A Additional Examples and Techniques Details

A.1 Examples of Data-Driven Optimization in Operations Research

We now give two canonical examples of stochastic optimization problems in operations research.

Example 2 (Multi-Product Newsvendor Problem)**.** The newsvendor problem has the objective function $c(\boldsymbol{w}, \boldsymbol{z}) = \boldsymbol{a}^\top (\boldsymbol{w} - \boldsymbol{z})^+ + \boldsymbol{d}^\top (\boldsymbol{z} - \boldsymbol{w})^+$, where for each $j \in [d_z]$: (1) $z^{(j)}$ is the customers' random demand of product j, ; (2) $w^{(j)}$ is the decision variable, the ordering quantity for product j; (3) $a^{(j)}$ is the holding cost for product j; (4) $d^{(j)}$ is the backlogging cost for product j and (5) the goal is to minimize the expected total cost.

Consider another classical problem in operations research, the portfolio optimization problem [Kallus and Mao, 2022, Grigas et al., 2021, Elmachtoub et al., 2023].

Example 3 (Portfolio Optimization)**.** Let $d_w = d_z + 1$ and denote the cost function as $c(\boldsymbol{w}, \boldsymbol{z}) = \gamma \left(\boldsymbol{w}^\top (\boldsymbol{z}, -1)\right)^2 + \exp\left(-\boldsymbol{w}^\top (\boldsymbol{z}, 0)\right)$. The decision $\boldsymbol{w}$ satisfies $\left(w^{(1)}, w^{(2)}, ..., w^{(d_w-1)}\right) \in \mathbb{R}^{d_w-1}$, denoting the investment fraction on products $1, 2, ...d_z$(i.e., $d_w - 1$) and $w^{(d_w)}$ is an auxiliary decision variable. The first component represents the risk (variance) of the portfolio and the second component represents the exponential utility of the portfolio.

A.2 Further Examples of Local Misspecification

Example 4 (Parametric Perturbation: Quadratic Mean Differentiability (QMD) family)**.** Suppose $P^n = P_{\boldsymbol{\theta}_0}^{\otimes n}$ for some fixed $\boldsymbol{\theta}_0$. Consider a sequence of vectors in $\mathbb{R}^{d_\theta}$, say $\{\boldsymbol{h}_n\}_{n=1}^\infty$. Suppose the joint distribution of $\{\boldsymbol{z}_i\}_{i=1}^n$, Q^n, is also in the parametric family, but is of the form $P_{\boldsymbol{\theta}_0+\boldsymbol{h}_n}^{\otimes n}$. If there exists a score function $\dot{\boldsymbol{\ell}}_{\boldsymbol{\theta}}(\boldsymbol{z}) : \mathcal{Z} \to \mathbb{R}^{d_\theta}$ with $\mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}(\boldsymbol{z})] = \mathbf{0}$ such that

$$\int \left(\sqrt{p_{\boldsymbol{\theta}_0+\boldsymbol{h}_n}} - \sqrt{p_{\boldsymbol{\theta}_0}} - \frac{1}{2}\dot{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}^\top \boldsymbol{h}_n \sqrt{p_{\boldsymbol{\theta}_0}}\right)^2 d\boldsymbol{z} = o(\|\boldsymbol{h}_n\|^2), \boldsymbol{h}_n \to \mathbf{0}.$$

In particular, in our framework, we focus on the case where $\boldsymbol{h}_n = \boldsymbol{h}/n^\alpha$ for a fixed vector $\boldsymbol{h}$. When $\alpha = 1/2$, van der Vaart [2000] shows that the likelihood ratio between Q^n and P^n satisfies:

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \boldsymbol{h}^\top \dot{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}(\boldsymbol{z}_i) - \frac{1}{2}\boldsymbol{h}^\top \boldsymbol{I} \boldsymbol{h} + o_{P^n}(\|\boldsymbol{h}\|^2),$$

where $\boldsymbol{I} := \mathbb{E}_{\boldsymbol{\theta}_0}[\dot{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0} \dot{\boldsymbol{\ell}}_{\boldsymbol{\theta}_0}^\top]$ denotes the Fisher information. In other words, under P^n

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} \xrightarrow{P^n} N(-\frac{1}{2}\boldsymbol{h}^\top \boldsymbol{I} \boldsymbol{h}, \boldsymbol{h}^\top \boldsymbol{I} \boldsymbol{h}),$$

where the limiting distribution is a Gaussian distribution with mean $-\frac{1}{2}\boldsymbol{h}^\top \boldsymbol{I} \boldsymbol{h}$ and variance $\boldsymbol{h}^\top \boldsymbol{I} \boldsymbol{h}$.

In the previous example, the ground truth distribution Q is still in $\{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ but is in the local neighbourhood of $P_{\boldsymbol{\theta}_0}$. The more common and interesting examples are when $Q \notin \{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ as discussed in examples below.

Example 5 (Semi-parametric Local Perturbation: Part I)**.** Suppose $P_{\boldsymbol{\theta}_0}$ is a given distribution, and $u(\boldsymbol{z}) : \mathcal{Z} \to \mathbb{R}$ is an unobserved random variable with $\mathbb{E}_{\boldsymbol{\theta}_0}[u] = 0$ and a finite variance $\mathbb{E}_{\boldsymbol{\theta}_0}[u^2]$. For a scalar t in a neighborhood of zero, we define the tilted distribution of $P_{\boldsymbol{\theta}_0}$, called Q_t, as

$$dQ_t(\boldsymbol{z}) = \frac{\exp(tu(\boldsymbol{z}))}{C_t} dP_{\boldsymbol{\theta}_0}(\boldsymbol{z})$$

where $C_t = \int \exp(tu(\boldsymbol{z}))dP_{\boldsymbol{\theta}_0}(\boldsymbol{z}) < \infty$ is a normalization constant. Clearly $Q_{t=0} = P_{\boldsymbol{\theta}_0}$.

Lemma 2 (Log Likelihood Ratio Property in Example 5)**.** *Under Definition 2, when $\alpha = 1/2$, i.e., $Q^n = Q_{1/\sqrt{n}}^{\otimes n}$, the log-likelihood ratio between Q^n and P^n satisfies:*

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n u(\boldsymbol{z}_i) - \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2] + o_{P^n}(1)\ .$$

This implies, under P^n, $\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} \xrightarrow{P^n} N(-\frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2], \mathbb{E}_{\boldsymbol{\theta}_0}[u^2])$.

Example 6 (Semi-parametric Local Perturbation: Part II). Consider the random variable $u(\boldsymbol{z}) : \mathcal{Z} \to \mathbb{R}$ with a zero mean, $\mathbb{E}_{\boldsymbol{\theta}_0}[u] = 0$, and finite second moment, say $\mathbb{E}_{\boldsymbol{\theta}_0}[u^2]$. Now we define the tilted distribution:

$$dQ_t(\boldsymbol{z}) = \frac{[1 + tu(\boldsymbol{z})]_+}{C_t} dP_{\boldsymbol{\theta}_0}(\boldsymbol{z}) \text{ where } C_t = \int [1 + tu(\boldsymbol{z})]_+ \, dP_{\boldsymbol{\theta}_0}(\boldsymbol{z}).$$

In particular, in our framework we focus on the case where $Q = Q_{1/n^\alpha}$ and $Q^n := Q_{1/n^\alpha}^{\otimes n}$. When $\alpha = 1/2$, by Duchi [2021], the log likelihood ratio satisfies

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n u(\boldsymbol{z}_i) - \frac{1}{2} \mathbb{E}_{\boldsymbol{\theta}_0}[u^2] + o_{P^n}(1).$$

In other words, under P^n,

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} \xrightarrow{P^n} N(-\frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2], \mathbb{E}_{\boldsymbol{\theta}_0}[u^2]).$$

Example 7 (Semi-parametric Local Perturbation: Part III). Consider the function $g : \mathbb{R} \to [-1, 1]$ be any three-times continuously differentiable function where $g(x) = x$ for $x \in [-1/2, 1/2]$ and $g' \geq 0$ and the first three derivatives of g are bounded. Consider the random variable $u(\boldsymbol{z}) : \mathcal{Z} \to \mathbb{R}$ with a zero mean $\mathbb{E}_{\boldsymbol{\theta}_0}[u] = \mathbf{0}$ and finite second moment, say $\mathbb{E}_{\boldsymbol{\theta}_0}[u^2]$. Now, for $\boldsymbol{t} \in \mathbb{R}$, we define the tilted distribution

$$dQ_t(\boldsymbol{z}) = \frac{1 + g(tu(\boldsymbol{z}))}{C_t} dP_{\boldsymbol{\theta}_0}(\boldsymbol{z}) \text{ where } C_t = 1 + \int g(tu(\boldsymbol{z}))dP_{\boldsymbol{\theta}_0}(\boldsymbol{z}).$$

In particular, in our framework we focus on the case where $Q = Q_{1/n^\alpha}$ and $Q^n := Q_{1/n^\alpha}^{\otimes n}$. When $\alpha = 1/2$, by Duchi and Ruan [2021], the log likelihood ratio satisfies the following Property:

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n u(\boldsymbol{z}_i) - \frac{1}{2} \mathbb{E}_{\boldsymbol{\theta}_0}[u^2] + o_{P^n}(1).$$

In other words, under P^n,

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} \xrightarrow{P^n} N(-\frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2], \mathbb{E}_{\boldsymbol{\theta}_0}[u^2]).$$

Example 8 (Semi-parametric Local Perturbation: Part IV (QMD Family)). Consider a scalar function $u(\boldsymbol{z}) : \mathcal{Z} \to \mathbb{R}$ with zero mean $\mathbb{E}_{\boldsymbol{\theta}_0}[u] = 0$ and finite second order moment $\mathbb{E}_{\boldsymbol{\theta}_0}[u^2]$. We define a tilted distribution Q_t for $t \in \mathbb{R}$ with probability density (mass) function q_t with respect to the dominated measure (note that Q_t is not necessarily in the parametric family $\{P_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$) with $q_{\mathbf{0}} = p_{\boldsymbol{\theta}_0}$. We further assume the quadratic mean differentiability

$$\int \left(\sqrt{q_t} - \sqrt{p_{\boldsymbol{\theta}_0}} - \frac{1}{2} tu\sqrt{p_{\boldsymbol{\theta}_0}} \right)^2 d\boldsymbol{z} = o(t^2).$$

Note that when $q_0 = p_{\boldsymbol{\theta}_0}$. In particular, in our framework we focus on the case where $Q = Q_{1/n^\alpha}$ and $Q^n := Q_{1/n^\alpha}^{\otimes n}$. When $\alpha = 1/2$, by Duchi [2021], we have

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n u(\boldsymbol{z}_i) - \frac{1}{2} \mathbb{E}_{\boldsymbol{\theta}_0}[u^2] + o_{P^n}(1).$$

Note that Example 8 is the most general version and includes Example 6-7 as particular examples under some mild assumptions.

A.3 Additional Technical Details

We introduce standard regularity assumptions for general M-estimation problems in asymptotic statistics [van der Vaart, 2000], which include our SAA, ETO, and IEO methods as examples.

Assumption 3 (Regularity Assumptions for M-estimation)**.** *Suppose the i.i.d. random variables $\{\boldsymbol{z}_i\}_{i=1}^n$ follows a distribution Q. Suppose the function $\boldsymbol{z} \to m_\zeta(\boldsymbol{z})$ is measurable with respect to $\boldsymbol{z}$ for all ζ and*

1. $\sup_\zeta \left|\frac{1}{n}\sum_{i=1}^n m_\zeta(\boldsymbol{z}_i) - \mathbb{E}_Q\left[m_\zeta(\boldsymbol{z})\right]\right| \xrightarrow{p} 0,$
2. *there exists* $\zeta^* = \operatorname{argmax}_\zeta \mathbb{E}_Q\left[m_\zeta(\boldsymbol{z})\right]$*, for all* $\varepsilon > 0$*,* $\sup_{\zeta:\|\zeta-\zeta^*\|} \mathbb{E}_Q\left[m_\zeta(\boldsymbol{z})\right] < \mathbb{E}_Q\left[m_{\zeta^*}(\boldsymbol{z})\right]$*,*
3. *the mapping $\zeta \to m_\zeta(\boldsymbol{z})$ is differentiable at ζ^* for Q-almost every $\boldsymbol{z}$ with derivative $\nabla_\zeta m_{\zeta^*}(\boldsymbol{z})$ and such that for every ζ_1 and ζ_2 in a neighbourhood of ζ^* and a measurable function K with $\mathbb{E}_Q[K(\boldsymbol{z})^2] < \infty$*
$$|m_{\zeta_1}(\boldsymbol{z}) - m_{\zeta_2}(\boldsymbol{z})| \leq K(\boldsymbol{z})\left\|\zeta_1 - \zeta_2\right\|.$$
4. *assume that the mapping $\zeta \to \mathbb{E}_Q[m_\zeta(\boldsymbol{z})]$ admits a second-order Taylor expansion at a point of maximum ζ^* with nonsigular symmetric second order matrix V_{ζ^*}.*

If the random sequence $\hat{\zeta}_n$ satisfies $\frac{1}{n}\sum_{i=1}^n m_{\hat{\zeta}_n}(\boldsymbol{z}_i) = \sup_\zeta \sum_{i=1}^n m_{\hat{\zeta}}(\boldsymbol{z}_i)$, then $\hat{\zeta}_n \xrightarrow{p} \zeta^*$ and

$$\sqrt{n}\left(\hat{\zeta}_n - \zeta^*\right) = -V_{\zeta^*}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \nabla_\zeta m_{\zeta^*}(\boldsymbol{z}_i) + o_Q(1).$$

Throughout this paper, we assume Assumption 3 holds.

- For SAA, consider $m_\zeta(\boldsymbol{z}) = -c(\boldsymbol{w}, \boldsymbol{z})$ with the parameter $\zeta = \boldsymbol{w}$.
- For ETO, consider $m_\zeta(\boldsymbol{z}) = \log p_{\boldsymbol{\theta}}(\boldsymbol{z})$ with parameter $\zeta = \boldsymbol{\theta}$.
- For IEO, consider $m_\zeta(\boldsymbol{z}) = c(\boldsymbol{w}_{\boldsymbol{\theta}}, \boldsymbol{z})$ with $\zeta = \boldsymbol{\theta}$.

When we say Assumption 3 holds, it means that Assumption 3 holds for the corresponding $m_\zeta(\boldsymbol{z})$ in SAA, ETO, and IEO.

Assumption 4 (Interchangeability)**.** *For any $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{w} \in \Omega$,*

$$\nabla_{\boldsymbol{\theta}} \int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}, \boldsymbol{z})^\top p_{\boldsymbol{\theta}}(\boldsymbol{z}) d\boldsymbol{z} = \int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}, \boldsymbol{z})^\top \nabla_{\boldsymbol{\theta}} p_{\boldsymbol{\theta}}(\boldsymbol{z}) d\boldsymbol{z},$$

$$\int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}, \boldsymbol{z}) p_{\boldsymbol{\theta}}(\boldsymbol{z}) d\boldsymbol{z}|_{\boldsymbol{w}=\boldsymbol{w}^*} = \nabla_{\boldsymbol{w}} \int c(\boldsymbol{w}, \boldsymbol{z}) p_{\boldsymbol{\theta}}(\boldsymbol{z}) d\boldsymbol{z}|_{\boldsymbol{w}=\boldsymbol{w}^*}$$

The interchangeability condition in Assumption 4 is a standard assumption in the Cramer-Rao bound [Bickel and Doksum, 2015]. A standard route to check the interchangeability condition is to use the dominated convergence theorem. For instance, we provide a way to check the first interchange equation. If $p_{\boldsymbol{\theta}}(\boldsymbol{z})$ is continuously differentiable with respect to $\boldsymbol{\theta}$, and there exists a real-valued function $q(\boldsymbol{z})$ such that $\int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}, \boldsymbol{z})^\top q(\boldsymbol{z}) d\boldsymbol{z} < +\infty$ and $\|\nabla_{\boldsymbol{\theta}} p_{\boldsymbol{\theta}}(\boldsymbol{z})\|_\infty \leq q(\boldsymbol{z})$, then we have $\nabla_{\boldsymbol{\theta}} \int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}, \boldsymbol{z})^\top p_{\boldsymbol{\theta}}(\boldsymbol{z}) d\boldsymbol{z} = \int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}, \boldsymbol{z})^\top \nabla_{\boldsymbol{\theta}} p_{\boldsymbol{\theta}}(\boldsymbol{z}) d\boldsymbol{z}$. Other sufficient conditions (more delicate but still based on the dominated convergence theorem) can be found in L'Ecuyer [1990], Asmussen and Glynn [2007], Glasserman [2004].

Next, we present some auxiliary lemmas that are helpful for deriving our theorems.

The first is a classic lemma in asymptotic statistics, called Le Cam's third lemma (Example 6.7 in van der Vaart [2000]).

Lemma 3 (Le Cam's third lemma)**.** *Let P^n and Q^n be sequences of probability measures on measurable spaces $(\Omega_n, \mathcal{F}_n)$ and let X_n be a sequence of random vectors. Suppose that*

$$\left(X_n, \log\frac{dQ^n}{dP^n}\right) \xrightarrow{P^n} N\left(\begin{pmatrix}\boldsymbol{\mu} \\ -\frac{1}{2}\sigma^2\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma} & \boldsymbol{\tau} \\ \boldsymbol{\tau}^\top & \sigma^2\end{pmatrix}\right),$$

then

$$X_n \xrightarrow{Q^n} N(\boldsymbol{\mu} + \boldsymbol{\tau}, \boldsymbol{\Sigma}).$$

We now state a auxiliary lemma about the directional differentiability of the optimal solutions to stochastic optimization problems.

Lemma 4 (Directional differentiability of optimal solutions: Part I)**.** *Consider the distribution $Q_t(\boldsymbol{z})$ in Definition 1. Let*

$$\boldsymbol{w}_t := \underset{\boldsymbol{w}\in\Omega}{\operatorname{argmin}}\, \mathbb{E}_{Q_t}\left[c(\boldsymbol{w}, \boldsymbol{z})\right].$$

Then

$$\lim_{t\to 0} \frac{1}{t}\left(\boldsymbol{w}_t - \boldsymbol{w}_0\right) = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})].$$

Equipped with the lemma above, we can get the convergence of $n^\alpha\left(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}\right)$ under the three locally misspecified regimes:

$$\lim_{n\to\infty} n^\alpha\left(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}\right) = \lim_{t\to 0}\frac{1}{t}\left(\boldsymbol{w}_t - \boldsymbol{w}_0\right) = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})].$$

Proof of Lemma 4. Note that $\boldsymbol{w}_n^*$ is the minimizer of $\mathbb{E}\left[c(\boldsymbol{w}, \boldsymbol{z})\right]$ under Q^n while $\boldsymbol{w}_{\boldsymbol{\theta}_0}$ under P^n. We will use the directional differentiablity of optimal solution to derive this fact.

We recall $Q_0 = P_{\boldsymbol{\theta}_0}$ with density $q_0(\boldsymbol{z}) = p_{\boldsymbol{\theta}_0}(\boldsymbol{z})$ and $dQ_t(\boldsymbol{z}) = dP_{\boldsymbol{\theta}_0}(\boldsymbol{z})\frac{\exp(tu(\boldsymbol{z}))}{C_t}$, $C_t = \int \exp(tu(\boldsymbol{z}))dP_{\boldsymbol{\theta}_0}(\boldsymbol{z})$. In this case, we denote $v(\boldsymbol{w}, Q_t)$ as $\mathbb{E}_{Q_t}[c(\boldsymbol{w}, \boldsymbol{z})]$, $\mathcal{G}(\boldsymbol{w}, t) := \nabla_{\boldsymbol{w}} v(\boldsymbol{w}, Q_t)$ and $\boldsymbol{w}_t := \operatorname{argmin} v(\boldsymbol{w}, Q_t)$. Note that $\mathcal{G}(\boldsymbol{w}_t, t) = 0$ for all t. By implicit function theorem,

$$\begin{aligned}
\lim_{t\to 0}\frac{1}{t}\left(\boldsymbol{w}_t - \boldsymbol{w}_0\right) &= -[\nabla_{\boldsymbol{w}}\mathcal{G}(\boldsymbol{w}_{\boldsymbol{\theta}_0}, 0)]^{-1}\frac{\partial}{\partial t}\nabla_{\boldsymbol{w}} v(\boldsymbol{w}_{\boldsymbol{\theta}_0}, Q_t)|_{t=0} \\
&= -\nabla_{\boldsymbol{w}\boldsymbol{w}}\mathbb{E}_{\boldsymbol{\theta}_0}[c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]\frac{\partial}{\partial t}\nabla_{\boldsymbol{w}} v(\boldsymbol{w}_{\boldsymbol{\theta}_0}, Q_t)|_{t=0} \\
&= -\boldsymbol{V}^{-1}\frac{\partial}{\partial t}\int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})dQ_t(\boldsymbol{z})|_{t=0} \\
&= -\boldsymbol{V}^{-1}\int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\frac{\partial}{\partial t}dQ_t(\boldsymbol{z})|_{t=0}.
\end{aligned}$$

From Definition 1, we know that for almost every $\boldsymbol{z}$, $\frac{\partial}{\partial t}\log q_t(\boldsymbol{z})|_{t=0} = u(\boldsymbol{z})$. Hence, we have for almost every $\boldsymbol{z}$,

$$\frac{\partial}{\partial t}q_t(\boldsymbol{z})\ \Big|_{t=0} = q_0(\boldsymbol{z})u(\boldsymbol{z}). \tag{3}$$

In conclusion,

$$\begin{aligned}
&-\boldsymbol{V}^{-1}\int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\frac{\partial}{\partial t}dQ_t(\boldsymbol{z})|_{t=0} \\
&= -\boldsymbol{V}^{-1}\int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\left\{q_0(\boldsymbol{z})u(\boldsymbol{z})\right\}d\boldsymbol{z}
\end{aligned}$$

Since

$$\int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\left[\int q_0(\boldsymbol{z})u(\boldsymbol{z})d\boldsymbol{z}\right]q_0(\boldsymbol{z})d\boldsymbol{z} = \left[\int q_0(\boldsymbol{z})u(\boldsymbol{z})d\boldsymbol{z}\right]\nabla_{\boldsymbol{w}}\mathbb{E}_{\boldsymbol{\theta}_0}[c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})] = 0$$

and

$$\int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})q_0(\boldsymbol{z})u(\boldsymbol{z})d\boldsymbol{z} = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})],$$

we have

$$-\boldsymbol{V}^{-1}\int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\frac{\partial}{\partial t}dQ_t(\boldsymbol{z})|_{t=0} = -\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})] = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})].$$

Therefore,

$$\lim_{n\to\infty}\sqrt{n}(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}) = \lim_{t\to 0}\frac{1}{t}\left(\boldsymbol{w}_t - \boldsymbol{w}_0\right) = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})].$$

More generally, under severely and mildly specified regime, we have further

$$\lim_{n\to\infty} n^\alpha(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}) = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})].$$

□

We note that Lemma 4 holds for Example 5. To be more specific,

$$q_t(\boldsymbol{z}) = \frac{\exp(tu(\boldsymbol{z}))}{C_t} q_0(\boldsymbol{z}) \text{ where } C_t = \int \exp(tu(\boldsymbol{z})) q_0(\boldsymbol{z}) d\boldsymbol{z}.$$

Therefore, the derivative of $q_t(\boldsymbol{z})$ with respect to t is

$$\begin{aligned} \frac{\partial}{\partial t} q_t(\boldsymbol{z}) = & \frac{q_0(\boldsymbol{z}) \exp(tu(\boldsymbol{z})) u(\boldsymbol{z}) \left[\int \exp(tu(\boldsymbol{z})) q_0(\boldsymbol{z}) d\boldsymbol{z}\right]}{\left[\int \exp(tu(\boldsymbol{z})) q_0(\boldsymbol{z}) d\boldsymbol{z}\right]^2} \\ & - \frac{\left[\int \exp(tu(\boldsymbol{z})) q_0(\boldsymbol{z}) u(\boldsymbol{z}) d\boldsymbol{z}\right] q_0(\boldsymbol{z}) \exp(tu(\boldsymbol{z}))}{\left[\int \exp(tu(\boldsymbol{z})) q_0(\boldsymbol{z}) d\boldsymbol{z}\right]^2}. \end{aligned}$$

At $t = 0$, since $\mathbb{E}_{\boldsymbol{\theta}_0}[u] = 0$, we have for almost every $\boldsymbol{z}$,

$$\frac{\partial}{\partial t} q_t(\boldsymbol{z}) \Big|_{t=0} = q_0(\boldsymbol{z}) u(\boldsymbol{z}) - \left[\int q_0(\boldsymbol{z}) u(\boldsymbol{z}) d\boldsymbol{z}\right] q_0(\boldsymbol{z}) = q_0(z) u(z).$$

It is also possible to extend the result to other examples under additional regularity assumptions.

For Example 6, the result still holds. Recall that

$$q_t(\boldsymbol{z}) = \frac{[1 + tu(\boldsymbol{z})]_+}{C_t} q_0(\boldsymbol{z}), \qquad C_t = \int [1 + tu(\boldsymbol{z})]_+ q_0(\boldsymbol{z}) d\boldsymbol{z}.$$

Hence,

$$\begin{aligned} \frac{\partial}{\partial t} q_t(\boldsymbol{z})|_{t=0} &= q_0(\boldsymbol{z}) \frac{C_t \frac{\partial}{\partial t} [1 + tu(\boldsymbol{z})]_+ - [1 + tu(\boldsymbol{z})]_+ \frac{\partial}{\partial t} C_t}{C_t^2} \Big|_{t=0} \\ &= q_0(\boldsymbol{z}) \left(\frac{\partial}{\partial t} [1 + tu(\boldsymbol{z})]_+ \Big|_{t=0} - \frac{\partial}{\partial t} C_t \Big|_{t=0} \right) \\ &= q_0(\boldsymbol{z}) \left(u(\boldsymbol{z}) \mathbb{1} \{1 + tu(\boldsymbol{z}) \geq 0\} \Big|_{t=0} - \int u(\boldsymbol{z}) \mathbb{1} \{1 + tu(\boldsymbol{z}) \geq 0\} \Big|_{t=0} q_0(\boldsymbol{z}) d\boldsymbol{z} \right) \\ &= q_0(\boldsymbol{z}) \left(u(\boldsymbol{z}) - \int u(\boldsymbol{z}) q_0(\boldsymbol{z}) d\boldsymbol{z} \right). \end{aligned}$$

The result is the same as (3) and the conclusion of Lemma 4 still holds.

For Example 7, the result still holds. Recall that

$$q_t(\boldsymbol{z}) = \frac{1 + g(tu(\boldsymbol{z}))}{C_t} q_0(\boldsymbol{z}), \qquad C_t = \int \left(1 + g(tu(\boldsymbol{z}))\right) q_0(\boldsymbol{z}) d\boldsymbol{z}.$$

Hence, by noting that $g'(0) = 1$,

$$\begin{aligned} \frac{\partial}{\partial t} q_t(\boldsymbol{z})|_{t=0} &= q_0(\boldsymbol{z}) \frac{C_t \frac{\partial}{\partial t} \left(1 + g(tu(\boldsymbol{z}))\right) - \left(1 + g(tu(\boldsymbol{z}))\right) \frac{\partial}{\partial t} C_t}{C_t^2} \Big|_{t=0} \\ &= q_0(\boldsymbol{z}) \left(\frac{\partial}{\partial t} \left(1 + g(tu(\boldsymbol{z}))\right) \Big|_{t=0} - \frac{\partial}{\partial t} C_t \Big|_{t=0} \right) \\ &= q_0(\boldsymbol{z}) \left(u(\boldsymbol{z}) g'(0) - \int u(\boldsymbol{z}) g'(0) q_0(\boldsymbol{z}) d\boldsymbol{z} \right) \\ &= q_0(\boldsymbol{z}) \left(u(\boldsymbol{z}) - \int u(\boldsymbol{z}) q_0(\boldsymbol{z}) d\boldsymbol{z} \right). \end{aligned}$$

The result is the same as (3) and the conclusion of Lemma 4 still holds.

We provide another auxiliary lemma similar to Lemma 4.

Lemma 5 (Directional differentiability of optimal solutions: Part II)**.** *Consider the distribution Q_t in Definition 1 where $Q_0 = P_{\boldsymbol{\theta}_0}$. We denote*

$$\begin{aligned} \boldsymbol{\theta}_t^{\mathrm{KL}} &:= \operatorname*{argmax}_{\boldsymbol{\theta} \in \Theta} \mathbb{E}_{Q_t}[\log p_{\boldsymbol{\theta}}(\boldsymbol{z})], \\ \boldsymbol{\theta}_t^* &:= \operatorname*{argmin}_{\boldsymbol{\theta} \in \Theta} \mathbb{E}_{Q_t}[c(\boldsymbol{w}_{\boldsymbol{\theta}}, \boldsymbol{z})]. \end{aligned}$$

Then we have

$$\nabla_t \boldsymbol{\theta}_t^{\mathrm{KL}} := \lim_{t\to 0} \frac{1}{t}\left(\boldsymbol{\theta}_t^{\mathrm{KL}} - \boldsymbol{\theta}_0\right) = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})],$$
$$\nabla_t \boldsymbol{\theta}_t^{*} := \lim_{t\to 0} \frac{1}{t}\left(\boldsymbol{\theta}_t^{*} - \boldsymbol{\theta}_0\right) = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})].$$

Proof of Lemma 5. We denote $v^{\mathrm{KL}}(\boldsymbol{\theta}, Q_t)$ as $\mathbb{E}_{Q_t}[\log p_{\boldsymbol{\theta}}(\boldsymbol{z})]$, $\mathcal{G}^{\mathrm{KL}}(\boldsymbol{\theta}, t) := \nabla_{\boldsymbol{\theta}} v^{\mathrm{KL}}(\boldsymbol{\theta}, Q_t)$ and $\boldsymbol{\theta}_t^{\mathrm{KL}} := \operatorname{argmin} v^{\mathrm{KL}}(\boldsymbol{\theta}, Q_t)$. Note that $\mathcal{G}^{\mathrm{KL}}(\boldsymbol{\theta}_t^{\mathrm{KL}}, t) = 0$ for all t. By implicit function theorem,

$$\begin{aligned}
\lim_{t\to 0} \frac{1}{t}\left(\boldsymbol{\theta}_t^{\mathrm{KL}} - \boldsymbol{\theta}_0\right) &= -[\nabla_{\boldsymbol{\theta}} \mathcal{G}^{\mathrm{KL}}(\boldsymbol{\theta}_0, 0)]^{-1} \frac{\partial}{\partial t} \nabla_{\boldsymbol{\theta}} v^{\mathrm{KL}}(\boldsymbol{\theta}_0, Q_t)|_{t=0} \\
&= -\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{\theta}_0}[\log p_{\boldsymbol{\theta}_0}(\boldsymbol{z})]^{-1} \frac{\partial}{\partial t} \nabla_{\boldsymbol{\theta}} v^{\mathrm{KL}}(\boldsymbol{\theta}_0, Q_t)|_{t=0} \\
&= \boldsymbol{I}^{-1} \frac{\partial}{\partial t} \int \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) dQ_t(\boldsymbol{z})|_{t=0} \\
&= \boldsymbol{I}^{-1} \int \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) \frac{\partial}{\partial t} dQ_t(\boldsymbol{z})|_{t=0}.
\end{aligned}$$

At $t = 0$, by (3),

$$\frac{\partial}{\partial t} q_t(\boldsymbol{z}) \Big|_{t=0} = q_0(\boldsymbol{z}) u(\boldsymbol{z}).$$

In conclusion,

$$\begin{aligned}
&\boldsymbol{I}^{-1} \int \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) \frac{\partial}{\partial t} dQ_t(\boldsymbol{z})|_{t=0} \\
=&\boldsymbol{I}^{-1} \int \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) \left\{q_0(\boldsymbol{z}) h^{\top} u(\boldsymbol{z})\right\} d\boldsymbol{z}.
\end{aligned}$$

Since

$$\int \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) \left[\int q_0(\boldsymbol{z}) u(\boldsymbol{z}) d\boldsymbol{z}\right] q_0(\boldsymbol{z}) d\boldsymbol{z} = \left[\int q_0(\boldsymbol{z}) u(\boldsymbol{z}) d\boldsymbol{z}\right] \mathbb{E}_{\boldsymbol{\theta}_0}\left[\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\right] = \mathbf{0}$$

and

$$\int \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) q_0(\boldsymbol{z}) u(\boldsymbol{z}) d\boldsymbol{z} = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z}) \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})],$$

we have

$$\lim_{t\to 0} \frac{1}{t}\left(\boldsymbol{\theta}_t^{\mathrm{KL}} - \boldsymbol{\theta}_0\right) = \boldsymbol{I}^{-1} \int \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) \frac{\partial}{\partial t} dQ_t(\boldsymbol{z})|_{t=0} = \boldsymbol{I}^{-1} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z}) \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})].$$

Similarly, we denote $v^*(\boldsymbol{\theta}, Q_t)$ as $\mathbb{E}_{Q_t}[c(\boldsymbol{w}_{\boldsymbol{\theta}}, \boldsymbol{z})]$, $\mathcal{G}^*(\boldsymbol{\theta}, t) := \nabla_{\boldsymbol{\theta}} v^*(\boldsymbol{\theta}, Q_t)$ and $\boldsymbol{\theta}_t^* := \operatorname{argmin} v^*(\boldsymbol{\theta}, Q_t)$. Note that $\mathcal{G}^*(\boldsymbol{\theta}_t^*, t) = 0$ for all t. By implicit function theorem,

$$\begin{aligned}
\lim_{t\to 0} \frac{1}{t}\left(\boldsymbol{\theta}_t^{*} - \boldsymbol{\theta}_0\right) &= -[\nabla_{\boldsymbol{\theta}} \mathcal{G}^*(\boldsymbol{\theta}_0, 0)]^{-1} \frac{\partial}{\partial t} \nabla_{\boldsymbol{\theta}} v^*(\boldsymbol{\theta}_0, Q_t)|_{t=0} \\
&= -\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \mathbb{E}_{\boldsymbol{\theta}_0}[c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]^{-1} \frac{\partial}{\partial t} \nabla_{\boldsymbol{\theta}} v^*(\boldsymbol{\theta}_0, Q_t)|_{t=0} \\
&= -\boldsymbol{\Phi}^{-1} \frac{\partial}{\partial t} \int \nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}) dQ_t(\boldsymbol{z})|_{t=0} \\
&= -\boldsymbol{\Phi}^{-1} \int \nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}) \frac{\partial}{\partial t} dQ_t(\boldsymbol{z})|_{t=0}.
\end{aligned}$$

At $t = 0$, by (3),

$$\frac{\partial}{\partial t} q_t(\boldsymbol{z}) \Big|_{t=0} = q_0(\boldsymbol{z}) u(\boldsymbol{z}).$$

In conclusion,

$$\begin{aligned}
&-\boldsymbol{\Phi}^{-1} \int \nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}) \frac{\partial}{\partial t} dQ_t(\boldsymbol{z})|_{t=0} \\
=&-\boldsymbol{\Phi}^{-1} \int \nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}) \left\{q_0(\boldsymbol{z}) u(\boldsymbol{z})\right\} d\boldsymbol{z}.
\end{aligned}$$

Since

$$\int \nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}) q_0(\boldsymbol{z}) u(\boldsymbol{z}) d\boldsymbol{z} = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z}) \nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})],$$

we have

$$\lim_{t \to 0} \frac{1}{t} (\boldsymbol{\theta}_t^* - \boldsymbol{\theta}_0) = -\boldsymbol{\Phi}^{-1} \int \nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}) \frac{\partial}{\partial t} dQ_t(\boldsymbol{z})|_{t=0} = -\boldsymbol{\Phi}^{-1} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z}) \nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})].$$

□

We remark that Lemma 5 also holds for Q_t in Example 6 and 7.

B Proofs

In this section, we supplement the proof of the results in this paper.

Proof of Theorem 6. We first notice the fact that, in the mildly misspecified regime, by defining $h_n = 1/(n^{\alpha - 1/2}) = o(1)$, we have

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h_n u(\boldsymbol{z}_i) - \frac{1}{2} \mathbb{E}_{\boldsymbol{\theta}_0}[u^2] h_n^2 + o_{P^n}(h_n) = o_{P^n}(1).$$

In the mild misspecified case, under P^n, we have a joint central limit theorem

$$\begin{bmatrix} \sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \\ \log \frac{dQ^n}{dP^n} \end{bmatrix} \xrightarrow{P^n} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z})) & 0 \\ 0 & 0 \end{bmatrix} \right).$$

Using LeCam's third lemma, we change the measure from P^n to Q^n and get that under Q^n,

$$\sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \xrightarrow{Q^n} N(0, \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z}))).$$

Using the same technique,

$$\begin{aligned} n^{\alpha}(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}) &\to \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z}) \mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})], \\ \sqrt{n}(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}) &\to \boldsymbol{0}. \end{aligned}$$

In conclusion,

$$\sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*) = \sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) - \sqrt{n}(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \xrightarrow{Q^n} N(0, \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z}))).$$

Let us now consider the regret. We use Taylor expansion of the regret with respect to $\boldsymbol{w}$ at $\boldsymbol{w}_n^*$ and note that $\nabla_{\boldsymbol{w}} v_n(\boldsymbol{w}_n^*) = 0$ for every n,

$$\begin{aligned} v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}_n^*) &= \frac{1}{2}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*)^\top \nabla_{\boldsymbol{w}\boldsymbol{w}} v_n(\boldsymbol{w}_n^*)(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*) + o_{Q^n}(\left\| \hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^* \right\|^2), \\ n(v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}_n^*)) &= \frac{1}{2} \sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*)^\top \nabla_{\boldsymbol{w}\boldsymbol{w}} v_n(\boldsymbol{w}_n^*) \sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*) + o_{Q^n}(1). \end{aligned}$$

By Assumption 2 that $\nabla_{\boldsymbol{w}\boldsymbol{w}} v_n(\boldsymbol{w}_n^*) \to \boldsymbol{V}$, the function $f : \Omega \to \mathbb{R}$ with $f(\cdot) := \frac{1}{2}(\cdot)^\top \boldsymbol{V}(\cdot)$ and function sequence $f_n : \Omega \to \mathbb{R}$ with $f_n(\cdot) := \frac{1}{2}(\cdot)^\top \nabla_{\boldsymbol{w}\boldsymbol{w}} v_n(\boldsymbol{w}_n^*)(\cdot)$ satisfy: for all sequence $\{\boldsymbol{w}_n\}_{n=1}^{\infty}$, if $\boldsymbol{w}_n \to \boldsymbol{w}$ for some $\boldsymbol{w} \in \Omega$, then $f_n(\boldsymbol{w}_n) \to f(\boldsymbol{w})$ since continuity is preserved under multiplication. Using the extended continuous mapping theorem (Theorem 1.11.1 in van der Vaart and Wellner [1996]), we have under Q^n,

$$n(v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}_n^*)) \xrightarrow{Q^n} \frac{1}{2} N^{\square} \boldsymbol{V} N^{\square}.$$

Moreover, ETO is stochastically dominated by IEO and IEO is stochastically dominated by SAA. □

Proof of Proposition 1. The asymptotic normality result is directly from van der Vaart [2000] by noting Lemma 1.

The asymptotic normality of SAA is by Proposition 2A of Elmachtoub et al. [2023]. For ETO and IEO, Proposition 2B and 2C of Elmachtoub et al. [2023] shows that

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{\text{ETO}} - \boldsymbol{\theta}_0\right) \xrightarrow{P^n} N(\mathbf{0}, \boldsymbol{I}^{-1}),$$
$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{\text{IEO}} - \boldsymbol{\theta}_0\right) \xrightarrow{P^n} N(\mathbf{0}, \boldsymbol{\Phi}^{-1}\operatorname{var}_{\boldsymbol{\theta}_0}\left(\nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\right)\boldsymbol{\Phi}^{-1}).$$

Regarding the notation, $\operatorname{var}_{\boldsymbol{\theta}_0}\left(\nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\right))$ is the variance of the random gradient $\nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}}, \boldsymbol{z})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, under the distribution $P_{\boldsymbol{\theta}_0}$. Note that the subscript $\boldsymbol{\theta}_0$ under the variance is not a variable here. Using the delta method, we have

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{\text{ETO}} - \boldsymbol{\theta}_0\right) \xrightarrow{P^n} N(\mathbf{0}, \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}_0}^\top \boldsymbol{I}^{-1} \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}_0}) = N(\mathbf{0}, \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{I}^{-1}\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}) = N(0, \operatorname{var}_{\boldsymbol{\theta}_0}(\text{IF}^{\text{ETO}}(\boldsymbol{z}))),$$
$$\begin{aligned}\sqrt{n}\left(\hat{\boldsymbol{\theta}}^{\text{IEO}} - \boldsymbol{\theta}_0\right) &\xrightarrow{P^n} N(\mathbf{0}, \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}_0}^\top \boldsymbol{\Phi}^{-1}\operatorname{var}_{\boldsymbol{\theta}_0}\left(\nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\right)\boldsymbol{\Phi}^{-1}\nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}_0}\right) \\ &= N(\mathbf{0}, \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\operatorname{var}_{\boldsymbol{\theta}_0}\left(\nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\right)\boldsymbol{\Phi}^{-1}\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\right) \\ &= N(0, \operatorname{var}_{\boldsymbol{\theta}_0}(\text{IF}^{\text{IEO}}(\boldsymbol{z}))).\end{aligned}$$

The inequality $\operatorname{var}_{\boldsymbol{\theta}_0}(\text{IF}^{\text{ETO}}(\boldsymbol{z})) \leq \operatorname{var}_{\boldsymbol{\theta}_0}(\text{IF}^{\text{IEO}}(\boldsymbol{z})) \leq \operatorname{var}_{\boldsymbol{\theta}_0}(\text{IF}^{\text{SAA}}(\boldsymbol{z}))$ is from Theorem 2 of Elmachtoub et al. [2023]. $\square$

Proof of Theorem 3. We use a different decomposition framework this time. We recall

$$\boldsymbol{\theta}_t^{\text{KL}} := \operatorname*{argmax}_{\boldsymbol{\theta}\in\Theta} \mathbb{E}_{Q_t}[\log p_{\boldsymbol{\theta}}(\boldsymbol{z})],$$
$$\boldsymbol{\theta}_t^* := \operatorname*{argmin}_{\boldsymbol{\theta}\in\Theta} \mathbb{E}_{Q_t}[c(\boldsymbol{w}_{\boldsymbol{\theta}}, \boldsymbol{z})],$$
$$\boldsymbol{w}_t^* := \operatorname*{argmin}_{\boldsymbol{w}\in\Omega} \mathbb{E}_{Q_t}[c(\boldsymbol{w}, \boldsymbol{z})].$$

We denote $t_n := 1/n^\alpha$. Note that here $\boldsymbol{w}_{t_n}^* = \boldsymbol{w}_n^*$ but generally $\boldsymbol{w}_{\boldsymbol{\theta}_{t_n}^{\text{KL}}} \neq \boldsymbol{w}_n^*$, $\boldsymbol{w}_{\boldsymbol{\theta}_{t_n}^*} \neq \boldsymbol{w}_n^*$. In this case,

$$\hat{\boldsymbol{w}}^{\text{ETO}} - \boldsymbol{w}_n^* = (\hat{\boldsymbol{w}}^{\text{ETO}} - \boldsymbol{w}_{\boldsymbol{\theta}_{t_n}^{\text{KL}}}) + (\boldsymbol{w}_{\boldsymbol{\theta}_{t_n}^{\text{KL}}} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) - (\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}),$$
$$\hat{\boldsymbol{w}}^{\text{IEO}} - \boldsymbol{w}_n^* = (\hat{\boldsymbol{w}}^{\text{IEO}} - \boldsymbol{w}_{\boldsymbol{\theta}_{t_n}^*}) + (\boldsymbol{w}_{\boldsymbol{\theta}_{t_n}^*} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) - (\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}),$$
$$\hat{\boldsymbol{w}}^{\text{SAA}} - \boldsymbol{w}_n^* = (\hat{\boldsymbol{w}}^{\text{SAA}} - \boldsymbol{w}_{t_n}^*) + (\boldsymbol{w}_{t_n}^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}) - (\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}).$$

We already show in Lemma 4 that

$$n^\alpha(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \to \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{SAA}}(\boldsymbol{z})].$$

Next we give a limit of the middle term, using Taylor expansion. For SAA, the middle term is equal to the third term. For ETO and IEO, $\boldsymbol{w}_{\boldsymbol{\theta}_0} = \boldsymbol{w}_{\boldsymbol{\theta}_t^*}|_{t=0}$ and $\boldsymbol{w}_{\boldsymbol{\theta}_0} = \boldsymbol{w}_{\boldsymbol{\theta}_t^{\text{KL}}}|_{t=0}$.

$$\boldsymbol{w}_{\boldsymbol{\theta}_t^*} - \boldsymbol{w}_{\boldsymbol{\theta}_0} := \nabla_t \boldsymbol{w}_{\boldsymbol{\theta}_t^*} + o(t) = \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}}^\top \nabla_t \boldsymbol{\theta}_t^* + o(t),$$
$$\boldsymbol{w}_{\boldsymbol{\theta}_t^{\text{KL}}} - \boldsymbol{w}_{\boldsymbol{\theta}_0} := \nabla_t \boldsymbol{w}_{\boldsymbol{\theta}_t^{\text{KL}}} + o(t) = \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}}^\top \nabla_t \boldsymbol{\theta}_t^{\text{KL}} + o(t).$$

By Lemma 5, we can get $\nabla_t \boldsymbol{\theta}_t^*$ and $\nabla_t \boldsymbol{\theta}_t^{\text{KL}}$ at $t = 0$:

$$\nabla_t \boldsymbol{\theta}_t^{\text{KL}} = \boldsymbol{I}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})],$$
$$\nabla_t \boldsymbol{\theta}_t^* = -\boldsymbol{\Phi}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})].$$

Moreover,

$$\nabla_t \boldsymbol{w}_{\boldsymbol{\theta}_t^{\text{KL}}} = \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}}^\top \nabla_t \boldsymbol{\theta}_t^{\text{KL}} = -\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{I}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}(u(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})) = \mathbb{E}_{\boldsymbol{\theta}_0}(u(\boldsymbol{z})\text{IF}^{\text{ETO}}(\boldsymbol{z})),$$
$$\nabla_t \boldsymbol{w}_{\boldsymbol{\theta}_t^*} = \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}}^\top \nabla_t \boldsymbol{\theta}_t^* = \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{\theta}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})] = \mathbb{E}_{\boldsymbol{\theta}_0}(u(\boldsymbol{z})\text{IF}^{\text{IEO}}(\boldsymbol{z})).$$

Finally, for the middle term,

$$n^{\alpha}(\boldsymbol{w}_{\boldsymbol{\theta}^*_{t_n}} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \to \mathbb{E}_{\boldsymbol{\theta}_0}(u(\boldsymbol{z})\text{IF}^{\text{IEO}}(\boldsymbol{z})),$$
$$n^{\alpha}(\boldsymbol{w}_{\boldsymbol{\theta}^{\text{KL}}_{t_n}} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \to \mathbb{E}_{\boldsymbol{\theta}_0}(u(\boldsymbol{z})\text{IF}^{\text{ETO}}(\boldsymbol{z})),$$
$$n^{\alpha}(\boldsymbol{w}^*_{t_n} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \to \mathbb{E}_{\boldsymbol{\theta}_0}(u(\boldsymbol{z})\text{IF}^{\text{SAA}}(\boldsymbol{z})).$$

For the first term, under Assumption 2 $(\hat{\boldsymbol{w}}^{\text{ETO}} - \boldsymbol{w}_{\boldsymbol{\theta}^{\text{KL}}_{t_n}})$, $(\hat{\boldsymbol{w}}^{\text{IEO}} - \boldsymbol{w}_{\boldsymbol{\theta}^*_{t_n}})$ and $(\hat{\boldsymbol{w}}^{\text{SAA}} - \boldsymbol{w}^*_{t_n})$ are all $O_{Q^n}(1/\sqrt{n})$, then

$$n^{\alpha}(\hat{\boldsymbol{w}}^{\text{ETO}} - \boldsymbol{w}_{\boldsymbol{\theta}^{\text{KL}}_{t_n}}) \xrightarrow{p} 0,$$
$$n^{\alpha}(\hat{\boldsymbol{w}}^{\text{IEO}} - \boldsymbol{w}_{\boldsymbol{\theta}^*_{t_n}}) \xrightarrow{p} 0,$$
$$n^{\alpha}(\hat{\boldsymbol{w}}^{\text{SAA}} - \boldsymbol{w}^*_{t_n}) \xrightarrow{p} 0.$$

The assumption is natural because it basically say the empirical part deviates from the expected part at the rate $O(1/\sqrt{n})$. When we multiply n^{α}, the term shrinks in probability to 0. In conclusion,

$$n^{\alpha}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}^*_n) \xrightarrow{p} \boldsymbol{b}^{\square}.$$

Let us now consider the regret. We use Taylor expansion of the regret with respect to $\boldsymbol{w}$ at $\boldsymbol{w}^*_n$ and note that $\nabla_{\boldsymbol{w}} v_n(\boldsymbol{w}^*_n) = 0$ for every n,

$$v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}^*_n) = \frac{1}{2}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}^*_n)^{\top}\nabla_{\boldsymbol{ww}} v_n(\boldsymbol{w}^*_n)(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}^*_n) + o_{Q^n}(\left\|\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}^*_n\right\|^2),$$
$$n^{2\alpha}(v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}^*_n)) = \frac{1}{2}n^{\alpha}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}^*_n)^{\top}\nabla_{\boldsymbol{ww}} v_n(\boldsymbol{w}^*_n)n^{\alpha}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}^*_n) + o_{Q^n}(1).$$

By Assumption 2 that $\nabla_{\boldsymbol{ww}} v_n(\boldsymbol{w}^*_n) \to \boldsymbol{V}$, the function $f : \Omega \to \mathbb{R}$ with $f(\cdot) := \frac{1}{2}(\cdot)^{\top}\boldsymbol{V}(\cdot)$ and function sequence $f_n : \Omega \to \mathbb{R}$ with $f_n(\cdot) := \frac{1}{2}(\cdot)^{\top}\nabla_{\boldsymbol{ww}} v_n(\boldsymbol{w}^*_n)(\cdot)$ satisfy: for all sequence $\{\boldsymbol{w}_n\}_{n=1}^{\infty}$, if $\boldsymbol{w}_n \to \boldsymbol{w}$ for some $\boldsymbol{w} \in \Omega$, then $f_n(\boldsymbol{w}_n) \to f(\boldsymbol{w})$ since continuity is preserved under mulptiplication. Using the extended continuous mapping theorem (Theorem 1.11.1 in van der Vaart and Wellner [1996]), we have under Q^n,

$$n^{2\alpha}(v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}^*_n)) \xrightarrow{p} \frac{1}{2}\left(\boldsymbol{b}^{\square}\right)^{\top}\boldsymbol{V}\boldsymbol{b}^{\square}.$$

□

Proof of Theorem 4. Recall the influence function of SAA, ETO, IEO:

$$\text{IF}^{\text{SAA}}(\boldsymbol{z}) = -\boldsymbol{V}^{-1}\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}),$$
$$\text{IF}^{\text{IEO}}(\boldsymbol{z}) = \boldsymbol{V}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}) = -\boldsymbol{V}^{-1}P_{\boldsymbol{\Sigma},\boldsymbol{V}}\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}),$$
$$\text{IF}^{\text{ETO}}(\boldsymbol{z}) = -\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{I}^{-1}\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) = -\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^{\top}]\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}).$$

For regret comparison, since $\boldsymbol{b}^{\text{SAA}} = \boldsymbol{0}$, we have $R^{\text{SAA}} = 0$. Also, $R^{\text{IEO}} \geq 0$ and $R^{\text{ETO}} \geq 0$.

By noting that $\boldsymbol{V} = \nabla_{\boldsymbol{ww}}\mathbb{E}_{\boldsymbol{\theta}_0}[c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]$, we observe that

$$\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{IEO}}(\boldsymbol{z})]^{\top}\boldsymbol{V}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{IEO}}(\boldsymbol{z})] = \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{IEO}}(\boldsymbol{z})]^{\top}\boldsymbol{V}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{SAA}}(\boldsymbol{z})].$$

This is because

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{IEO}}(\boldsymbol{z})]^{\top}\boldsymbol{V}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{IEO}}(\boldsymbol{z})] \\
=&\left(\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]^{\top}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\right)\cdot\boldsymbol{V}\cdot \\
&\left(\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]\right) \\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]^{\top}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})] \\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]^{\top}\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}^{\top}\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})] \\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{SAA}}(\boldsymbol{z})]^{\top}\boldsymbol{V}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{IEO}}(\boldsymbol{z})] \\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{IEO}}(\boldsymbol{z})]^{\top}\boldsymbol{V}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\text{IF}^{\text{SAA}}(\boldsymbol{z})].
\end{aligned}$$

Now let us prove $R^{\mathrm{ETO}} - R^{\mathrm{IEO}} \geq 0$.

$$
\begin{aligned}
&2R^{\mathrm{ETO}}\\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})(\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z}) - \mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z}))]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})(\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z}) - \mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z}))]\\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})]\\
&- 2\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})] + \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})]\\
&2R^{\mathrm{IEO}}\\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})(\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z}) - \mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z}))]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})(\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z}) - \mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z}))]\\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})]\\
&- 2\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})] + \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})]\\
=&- \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})] + \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})].
\end{aligned}
$$

Hence,

$$
\begin{aligned}
&2R^{\mathrm{ETO}} - 2R^{\mathrm{IEO}}\\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})]\\
&- 2\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})] + \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})]^\top \boldsymbol{V} \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})]\\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})]^\top \boldsymbol{I}^{-1}\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{I}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})]\\
&+ 2\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})]^\top \boldsymbol{I}^{-1}\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]\\
&+ \mathbb{E}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]^\top \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]\\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})]^\top \boldsymbol{I}^{-1}\left(\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)\boldsymbol{I}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})]\\
&+ 2\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})]^\top \boldsymbol{I}^{-1}\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]\\
&+ \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]^\top \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]\\
=&\left\|\left(\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)^{1/2}\boldsymbol{I}^{-1}\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})] - \left(\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)^{-1/2}\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]\right\|^2\\
\geq&0.
\end{aligned}
$$

The last equality is from the fact that

$$
\boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} - 2\boldsymbol{x}^\top \boldsymbol{y} + \boldsymbol{y}^\top \boldsymbol{A}^{-1}\boldsymbol{y} = \left(\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x} - \boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{y}\right)^\top \left(\boldsymbol{A}^{\frac{1}{2}}\boldsymbol{x} - \boldsymbol{A}^{-\frac{1}{2}}\boldsymbol{y}\right).
$$

In conclusion, we have

$$
R^{\mathrm{ETO}} \geq R^{\mathrm{IEO}} \geq R^{\mathrm{SAA}} = 0.
$$

By the definition of $\boldsymbol{b}^{\square}$ and $R^{\square}$, we know $\left\|\boldsymbol{b}^{\square}\right\|_{\boldsymbol{V}} = \sqrt{2R^{\square}}$. Hence, by the monotonicity of square root function, we have $0 = \left\|\boldsymbol{b}^{\mathrm{SAA}}\right\|_{\boldsymbol{V}} \leq \left\|\boldsymbol{b}^{\mathrm{IEO}}\right\|_{\boldsymbol{V}} \leq \left\|\boldsymbol{b}^{\mathrm{ETO}}\right\|_{\boldsymbol{V}}$. □

Proof of Theorem 5. Part (i): We note that when $u(\boldsymbol{z}) = \boldsymbol{\beta}^\top s_{\boldsymbol{\theta}_0}(\boldsymbol{z})$ for some $\boldsymbol{\beta} \in \mathbb{R}^{d_\theta}$,

$$
\begin{aligned}
&\boldsymbol{b}^{\mathrm{ETO}}\\
=&\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})] - \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})]\\
=&\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\left(\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\right)^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})] - \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]\\
=&\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\left(\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\right)^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{\beta}^\top \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})] - \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{\beta}^\top \boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})]\\
=&\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\left(\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\right)^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top \boldsymbol{\beta}] - \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top \boldsymbol{\beta}]\\
=&\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\left(\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\right)^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\boldsymbol{\beta} - \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\boldsymbol{\beta}\\
=&\boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\boldsymbol{\beta} - \boldsymbol{V}^{-1}\mathbb{E}_{\boldsymbol{\theta}_0}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top]\boldsymbol{\beta}\\
=&\boldsymbol{0}.
\end{aligned}
$$

□

To prove Theorem 1, we need a useful result here. When $\alpha = 1/2$, we can show that the log-likelihood ratio is asymptotically normal characterized by the mean and variance of the perturbation direction. This result is used to convert the asymptotics in P^n to Q^n by conducting a change of measure from P^n to Q^n, and also contributes to the overall asymptotically normal limit of the decision that encompasses the bias term. It will also be leveraged later to prove results in the mild misspecification case.

Lemma 6 (Log Likelihood Ratio Property in Definition 1[Duchi, 2021])**.** *Under Definition 2, when $\alpha = 1/2$, i.e., $Q^n = Q_{1/\sqrt{n}}^{\otimes n}$, the log-likelihood ratio between Q^n and P^n satisfies:*

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} u(\boldsymbol{z}_i) - \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2] + o_{P^n}(1).$$

Proof of Theorem 1. By Lemma 2 and Proposition 1, under P^n, we have a joint central limit theorem

$$\begin{bmatrix} \sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \\ \log \frac{dQ^n}{dP^n} \end{bmatrix} \xrightarrow{P^n} N\left(\begin{bmatrix} 0 \\ -\frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}_0}(u^2) \end{bmatrix}, \begin{bmatrix} \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z})) & \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\square}(\boldsymbol{z})] \\ \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\square}(\boldsymbol{z})^\top] & \mathbb{E}_{\boldsymbol{\theta}_0}(u^2) \end{bmatrix} \right).$$

Using LeCam's third lemma, we change the measure from P^n to Q^n and get that under Q^n,

$$\sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \xrightarrow{Q^n} N(\mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\square}(\boldsymbol{z})], \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z}))).$$

Next, by Lemma 4 (note that this is not a stochastic convergence but deterministic sequence convergence)

$$\sqrt{n}(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \to \mathbb{E}_{\boldsymbol{\theta}_0}[u(\boldsymbol{z})\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})].$$

In conclusion,

$$\sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*) = \sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_{\boldsymbol{\theta}_0}) - \sqrt{n}(\boldsymbol{w}_n^* - \boldsymbol{w}_{\boldsymbol{\theta}_0}) \xrightarrow{Q^n} N(\boldsymbol{b}^{\square}, \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z}))).$$

Let us now consider the regret. We use Taylor expansion of the regret with respect to $\boldsymbol{w}$ at $\boldsymbol{w}_n^*$ and note that $\nabla_{\boldsymbol{w}} v_n(\boldsymbol{w}_n^*) = 0$ for every n,

$$v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}_n^*) = \frac{1}{2}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*)^\top \nabla_{\boldsymbol{ww}} v_n(\boldsymbol{w}_n^*)(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*) + o_{Q^n}(\left\|\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*\right\|^2),$$
$$n(v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}_n^*)) = \frac{1}{2}\sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*)^\top \nabla_{\boldsymbol{ww}} v_n(\boldsymbol{w}_n^*)\sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*) + o_{Q^n}(1).$$

By Assumption 2 that $\nabla_{\boldsymbol{ww}} v_n(\boldsymbol{w}_n^*) \to \boldsymbol{V}$, the function $f : \Omega \to \mathbb{R}$ with $f(\cdot) := \frac{1}{2}(\cdot)^\top \boldsymbol{V}(\cdot)$ and function sequence $f_n : \Omega \to \mathbb{R}$ with $f_n(\cdot) := \frac{1}{2}(\cdot)^\top \nabla_{\boldsymbol{ww}} v_n(\boldsymbol{w}_n^*)(\cdot)$ satisfy: for all sequence $\{\boldsymbol{w}_n\}_{n=1}^{\infty}$, if $\boldsymbol{w}_n \to \boldsymbol{w}$ for some $\boldsymbol{w} \in \Omega$, then $f_n(\boldsymbol{w}_n) \to f(\boldsymbol{w})$ since continuity is preserved under mulptiplication. Using the extended continuous mapping theorem (Theorem 1.11.1 in van der Vaart and Wellner [1996]), we have under Q^n,

$$n(v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}_n^*)) \xrightarrow{Q^n} \frac{1}{2} N^{\square} \boldsymbol{V} N^{\square}.$$

$\square$

Proof of Theorem 2. Recall that:

$$\sqrt{n}(\hat{\boldsymbol{w}}^{\square} - \boldsymbol{w}_n^*) \xrightarrow{Q^n} N^{\square} := N(\boldsymbol{b}^{\square}, \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z}))).$$

$$n(v_n(\hat{\boldsymbol{w}}^{\square}) - v_n(\boldsymbol{w}_n^*)) \xrightarrow{Q^n} \mathbb{G}^{\square} := \frac{1}{2}(N^{\square})^\top \boldsymbol{V} N^{\square}.$$

By denoting $\boldsymbol{b}^{\square}$ as $\mathbb{E}_{\boldsymbol{\theta}_0}(u(\boldsymbol{z})(\mathrm{IF}^{\square}(\boldsymbol{z}) - \mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})))$, we can rewrite $\mathbb{G}^{\square}$ as

$$\begin{aligned} &\mathbb{G}^{\square} \\ =&\frac{1}{2}\left(N(0, \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z}))) - \boldsymbol{b}^{\square}\right)^\top \boldsymbol{V} \left(N(0, \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z}))) - \boldsymbol{b}^{\square}\right) \\ =&\frac{1}{2}\left[N(0, \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z})))^\top \boldsymbol{V} N(0, \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z}))) - 2\left(\boldsymbol{b}^{\square}\right)^\top \boldsymbol{V} N(0, \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z}))) + \left(\boldsymbol{b}^{\square}\right)^\top \boldsymbol{V} \boldsymbol{b}^{\square}\right]. \end{aligned}$$

By taking the expectation, the cross term is zero. Hence,

$$\mathbb{E}\left(\mathbb{G}^{\square}\right) = \frac{1}{2}\left[\mathbb{E}\left[N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z})))^\top \boldsymbol{V} N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\square}(\boldsymbol{z})))\right] + \left(\boldsymbol{b}^{\square}\right)^\top \boldsymbol{V}\boldsymbol{b}^{\square}\right].$$

Since $\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})) \leq \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})) \leq \mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z}))$, we know the stochastic dominance of the SAA, IEO and ETO, and their corresponding expectation.

$$\begin{aligned} & N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})))^\top \boldsymbol{V} N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z}))) \\ \preceq_{\mathrm{st}} & N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})))^\top \boldsymbol{V} N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z}))) \\ \preceq_{\mathrm{st}} & N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})))^\top \boldsymbol{V} N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z}))) \end{aligned}$$

and

$$\begin{aligned} & \mathbb{E}\left[N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})))^\top \boldsymbol{V} N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})))\right] \\ \leq & \mathbb{E}\left[N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})))^\top \boldsymbol{V} N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z})))\right] \\ \leq & \mathbb{E}\left[N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})))^\top \boldsymbol{V} N(0,\mathrm{var}_{\boldsymbol{\theta}_0}(\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})))\right]. \end{aligned}$$

From pervious analysis, we already know

$$\left(\boldsymbol{b}^{\mathrm{ETO}}\right)^\top \boldsymbol{V}\boldsymbol{b}^{\mathrm{ETO}} \geq \left(\boldsymbol{b}^{\mathrm{IEO}}\right)^\top \boldsymbol{V}\boldsymbol{b}^{\mathrm{IEO}} \geq \left(\boldsymbol{b}^{\mathrm{SAA}}\right)^\top \boldsymbol{V}\boldsymbol{b}^{\mathrm{SAA}}.$$

Therefore, $\mathbb{E}(\mathbb{G}^{\square})$ consist of two terms. For the first term, ETO is less than IEO, and IEO is less than SAA. For the second term, the direction is flipped. $\square$

Proposition 1 was essentially established by Elmachtoub et al. [2023], but here, we express the asymptotic behaviors of solutions more explicitly in terms of influence functions. Moreover, these more explicit expressions arise from a new projection interpretation of influence functions that allows us to describe the performances geometrically, providing another perspective different from Elmachtoub et al. [2023].

To this end, let $\boldsymbol{P}$ be the projection matrix onto the column span of $\boldsymbol{\Sigma}$ with respect to the norm $\|\boldsymbol{x}\|_{\boldsymbol{V}^{-1}}$, i.e.,

$$\boldsymbol{P}\boldsymbol{x} = \underset{\boldsymbol{y}:\boldsymbol{y}\in\mathrm{col}(\boldsymbol{\Sigma})}{\mathrm{argmin}} \|\boldsymbol{y}-\boldsymbol{x}\|^2_{\boldsymbol{V}^{-1}},$$

which has a closed-form expression $\boldsymbol{P} = \boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\right)^{-1}\boldsymbol{\Sigma}^\top\boldsymbol{V}^{-1}$. Second, define the functional $\mathcal{T} : L_2(P_{\boldsymbol{\theta}_0})^{d_w} \to L_2(P_{\boldsymbol{\theta}_0})^{d_w}$ as the projection operator onto the linear function subspace $\{\boldsymbol{A}\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) : \boldsymbol{A}\in\mathbb{R}^{d_w\times d_\theta}\}$, i.e., for a square integrable function $\boldsymbol{f}(\boldsymbol{z}) : \mathcal{Z}\to\mathbb{R}^{d_w}$,

$$\mathcal{T}\boldsymbol{f} = \underset{\boldsymbol{g}:\boldsymbol{g}=\boldsymbol{A}\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})}{\mathrm{argmin}} \int \|\boldsymbol{f}(\boldsymbol{z})-\boldsymbol{g}(\boldsymbol{z})\|^2 p_{\boldsymbol{\theta}_0}(\boldsymbol{z})d\boldsymbol{z}.$$

Theorem 7. *Under Assumptions 1, 3 and 4, the influence functions of* IEO *and* ETO *have the following projection interpretation.*

1. $\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z}) = -\boldsymbol{V}^{-1}\boldsymbol{P}\nabla_{\boldsymbol{w}}c(\boldsymbol{w}_{\boldsymbol{\theta}_0},\boldsymbol{z})$,
2. $\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z}) = -\boldsymbol{V}^{-1}\mathcal{T}\nabla_{\boldsymbol{w}}c(\boldsymbol{w}_{\boldsymbol{\theta}_0},\boldsymbol{z})$.

The above theorem points out that the influence functions of IEO and ETO are essentially projections of that of SAA, either in vector or function spaces, shedding light on the ordering of their variances by the contraction properties of projections.

Proof of Theorem 7. The fact that

$$\mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z}) = -\boldsymbol{V}^{-1}\boldsymbol{P}\nabla_{\boldsymbol{w}}c(\boldsymbol{w}_{\boldsymbol{\theta}_0},\boldsymbol{z})$$

is because

$$\begin{aligned} \mathrm{IF}^{\mathrm{IEO}}(\boldsymbol{z}) &= \boldsymbol{V}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Phi}^{-1}\nabla_{\boldsymbol{\theta}}c(\boldsymbol{w}_{\boldsymbol{\theta}_0},\boldsymbol{z}) \\ &= \boldsymbol{V}^{-1}\boldsymbol{\Sigma}(\nabla_{\boldsymbol{\theta}}\boldsymbol{w}_{\boldsymbol{\theta}_0}\boldsymbol{V}\nabla_{\boldsymbol{\theta}}\boldsymbol{w}_{\boldsymbol{\theta}_0}^\top)^{-1}\nabla_{\boldsymbol{\theta}}\boldsymbol{w}_{\boldsymbol{\theta}_0}\nabla_{\boldsymbol{w}}c(\boldsymbol{w}_{\boldsymbol{\theta}_0},\boldsymbol{z}) \\ &= \boldsymbol{V}^{-1}\boldsymbol{\Sigma}(\boldsymbol{\Sigma}^\top\boldsymbol{V}^{-1}\boldsymbol{V}\boldsymbol{V}^{-1}\boldsymbol{\Sigma})^{-1}\left(-\boldsymbol{\Sigma}^\top\boldsymbol{V}^{-1}\right)\nabla_{\boldsymbol{w}}c(\boldsymbol{w}_{\boldsymbol{\theta}_0},\boldsymbol{z}) \\ &= -\boldsymbol{V}^{-1}\boldsymbol{P}_{\boldsymbol{\Sigma},\boldsymbol{V}}\nabla_{\boldsymbol{w}}c(\boldsymbol{w}_{\boldsymbol{\theta}_0},\boldsymbol{z}). \end{aligned}$$

We then show the relationship between $\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z})$ and $\mathrm{IF}^{\mathrm{SAA}}(\boldsymbol{z})$. Let $\mathcal{T}: L_2(p_{\boldsymbol{\theta}_0})^{d_w} \to L_2(p_{\boldsymbol{\theta}_0})^{d_\theta}$ be the projection matrix on the linear function subspace $\{\boldsymbol{A}\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z}) : \boldsymbol{A} \in \mathbb{R}^{d_w \times d_\theta}\}$, i.e., for general function $\boldsymbol{f}(\boldsymbol{z}) : \mathcal{Z} \to \mathbb{R}^{d_\theta}$,

$$\mathcal{T}\boldsymbol{f} = \underset{\boldsymbol{g}:\boldsymbol{g}=\boldsymbol{A}\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})}{\operatorname{argmin}} \int \|\boldsymbol{f}(\boldsymbol{z}) - \boldsymbol{g}(\boldsymbol{z})\|^2 p_{\boldsymbol{\theta}_0}(\boldsymbol{z}) d\boldsymbol{z}.$$

The influence function of ETO is also a projection, i.e.,

$$\mathrm{IF}^{\mathrm{ETO}}(\boldsymbol{z}) = \boldsymbol{V}^{-1}\mathcal{T}\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}).$$

The reason is as follows. For ETO, it suffices to prove the following fact:

$$\mathcal{T}\boldsymbol{f} = \mathbb{E}_{\boldsymbol{\theta}_0}(\boldsymbol{f}\boldsymbol{s}_{\boldsymbol{\theta}_0}^\top)\boldsymbol{I}^{-1}\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})$$

since $\boldsymbol{\Sigma} = \mathbb{E}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})]$. To prove the fact, we need to show that $\boldsymbol{A}^* = \mathbb{E}_{\boldsymbol{\theta}_0}(\boldsymbol{f}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top)\boldsymbol{I}^{-1}$ is the minimizer of the optimization problem

$$\min_{\boldsymbol{A}\in\mathbb{R}^{d_w\times d_\theta}} \int \|\boldsymbol{f}(\boldsymbol{z}) - \boldsymbol{A}\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\|^2 p_{\boldsymbol{\theta}_0}(\boldsymbol{z}) d\boldsymbol{z}.$$

Since this is essentially a quadratic optimization problem, the stationary point is the global minimum. Denote the objective function $h(\boldsymbol{A})$ and we require $\nabla_{\boldsymbol{A}} h(\boldsymbol{A}^*) = 0$. In other words, for all $\tilde{i}, \tilde{j}$, $\partial h(\boldsymbol{A})/\partial A_{\tilde{i},\tilde{j}} = 0$. For simplicity, we write $p_{\boldsymbol{\theta}_0}(\boldsymbol{z})$ as $p(\boldsymbol{z})$ and $\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})$ as $s(\boldsymbol{z})$. Note that

$$\begin{aligned} h(\boldsymbol{A}) &= \int_{\boldsymbol{z}\in\mathcal{Z}} \sum_{i=1}^{d_w} (f_i(\boldsymbol{z}) - \sum_{j=1}^{d_\theta} A_{ij}s_j(\boldsymbol{z}))^2 p(\boldsymbol{z}) d\boldsymbol{z} \\ &= \sum_{i=1}^{d_w} \int_{\boldsymbol{z}\in\mathcal{Z}} \left[f_i(\boldsymbol{z})^2 + (\sum_{j=1}^{d_\theta} A_{ij}s_j(\boldsymbol{z}))^2 - 2f_i(\boldsymbol{z})\sum_{j=1}^{d_\theta} A_{ij}s_j(\boldsymbol{z}) \right] p(\boldsymbol{z}) d\boldsymbol{z} \end{aligned}$$

We have

$$\partial h(\boldsymbol{A}^*)/\partial A_{\tilde{i},\tilde{j}} = \int_{\boldsymbol{z}\in\mathcal{Z}} \left[-2f_{\tilde{i}}(\boldsymbol{z})s_{\tilde{j}}(\boldsymbol{z}) + \left[2\sum_{j=1}^{d_\theta} A_{\tilde{i}j}s_j(\boldsymbol{z}) \right] \right] p(\boldsymbol{z}) d\boldsymbol{z} = 0.$$

For all $\tilde{i}, \tilde{j}$, we have

$$\int_{\boldsymbol{z}\in\mathcal{Z}} f_{\tilde{i}}(\boldsymbol{z})s_{\tilde{j}}(\boldsymbol{z})p(\boldsymbol{z})d\boldsymbol{z} = \int_{\boldsymbol{z}\in\mathcal{Z}} \sum_{j=1}^{d_\theta} A_{\tilde{i}j}s_j(\boldsymbol{z})s_{\tilde{j}}(\boldsymbol{z})p(\boldsymbol{z})d\boldsymbol{z}.$$

Writing in a matrix form, the left hand side is $\mathbb{E}_{\boldsymbol{\theta}_0}(\boldsymbol{f}(\boldsymbol{z})s(\boldsymbol{z})^\top)$. The write hand side is $\mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{A}^* s(\boldsymbol{z})s(\boldsymbol{z})^\top] = \boldsymbol{A}^*\mathbb{E}_{\boldsymbol{\theta}_0}(s(\boldsymbol{z})s(\boldsymbol{z})^\top) = \boldsymbol{A}^*\boldsymbol{I}$. In conclusion, $\boldsymbol{A}^* = \mathbb{E}(\boldsymbol{f}(\boldsymbol{z})s(\boldsymbol{z})^\top)\boldsymbol{I}^{-1}$ and $\mathcal{T}\boldsymbol{f} = \mathbb{E}(\boldsymbol{f}(\boldsymbol{z})\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})^\top)\boldsymbol{I}^{-1}\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})$. □

Proof of Lemma 1. The first identity follows from

$$\begin{aligned} \boldsymbol{\Sigma}^\top &= \nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{w}} v(\boldsymbol{w}, \boldsymbol{\theta})|_{\boldsymbol{w}=\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \nabla_{\boldsymbol{\theta}}\mathbb{E}_{\boldsymbol{\theta}}[\nabla_{\boldsymbol{w}} c(\boldsymbol{w}, \boldsymbol{z})]|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0, \boldsymbol{w}=\boldsymbol{w}_{\boldsymbol{\theta}_0}} \\ &= \nabla_{\boldsymbol{\theta}} \int \nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}) p_{\boldsymbol{\theta}_0}(\boldsymbol{z}) d\boldsymbol{z} \\ &= \int \nabla_{\boldsymbol{\theta}} p_{\boldsymbol{\theta}_0}(\boldsymbol{z}) \nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})^\top d\boldsymbol{z} \\ &= \int (\nabla_{\boldsymbol{\theta}} \log p_{\boldsymbol{\theta}_0}(\boldsymbol{z}))\, p_{\boldsymbol{\theta}_0}(\boldsymbol{z})\, (\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}))^\top d\boldsymbol{z} \\ &= \mathbb{E}_{\boldsymbol{\theta}_0}[\boldsymbol{s}_{\boldsymbol{\theta}_0}(\boldsymbol{z})\, (\nabla_{\boldsymbol{w}} c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z}))^\top]. \end{aligned}$$

For the second identity, by implicit function theorem and applying Barratt [2018], we can prove the first identity

$$\begin{aligned} \mathbf{0} &= \nabla_{\boldsymbol{ww}} v(\boldsymbol{w}, \boldsymbol{\theta}_0)|_{\boldsymbol{w}=\boldsymbol{w}_{\boldsymbol{\theta}_0}} \left(\nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\right)^\top + \nabla_{\boldsymbol{w}} \nabla_{\boldsymbol{\theta}} v(\boldsymbol{w}, \boldsymbol{\theta})|_{\boldsymbol{w}=\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0}, \\ \Rightarrow \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} &= -\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{w}} v(\boldsymbol{w}, \boldsymbol{\theta})|_{\boldsymbol{w}=\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{\theta}=\boldsymbol{\theta}_0} \cdot \nabla_{\boldsymbol{ww}} v(\boldsymbol{w}, \boldsymbol{\theta}_0)^{-1}|_{\boldsymbol{w}=\boldsymbol{w}_{\boldsymbol{\theta}_0}}, \\ &= -\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}. \end{aligned}$$

The third identity follows since

$$\begin{aligned} \boldsymbol{\Phi} &= \nabla_{\boldsymbol{\theta\theta}} \mathbb{E}_{\boldsymbol{\theta}_0} \left[c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\right] \\ &= \nabla_{\boldsymbol{\theta}} \left(\nabla_{\boldsymbol{w}} \mathbb{E}_{\boldsymbol{\theta}_0} \left[c(\boldsymbol{w}_{\boldsymbol{\theta}}, \boldsymbol{z})\right] \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}}^\top\right)|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}} \nabla_{\boldsymbol{ww}} \mathbb{E}_{\boldsymbol{\theta}_0} \left[c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\right] \nabla_{\boldsymbol{\theta}} \boldsymbol{w}_{\boldsymbol{\theta}}^\top \\ &= \left(-\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\right) \boldsymbol{V} \left(-\boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1}\right)^\top \\ &= \boldsymbol{\Sigma}^\top \boldsymbol{V}^{-1} \boldsymbol{\Sigma} \end{aligned}$$

by noting that $\nabla_{\boldsymbol{w}} \mathbb{E}_{\boldsymbol{\theta}_0} \left[c(\boldsymbol{w}_{\boldsymbol{\theta}_0}, \boldsymbol{z})\right] = \mathbf{0}$ since $\boldsymbol{w}_{\boldsymbol{\theta}_0}$ is the minimizer of the function $\boldsymbol{w} \to \mathbb{E}_{\boldsymbol{\theta}_0} \left[c(\boldsymbol{w}, \boldsymbol{z})\right]$.

□

Proof of Lemma 2.

$$\log \frac{dQ^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)}{dP^n(\boldsymbol{z}_1, ..., \boldsymbol{z}_n)} = \log \prod_{i=1}^n \frac{\exp(u(\boldsymbol{z}_i)/\sqrt{n})}{C_{1/\sqrt{n}}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n u(\boldsymbol{z}_i) - n \log C_{1/\sqrt{n}}.$$

It now suffices to show that $n \log C_{1/\sqrt{n}} = \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2] + o_{P^n}(1)$. From the definition of C_t, we know

$$C_t = \int \exp(tu(\boldsymbol{z})) dP_{\boldsymbol{\theta}_0}(\boldsymbol{z}).$$

Taking the derivative, we have

$$\left(C_t\right)'|_{t=0} = \int \exp(tu(\boldsymbol{z})) u(\boldsymbol{z}) dP_{\boldsymbol{\theta}_0}(\boldsymbol{z})|_{t=0} = \mathbb{E}_{\boldsymbol{\theta}_0}(u(\boldsymbol{z})) = 0.$$

Taking the second order derivative, we have

$$\left(C_t\right)''|_{t=0} = \int \exp(tu(\boldsymbol{z})) u(\boldsymbol{z}) u(\boldsymbol{z}) dP_{\boldsymbol{\theta}_0}(\boldsymbol{z})|_{t=0} = \mathbb{E}_{\boldsymbol{\theta}_0}(u^2).$$

By Talor expansion, we have

$$C_t = 1 + \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2] t^2 + o(t^2)$$

In conclusion,

$$\begin{aligned} n \log C_{1/\sqrt{n}} &= n \log \left(1 + \frac{1}{2}\frac{1}{\sqrt{n}}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2]\frac{1}{\sqrt{n}} + o\left(\frac{1}{n}\right)\right) \quad = n \left(\frac{1}{2}\frac{1}{\sqrt{n}}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2]\frac{1}{\sqrt{n}} + o\left(\frac{1}{n}\right)\right) \\ &= \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}_0}[u^2] + o(1). \end{aligned}$$

□