Supplementary material for

Logistic Regression for Massive Data with Rare Events

In this section, we prove all theoretical results in the paper. To facilitate the presentation of the proofs, denote

$$a_n = \sqrt{ne^{\alpha_{nt}}}.$$

The condition that $\mathbb{E}(e^{t\|\mathbf{x}\|}) < \infty$ for any $t > 0$ implies that

$$\mathbb{E}(e^{t_1\|\mathbf{x}\|}\|\mathbf{z}\|^{t_2}) < \infty, \tag{S.1}$$

for any $t_1 > 0$ and $t_2 > 0$, and we will use this result multiple times in the proof. The inequality in (S.1) is true because for any $t_1 > 0$ and $t_2 > 0$, we can choose $t > t_1$ and $k > t_2$ so that

$$e^{t\|\mathbf{x}\|} \geq e^{-t}e^{t\|\mathbf{z}\|} = e^{-t}e^{t_1\|\mathbf{z}\|}e^{(t-t_1)\|\mathbf{z}\|} \geq \frac{(t-t_1)^k e^{-t}}{k!}e^{t_1\|\mathbf{x}\|}\|\mathbf{z}\|^k \geq \frac{(t-t_1)^k e^{-t}}{k!}e^{t_1\|\mathbf{x}\|}\|\mathbf{z}\|^{t_2},$$

with probability one.

S.1 Proof of Theorem 1

Proof of Theorem 1. The estimator $\hat{\boldsymbol{\theta}}$ is the maximizer of

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n}\big[(\alpha + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})y_i - \log\{1 + \exp(\alpha + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\}\big], \tag{S.2}$$

so $\mathbf{u}_n = a_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{nt})$ is the maximizer of

$$\gamma(\mathbf{u}) = \ell(\boldsymbol{\theta}_{nt} + a_n^{-1}\mathbf{u}) - \ell(\boldsymbol{\theta}_{nt}). \tag{S.3}$$

By Taylor's expansion,

$$\gamma(\mathbf{u}) = a_n^{-1}\mathbf{u}^{\mathrm{T}}\dot{\ell}(\boldsymbol{\theta}_{nt}) + 0.5a_n^{-2}\sum_{i=1}^{n}\phi_i(\boldsymbol{\theta}_{nt} + a_n^{-1}\acute{\mathbf{u}})(\mathbf{z}_i^{\mathrm{T}}\mathbf{u})^2, \tag{S.4}$$

where $\phi_i(\boldsymbol{\theta}) = p_i(\alpha, \boldsymbol{\beta})\{1 - p_i(\alpha, \boldsymbol{\beta})\}$,

$$\dot{\ell}(\boldsymbol{\theta}) = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n}\{y_i - p_i(\boldsymbol{\theta})\}\mathbf{z}_i = \sum_{i=1}^{n}\{y_i - p_i(\alpha, \boldsymbol{\beta})\}\mathbf{z}_i$$

is the gradient of $\ell(\boldsymbol{\theta})$, and $\acute{\mathbf{u}}$ lies between $\mathbf{0}$ and $\mathbf{u}$. If we can show that

$$a_n^{-1}\dot{\ell}(\boldsymbol{\theta}_{nt}) \longrightarrow \mathbb{N}\big(\mathbf{0},\ \mathbf{M}_f\big), \tag{S.5}$$

in distribution, and for any $\mathbf{u}$,

$$a_n^{-2}\sum_{i=1}^{n}\phi_i(\boldsymbol{\theta}_{nt}+a_n^{-1}\acute{\mathbf{u}})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \longrightarrow \mathbf{M}_f, \tag{S.6}$$

in probability, then from the Basic Corollary in page 2 of Hjort & Pollard (2011), we know that $a_n(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_{nt})$, the maximizer of $\gamma(\mathbf{u})$, satisfies that

$$a_n(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_{nt}) = \mathbf{M}_f^{-1}\times a_n^{-1}\dot{\ell}(\boldsymbol{\theta}_{nt}) + o_P(1). \tag{S.7}$$

Slutsky's theorem together with (S.5) and (S.7) implies the result in Theorem 1. We prove (S.5) and (S.6) in the following.

Note that

$$\dot{\ell}(\boldsymbol{\theta}_{nt}) = \sum_{i=1}^{n}\big\{y_i - p_i(\alpha_{nt},\boldsymbol{\beta}_t)\big\}\mathbf{z}_i, \tag{S.8}$$

is a summation of i.i.d. quantities. Since $\alpha_{nt}\to-\infty$ as $n\to\infty$, the distribution of $\{y-p(\alpha_{nt},\boldsymbol{\beta}_t)\}\mathbf{z}$ depends on n, we need to use a central limit theorem for triangular arrays. The Lindeberg-Feller central limit theorem (see, Section *2.8 of van der Vaart, 1998) is appropriate.

We exam the mean and variance of $a_n^{-1}\dot{\ell}(\boldsymbol{\theta}_{nt})$. For the mean, from the fact that

$$\mathbb{E}[\{y_i - p_i(\alpha_{nt},\boldsymbol{\beta}_t)\}\mathbf{z}_i] = \mathbb{E}[\mathbb{E}\{y_i - p_i(\alpha_{nt},\boldsymbol{\beta}_t)|\mathbf{z}_i\}\mathbf{z}_i] = \mathbf{0},$$

we know that $\mathbb{E}\{a_n^{-1}\dot{\ell}(\boldsymbol{\theta}_{nt})\} = \mathbf{0}$.

For the variance,

$$\begin{aligned}\mathbb{V}\{a_n^{-1}\dot{\ell}(\boldsymbol{\theta}_{nt})\} &= a_n^{-2}\sum_{i=1}^{n}\mathbb{V}[\{y_i - p_i(\alpha_{nt},\boldsymbol{\beta}_t)\}\mathbf{z}_i] = a_n^{-2}n\mathbb{E}\{\phi(\boldsymbol{\theta}_{nt})\mathbf{z}\mathbf{z}^{\mathrm{T}}\}\\ &= a_n^{-2}n\mathbb{E}\bigg\{\frac{e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}}{(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2}\bigg\} = \mathbb{E}\bigg\{\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}}{(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2}\bigg\}.\end{aligned}$$

Note that

$$\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}}{(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2} \longrightarrow e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}},$$

almost surely, and

$$\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^2}{(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2} \le e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^2 \quad \text{with} \quad \mathbb{E}(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^2) \le \infty.$$

Thus, from the dominated convergence theorem,

$$\mathbb{V}\{a_n^{-1}\dot{\ell}(\boldsymbol{\theta}_{nt})\} = \mathbb{E}\bigg\{\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}}{(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2}\bigg\} \longrightarrow \mathbb{E}\big(e^{\boldsymbol{\theta}_{nt}^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\big).$$

Now we check the Lindeberg-Feller condition. For any $\epsilon > 0$,

$$\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}\Big[\|\{y_i - p_i(\alpha_{nt},\boldsymbol{\beta}_t)\}\mathbf{z}_i\|^2 I(\|\{y_i - p_i(\alpha_{nt},\boldsymbol{\beta}_t)\}\mathbf{z}_i\| > a_n\epsilon)\Big]\\
&= n\mathbb{E}\Big[\|\{y - p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\|^2 I(\|\{y - p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\| > a_n\epsilon)\Big]\\
&= n\mathbb{E}\big[p(\boldsymbol{\theta}_{nt})\{1-p(\boldsymbol{\theta}_{nt})\}^2\|\mathbf{z}\|^2 I(\|\{1-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\| > a_n\epsilon)\big]\\
&\quad + n\mathbb{E}\big[\{1-p(\boldsymbol{\theta}_{nt})\}\{p(\boldsymbol{\theta}_{nt})\}^2\|\mathbf{z}\|^2 I(\|p(\boldsymbol{\theta}_{nt})\mathbf{z}\| > a_n\epsilon)\big]\\
&\le n\mathbb{E}\big[p(\boldsymbol{\theta}_{nt})\|\mathbf{z}\|^2 I(\|\mathbf{z}\| > a_n\epsilon)\big] + n\mathbb{E}\big[\{p(\boldsymbol{\theta}_{nt})\}^2\|\mathbf{z}\|^2 I(\|p(\boldsymbol{\theta}_{nt})\mathbf{z}\| > a_n\epsilon)\big]\\
&\le a_n^2\mathbb{E}\{e^{\|\beta_t\|\|\mathbf{x}\|}\|\mathbf{z}\|^2 I(\|\mathbf{z}\| > a_n\epsilon)\} + a_n^2\mathbb{E}\{e^{\|\beta_t\|\|\mathbf{x}\|}\|\mathbf{z}\|^2 I(\|\mathbf{z}\| > a_n\epsilon)\}\\
&= o(a_n^2),
\end{aligned}$$

where the last step is from the dominated convergence theorem. Thus, applying the Lindeberg-Feller central limit theorem (Section *2.8 of van der Vaart, 1998), we finish the proof of (S.5).

The last step is to prove (S.6). We first show that

$$\begin{aligned}
&\bigg|a_n^{-2}\sum_{i=1}^{n}\phi_i(\boldsymbol{\theta}_{nt} + a_n^{-1}\acute{\mathbf{u}})\|\mathbf{z}_i\|^2 - a_n^{-2}\sum_{i=1}^{n}\phi_i(\boldsymbol{\theta}_{nt})\|\mathbf{z}_i\|^2\bigg|\\
&\le a_n^{-2}\sum_{i=1}^{n}\big|\phi_i(\boldsymbol{\theta}_{nt} + a_n^{-1}\acute{\mathbf{u}}) - \phi_i(\boldsymbol{\theta}_{nt})\big|\|\mathbf{z}_i\|^2\\
&\le \|a_n^{-1}\acute{\mathbf{u}}\|a_n^{-2}\sum_{i=1}^{n}p_i(\boldsymbol{\theta}_{nt} + a_n^{-1}\breve{\mathbf{u}})\|\mathbf{z}_i\|^3\\
&= \frac{\|a_n^{-1}\acute{\mathbf{u}}\|}{n}\sum_{i=1}^{n}\frac{e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t + a_n^{-1}\breve{\mathbf{u}}^{\mathrm{T}}\mathbf{z}_i}}{\{1+e^{\boldsymbol{\theta}_{nt}^{\mathrm{T}}\mathbf{z}_i + a_n^{-1}\breve{\mathbf{u}}^{\mathrm{T}}\mathbf{z}_i}\}^2}\|\mathbf{z}_i\|^3\\
&\le \frac{\|a_n^{-1}\acute{\mathbf{u}}\|}{n}\sum_{i=1}^{n}e^{(\|\boldsymbol{\beta}_t\|+\|\mathbf{u}\|)(1+\|\mathbf{x}_i\|)}\|\mathbf{z}_i\|^3 = o_P(1).
\end{aligned}\tag{S.9}$$

Here, the last inequality in (S.9) is because $\breve{\mathbf{u}}$ lies between $\mathbf{0}$ and $\acute{\mathbf{u}}$, and thus $\|a_n^{-1}\breve{\mathbf{u}}\| \le \|\mathbf{u}\|$ for $a_n \ge 1$.

To finish the proof, we only need to prove that

$$a_n^{-2}\sum_{i=1}^{n}\phi_i(\boldsymbol{\theta}_{nt})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \longrightarrow \mathbb{E}(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}),\tag{S.10}$$

in probability. This is done by noting that

$$a_n^{-2}\sum_{i=1}^{n}\phi_i(\boldsymbol{\theta}_{nt})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} = \frac{1}{ne^{\alpha_{nt}}}\sum_{i=1}^{n}\frac{e^{\boldsymbol{\theta}_{nt}^{\mathrm{T}}\mathbf{z}_i}}{(1+e^{\boldsymbol{\theta}_{nt}^{\mathrm{T}}\mathbf{z}_i})^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\tag{S.11}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\frac{e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{(1+e^{\boldsymbol{\theta}_{nt}^{\mathrm{T}}\mathbf{z}_i})^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}=\mathbb{E}(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}})+o_P(1), \tag{S.12}$$

by Proposition 1 of Wang (2019). □

S.2 Proof of Theorem 2

Proof of Theorem 2. The estimator $\hat{\boldsymbol{\theta}}^{\mathrm{w}}_{\mathrm{under}}$ is the maximizer of $\ell^{\mathrm{w}}_{\mathrm{under}}(\boldsymbol{\theta})$ defined in (10), so $\sqrt{a_n}(\hat{\boldsymbol{\theta}}^{\mathrm{w}}_{\mathrm{under}}-\theta_t)$ is the maximizer of $\gamma^{\mathrm{w}}_{\mathrm{under}}(\mathbf{u})=\ell^{\mathrm{w}}_{\mathrm{under}}(\boldsymbol{\theta}_{nt}+a_n^{-1}\mathbf{u})-\ell^{\mathrm{w}}_{\mathrm{under}}(\boldsymbol{\theta}_{nt})$. By Taylor's expansion,

$$\gamma^{\mathrm{w}}_{\mathrm{under}}(\mathbf{u})=\frac{1}{a_n}\mathbf{u}^{\mathrm{T}}\dot{\ell}^{\mathrm{w}}_{\mathrm{under}}(\boldsymbol{\theta}_{nt})+\frac{1}{2a_n^2}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\phi_i(\boldsymbol{\theta}_{nt}+a_n^{-1}\acute{\mathbf{u}})(\mathbf{z}_i^{\mathrm{T}}\mathbf{u})^2, \tag{S.13}$$

where

$$\dot{\ell}^{\mathrm{w}}_{\mathrm{under}}(\boldsymbol{\theta})=\frac{\partial\ell^{\mathrm{w}}_{\mathrm{under}}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}=\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\{y_i-p_i(\boldsymbol{\theta})\}\mathbf{z}_i=\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\{y_i-p_i(\alpha,\boldsymbol{\beta})\}\mathbf{z}_i$$

is the gradient of $\ell^{\mathrm{w}}_{\mathrm{under}}(\boldsymbol{\theta})$, and $\acute{\mathbf{u}}$ lies between $\mathbf{0}$ and $\mathbf{u}$. Similarly to the proof of Theorem 1, we only need to show that

$$a_n^{-1}\dot{\ell}^{\mathrm{w}}_{\mathrm{under}}(\boldsymbol{\theta}_{nt})\longrightarrow\mathbb{N}\Big[\mathbf{0},\ \mathbb{E}\big\{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}(1+ce^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})\mathbf{z}\mathbf{z}^{\mathrm{T}}\big\}\Big], \tag{S.14}$$

in distribution, and for any $\mathbf{u}$,

$$a_n^{-2}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\phi_i(\boldsymbol{\theta}_{nt}+a_n^{-1}\acute{\mathbf{u}})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\longrightarrow\mathbb{E}\big(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\big), \tag{S.15}$$

in probability.

We prove (S.14) first. Recall that $\mathcal{D}_n$ is the full data set and $\delta_i=y_i+(1-y_i)I(u_i\le\pi_0)$, satisfying that

$$\pi_i=\mathbb{E}(\delta_i|\mathcal{D}_n)=y_i+(1-y_i)\pi_0=\pi_0+(1-\pi_0)y_i.$$

We notice that

$$\mathbb{E}(\delta_i|\mathbf{z}_i)=p_i(\alpha_{nt},\boldsymbol{\beta}_t)+\{1-p_i(\alpha_{nt},\boldsymbol{\beta}_t)\}\pi_0=\pi_0+(1-\pi_0)p_i(\alpha_{nt},\boldsymbol{\beta}_t).$$

Let $\eta_i=\frac{\delta_i}{\pi_i}\{y_i-p_i(\boldsymbol{\theta}_{nt})\}\mathbf{z}_i$, we know that η_i, $i=1,...,n$, are i.i.d., with the underlying distribution of η_i being dependent on n. From direct calculation, we have

$$\begin{aligned}
&\mathbb{E}(\eta_i|\mathbf{z}_i)=\mathbf{0},\qquad\text{and}\\
&\mathbb{V}(\eta_i|\mathbf{z}_i)=\mathbb{E}\left[\frac{\{y_i-p_i(\boldsymbol{\theta}_{nt})\}^2}{\pi_0+y_i(1-\pi_0)}\middle|\mathbf{z}_i\right]\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\\
&\qquad=\big[p_i(\boldsymbol{\theta}_{nt})\{1-p_i(\boldsymbol{\theta}_{nt})\}^2+\pi_0^{-1}\{1-p_i(\boldsymbol{\theta}_{nt})\}\{p_i(\boldsymbol{\theta}_{nt})\}^2\big]\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\\
&\qquad=\big\{1-p_i(\boldsymbol{\theta}_{nt})+\pi_0^{-1}p_i(\boldsymbol{\theta}_{nt})\big\}p_i(\boldsymbol{\theta}_{nt})\{1-p_i(\boldsymbol{\theta}_{nt})\}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1+\pi_0^{-1}e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{(1+e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})^2} p_i(\boldsymbol{\theta}_{nt})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \\
&\le e^{\alpha_{nt}}(1+\pi_0^{-1}e^{\alpha_{nt}}e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}.
\end{aligned}
$$

Thus, by the dominated convergence theorem, we obtain that

$$
\mathbb{V}(\eta_i) = \mathbb{E}\{\mathbb{V}(\eta_i|\mathbf{z}_i)\} = e^{\alpha_{nt}}\mathbb{E}\Big\{e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}(1+ce^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\Big\}\{1+o(1)\}. \tag{S.16}
$$

Now we check the Lindeberg-Feller condition (Section *2.8 of van der Vaart, 1998). For simplicity, let $\pi = \pi_0 + (1-\pi_0)y$ and $\delta = y + (1-y)I(u \le \pi)$, where $u \sim \mathbb{U}(0,1)$. For any $\epsilon > 0$,

$$
\begin{aligned}
&\sum_{i=1}^{n}\mathbb{E}\big\{\|\eta_i\|^2 I(\|\eta_i\| > a_n\epsilon)\big\} \\
=& n\mathbb{E}\big[\|\pi^{-1}\delta\{y-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\|^2 I(\|\pi^{-1}\delta\{y-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\| > a_n\epsilon)\big] \\
=& \pi_0 n\mathbb{E}\big[\|\pi^{-1}\{y-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\|^2 I(\|\pi^{-1}\{y-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\| > a_n\epsilon)\big] \\
&+ (1-\pi_0)n\mathbb{E}\big[\pi^{-1}\|y\{y-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\|^2 I(\|\pi^{-1}y\{y-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\| > a_n\epsilon)\big] \\
=& \pi_0 n\mathbb{E}\big[p(\boldsymbol{\theta}_{nt})\|\{1-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\|^2 I(\|\{1-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\| > a_n\epsilon)\big] \\
&+ \pi_0^{-1}n\mathbb{E}\big[\{1-p(\boldsymbol{\theta}_{nt})\}\|p(\boldsymbol{\theta}_{nt})\mathbf{z}\|^2 I(\pi_0^{-1}\|p(\boldsymbol{\theta}_{nt})\mathbf{z}\| > a_n\epsilon)\big] \\
&+ (1-\pi_0)n\mathbb{E}\big[p(\boldsymbol{\theta}_{nt})\|\{1-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\|^2 I(\|\{1-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\| > a_n\epsilon)\big] \\
\le& n\mathbb{E}\big\{p(\boldsymbol{\theta}_{nt})\|\mathbf{z}\|^2 I(\|\mathbf{z}\| > a_n\epsilon)\big\} + n\pi_0^{-1}\mathbb{E}\big\{\|p(\boldsymbol{\theta}_{nt})\mathbf{z}\|^2 I(\|\pi_0^{-1}p(\boldsymbol{\theta}_{nt})\mathbf{z}\| > a_n\epsilon)\big\} \\
\le& ne^{\alpha_{nt}}\mathbb{E}\big\{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^2 I(\|\mathbf{z}\| > a_n\epsilon)\big\} + n\pi_0^{-1}e^{2\alpha_{nt}}\mathbb{E}\big\{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^2 I(\pi_0^{-1}e^{\alpha_{nt}}e^{\alpha_{nt}}\|\mathbf{z}\| > a_n\epsilon)\big\} \\
=& o(ne^{\alpha_{nt}}) = o(a_n^2),
\end{aligned}
$$

where the second last step is from the dominated convergence theorem and the facts that $a_n \to \infty$ and $\lim_{n\to\infty} e^{\alpha}/\pi_0 = c < \infty$. Thus, applying the Lindeberg-Feller central limit theorem (Section *2.8 of van der Vaart, 1998) finishes the proof of (S.14).

Now we prove (S.15). By direct calculation, we first notice that

$$
\Delta_1 \equiv a_n^{-2}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\phi_i(\boldsymbol{\theta}_{nt})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\{y_i+(1-y_i)I(u_i \le \pi_0)\}e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{\pi_i(1+e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \tag{S.17}
$$

has a mean of

$$
\mathbb{E}(\Delta_1) = \mathbb{E}\bigg\{\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2}\mathbf{z}\mathbf{z}^{\mathrm{T}}\bigg\} = \mathbb{E}\big(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\big) + o(1), \tag{S.18}
$$

where the last step is by the dominated convergence theorem. In addition, the variance of each component of Δ_1 is bounded by

$$
\frac{1}{n}\mathbb{E}\bigg\{\frac{e^{2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^4}{\pi(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^4}\bigg\} \le \frac{\mathbb{E}(e^{2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^4)}{n\pi_0} = o(1), \tag{S.19}
$$

where the last step is because $ne^{\alpha_{nt}} \to \infty$ and $e^{\alpha_{nt}}/\pi_0 \to c < \infty$ imply that $n\pi_0 \to \infty$. From (S.18) and (S.19), Chebyshev's inequality implies that $\Delta_1 \to \mathbb{E}\big(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\big)$ in probability. Notice that

$$\begin{aligned}
&\left|a_n^{-2}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\phi_i(\boldsymbol{\theta}_{nt}+a_n^{-1}\acute{\mathbf{u}})\|\mathbf{z}_i\|^2 - a_n^{-2}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}\phi_i(\boldsymbol{\theta}_{nt})\|\mathbf{z}_i\|^2\right| \\
&\le \|a_n^{-1}\acute{\mathbf{u}}\|a_n^{-2}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}p_i(\boldsymbol{\theta}_{nt}+a_n^{-1}\breve{\mathbf{u}})\|\mathbf{z}_i\|^3 \\
&\le \|a_n^{-1}\acute{\mathbf{u}}\| \times \frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i}{\pi_i}e^{(\|\boldsymbol{\beta}_t\|+\|\mathbf{u}\|)\|\mathbf{z}_i\|}\|\mathbf{z}_i\|^3 \equiv \|a_n^{-1}\acute{\mathbf{u}}\| \times \Delta_2.
\end{aligned}$$

Since $\|a_n^{-1}\acute{\mathbf{u}}\| \to 0$, to finish the proof of (S.15), we only need to prove that Δ_2 is bounded in probability. Using an approach similar to (S.18) and (S.19), we can show that Δ_2 has a mean that is bounded and a variance that converges to zero.

□

S.3 Proof of Theorem 3

Proof of Theorem 3. If we use Υ_{bc} to denote the under-sampled objective function shifted by $\mathbf{b}$, i.e., $\Upsilon_{bc}(\boldsymbol{\theta}) = \ell^{\mathrm{u}}_{\mathrm{under}}(\boldsymbol{\theta}-\mathbf{b})$, then the estimator $\hat{\boldsymbol{\theta}}^{\mathrm{ubc}}_{\mathrm{under}}$ is the maximizer of

$$\Upsilon_{bc}(\boldsymbol{\theta}) = \sum_{i=1}^{n}\delta_i\big[(\boldsymbol{\theta}-\mathbf{b})^{\mathrm{T}}\mathbf{z}_i y_i - \log\{1+e^{(\boldsymbol{\theta}-\mathbf{b})^{\mathrm{T}}\mathbf{z}_i}\}\big]. \tag{S.20}$$

We notice that $\sqrt{a_n}(\hat{\boldsymbol{\theta}}^{\mathrm{ubc}}_{\mathrm{under}} - \boldsymbol{\theta}_{nt})$ is the maximizer of $\gamma_{bc}(\mathbf{u}) = \Upsilon_{bc}(\boldsymbol{\theta}_{nt}+a_n^{-1}\mathbf{u}) - \Upsilon_{bc}(\boldsymbol{\theta}_{nt})$. By Taylor's expansion,

$$\gamma_p(\mathbf{u}) = \frac{1}{a_n}\mathbf{u}^{\mathrm{T}}\dot{\Upsilon}_{bc}(\boldsymbol{\theta}_{nt}) + \frac{1}{2a_n^2}\sum_{i=1}^{n}\delta_i\phi_i(\boldsymbol{\theta}_{nt}-\mathbf{b}+a_n^{-1}\acute{\mathbf{u}})(\mathbf{z}_i^{\mathrm{T}}\mathbf{u})^2, \tag{S.21}$$

where

$$\dot{\Upsilon}_{bc}(\boldsymbol{\theta}) = \frac{\partial\Upsilon_{bc}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \sum_{i=1}^{n}\delta_i\{y_i - p_i(\boldsymbol{\theta}_{nt}-\mathbf{b})\}\mathbf{z}_i = \sum_{i=1}^{n}\delta_i\{y_i - p_i(\alpha_{nt}-b,\boldsymbol{\beta}_t)\}\mathbf{z}_i$$

is the gradient of $\Upsilon_{bc}(\boldsymbol{\theta})$, and $\acute{\mathbf{u}}$ lies between $\mathbf{0}$ and $\mathbf{u}$.

Similarly to the proof of Theorem 1, we only need to show that

$$a_n^{-1}\dot{\Upsilon}_{bc}(\boldsymbol{\theta}_{nt}) \longrightarrow \mathbb{N}\left\{\mathbf{0},\ \mathbb{E}\left(\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}}{1+ce^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}\right)\right\}, \tag{S.22}$$

in distribution, and for any $\mathbf{u}$,

$$a_n^{-2}\sum_{i=1}^{n}\delta_i\phi_i(\boldsymbol{\theta}_{nt}-\mathbf{b}+a_n^{-1}\acute{\mathbf{u}})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \longrightarrow \mathbb{E}\left(\frac{e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{1+ce^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\right) \tag{S.23}$$

in probability.

We prove (S.22) first. Define $\eta_{ui} = \delta_i\{y_i - p_i(\alpha_{nt} - b, \boldsymbol{\beta}_t)\}\mathbf{z}_i$. We have that

$$\begin{aligned}\mathbb{E}(\eta_{ui}|\mathbf{z}_i) &= \mathbb{E}[\{\pi_0 + y_i(1-\pi_0)\}\{y_i - p_i(\alpha_{nt} - b, \boldsymbol{\beta}_t)\}|\mathbf{z}_i]\mathbf{z}_i \\ &= [p_i(\alpha_{nt}, \boldsymbol{\beta}_t)\{1 - p_i(\alpha_{nt} - b, \boldsymbol{\beta}_t)\} - \pi_0\{1 - p_i(\alpha_{nt}, \boldsymbol{\beta}_t)\}\{p_i(\alpha_{nt} - b, \boldsymbol{\beta}_t)\}]\mathbf{z}_i = 0,\end{aligned}$$

which implies that $\mathbb{E}(\eta_{ui}) = \mathbf{0}$. For the conditional variance

$$\begin{aligned}\mathbb{V}(\eta_{ui}|\mathbf{z}_i) &= \mathbb{E}[\{\pi_0 + y_i(1-\pi_0)\}\{y_i - p_i(\alpha_{nt} - b, \boldsymbol{\beta}_t)\}^2|\mathbf{z}_i]\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \\ &= \big[p_i(\alpha_{nt}, \boldsymbol{\beta}_t)\{1 - p_i(\alpha_{nt} - b, \boldsymbol{\beta}_t)\}^2 + \pi_0\{1 - p_i(\alpha_{nt}, \boldsymbol{\beta}_t)\}\{p_i(\alpha_{nt} - b, \boldsymbol{\beta}_t)\}^2\big]\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \\ &= \frac{e^{\alpha_{nt} + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t} + \pi_0 e^{2(\alpha_{nt} - b_0 + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t)}}{(1 + e^{\alpha_{nt} + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})(1 + e^{\alpha_{nt} - b_0 + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \\ &= \frac{e^{\alpha_{nt} + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{1 + e^{\alpha_{nt} - b_0 + \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}\{1 - p_i(\alpha_{nt}, \boldsymbol{\beta}_t)\}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \le e^{\alpha_{nt}} e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}},\end{aligned}$$

where $e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}$ is integrable. Thus, by the dominated convergence theorem, $\mathbb{V}(\eta_{ui})$ satisfies that

$$\mathbb{V}(\eta_{ui}) = \mathbb{E}\{\mathbb{V}(\eta_{ui}|\mathbf{z}_i)\} = e^{\alpha_{nt}}\mathbb{E}\left(\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{1 + ce^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}\right)\{1 + o(1)\}. \tag{S.24}$$

Therefore, we have

$$a_n^{-2}\sum_{i=1}^n \mathbb{V}(\eta_{ui}) \longrightarrow \mathbb{E}\left(\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{1 + ce^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\right). \tag{S.25}$$

Now we check the Lindeberg-Feller condition. For any $\epsilon > 0$,

$$\begin{aligned}&\sum_{i=1}^n \mathbb{E}\big\{\|\eta_{ui}\|^2 I(\|\eta_{ui}\| > a_n\epsilon)\big\} \\ =& n\mathbb{E}\big[\|\delta\{y - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\|^2 I(\|\delta\{y - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\| > a_n\epsilon)\big] \\ =& \pi_0 n\mathbb{E}\big[\|\{y - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\|^2 I(\|\{y - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\| > a_n\epsilon)\big] \\ &+ (1-\pi_0)n\mathbb{E}\big[\|y\{y - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\|^2 I(\|y\{y - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\| > a_n\epsilon)\big] \\ =& \pi_0 n\mathbb{E}\big[p(\boldsymbol{\theta}_{nt})\|\{1 - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\|^2 I(\|\{1 - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\| > a_n\epsilon)\big] \\ &+ \pi_0 n\mathbb{E}\big[\{1 - p(\boldsymbol{\theta}_{nt})\}\|p(\boldsymbol{\theta}_{nt} - \mathbf{b})\mathbf{z}\|^2 I(\|p(\boldsymbol{\theta}_{nt} - \mathbf{b})\mathbf{z}\| > a_n\epsilon)\big] \\ &+ (1-\pi_0)n\mathbb{E}\big[p(\boldsymbol{\theta}_{nt})\|\{1 - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\|^2 I(\|\{1 - p(\boldsymbol{\theta}_{nt} - \mathbf{b})\}\mathbf{z}\| > a_n\epsilon)\big] \\ \le& n\mathbb{E}\big\{p(\boldsymbol{\theta}_{nt})\|\mathbf{z}\|^2 I(\|\mathbf{z}\| > a_n\epsilon)\big\} + \pi_0 n\mathbb{E}\big[\|p(\boldsymbol{\theta}_{nt} - \mathbf{b})\mathbf{z}\|^2 I(\|\mathbf{z}\| > a_n\epsilon)\big] \\ \le& ne^{\alpha_{nt}}\mathbb{E}\big\{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^2 I(\|\mathbf{z}\| > a_n\epsilon)\big\} + \pi_0^{-1} ne^{2\alpha_{nt}}\mathbb{E}\big\{e^{2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^2 I(\|\mathbf{z}\| > a_n\epsilon)\big\} \\ =& o(ne^{\alpha_{nt}}) = o(a_n^2),\end{aligned}$$

where the second last step is from the dominated convergence theorem. Thus, applying the Lindeberg-Feller central limit theorem (Section *2.8 of van der Vaart, 1998) finishes the proof of (S.22).

No we prove (S.23). First, letting

$$\Delta_3 \equiv a_n^{-2}\sum_{i=1}^n \delta_i\phi_i(\boldsymbol{\theta}_{nt}-\mathbf{b})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} = \frac{1}{n}\sum_{i=1}^n \frac{\{y_i+(1-y_i)I(u_i\le\pi_0)\}e^{-b_0+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{\{1+e^{\alpha_{nt}-b_0+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}\}^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}, \tag{S.26}$$

the mean of Δ_3 satisfies that

$$\mathbb{E}(\Delta_3) = \mathbb{E}\left[\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{\{1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\}\{1+e^{\alpha_{nt}-b_0+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\right] = \mathbb{E}\left(\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{1+ce^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\right) + o(1), \tag{S.27}$$

by the dominated convergence theorem, and the variance of each component of Δ_3 is bounded by

$$\frac{1}{n}\mathbb{E}\left[\frac{\{y+(1-y)I(u\le\pi_0)\}e^{-2b_0+2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{\{1+e^{\alpha_{nt}-b_0+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\}^4}\|\mathbf{z}\|^4\right] \le \frac{\mathbb{E}(e^{2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^4)}{n\pi_0} = o(1). \tag{S.28}$$

Thus, Chebyshev's inequality implies that

$$\Delta_3 \longrightarrow \mathbb{E}\left(\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{1+ce^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\right), \tag{S.29}$$

in probability. Furthermore,

$$\begin{aligned}
&\left|a_n^{-2}\sum_{i=1}^n \delta_i\phi_i(\boldsymbol{\theta}_{nt}-\mathbf{b}+a_n^{-1}\acute{\mathbf{u}})\|\mathbf{z}_i\|^2 - a_n^{-2}\sum_{i=1}^n \delta_i\phi_i(\boldsymbol{\theta}_{nt}-\mathbf{b})\|\mathbf{z}_i\|^2\right|\\
&\le \|a_n^{-1}\acute{\mathbf{u}}\|a_n^{-2}\sum_{i=1}^n \delta_i p_i(\boldsymbol{\theta}_{nt}-\mathbf{b}+a_n^{-1}\breve{\mathbf{u}})\|\mathbf{z}_i\|^3\\
&\le \frac{\|a_n^{-1}\acute{\mathbf{u}}\|}{n}\sum_{i=1}^n \frac{\delta_i}{\pi_0}e^{(\|\boldsymbol{\beta}_t\|+\|\mathbf{u}\|)(1+\|\mathbf{x}_i\|)}\|\mathbf{z}_i\|^3 \equiv \|a_n^{-1}\acute{\mathbf{u}}\|\times\Delta_4 = o_P(1),
\end{aligned} \tag{S.30}$$

where the last step is because Δ_4 is bounded in probability due to the fact that it has a mean that is bounded and a variance that converges to zero. Combing (S.29) and (S.30), (S.23) follows. □

S.4 Proof of Proposition 1

Proof of Proposition 1. Let

$$\mathbf{g} = \frac{1}{\sqrt{h}}\{\mathbb{E}(h^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}})\}^{-1}\mathbf{v} - \sqrt{h}\{\mathbb{E}(\mathbf{v}\mathbf{v}^{\mathrm{T}})\}^{-1}\mathbf{v}.$$

Since $\mathbf{g}\mathbf{g}^{\mathrm{T}} \ge \mathbf{0}$, we have

$$\mathbf{0} \le \mathbb{E}(\mathbf{g}\mathbf{g}^{\mathrm{T}}) = \{\mathbb{E}(\mathbf{v}\mathbf{v}^{\mathrm{T}})\}^{-1}\mathbb{E}(h\mathbf{v}\mathbf{v}^{\mathrm{T}})\{\mathbb{E}(\mathbf{v}\mathbf{v}^{\mathrm{T}})\}^{-1} - \{\mathbb{E}(h^{-1}\mathbf{v}\mathbf{v}^{\mathrm{T}})\}^{-1},$$

which finishes the proof. □

S.5 Proof of Theorem 4

Proof of Theorem 4. The estimator $\hat{\boldsymbol{\theta}}^{\mathrm{w}}_{\mathrm{over}}$ is the maximizer of (20), so $\sqrt{a_n}(\hat{\boldsymbol{\theta}}^{\mathrm{w}}_{\mathrm{over}} - \theta_t)$ is the maximizer of $\gamma^{\mathrm{w}}_{\mathrm{over}}(\mathbf{u}) = \ell^{\mathrm{w}}_{\mathrm{over}}(\boldsymbol{\theta}_{nt} + a_n^{-1}\mathbf{u}) - \ell^{\mathrm{w}}_{\mathrm{over}}(\boldsymbol{\theta}_{nt})$. By Taylor's expansion,

$$\gamma^{\mathrm{w}}_{\mathrm{over}}(\mathbf{u}) = \frac{1}{a_n}\mathbf{u}^{\mathrm{T}}\dot{\ell}^{\mathrm{w}}_{\mathrm{over}}(\boldsymbol{\theta}_{nt}) + \frac{1}{2a_n^2}\sum_{i=1}^{n}\frac{\tau_i}{w_i}\phi_i(\boldsymbol{\theta}_{nt} + a_n^{-1}\acute{\mathbf{u}})(\mathbf{z}_i^{\mathrm{T}}\mathbf{u})^2, \tag{S.31}$$

where

$$\dot{\ell}^{\mathrm{w}}_{\mathrm{over}}(\boldsymbol{\theta}) = \frac{\partial \ell^{\mathrm{w}}_{\mathrm{over}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n}\frac{\tau_i}{w_i}\{y_i - p_i(\boldsymbol{\theta}_{nt})\}\mathbf{z}_i = \sum_{i=1}^{n}\frac{\tau_i}{w_i}\{y_i - p_i(\alpha_{nt} - b, \boldsymbol{\beta}_t)\}\mathbf{z}_i$$

is the gradient of $\ell^{\mathrm{w}}_{\mathrm{over}}(\boldsymbol{\theta})$, and $\acute{\mathbf{u}}$ lies between $\mathbf{0}$ and $\mathbf{u}$. Similarly to the proof of Theorem 1, we only need to show that

$$a_n^{-1}\dot{\ell}^{\mathrm{w}}_{\mathrm{over}}(\boldsymbol{\theta}_{nt}) \longrightarrow \mathbb{N}\Big\{\mathbf{0},\ \frac{(1+\lambda)^2+\lambda}{(1+\lambda)^2}\mathbb{E}\big(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\big)\Big\}, \tag{S.32}$$

in distribution, and for any $\mathbf{u}$,

$$a_n^{-2}\sum_{i=1}^{n}\frac{\tau_i}{w_i}\phi_i(\boldsymbol{\theta}_{nt} + a_n^{-1}\acute{\mathbf{u}})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \longrightarrow \mathbb{E}\big(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\big), \tag{S.33}$$

in probability.

We prove (S.32) first. Denote $\eta_{owi} = \tau_i w_i^{-1}\{y_i - p_i(\boldsymbol{\theta}_{nt})\}\mathbf{z}_i$, so η_{owi}, $i = 1, ..., n$, are i.i.d. with the underlying distribution of η_{owi} being dependent on n. From direct calculation, we have

$$\begin{aligned}
&\mathbb{E}(\eta_{owi}|\mathbf{z}_i) = \mathbf{0}, \quad \text{and}\\
&\mathbb{V}(\eta_{owi}|\mathbf{z}_i) = \mathbb{E}\bigg[\frac{\{y_i(3\lambda_n + \lambda_n^2) + 1\}\{y_i - p_i(\boldsymbol{\theta}_{nt})\}^2}{(1+\lambda_n y_i)^2}\bigg|\mathbf{z}_i\bigg]\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\\
&\qquad = \bigg[p_i(\boldsymbol{\theta}_{nt})\{1 - p_i(\boldsymbol{\theta}_{nt})\}^2\frac{(1+\lambda_n)^2+\lambda_n}{(1+\lambda_n)^2} + \{1 - p_i(\boldsymbol{\theta}_{nt})\}\{p_i(\boldsymbol{\theta}_{nt})\}^2\bigg]\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\\
&\qquad = \frac{(1+\lambda_n)^2+\lambda_n}{(1+\lambda_n)^2}e^{\alpha_{nt}}e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\{1 + o_P(1)\},
\end{aligned}$$

where the $o_P(1)$ is bounded. Thus, by the dominated convergence theorem, we obtain that

$$\mathbb{V}(\eta_{owi}) = \frac{(1+\lambda)^2+\lambda}{(1+\lambda)^2}e^{\alpha_{nt}}\mathbb{E}\big(e^{\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_t}\mathbf{z}\mathbf{z}^{\mathrm{T}}\big)\{1 + o(1)\}.$$

Now we check the Lindeberg-Feller condition (Section *2.8 of van der Vaart, 1998). Let $w = 1 + \lambda_n y$ and $\tau = yv + 1$, where $v \sim \mathbb{POI}(\lambda_n)$. For any $\epsilon > 0$,

$$\sum_{i=1}^{n}\mathbb{E}\big[\|\eta_{owi}\|^2 I(\|\eta_{owi}\| > a_n\epsilon)\big]$$

$$
\begin{aligned}
&= n\mathbb{E}\big[\|w^{-1}\tau\{y-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\|^2 I(\|w^{-1}\tau\{y-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\| > a_n\epsilon)\big]\\
&\le \frac{n}{a_n\epsilon}\mathbb{E}\big[\|w^{-1}\tau\{y-p(\boldsymbol{\theta}_{nt})\}\mathbf{z}\|^3\\
&= \frac{n}{a_n\epsilon}\mathbb{E}\bigg[\frac{(1+vy)^3}{(1+\lambda_n y)^3}\{y-p(\boldsymbol{\theta}_{nt})\}^3\|\mathbf{z}\|^3\bigg]\\
&\le \frac{n}{a_n\epsilon}\frac{1+7\lambda_n+6\lambda_n^2+\lambda_n^3}{(1+\lambda_n)^3}\mathbb{E}\{p(\boldsymbol{\theta}_{nt})\|\mathbf{z}\|^3\} + \frac{n}{a_n\epsilon}\mathbb{E}[\{p(\boldsymbol{\theta}_{nt})\}^3\|\mathbf{z}\|^3]\\
&\le \frac{a_n}{\epsilon}\frac{1+7\lambda_n+6\lambda_n^2+\lambda_n^3}{(1+\lambda_n)^3}\mathbb{E}(e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}\|\mathbf{z}\|^3) + \frac{a_n e^{2\alpha_{nt}}}{\epsilon}\mathbb{E}(e^{3\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}\|\mathbf{z}\|^3) = o(a_n^2).
\end{aligned}
$$

Thus, applying the Lindeberg-Feller central limit theorem (Section *2.8 of van der Vaart, 1998) finishes the proof of (S.32).

Now we prove (S.33). Let

$$
\Delta_5 \equiv a_n^{-2}\sum_{i=1}^n \frac{\tau_i}{w_i}\phi_i(\boldsymbol{\theta}_{nt})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} = \frac{1}{n}\sum_{i=1}^n \frac{\tau_i}{w_i}\frac{e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{(1+e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}.
$$

Since

$$
\mathbb{E}(\Delta_5) = \mathbb{E}\bigg\{\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2}\mathbf{z}\mathbf{z}^{\mathrm{T}}\bigg\} = \mathbb{E}\big(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\big) + o(1),
$$

by the dominated convergence theorem, and each component of Δ_5 has a variance that is bounded by

$$
\frac{1}{n}\mathbb{E}\bigg\{\frac{2e^{2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^4}{(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^4}\bigg\} \le \frac{2\mathbb{E}(e^{2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^4)}{n} = o(1),
$$

applying Chebyshev's inequality gives that

$$
\Delta_5 \longrightarrow \mathbb{E}\big(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\big),
$$

in probability. Thus, (S.33) follows from the fact that

$$
\begin{aligned}
&\bigg|a_n^{-2}\sum_{i=1}^n \frac{\tau_i}{w_i}\phi_i(\boldsymbol{\theta}_{nt}+a_n^{-1}\acute{\mathbf{u}})\|\mathbf{z}_i\|^2 - a_n^{-2}\sum_{i=1}^n \frac{\tau_i}{w_i}\phi_i(\boldsymbol{\theta}_{nt})\|\mathbf{z}_i\|^2\bigg|\\
&\le \|a_n^{-1}\acute{\mathbf{u}}\|a_n^{-2}\sum_{i=1}^n \frac{\tau_i}{w_i}p_i(\boldsymbol{\theta}_{nt}+a_n^{-1}\breve{\mathbf{u}})\|\mathbf{z}_i\|^3\\
&\le \frac{\|a_n^{-1}\acute{\mathbf{u}}\|}{n}\sum_{i=1}^n \frac{\tau_i}{w_i}e^{(\|\boldsymbol{\beta}_t\|+\|\mathbf{u}\|)\|\mathbf{z}_i\|}\|\mathbf{z}_i\|^3 = o_P(1),
\end{aligned}
$$

where the last step is because $n^{-1}\sum_{i=1}^n \tau_i w_i^{-1}e^{(\|\boldsymbol{\beta}_t\|+\|\mathbf{u}\|)\|\mathbf{z}_i\|}\|\mathbf{z}_i\|^3$ has a bounded mean and a bounded variance and thus it is bounded in probability. □

S.6 Proof of Theorem 5

Proof of Theorem 5. The over-sampled estimator $\hat{\boldsymbol{\theta}}_{\text{over}}^{\text{ubc}}$ is the maximizer of

$$\Upsilon_{oc}(\boldsymbol{\theta}) = \frac{1}{1+\lambda_n}\sum_{i=1}^{n}\tau_i\big[(\boldsymbol{\theta}+\mathbf{b}_o)^{\mathrm{T}}\mathbf{z}_i y_i - \log\{1+e^{\mathbf{z}_i^{\mathrm{T}}(\boldsymbol{\theta}+\mathbf{b}_o)}\}\big]. \tag{S.34}$$

Thus, $\sqrt{a_n}(\hat{\boldsymbol{\theta}}_{\text{over}}^{\text{ubc}} - \boldsymbol{\theta}_{nt})$ is the maximizer of $\gamma_{oc}(\mathbf{u}) = \Upsilon_{oc}(\boldsymbol{\theta}_{nt} + a_n^{-1}\mathbf{u}) - \Upsilon_{oc}(\boldsymbol{\theta}_{nt})$. By Taylor's expansion,

$$\gamma_{oc}(\mathbf{u}) = \frac{1}{a_n}\mathbf{u}^{\mathrm{T}}\dot{\Upsilon}_{oc}(\boldsymbol{\theta}_{nt}) + \frac{1}{2a_n^2(1+\lambda_n)}\sum_{i=1}^{n}\tau_i\phi_i(\boldsymbol{\theta}_{nt}+\mathbf{b}_o+a_n^{-1}\acute{\mathbf{u}})(\mathbf{z}_i^{\mathrm{T}}\mathbf{u})^2, \tag{S.35}$$

where

$$\dot{\Upsilon}_{oc}(\boldsymbol{\theta}) = \frac{\partial\Upsilon_{oc}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{1}{1+\lambda_n}\sum_{i=1}^{n}\tau_i\{y_i - p_i(\alpha_{nt}+b_{o0},\boldsymbol{\beta}_t)\}\mathbf{z}_i$$

is the gradient of $\Upsilon_{oc}(\boldsymbol{\theta})$, and $\acute{\mathbf{u}}$ lies between $\mathbf{0}$ and $\mathbf{u}$.

Similarly to the proof of Theorem 1, we only need to show that

$$a_n^{-1}\dot{\Upsilon}_{oc}(\boldsymbol{\theta}_{nt}) \longrightarrow \mathbb{N}\bigg[\mathbf{0},\ \frac{(1+\lambda)^2+\lambda}{(1+\lambda)^2}\ \mathbb{E}\bigg\{\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{(1+c_o e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2}\mathbf{z}\mathbf{z}^{\mathrm{T}}\bigg\}\bigg], \tag{S.36}$$

in distribution, and for any $\mathbf{u}$,

$$\frac{1}{a_n^2(1+\lambda_n)}\sum_{i=1}^{n}\tau_i\phi_i(\boldsymbol{\theta}_{nt}+\mathbf{b}_o+a_n^{-1}\acute{\mathbf{u}})\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} \longrightarrow \mathbb{E}\bigg(\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{1+c_o e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\bigg), \tag{S.37}$$

in probability.

We prove (S.36) first. Let $\eta_{obi} = (1+\lambda_n)^{-1}\tau_i\{y_i - p_i(\alpha_{nt}+b_{o0},\boldsymbol{\beta}_t)\}\mathbf{z}_i$. We have that

$$\begin{aligned}(1+\lambda_n)\mathbb{E}(\eta_{obi}|\mathbf{z}_i) &= \mathbb{E}[(1+\lambda_n y_i)\{y_i - p_i(\alpha_{nt}+b_{o0},\boldsymbol{\beta}_t)\}|\mathbf{z}_i]\mathbf{z}_i\\ &= [p_i(\alpha_{nt},\boldsymbol{\beta}_t)(1+\lambda_n)\{1-p_i(\alpha_{nt}+b_{o0},\boldsymbol{\beta}_t)\}\\ &\quad - \{1-p_i(\alpha_{nt},\boldsymbol{\beta}_t)\}\{p_i(\alpha_{nt}+b_{o0},\boldsymbol{\beta}_t)\}]\mathbf{z}_i = 0,\end{aligned}$$

which implies that $\mathbb{E}(\eta_{obi}) = \mathbf{0}$. For the conditional variance

$$\begin{aligned}&(1+\lambda_n)^2\mathbb{V}(\eta_{obi}|\mathbf{z}_i)\\ &= \mathbb{E}[\{1+3\lambda_n y_i + \lambda_n^2 y_i\}\{y_i - p_i(\alpha_{nt}+b_{o0},\boldsymbol{\beta}_t)\}^2|\mathbf{z}_i]\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\\ &= \big[p_i(\alpha_{nt},\boldsymbol{\beta}_t)(1+3\lambda_n+\lambda_n^2)\{1-p_i(\alpha_{nt}+b_{o0},\boldsymbol{\beta}_t)\}^2\\ &\quad + \{1-p_i(\alpha_{nt},\boldsymbol{\beta}_t)\}\{p_i(\alpha_{nt}+b_{o0},\boldsymbol{\beta}_t)\}^2\big]\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\\ &= \frac{(1+3\lambda_n+\lambda_n^2)e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t} + e^{2(\alpha_{nt}+b_{o0}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t)}}{(1+e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})(1+e^{\alpha_{nt}+b_{o0}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\\ &= \frac{(1+3\lambda_n+\lambda_n^2)e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{(1+e^{\alpha_{nt}+b_{o0}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})^2}\frac{1+\frac{1+2\lambda_n+\lambda_n^2}{1+3\lambda_n+\lambda_n^2}e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{1+e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\end{aligned}$$

$$= \frac{(1+3\lambda_n+\lambda_n^2)e^{\alpha_{nt}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{(1+e^{\alpha_{nt}+b_{o0}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\{1+o_P(1)\}$$
$$= e^{\alpha_{nt}}(1+3\lambda_n+\lambda_n^2)\frac{e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{(1+c_o e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t})^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}\{1+o_P(1)\},$$

where the $o_P(1)$'s above are all bounded and the last step is because $(1+\lambda_n)e^{\alpha_{nt}} \to c_o$. Thus, by the dominated convergence theorem, $\mathbb{V}(\eta_{obi})$ satisfies that

$$\mathbb{V}(\eta_{obi}) = e^{\alpha_{nt}}\frac{(1+\lambda)^2+\lambda}{(1+\lambda)^2}\mathbb{E}\bigg\{\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{(1+c_o e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2}\bigg\}\{1+o(1)\}, \tag{S.38}$$

which indicates that

$$\frac{1}{a_n^2}\sum_{i=1}^n \mathbb{V}(\eta_{obi}) \longrightarrow \frac{(1+\lambda)^2+\lambda}{(1+\lambda)^2}\ \mathbb{E}\bigg\{\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{(1+c_o e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2}\mathbf{z}\mathbf{z}^{\mathrm{T}}\bigg\}. \tag{S.39}$$

Now we check the Lindeberg-Feller condition. Recall that $\tau = yv+1$, where $v \sim \mathbb{POI}(\lambda_n)$. We can show that $\mathbb{E}\{(1+v)^3\} < 2(1+\lambda_n)^3$. For any $\epsilon > 0$,

$$\begin{aligned}
&a_n\epsilon(1+\lambda_n)^3\sum_{i=1}^n \mathbb{E}\big\{\|\eta_{obi}\|^2 I(\|\eta_{obi}\| > a_n\epsilon)\big\} \le (1+\lambda_n)^3\sum_{i=1}^n \mathbb{E}(\|\eta_{obi}\|^3)\\
&= n\mathbb{E}\big[\|\tau^3\{y-p(\boldsymbol{\theta}_{nt}+\mathbf{b}_o)\}\mathbf{z}\|^3\big]\\
&= n\mathbb{E}\big[p(\boldsymbol{\theta}_{nt})(1+v)^3\|\{1-p(\boldsymbol{\theta}_{nt}+\mathbf{b}_o)\}\mathbf{z}\|^3\big] + n\mathbb{E}\big[\{1-p(\boldsymbol{\theta}_{nt})\}\|p(\boldsymbol{\theta}_{nt}+\mathbf{b}_o)\mathbf{z}\|^3\big]\\
&\le 2n(1+\lambda_n)^3\mathbb{E}\big\{p(\boldsymbol{\theta}_{nt})\|\mathbf{z}\|^3\big\} + n\mathbb{E}\big\{\|p(\boldsymbol{\theta}_{nt}+\mathbf{b}_o)\mathbf{z}\|^3\big\}\\
&\le 2n(1+\lambda_n)^3 e^{\alpha_{nt}}\mathbb{E}\big(e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^3\big) + n(1+\lambda_n)^3 e^{3\alpha_{nt}}\mathbb{E}\big(e^{3\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^3\big)\\
&= (1+\lambda_n)^3 O(a_n^2).
\end{aligned}$$

This indicates that $a_n^{-2}\sum_{i=1}^n \mathbb{E}\{\|\eta_{obi}\|^2 I(\|\eta_{obi}\| > a_n\epsilon)\} = o(1)$, and thus the Lindeberg-Feller condition holds. Applying the Lindeberg-Feller central limit theorem (Section *2.8 of van der Vaart, 1998) finishes the proof of (S.36).

No we prove (S.37). Let

$$\Delta_6 \equiv \frac{1}{a_n^2(1+\lambda_n)}\sum_{i=1}^n \tau_i\phi_i(\boldsymbol{\theta}_{nt}+\mathbf{b}_o)\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}} = \frac{1}{n}\sum_{i=1}^n \frac{(1+v_i y_i)e^{\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}}{\{1+e^{\alpha_{nt}+b_{o0}+\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}_t}\}^2}\mathbf{z}_i\mathbf{z}_i^{\mathrm{T}}. \tag{S.40}$$

Note that

$$\mathbb{E}(\Delta_6) = \mathbb{E}\bigg\{\frac{(1+\lambda_n y)e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{(1+e^{\alpha_{nt}+b_{o0}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})^2}\mathbf{z}\mathbf{z}^{\mathrm{T}}\bigg\} \tag{S.41}$$
$$= \mathbb{E}\bigg\{\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{(1+e^{\alpha_{nt}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})(1+e^{\alpha_{nt}+b_{o0}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}})}\mathbf{z}\mathbf{z}^{\mathrm{T}}\bigg\} \tag{S.42}$$
$$= \mathbb{E}\bigg(\frac{e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{1+c_o e^{\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}\mathbf{z}\mathbf{z}^{\mathrm{T}}\bigg) + o(1), \tag{S.43}$$

by the dominated convergence theorem, and the variance of each component of Δ_6 is bounded by

$$\begin{aligned}
&\frac{1}{n}\mathbb{E}\bigg[\frac{(1+vy)^2e^{2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{\{1+e^{\alpha_{nt}+b_{o0}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\}^4}\|\mathbf{z}\|^4\bigg]\\
&=\frac{1}{n}\mathbb{E}\bigg[\frac{\{1+(3\lambda_n+\lambda_n^2)p(\boldsymbol{\theta}_{nt})\}e^{2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}}{\{1+e^{\alpha_{nt}+b_{o0}+\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\}^4}\|\mathbf{z}\|^4\bigg]\\
&\le\frac{\mathbb{E}(e^{2\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^4)}{n}+\frac{e^{\alpha_{nt}}(3\lambda_n+\lambda_n^2)}{n}\mathbb{E}\big(e^{3\boldsymbol{\beta}_t^{\mathrm{T}}\mathbf{x}}\|\mathbf{z}\|^4\big)=o(1),
\end{aligned}$$

where the last step is because $n^{-1}e^{\alpha_{nt}}\lambda_n^2=(e^{\alpha_{nt}}\lambda_n)^2a_n^{-2}\to 0$ and both expectations are finite. Therefore, Chebyshev's inequality implies that $\Delta_6\to 0$ in probability. Thus, (S.37) follows from the fact that

$$\begin{aligned}
&\bigg|\frac{1}{a_n^2(1+\lambda_n)}\sum_{i=1}^{n}\tau_i\phi_i(\boldsymbol{\theta}_{nt}+\mathbf{b}_o+a_n^{-1}\acute{\mathbf{u}})\|\mathbf{z}_i\|^2-\frac{1}{a_n^2(1+\lambda_n)}\sum_{i=1}^{n}\tau_i\phi_i(\boldsymbol{\theta}_{nt}+\mathbf{b}_o)\|\mathbf{z}_i\|^2\bigg|\\
&\le\frac{\|a_n^{-1}\acute{\mathbf{u}}\|}{a_n^2(1+\lambda_n)}\sum_{i=1}^{n}\tau_ip_i(\boldsymbol{\theta}_{nt}+\mathbf{b}_o+a_n^{-1}\breve{\mathbf{u}})\|\mathbf{z}_i\|^3\\
&\le\frac{\|a_n^{-1}\acute{\mathbf{u}}\|}{n}\sum_{i=1}^{n}(1+v_iy_i)e^{(\|\boldsymbol{\beta}_t\|+\|\mathbf{u}\|)\|\mathbf{z}_i\|}\|\mathbf{z}_i\|^3=o_P(1),
\end{aligned}$$

where the last step is from the fact that $n^{-1}\sum_{i=1}^{n}(1+v_iy_i)e^{(\|\boldsymbol{\beta}_t\|+\|\mathbf{u}\|)\|\mathbf{z}_i\|}\|\mathbf{z}_i\|^3$ has a bounded mean and a bounded variance, and an application of Chebyshev's inequality. □

References

Hjort, N. L. and Pollard, D. Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*, 2011.

van der Vaart, A. *Asymptotic Statistics*. Cambridge University Press, London, 1998.

Wang, H. More efficient estimation for logistic regression with optimal subsamples. *Journal of Machine Learning Research*, 20(132):1–59, 2019.