_{p,q}(\mathbf{W})|\mathbf{E}] + o_{\mathbb{P}}(1)$, implying $\|\mathbf{D}_n - \mathbf{D}\|_2 = o_{\mathbb{P}}(1)$ and $\lambda_{\min}(\mathbf{D}_n) - \lambda_{\min}(\mathbf{D}) = o_{\mathbb{P}}(1)$, making $\lambda_{\min}(\mathbf{D}_n) \gtrsim_{\mathbb{P}} 1$. Take

$$\boldsymbol{\Upsilon}_n := \frac{1}{nd_n^2} \sum_{i=1}^{n} \mathbf{r}\Big(\frac{\eta_i}{d_n}\Big)\mathbf{r}\Big(\frac{\eta_i}{d_n}\Big)^\top K^2\Big(\frac{\eta_i}{d_n}\Big)\mathbb{V}[Y_i|\eta_i].$$

Hence variance can be bounded by

$$\mathbb{V}[\widehat{\gamma}_0|\mathbf{E}, \mathbf{W}] = \mathbf{e}_0^{\mathbf{T}}\mathbf{D}_n^{-1}\boldsymbol{\Upsilon}_n\mathbf{D}_n^{-1}\mathbf{e}_0 \lesssim (nd_n)^{-1}, \tag{SA-39}$$

$$\mathbb{V}[\widehat{\gamma}_1|\mathbf{E}, \mathbf{W}] = n^{1/2}\mathbf{e}_1^{\mathbf{T}}\mathbf{D}_n^{-1}\boldsymbol{\Upsilon}_n\mathbf{D}_n^{-1}\mathbf{e}_1 \lesssim n^{1/2}(nd_n^3)^{-1} = n^{-1/2}d_n^{-3}. \tag{SA-40}$$

By similar argument as in Case 1, assume $d_n \gg 1$, we can show

$$\mathbb{E}[\widehat{\gamma}_0|\mathbf{E}] - \gamma_0 = O(n^{-1/4} + n^{-1/2}d_n^2), \qquad \mathbb{E}[\widehat{\gamma}_1|\mathbf{E}] - \gamma_1 = O(n^{-1/4}d_n^2).$$

Hence if we choose d_n such that $1 \ll d_n \ll n^{1/8}$, then $(\widehat{\gamma}_0, \widehat{\gamma}_1)$ is a consistent estimator for (γ_0, γ_1). The only assumption we made for the existence of such a d_n is $\liminf_{n\to\infty} n\rho_n^2 \geq c$ for some $c > 0$.

Case 2.2: $n\rho_n^2 = o(1)$ Take $\eta_i := \sqrt{n\rho_n}(\frac{M_i}{N_i} - \pi)$, $d_n = \sqrt{n\rho_n}h_n$. By similar decomposition based on latent variables, we can show if $n\rho_n \to \infty$ as $n \to \infty$, then there exists h_n such that $(\widehat{\gamma}_0, \widehat{\gamma}_1)$ is a consistent estimator for (γ_0, γ_1).

References

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

[2] Norman Bleistein and Richard A Handelsman. *Asymptotic expansions of integrals*. Ardent Media, 1975.

[3] Sourav Chatterjee. Spin glasses and stein's method. *Probability Theory and Related Fields*, 148(3):567–600, 2010.

[4] Yuval Dagan, Constantinos Daskalakis, Nishanth Dikkala, and Anthimos Vardis Kandiros. Learning ising models from one or multiple samples. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 161–168, 2021.

[5] Peter Eichelsbacher and Matthias Loewe. Stein's Method for Dependent Random Variables Occuring in Statistical Mechanics. *Electronic Journal of Probability*, 15:962 – 988, 2010.

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

[7] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[8] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge university press, 2019.