Supplementary material for “Nearly optimal two-step Poisson sampling and empirical likelihood weighting estimation for M-estimation with big data”

Yan Fan*
School of Statistics and Information, Shanghai University of International Business and Economics, Shanghai, 201620, China
and
Yang Liu*†
School of Mathematical Sciences,
Soochow University, Suzhou, 215006, China
and
Yukun Liu†
KLATASDS-MOE, School of Statistics,
East China Normal University, Shanghai, 200062, China
and
Jing Qin
National Institute of Allergy and Infectious Diseases,
National Institutes of Health, Frederick, Maryland 20892, USA

Abstract

This supplementary material consists of three sections. In Section 1, we present the assumptions and conclusions in the main paper. Section 2 contains proofs for the corresponding results, namely Lemma 1 and Theorems 1 and 2. Section 3 contains additional simulation results on the performances of our sample size determination methods.

*The first two authors contribute equally to this paper.
†Corresponding author: liuyangecnu@163.com, ykliu@sfs.ecnu.edu.cn

1 Assumptions and conclusions in the main paper

Suppose that the big data consist of N observations $\{Z_i\}_{i=1}^N$, which are independent and identically distributed (i.i.d.) copies from a population Z with an unknown cumulative distribution function F. Parametric models indexed by a q-dimensional parameter θ are usually imposed to extract information from data. Let $\ell(z, \theta)$ be a user-specific convex loss function that quantifies the lack-of-fit of a parametric model indexed by a parameter θ based on an observation z. The average loss or risk function is $R(\theta) = \mathbb{E}\{\ell(Z, \theta)\} = \int \ell(z, \theta) dF(z)$. We define the parameter of interest θ_0 to be the risk minimizer (Shen *et al.*, 2021)

$$\theta_0 = \arg\min_{\theta} R(\theta). \tag{1}$$

This setup includes many common problems as special cases. When Z is a scalar, the true parameter value θ_0 is the mean or median of Z if $\ell(z, \theta) = (z - \theta)^2$ or $|z - \theta|$. When $Z = (Y, X^\top)^\top$, θ_0 may be the population-level regression coefficient in the generalized linear regression, least-squares regression, quantile regression, and expectile regression models under the specification of $\ell(z; \theta)$ given in Table 1.

Assumption 1. *The N random vectors (Z_i, D_{i1}, D_{i2}) $(i = 1, \ldots, N)$ are independent and identically distributed (i.i.d.) copies of $(Z, D_{(1)}, D_{(2)})$. Suppose that the distribution $F(z)$ of Z is nondegenerate, $\mathbb{E}(D_{(1)}|Z) = \mathbb{E}(D_{(1)}) = \alpha_{10}$, $\mathbb{E}(D_{(2)}|Z) = \pi(Z)$ and $\alpha_{20} = \mathbb{E}(D_{(2)}) = \mathbb{E}\{\pi(Z)\}$.*

Let $D = I(D_{(1)} + D_{(2)} > 0)$ and thus $\mathbb{E}(D) = 1 - \{1 - \mathbb{E}(D_{(1)})\}\{1 - \mathbb{E}(D_{(2)})\} = 1 - (1 - \alpha_{10})(1 - \alpha_{20})$. Under Assumption 1, given the datum Z, the conditional probability of being sampled is $\varphi(Z) = \mathbb{E}(D|Z) = 1 - (1 - \alpha_{10})\{1 - \pi(Z)\}$.

Table 1: Loss functions and matrices V under commonly-used regression models. Here $\ddot{a}(x)$ is the second derivative of $a(x)$ and the function $f(y \mid x)$ is the conditional density function of Y given $X = x$.

Regression model	$\ell(z;\theta)$	V
Generalized linear	$-yx^\top\theta + a(x^\top\theta) - \log\{b(y)\}$	$\mathbb{E}\left[XX^\top\ddot{a}(X^\top\theta_0\right]$
Poisson	$-yx^\top\theta + \exp(x^\top\theta) + \log(y!)$	$\mathbb{E}\{XX^\top\exp(X^\top\theta_0)\}$
Logistic	$-yx^\top\theta + \log\{1 + \exp(x^\top\theta)\}$	$\mathbb{E}\left[XX^\top\frac{\exp(x^\top\theta_0)}{\{1+\exp(x^\top\theta_0)\}^2}\right]$
Least squares	$(y - x^\top\theta)^2$	$\mathbb{E}(XX^\top)$
Quantile	$(y - x^\top\theta)\{\tau - I(y - x^\top\theta < 0)\}$	$\mathbb{E}\{XX^\top f(X^\top\theta_0 \mid X)\}$
Expectile	$(y - x^\top\theta)^2\lvert\tau - I(y - x^\top\theta < 0)\rvert$	$\mathbb{E}\{XX^\top\lvert\tau - I(Y \leq X^\top\theta_0)\rvert\}$

Assumption 2. *Suppose that $\ell(z,\theta)$ is a loss function and convex with respect to θ, and that $\ell(z,\theta_0+t) = \ell(z,\theta_0) + \dot{\ell}(z)^\top t + \xi(z,t)$ holds in a neighborhood of $t=0$, where $\dot{\ell}(z)$ satisfies $\mathbb{E}\{\dot{\ell}(Z)\} = 0$ and $B_{\dot{\ell}\dot{\ell}} = \mathbb{E}\{\dot{\ell}(Z)\dot{\ell}^\top(Z)/\varphi(Z)\}$ is finite, and $\xi(z,t)$ satisfies $\mathbb{E}\{\xi(Z,t)\} = (1/2)t^\top V t + o(\|t\|^2)$ and $\mathbb{E}\{\xi^2(Z,t)\} = o(\|t\|^2)$ for a positive definite matrix V as $\|t\| \to 0$.*

Lemma 1. *Suppose that Assumptions 1 and 2 are satisfied and that $\alpha_{10}, \alpha_{20} \in (0,1)$ are fixed quantities. As N goes to infinity, $\sqrt{N}(\hat{\theta}_{\rm IPW} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\rm IPW})$, where $\xrightarrow{d}$ stands for "converge in distribution to" and $\Sigma_{\rm IPW} = V^{-1}B_{\dot{\ell}\dot{\ell}}V^{-1}$.*

Theorem 1. *Suppose that Assumptions 1 and 2 hold, $B_{hh} = \mathbb{E}\{h_e(Z)h_e^\top(Z)/\varphi(Z)\}$ is positive definite and that $\alpha_{10}, \alpha_{20} \in (0,1)$ are fixed and known. As N goes to infinity,*

(a) $\hat{\theta}_{\rm ELW}$ is consistent to θ_0, and $\sqrt{N}(\hat{\theta}_{\rm ELW} - \theta_0) = -V^{-1} \cdot N^{1/2}\sum_{i=1}^N \hat{p}_i \dot{\ell}(Z_i) + o_p(1)$;

(b) $\sqrt{N}(\hat{\theta}_{\rm ELW} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\rm ELW})$ with $\Sigma_{\rm ELW} = V^{-1}(B_{\dot{\ell}\dot{\ell}} - B_{\dot{\ell}h}B_{hh}^{-1}B_{\dot{\ell}h}^\top)V^{-1}$, where $B_{\dot{\ell}h} = \mathbb{E}\{\dot{\ell}(Z)h_e^\top(Z)/\varphi(Z)\}$ and $B_{\dot{\ell}\dot{\ell}} = \mathbb{E}\{\dot{\ell}(Z)\dot{\ell}^\top(Z)/\varphi(Z)\}$;

(c) If the auxiliary information defined by $\sum_{i=1}^N p_i h(Z_i) = 0$ is ignored, then $\sqrt{N}(\hat{\theta}_{\rm ELW} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\rm ELW0})$, where $\Sigma_{\rm ELW0} = V^{-1}\{B_{\dot{\ell}\dot{\ell}} - (B_{\dot{\ell}1}B_{\dot{\ell}1}^\top)/(B_{11} - \alpha_0^{-1})\}V^{-1}$ and $B_{\dot{\ell}1} = \mathbb{E}\{\dot{\ell}(Z)/\varphi(Z)\}$.

Thus far, we have assumed that the overall sampling fraction of the big data is nonnegligible, i.e. $\alpha_0 \in (0,1)$. When the volume of the big data is huge, it is reasonable to assume that the sampling fraction may be negligible.

Assumption 3. *Suppose that $\pi(z)$ depends on N and is written as $\pi_N(z)$, there exist a positive sequence $\{b_N\}_{N=1}^\infty$, a positive function $0 < \pi_*(Z) \le 1$, and a positive constant α_{1*} such that $b_N \to \infty$, $b_N/N \to 0$, $b_N\pi_N(Z) \to \pi_*(Z)$, and $b_N\alpha_{10} \to \alpha_{1*}$ as $N \to \infty$.*

Under Assumption 3, we have $b_N\alpha_{20} = \mathbb{E}\{b_N\pi_N(Z)\} \to \alpha_{2*} = \mathbb{E}\{\pi_*(Z)\}$ as $N \to \infty$. Define $\alpha_0 = b_N\alpha_{10} + b_N\alpha_{20}$ and $\varphi(Z) = b_N\alpha_{10} + b_N\pi_N(Z)$. Then, α_0 and $\varphi(Z)$ converge to $\alpha_* = \alpha_{1*} + \alpha_{2*}$ and $\varphi_*(Z) = \alpha_{1*} + \pi_*(Z)$, respectively. Because α_{10} and the $\pi(Z_i)$ are prespecified, the log-likelihood under Assumption 3, up to a constant not depending on the unknown parameters p_i, is equal to $\sum_{i=1}^N D_i \log(p_i)$. Besides the constraints $p_i \geq 0$ and $\sum_{i=1}^N p_i = 1$, p_i's in this situation should satisfy $\sum_{i=1}^N p_i h_{e*}(Z) = 0$, where $h_{e*}(Z) = (\varphi_*(Z) - \alpha_*, h^\top(Z))^\top$.

The maximum empirical likelihood estimator of p_i is $\hat{p}_{i*} = n^{-1}\{1 + \lambda_*^\top h_{e*}(Z_i)\}$, where λ is the solution to

$$\frac{1}{n}\sum_{i=1}^{n} \frac{h_{e*}(Z_i)}{1 + \lambda_*^\top h_{e*}(Z_i)} = 0. \tag{2}$$

The empirical likelihood weighting (ELW) estimator of θ_0 is $\hat{\theta}_{\mathrm{ELW}} = \arg\min_\theta \sum_{i=1}^n \hat{p}_{i*}\ell(Z_i, \theta)$.

Theorem 2. *Suppose that Assumptions 1–3 hold, the distribution of Z is nondegenerate, and that $C_{hh*} = \mathbb{E}\{h_{e*}(Z)h_{e*}^\top(Z)/\varphi_*(Z)\}$ is positive definite. As N goes to infinity, $\sqrt{N/b_N}(\hat{\theta}_{\mathrm{ELW}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathrm{ELW}*})$ and $\sqrt{N/b_N}(\hat{\theta}_{\mathrm{IPW}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\mathrm{IPW}*})$, where $\Sigma_{\mathrm{ELW}*} = V^{-1}(C_{\dot{\ell}\dot{\ell}*} - C_{\dot{\ell}h*}C_{hh*}^{-1}C_{\dot{\ell}h*}^\top)V^{-1}$ and $\Sigma_{\mathrm{IPW}*} = V^{-1}C_{\dot{\ell}\dot{\ell}*}V^{-1}$ with $C_{\dot{\ell}h*} = \mathbb{E}\{\dot{\ell}(Z)h_{e*}^\top(Z)/\varphi_*(Z)\}$ and $C_{\dot{\ell}\dot{\ell}*} = \mathbb{E}\{\dot{\ell}(Z)\dot{\ell}^\top(Z)/\varphi_*(Z)\}$.*

2 Proofs of Lemma 1 and Theorems 1–2

Before presenting the proofs of Theorems, we give some important lemmas which can ease the burden of proofs. The Lemma S1 characterizes the asymptotic behavior of the minimizer of a convex loss function. Lemma S2 states that the proposed ELW method is well defined with probability approaching one. Lemmas S3 and S4 investigate the large-sample properties of the Lagrange multipliers.

Lemma S1 (Basic corollary of Hjort and Pollard (2011))**.** *Suppose $A_N(s)$ is convex, and can be represented as*

$$A_N(s) = \frac{1}{2}s^\top V s + U_N^\top s + C_N + r_N(s),$$

where V is symmetric and positive definite, U_N is stochastically bounded, C_N is arbitrary, and $r_N(s)$ goes to zero in probability for each s. Let $\alpha_N = \arg\min_s A_N(s)$ and

$$\beta_N = \arg\min_s \left(\frac{1}{2}s^\top V s + U_N^\top s + C_N\right) = -V^{-1}U_N.$$

Then $\alpha_N = \beta_N + o_p(1)$ as N goes to infinty. Furthermore, if $U_N \xrightarrow{d} U$, then $\alpha_N \xrightarrow{d} -V^{-1}U$.

Lemma S2. *For a function h, let $(\hat{p}_1, \ldots, \hat{p}_N)$ be the maximizer of $\sum_{i=1}^N D_i \log(p_i)$ under the constraints $p_i \geq 0, \sum_{i=1}^N p_i = 1$, and $\sum_{i=1}^N p_i h_e(Z_i) = 0$, where $h_e(Z) = (\varphi(Z) - \alpha_0,\ h^\top(Z))^\top$. If $\mathbb{V}\mathrm{ar}(h_e(Z)|D = 1)$ is positive definite, then* $\lim_{N\to\infty} P(\hat{p}_1, \ldots, \hat{p}_N$ *are well defined*$) = 1$.

Proof. It is clear that $\hat{p}_i = 0$ if $D_i = 0$. Let P_N denote the empirical probability measure based on Z_i's with $D_i = 1$, and Ω be the set of unit vector of the same dimension as $h_e(Z_i)$. It can be verified that

$$\begin{aligned} E_N &= \{\text{those } \hat{p}_i \text{ with } D_i = 1 \text{ are well defined}\} \\ &= \{\inf_{u\in\Omega} P_N(u^\top h_e(Z) > 0) > 0\}. \end{aligned}$$

By a generalization of the Glivenko-Cantelli theorem to uniform convergence over half spaces,

$$\sup_{u\in\Omega} |P_N(u^\top h_e(Z) > 0) - P(u^\top h_e(Z) > 0|D = 1)| \to 0 \quad a.s. \tag{3}$$

Under the condition that $\mathbb{V}\text{ar}(h_e(Z)|D = 1)$ is positive definite, Lemma 2 of Owen (1990) indicates that

$$\epsilon_0 = \inf_{u \in \Omega} P(u^\top h_e(Z) > 0 | D = 1) > 0.$$

This together with (3) implies

$$P(E_N) \geq P(\inf_{u \in \Omega} P_N(u^\top h_e(Z) > 0) \geq \epsilon_0/2) \to 1$$

as $N \to \infty$. This proves Lemma 2. $\square$

Lemma S3. *Suppose* $\alpha_0 \in (0, 1)$ *and* $\mathbb{E}\{h_e(Z)h_e^\top(Z)/\varphi(Z)\}$ *is positive definite. Then* $\hat{\lambda} - \lambda_0 = O_p(N^{-1/2})$*, where* $\lambda_0 = (\alpha_0^{-1}, 0, \ldots, 0)^\top$.

Proof. The Equation (5) defining $\hat{\lambda}$ is equivalent to

$$0 \quad = \quad \frac{1}{N} \sum_{i=1}^N \frac{\alpha_0 D_i}{\varphi(Z_i)} \cdot \frac{h_e(Z_i)}{1 + \frac{\alpha_0}{\varphi(Z_i)}(\hat{\lambda} - \lambda_0)^\top h_e(Z_i)}. \tag{4}$$

Write $\hat{\lambda} - \lambda_0 = \rho u$, where $\rho = \|\hat{\lambda} - \lambda_0\|$ and u is a unit vector. Multiplying both sides of the above equation by $u^\top$ from left gives

$$\begin{aligned} 0 \quad &= \quad \frac{1}{N} \sum_{i=1}^N \frac{\alpha_0 D_i}{\varphi(Z_i)} \cdot \frac{u^\top h_e(Z_i)}{1 + \alpha_0 \rho u^\top h_e(Z_i)/\varphi(Z_i)} \\ &= \quad \frac{1}{N} \sum_{i=1}^N \frac{\alpha_0 D_i}{\varphi(Z_i)} \cdot \left\{ u^\top h_e(Z_i) - \frac{(u^\top h_e(Z_i))^2/\varphi(Z_i)}{1 + \alpha_0 \rho u^\top h_e(Z_i)/\varphi(Z_i)} \alpha_0 \rho \right\}. \end{aligned}$$

This equation is further equivalent to

$$\left| \frac{1}{N} \sum_{i=1}^N \frac{D_i u^\top h_e(Z_i)}{\varphi(Z_i)} \right| \quad = \quad \frac{1}{N} \sum_{i=1}^N \frac{D_i}{\varphi(Z_i)} \frac{(u^\top h_e(Z_i))^2/\varphi(Z_i)}{1 + \alpha_0 \rho u^\top h_e(Z_i)/\varphi(Z_i)} \alpha_0 \rho.$$

Because $B_{hh} = \mathbb{E}\{h_e(Z)h_e^\top(Z)/\varphi(Z)\}$ is positive definite, it follows Lemma 3 of Owen (1990) and the Cauchy-Schwartz inequality that

$$\max_{1 \leq i \leq N} \{D_i \|h_e(Z_i)\|/\varphi(Z_i)\} = o_p(N^{1/2})$$

and $\xi_N = \max_{1\le i\le N}\{|u^\top D_i h_e(Z_i)|/\varphi(Z_i)\} = o_p(N^{1/2})$. Therefore

$$\left|\frac{1}{N}\sum_{i=1}^N \frac{D_i u^\top h_e(Z_i)}{\varphi(Z_i)}\right| \quad \ge \quad \frac{1}{1+\alpha_0\rho\xi_N}\cdot u^\top \frac{1}{N}\sum_{i=1}^N \frac{D_i}{\varphi(Z_i)}\frac{h_e(Z_i)h_e^\top(Z_i)}{\varphi(Z_i)} u\cdot \alpha_0\rho.$$

Because the left-hand side is equal to $O_p(N^{-1/2})$ and $\mathbb{E}[Dh_e(Z)h_e^\top(Z)/\{\varphi(Z)\}^2] = \mathbb{E}[h_e(Z)h_e^\top(Z)/\varphi(Z)]$ is positive definite, we conclude that $\rho = O_p(N^{-1/2})$.

□

Lemma S4. *Suppose* $\alpha_0 \in (0,1)$ *and* $B_{hh} = \mathbb{E}\{h_e(Z)h_e^\top(Z)/\varphi(Z)\}$ *is positive definite. Then*

$$\hat\lambda - \lambda_0 \quad = \quad \alpha_0^{-1}B_{hh}^{-1}\frac{1}{N}\sum_{i=1}^N \frac{D_i h_e(Z_i)}{\varphi(Z_i)} + o_p(N^{-1/2}).$$

Proof. Applying the second-order Taylor expansion to Equation (4) gives

$$0 \quad = \quad \frac{1}{N}\sum_{i=1}^N \frac{D_i h_e(Z_i)}{\varphi(Z_i)}\cdot\left\{1 - \frac{\alpha_0}{\varphi(Z_i)}(\hat\lambda-\lambda_0)^\top h_e(Z_i)\right\} + R_N,$$

where $R_N = o_p(\|\hat\lambda - \lambda_0\|)$. This lemma is proved by noting that

$$\frac{1}{N}\sum_{i=1}^N \frac{D_i h_e(Z_i)h_e^\top(Z_i)}{\{\pi(Z_i)\}^2} = B_{hh} + o_p(1).$$

□

2.1 Proof of Lemma 1

We prove the asymptotic normality of the IPW estimator using Lemma S1. Define

$$A_N(s) = \sum_{i=1}^N \frac{D_i}{\varphi(Z_i)}\left\{\ell(Z_i,\theta_0 + s/\sqrt{N}) - \ell(Z_i,\theta_0)\right\}.$$

Under Assumption 2, $A_N(s)$ is convex with respect to s, and its minimum is attained at $\sqrt{N}(\hat\theta_{\mathrm{IPW}} - \theta_0)$.

Using the representation of $\ell(z, \theta_0 + t)$, we have

$$A_N(s) = U_N^\top s + r_N(s) + \frac{1}{2} s^\top V s + r_{N,0}(s),$$

where

$$\begin{aligned} U_N &= N^{-1/2} \sum_{i=1}^{N} \frac{D_i}{\varphi(Z_i)} \dot{\ell}(Z_i), \\ r_N(s) &= \sum_{i=1}^{N} \left\{ \frac{D_i}{\varphi(Z_i)} \xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z_i, s/\sqrt{N}) \right\}, \\ r_{N,0}(s) &= N\mathbb{E}\xi(Z, s/\sqrt{N}) - \frac{1}{2} s^\top V s. \end{aligned}$$

As assumed, $r_{N,0}(s) = N \cdot o(\|s\|^2/N) \to 0$ for fixed s. For $r_N(s)$, because (D_i, Z_i)'s are i.i.d. as (D, Z) and $\mathbb{V}\text{ar}(T) = \mathbb{V}\text{ar}\{\mathbb{E}(T|Z)\} + \mathbb{E}\{\mathbb{V}\text{ar}(T|Z)\}$ for any variable T, we have

$$\begin{aligned} \mathbb{V}\text{ar}\{r_N(s)\} &= N\mathbb{V}\text{ar}\left\{ \frac{D}{\varphi(Z)} \xi(Z, s/\sqrt{N}) \right\} \\ &= N\mathbb{V}\text{ar}\left\{ \xi(Z, s/\sqrt{N}) \right\} + N\mathbb{E}\left\{ \frac{1 - \varphi(Z)}{\varphi(Z)} \xi^2(Z, s/\sqrt{N}) \right\} \\ &\leq N\mathbb{E}\left\{ \frac{\xi^2(Z, s/\sqrt{N})}{\varphi(Z)} \right\}, \end{aligned}$$

for each fixed s. Further it follows from $\varphi(Z) \geq \alpha_{10}$ that

$$\mathbb{V}\text{ar}\{r_N(s)\} \leq N\alpha_{10}^{-1}\mathbb{E}\left\{ \xi^2(Z, s/\sqrt{N}) \right\} = N \cdot o(N^{-1}) = o(1).$$

Applying the Markov's inequality, this together with $\mathbb{E}\{r_N(s)\} = 0$ implies $r_N(s) = o_p(1)$. By Lemma S1, we have

$$\sqrt{N}(\hat{\theta}_{\text{IPW}} - \theta_0) = -V^{-1}U_N + o_p(1).$$

Because $B_{\dot{\ell}\dot{\ell}} = \mathbb{E}\{\dot{\ell}(Z)\dot{\ell}^\top(Z)/\pi(Z)\}$ is finite, by the central limit theorem, we have $U_N \xrightarrow{d} \mathcal{N}(0, B_{\dot{\ell}\dot{\ell}})$. Accordingly $\sqrt{N}(\hat{\theta}_{\text{IPW}} - \theta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\text{IPW}})$. This completes the proof of Lemma 1.

2.2 Proof of Theorem 1

By Lemma S2, $\hat{p}_i$'s are all well defined with probability approaching one. Without loss of generality, we assume that they are well defined. When they are well defined,

$$\hat{p}_i = \frac{1}{n} \cdot \frac{D_i}{1 + \hat{\lambda}^\top h_e(Z_i)},$$

where $n = \sum_{i=1}^N D_i$ and $\hat{\lambda}$ satisfies

$$\sum_{i=1}^N \frac{D_i h_e(Z_i)}{1 + \hat{\lambda}^\top h_e(Z_i)} = 0. \tag{5}$$

Before giving the proofs, we begin by deriving an approximate of $\hat{p}_i$. Let $\delta_1 = n/N - \alpha_0$ and $\delta_2 = \hat{\lambda} - \lambda_0$. By the central limit theorem and Lemma S3, it follows that $\delta_1 = O_p(N^{-1/2})$ and $\delta_2 = O_p(N^{-1/2})$. Applying the second-order Taylor expansion to p_i with respect to (δ_1, δ_2) at zero, we have

$$\begin{aligned}
\hat{p}_i &= \frac{\alpha_0}{N(n/N)} \cdot \frac{D_i}{\varphi(Z_i)} \frac{1}{1 + \alpha_0(\hat{\lambda} - \lambda_0)^\top h_e(Z_i)/\varphi(Z_i)} \\
&= \frac{1}{N(1 + \delta_1/\alpha_0)} \cdot \frac{D_i}{\varphi(Z_i)} \frac{1}{1 + \alpha_0 \delta_2^\top h_e(Z_i)/\varphi(Z_i)} \\
&= \frac{1}{N}\left\{1 - \frac{\delta_1}{\alpha_0} + \frac{2\alpha_0\delta_1^2}{(\alpha_0 + \xi_1)^3}\right\} \times \frac{D_i}{\varphi(Z_i)} \times \left[1 - \frac{\alpha_0 \delta_2^\top h_e(Z_i)}{\varphi(Z_i)}\right. \\
&\qquad \left. + \frac{2}{\{1 + \alpha_0 \xi_2^\top h_e(Z_i)/\varphi(Z_i)\}^3}\left\{\frac{\alpha_0\delta_2^\top h_e(Z_i)}{\varphi(Z_i)}\right\}^2\right],
\end{aligned}$$

where the first equation uses the definition of $\lambda_0 = (\alpha_0^{-1}, 0, \ldots, 0)^\top$, ξ_1 lies between 0 and δ_1, and ξ_2 lies between 0 and δ_2. Further we have

$$\hat{p}_i = \frac{D_i}{N\varphi(Z_i)}\left\{1 - \frac{\delta_1}{\alpha_0} - \frac{\alpha_0\delta_2^\top h_e(Z_i)}{\varphi(Z_i)}\right\} + \delta_{Ni}, \tag{6}$$

where

$$\begin{aligned}
\delta_{Ni} &= \frac{D_i}{N}\frac{2\alpha_0\delta_1^2}{(\alpha_0 + \xi_1)^3}\left[\frac{1}{\varphi(Z_i)} - \frac{\alpha_0\delta_2^\top h_e(Z_i)}{\{\varphi(Z_i)\}^2} + \frac{2}{\{1 + \alpha_0\xi_2^\top h_e(Z_i)/\varphi(Z_i)\}^3}\frac{\{\alpha_0\delta_2^\top h_e(Z_i)\}^2}{\{\varphi(Z_i)\}^3}\right] \\
&\quad + \frac{D_i}{N}\left(1 - \frac{\delta_1}{\alpha_0}\right) \times \frac{2}{\{1 + \alpha_0\xi_2^\top h_e(Z_i)/\varphi(Z_i)\}^3}\frac{\{\alpha_0\delta_2^\top h_e(Z_i)\}^2}{\{\varphi(Z_i)\}^3} \\
&\quad + \frac{D_i}{N}\frac{\delta_1}{\alpha_0} \times \frac{\alpha_0\delta_2^\top h_e(Z_i)}{\{\varphi(Z_i)\}^2}.
\end{aligned}$$

2.2.1 Proof of Result (a) in Theorem 1

Consider the convex function

$$A_N(s) = N\sum_{i=1}^N \hat{p}_i\{\ell(Z_i, \theta_0 + s/\sqrt{N}) - \ell(Z_i, \theta_0)\}.$$

Clearly, it attains its minimal value at $s = \sqrt{N}(\hat{\theta}_{\text{ELW}} - \theta_0)$. Note that $N\mathbb{E}\xi(Z, s/\sqrt{N}) = s^\top Vs/2 + r_{N,0}(s)$, where $r_{N,0}(s) = No(\|s\|^2/N) \to 0$ for fixed s. Using the representation of $\ell(z, \theta)$ in Assumption 2, we have

$$\begin{aligned} A_N(s) &= N\sum_{i=1}^N \hat{p}_i\{\dot{\ell}(Z_i)^\top s/\sqrt{N} + \xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z_i, s/\sqrt{N})\} + N\mathbb{E}\{\xi(Z, s/\sqrt{N})\} \\ &= U_N^\top s + r_N(s) + \frac{1}{2}s^\top Vs + r_{N,0}(s), \end{aligned}$$

where

$$U_N = N^{1/2}\sum_{i=1}^N \hat{p}_i\dot{\ell}(Z_i) \quad \text{and} \quad r_N(s) = N\sum_{i=1}^N \hat{p}_i\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z_i, s/\sqrt{N})\}.$$

If we can prove that $r_N(s) = o_p(1)$ for each s, then the Result (a) of Theorem 1 follows by Lemma S1. In fact, using the approximate of $\hat{p}_i$ in (6) implies

$$\begin{aligned} r_N(s) &= N\sum_{i=1}^N \hat{p}_i\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})\} \\ &= \sum_{i=1}^N \frac{D_i}{\varphi(Z_i)}\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})\} \\ &\quad -\frac{\delta_1}{\alpha_0}\sum_{i=1}^N \frac{D_i}{\varphi(Z_i)}\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})\} \\ &\quad -\alpha_0\sum_{i=1}^N D_i\frac{\delta_2^\top h_e(Z_i)}{\{\varphi(Z_i)\}^2}\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})\} \\ &\quad +N\sum_{i=1}^N\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})\}\delta_{Ni} \\ &=: \sum_{j=1}^4 r_{Nj}(s). \end{aligned}$$

We shall prove that $r_{Nj}(s) = o_p(1)$ for $1 \le j \le 4$ and each fixed s.

In the proof of Lemma 1, we have shown that $\mathbb{E}(r_{N1}(s)) = 0$ and $\mathbb{V}\text{ar}(r_{N1}(s)) = N\mathbb{V}\text{ar}\{D\xi(Z, s/\sqrt{N})/\varphi(Z)\} = o(1)$ for fixed s, which implies that $r_{N1}(s) = o_p(1)$. It follows from $\delta_1 = O_p(N^{-1/2})$ that $r_{N2}(s) = o_p(1)r_{N1}(s) = o_p(1)$ for fixed s.

For $r_{N3}(s)$, by the Cauchy-Schwarz inequality, we have

$$\begin{aligned} |r_{N3}(s)| &= \alpha_0 \left| \sum_{i=1}^{N} D_i \frac{\delta_2^\top h_e(Z_i)}{\varphi(Z_i)} \cdot \frac{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})}{\varphi(Z_i)} D_i \right| \\ &\le \alpha_0 \sqrt{\sum_{i=1}^{N} \frac{D_i\{\delta_2^\top h_e(Z_i)\}^2}{\{\varphi(Z_i)\}^2} \times \sum_{i=1}^{N} \frac{D_i\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z_i, s/\sqrt{N})\}^2}{\{\varphi(Z_i)\}^2}}. \end{aligned}$$

Since $\varphi(Z) \ge \alpha_{10}$, it follows that

$$\begin{aligned} & \mathbb{E}\left[\sum_{i=1}^{N} \frac{D_i}{\{\varphi(Z_i)\}^2}\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z_i, s/\sqrt{N})\}^2\right] \\ &= N\mathbb{E}\left[\frac{\{\xi(Z, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})\}^2}{\varphi(Z)}\right] \\ &\le N\alpha_{10}^{-1}\mathbb{E}\left\{\xi^2(Z, s/\sqrt{N})\right\} = N \cdot o(N^{-1}) = o(1). \end{aligned}$$

Using the Markov's inequality, we get

$$\sum_{i=1}^{N} \frac{D_i}{(\varphi(Z_i))^2}\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z_i, s/\sqrt{N})\}^2 = o_p(1).$$

This together with

$$\sum_{i=1}^{N} \frac{D_i\{\delta_2^\top h_e(Z_i)\}^2}{\{\varphi(Z_i)\}^2} \le \|\delta_2\|^2 \cdot \sum_{i=1}^{N} \frac{D_i\|h_e(Z_i)\|^2}{\{\varphi(Z_i)\}^2} = O_p(1)$$

implies that $|r_{N3}(s)| = \sqrt{O_p(1) \times o_p(1)} = o_p(1)$.

For $r_{N4}(s)$, it can be verified that δ_{Ni} in (6) is asymptotically equal to

$$\delta_{Ni} = \frac{D_i}{N\varphi(Z_i)} \cdot \delta_{Ni1}\{1 + \delta_{Ni2}\},$$

where $\max_i \delta_{Ni2} = o_p(1)$ and

$$\delta_{Ni1} = \frac{\delta_1^2}{\alpha_0^2} + \frac{D_i\{\alpha_0 \delta_2^\top h_e(Z_i)\}^2}{\{\varphi(Z_i)\}^2}.$$

Then, using the Cauchy-Schwarz inequality implies

$$\begin{aligned} |r_{N4}(s)| &\le \left| \sum_{i=1}^N \{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})\} \frac{D_i}{\varphi(Z_i)} \cdot \delta_{Ni1}\{1 + \delta_{Ni2}\} \right| \\ &\le \{1 + o_p(1)\} \cdot \sqrt{\sum_{i=1}^N \frac{D_i\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})\}^2}{\{\varphi(Z_i)\}^2} \cdot \sum_{i=1}^N \delta_{Ni1}^2}. \end{aligned}$$

We have proved that $\sum_{i=1}^N D_i\{\xi(Z_i, s/\sqrt{N}) - \mathbb{E}\xi(Z, s/\sqrt{N})\}^2/\{\varphi(Z_i)\}^2 = o_p(1)$, and it can be shown that $\sum_{i=1}^N \delta_{Ni1}^2 = O_p(1)$. Therefore $r_{N4}(s) = o_p(1)$.

In summary, we prove that $r_N(s) = \sum_{j=1}^4 r_{Nj}(s) = o_p(1)$. This finishes the proof of Result (a) of Theorem 1.

2.3 Proof of Results (b) and (c) of Theorem 1

Result (a) indicates that $\sqrt{N}(\hat{\theta}_{\text{ELW}} - \theta_0) = -V^{-1} \cdot N^{1/2} \sum_{i=1}^N \hat{p}_i \dot{\ell}(Z_i) + o_p(1)$. To prove result (b), we need to prove the asymptotical normality of $U_N = N^{1/2} \sum_{i=1}^N \hat{p}_i \dot{\ell}(Z_i)$.

In the proof of Result (a), we have shown that

$$\hat{p}_i = \frac{D_i}{N\varphi(Z_i)} \left\{ 1 - \frac{\delta_1}{\alpha_0} - \frac{\alpha_0 \delta_2^\top h_e(Z_i)}{\varphi(Z_i)} + \tau_{Ni} \right\},$$

where

$$\tau_{Ni} = \left[\frac{\delta_1^2}{\alpha_0^2} + \frac{D_i\{\alpha_0 \delta_2^\top h_e(Z_i)\}^2}{\{\varphi(Z_i)\}^2} \right] \cdot \{1 + o_p(1)\}.$$

It then follows that

$$
\begin{aligned}
U_N &= N^{-1/2}\sum_{i=1}^{N}\frac{D_i}{\varphi(Z_i)}\dot{\ell}(Z_i) - N^{-1/2}\frac{\delta_1}{\alpha_0}\sum_{i=1}^{N}\frac{D_i}{\varphi(Z_i)}\dot{\ell}(Z_i) \\
&\quad -\alpha_0\delta_2^{\top}N^{-1/2}\sum_{i=1}^{N}\frac{D_i h_e(Z_i)}{\{\varphi(Z_i)\}^2}\dot{\ell}(Z_i) + N^{-1/2}\sum_{i=1}^{N}\frac{D_i\tau_{Ni}}{\varphi(Z_i)}\dot{\ell}(Z_i) \\
&= N^{-1/2}\sum_{i=1}^{N}\frac{D_i}{\varphi(Z_i)}\dot{\ell}(Z_i) - N^{-1/2}\times\alpha_0\sum_{i=1}^{N}\dot{\ell}(Z_i)\frac{D_i\delta_2^{\top}h_e(Z_i)}{\{\varphi(Z_i)\}^2} + o_p(1) \\
&= N^{-1/2}\sum_{i=1}^{N}\frac{D_i}{\varphi(Z_i)}\dot{\ell}(Z_i) - N^{-1/2}B_{\dot{\ell}h}B_{hh}^{-1}\sum_{i=1}^{N}\frac{D_i h_e(Z_i)}{\varphi(Z_i)} + o_p(1) \\
&= N^{-1/2}\sum_{i=1}^{N}\frac{D_i}{\varphi(Z_i)}\{\dot{\ell}(Z_i) - B_{\dot{\ell}h}B_{hh}^{-1}h_e(Z_i)\} + o_p(1),
\end{aligned}
$$

where $B_{\dot{\ell}h} = \mathbb{E}\{\dot{\ell}(Z)h_e^{\top}(Z)/\varphi(Z)\}$ and we have used the approximate of $\delta_2 = \hat{\lambda} - \lambda_0$ given in Lemma S4.

Thus $U_N \xrightarrow{d} \mathcal{N}(0, B_{\dot{\ell}\dot{\ell}} - B_{\dot{\ell}h}B_{hh}^{-1}B_{\dot{\ell}h}^{\top})$. This together with Result (a) implies

$$
\sqrt{N}(\hat{\theta}_{\text{ELW}} - \theta_0) = -V^{-1}U_N + o_p(1) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\text{ELW}}),
$$

where $\Sigma_{\text{ELW}} = V^{-1}\left(B_{\dot{\ell}\dot{\ell}} - B_{\dot{\ell}h}B_{hh}^{-1}B_{\dot{\ell}h}^{\top}\right)V^{-1}$. This proves Result (b).

Result (c) is a direct corollary of Result (b) by noticing that when the auxiliary information is ignored,

$$
\begin{aligned}
B_{\dot{\ell}h} &= \mathbb{E}\left[\frac{\dot{\ell}(Z)\{\varphi(Z)-\alpha_0\}}{\varphi(Z)}\right] = -\alpha_0 B_{\dot{\ell}1}, \\
B_{hh} &= \mathbb{E}\left[\frac{\{\varphi(Z)-\alpha_0\}^2}{\varphi(Z)}\right] = \alpha_0^2 B_{11} - \alpha_0,
\end{aligned}
$$

where $B_{\dot{\ell}1} = \mathbb{E}\{\dot{\ell}(Z)/\varphi(Z)\}$ and $B_{11} = \mathbb{E}\{1/\varphi(Z)\}$.

2.4 Proof of Theorem 2

The proof is similar to that of Theorem 1 and hence is omitted.

3 Additional simulation results

Table 2 presents the full-data-based maximum likelihood estimator of the regression coefficients of the Poisson regression models.

Table 2: Full-data-based maximum likelihood estimates of regression coefficients.

Data	Intercept	X_1	X_2	X_3	X_4	X_5	X_6
Bike sharing	5.136	0.014	0.352	-0.262	0.024		
Hospital length-of-stay	1.280	0.110	0.198	0.227	0.251	0.243	0.205

In Section 4 of the main paper, we have presented two sample size determination methods, M1 and M2, under the target precision requirements (R1) and (R2), respectively. Both the M1 and M2 methods depend on the proposed sampling and estimation strategy (ELW or ELWAI). When no auxiliary information is taken into consideration, we use ELW to denote the toolkit of the proposed ELW estimation method, together with the corresponding nearly optimal sampling plan and the corresponding required sample size determination method given a target precision requirement (R1 or R2). Let ELWAI denote the counterpart when auxiliary information (big-data sample mean) is taken into consideration. Given a precision requirement, we may wonder whether the ELW and ELWAI toolkits produce desirable estimates with the promised precision.

To this end, we fix the first step sample size to $n_{10} = 200$ and set the ideal second step sample size to be $n_{20} = N(n_0 - n_{10})/(N - n_{10})$ in the second step sampling, where n_0 is the root of equation (4.2) in the main paper under requirement (R1) or equation (4.3) in the main paper under requirement (R2) with $a = 5\%$ or coverage level 95%. 10 distinct values of C_0 in requirement (R1) and d_0 in requirement (R2) are considered such that n_{20} ranges from 300 to 1000. We generate subsamples from the two real datasets for each choice of C_0 or d_0. For a generic estimator $\breve{\theta}$, we calculate the ratios of the simulated MSEs to the target C_0 under requirement (R1) and the simulated coverage probabilities of $\{\theta : \|\breve{\theta} - \theta\| \le d_0\}$ under requirement (R2) based on 500 simulated subsamples. Figure 1 presents the plots of MSE-to-C_0 ratios versus C_0 and coverage probabilities versus d_0.

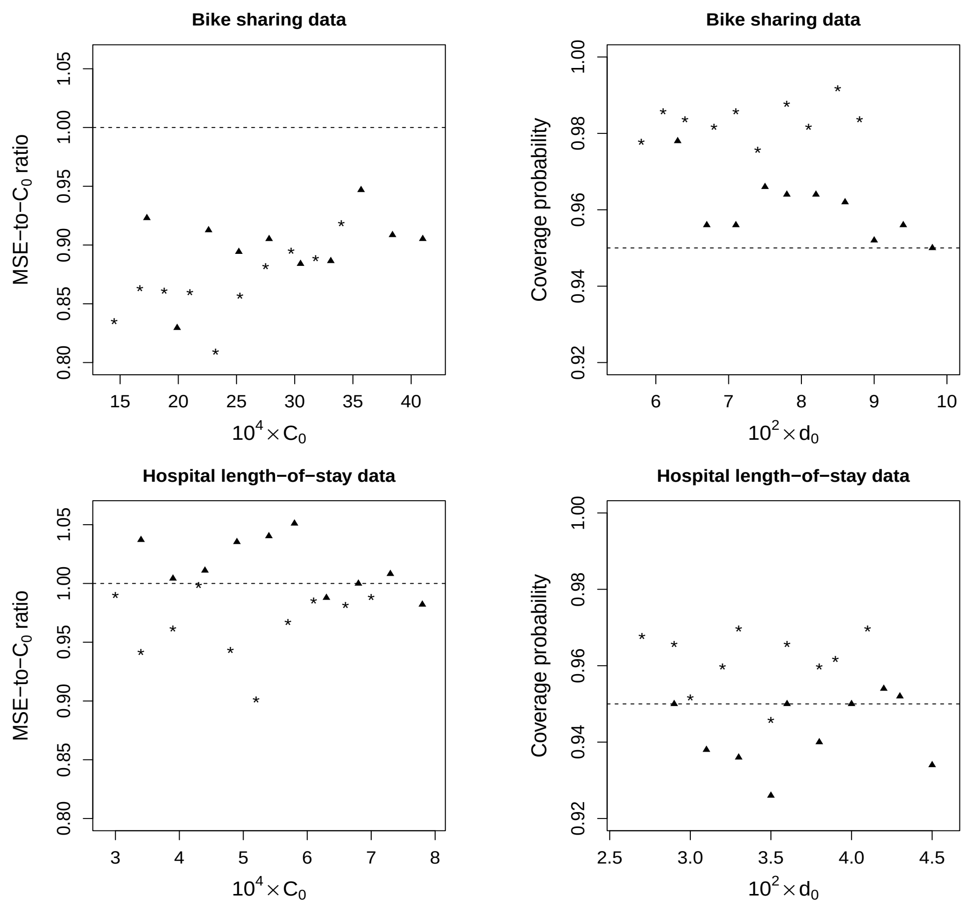

Figure 1: Simulated MSE-to-C_0 ratios under requirement (R1) and coverage probabilities under requirement (R2) with $a = 5\%$, where data were generated from the bike sharing data and hospital length-of-stay data. ELW: ▲; ELWAI: ⋆.

Under requirement (R1), the MSE-to-C_0 ratios of the ELW and ELWAI methods are all no greater than 1.05 for various choices of C_0 and both real data, and in particular, they are no greater than 0.95 for the bike sharing data. This indicates that the proposed sample size determination method M1 does fulfill its promise and gives desirable sample sizes such that the corresponding ELW or ELWAI estimator meets the target precision (R1). Under requirement (R2), the coverage probabilities of the ELW and ELWAI methods are all no less than the target level 95% for the bike sharing data and all choices of d_0 and are no less than 93% for the hospital length-of-stay data and most choice of d_0. This confirms that the proposed sample size determination method M2 also does fulfill its promise and gives desirable sample sizes such that the corresponding ELW or ELWAI estimator meets the target precision (R2). In the meantime, the ELWAI method has smaller MSE-to-C_0 ratios than the ELW method in most cases, which may shows again that the former is more efficient than the latter. This probably also explains the larger coverage probabilities of ELWAI than ELW. Overall, the ELW and ELWAI toolkits do produce desirable estimates with promised precision.

References

Hjort, N. L., and Pollard D. (2011). Asymptotics for minimisers of convex processes. ArXiv:1107.3806.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *The Annals of Statistics*, **18** (1), 90–120.

Shen, X., Chen, K., and Yu, W. (2021). Surprise sampling: Improving and extending the local case-control sampling. *Electronic Journal of Statistics*, **15**(1), 2454–2482.