EFFICIENT ESTIMATION USING PROJECTION: RECAP

- **Score function:** In parametric model, $f_Z(z;\theta)$, the *score function* $S_\theta(z,\theta_0)$ is defined by

$$S_\theta(z,\theta_0) = \frac{\partial}{\partial\theta}\{\log f_Z(z;\theta)\}_{\theta=\theta_0} = \begin{bmatrix} S_\psi(z,\theta_0) \\ S_\beta(z,\theta_0) \end{bmatrix} \qquad \begin{matrix} q\times 1 & \text{parameter of interest} \\ r \times 1 & \text{nuisance parameter} \end{matrix}$$

  NB: *Tsiatis uses β as the parameter of interest, and η as the nuisance parameter*

- **Influence function and asymptotic linearity:** An estimator $\widehat{\psi}_n$ of parameter ψ_0 is *asymptotically linear* if there exists a q-dimensional function, $\varphi(Z)$, with

$$\mathbb{E}[\varphi(Z)] = \mathbf{0}_q \qquad \mathbb{E}[\varphi(Z)\{\varphi(Z)\}^\top] < \infty, \text{ nonsingular}$$

  such that

$$\sqrt{n}(\widehat{\psi}_n - \psi_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \varphi(Z_i) + o_p(1). \tag{1}$$

  Then as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \varphi(Z_i) \xrightarrow{d} Normal\left(\mathbf{0}_q, \mathbb{E}[\varphi(Z)\{\varphi(Z)\}^\top]\right) \tag{2}$$

  it follows that $\sqrt{n}(\widehat{\psi}_n - \psi_0)$ also has this asymptotic distribution. If such a $\varphi(.)$ exists it is termed the *influence function* for the estimator, and it is *unique* (see Tsiatis Theorem 3.1, p23: it is straightforward to show that if the representation in equation (1) holds for two different influence functions, then those two influence functions are almost surely equal).

- **Influence function examples:**
  (a) **Likelihood estimation:** In *likelihood estimation*, we have the log-likelihood

$$\ell_n(\theta) = \sum_{i=1}^n \log f_Z(z_i;\theta) = \sum_{i=1}^n \ell(z_i,\theta).$$

  say. Under regularity conditions, by the *mean-value theorem*,

$$\dot{\ell}_n(\theta) = \dot{\ell}_n(\theta_0) + \ddot{\ell}_n(\theta')(\theta - \theta_0)$$

  where $\|\theta' - \theta_0\| < \|\theta - \theta_0\|$, and where

$$\dot{\ell}_n(\theta) = \frac{\partial \ell_n(\theta)}{\partial\theta} \qquad \ddot{\ell}_n(\theta) = \frac{\partial^2 \ell_n(\theta)}{\partial\theta\partial\theta^\top}.$$

  are $(p\times 1)$ and $(p \times p)$ respectively. Evaluating at $\theta = \widehat{\theta}_n$, and noting that $\dot{\ell}_n(\widehat{\theta}_n) = \mathbf{0}_p$, we have on rearrangement and multiplying through by $1/\sqrt{n}$ that

$$\left\{-\frac{1}{n}\ddot{\ell}_n(\theta')\right\}\sqrt{n}(\widehat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}}\dot{\ell}_n(\theta_0)$$

  where $\|\theta' - \theta_0\| < \|\widehat{\theta}_n - \theta_0\|$. As $n \longrightarrow \infty$, we have for the random variable version

$$\left\{\frac{1}{n}\ddot{\ell}_n(\theta')\right\} \xrightarrow{p} \mathbb{E}\left[\left.\frac{\partial^2 \log f_Z(Z;\theta)}{\partial\theta\partial\theta^\top}\right|_{\theta=\theta_0}\right] = \mathbb{E}\left[\ddot{\ell}(Z,\theta_0)\right]$$

  say, as $\widehat{\theta}_n \xrightarrow{p} \theta_0$. Therefore

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \left\{-\mathbb{E}\left[\ddot{\ell}(Z,\theta_0)\right]\right\}^{-1}\frac{1}{\sqrt{n}}\dot{\ell}_n(\theta_0) + o_p(1)$$

  which yields the influence function

$$\varphi(Z) \equiv \varphi(Z,\theta_0) = \left\{-\mathbb{E}\left[\ddot{\ell}(Z,\theta_0)\right]\right\}^{-1}\dot{\ell}(Z,\theta_0) = \mathcal{J}^{-1}(\theta_0)\dot{\ell}(Z,\theta_0) \equiv \mathcal{J}^{-1}(\theta_0)S_\theta(Z,\theta_0).$$

(b) m**-estimation:** In m-estimation, we replace the score equation for θ by the general form

$$\sum_{i=1}^{n} m(Z_i, \theta) = \mathbf{0}_p$$

for function $m(.,.)$, with

$$\begin{aligned} &\text{(i)} \quad \mathbb{E}[m(Z, \theta)] = \mathbf{0}_p \\ &\text{(ii)} \quad \mathbb{E}[\{m(Z, \theta)\}^\top m(Z, \theta)] < \infty \\ &\text{(iii)} \quad \mathbb{E}[m(Z, \theta)\{m(Z, \theta)\}^\top] \text{ nonsingular} \end{aligned}$$

for all possible data generating θ. Using the same mean-value theorem expansion as in the likelihood case, we have

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_m(Z_i) + \mathrm{o}_p(1)$$

for the influence function associated with m

$$\varphi_m(Z) = \{-\mathbb{E}\left[\dot{m}(Z, \theta_0)\right]\}^{-1} m(Z, \theta_0).$$

- **Asymptotic variance**: For the more general m-estimation case, by the CLT and elementary results for the Normal distribution, we have that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \longrightarrow Normal(\mathbf{0}_p, \mathcal{J}^{-1}\mathcal{I}\mathcal{J}^{-\top})$$

where

$$\mathcal{I} \equiv \mathcal{I}(\theta_0) = \mathbb{E}[m(Z, \theta_0)\{m(Z, \theta_0)\}^\top] \qquad \mathcal{J} \equiv \mathcal{J}(\theta_0) = -\mathbb{E}[\dot{m}(Z, \theta_0)]. \tag{3}$$

These $p \times p$ matrices are typically estimated by

$$\widehat{I}_n = \frac{1}{n} \sum_{i=1}^{n} m(Z_i, \widehat{\theta}_n)\{m(Z_i, \widehat{\theta}_n)\}^\top \qquad \widehat{J}_n = -\frac{1}{n} \sum_{i=1}^{n} \dot{m}(Z_i, \widehat{\theta}_n).$$

In the likelihood case, we have

$$\mathrm{Var}[\varphi(Z)] = \mathcal{J}^{-1}(\theta_0)\mathrm{Var}[\dot{\ell}(Z, \theta_0)]\mathcal{J}^{-1}(\theta_0)$$

where, under standard likelihood theory,

$$\mathcal{I}(\theta_0) = \mathrm{Var}[\dot{\ell}(Z, \theta_0)] = \mathbb{E}\left[\dot{\ell}(Z, \theta_0)\{\dot{\ell}(Z, \theta_0)\}^\top\right] \equiv -\mathbb{E}\left[\ddot{\ell}(Z, \theta_0)\right]$$

that is $\mathcal{I}(\theta_0) = \mathcal{J}(\theta_0)$ so that

$$\mathrm{Var}[\varphi(Z)] = \left\{\mathbb{E}\left[\dot{\ell}(Z, \theta_0)\{\dot{\ell}(Z, \theta_0)\}^\top\right]\right\}^{-1} = \mathcal{I}^{-1}(\theta_0)$$

which is the asymptotic variance of $\widehat{\theta}_n$.

- **Differentiable parameter:** Consider a q-dimensional parameter of interest $\Psi(\theta)$, and let

$$\Gamma(\theta) = \frac{\partial \Psi(\theta)}{\partial \theta^\top} \qquad q \times p. \tag{4}$$

We say that $\Psi(\theta)$ is a *differentiable parameter*: in the parametric model, this is differentiability in the ordinary calculus sense.

- **Key Theorem:** see Tsiatis Theorem 3.2 and Corollary 1 (pp 28–37). Suppose that estimator $\widehat{\Psi}_n$ is *regular and asymptotically linear* (RAL) with influence function $\varphi(.)$ such that $\mathbb{E}[\{\varphi(Z)\}^\top \varphi(Z)] < \infty$. Then
$$\mathbb{E}[\varphi(Z)\,\{S_\theta(Z,\theta_0)\}^\top] = \Gamma(\theta_0).$$
Special case: If θ is partitioned $\theta = (\psi^\top, \beta^\top)^\top$ and $\Psi(\theta) \equiv \psi$ (ie just taking the first q components of θ), then by differentiation we have that
$$\Gamma(\theta_0) = \begin{bmatrix} \mathbf{I}_{q\times q} & \mathbf{0}_{q\times r} \\ \mathbf{0}_{r\times q} & \mathbf{0}_{r\times r} \end{bmatrix}$$
and so
  (i) $\mathbb{E}[\varphi(Z)\,\{S_\psi(Z,\theta_0)\}^\top] = \mathbf{I}_{q\times q}$, and
  (ii) $\mathbb{E}[\varphi(Z)\,\{S_\beta(Z,\theta_0)\}^\top] = \mathbf{0}_{q\times r}$.

That is, $\varphi(Z)$ is *orthogonal* to $S_\beta(Z,\theta_0)$; note that this is "orthogonality" in the *uncorrelatedness* sense rather than the Hilbert space (geometric) sense; all elements of $\varphi(Z)$ are **uncorrelated** with all elements of $S_\beta(Z,\theta_0)$.

Notes:
  (a) The Theorem applies to m-estimators (see Tsiatis pp 33–34).
  (b) The converse of the result is also true: **if** an influence function $\varphi(X)$ satisfies the Theorem, **then** it is the (unique) influence function for an RAL estimator (see Tsiatis pp 38–41).
  (c) Part (ii) of the Theorem implies that if $\mathbf{B}$ is an arbitrary deterministic $q \times r$ matrix, then
$$\mathbb{E}\left[\{\varphi(Z)\}^\top \{\mathbf{B} S_\beta(Z,\theta_0)\}\right] = \sum_{j=1}^{q}\sum_{k=1}^{r} b_{jk}\mathbb{E}[\varphi_j(Z) S_{\beta_k}(Z,\theta_0)] = 0$$
and hence $\varphi(Z)$ is orthogonal (in the Hilbert space sense) to all such $\mathbf{B}S_\beta(Z,\theta_0)$. Because $\mathbf{B}$ is arbitrary, we can therefore say that $\varphi(Z)$ is orthogonal to the space, Λ, constructed as
$$\Lambda = \{\mathbf{B}S_\beta(Z,\theta_0) : \mathbf{B} \text{ an arbitrary } q\times r \text{ matrix}\}.$$
Thus if $\varphi(Z)$ is an RAL estimator for ψ, it must be an element in the space that orthogonal to Λ, which we denote $\varphi(X) \in \Lambda^\perp$.
  (d) **Tangent space:** Extending the above ideas, we can also consider the tangent space related to the entire θ vector, $\mathcal{T}$, namely
$$\mathcal{T} = \{\mathbf{B}S_\theta(Z,\theta_0) : \mathbf{B} \text{ an arbitrary } q\times p \text{ matrix}\}$$
that is, the space constructed by taking matrix multiples of the $p \times 1$ score vector $S_\theta(Z,\theta_0)$. This also defines a linear subspace of $\mathcal{H}_q$, and as by direct calculation
$$\mathbf{B}S_\theta(Z,\theta_0) \equiv [\mathbf{B}_1\ \mathbf{B}_2]\begin{bmatrix} S_\psi(Z,\theta_0) \\ S_\beta(Z,\theta_0)\end{bmatrix} = \mathbf{B}_1 S_\psi(Z,\theta_0) + \mathbf{B}_2 S_\beta(Z,\theta_0)$$
say, we observe that elements in $\mathcal{T}$ can be represented as the sum of elements in the set
$$\mathcal{T}_\psi = \{\mathbf{B}S_\psi(Z,\theta_0) : \mathbf{B} \text{ an arbitrary } q\times q \text{ matrix}\}$$
and elements in Λ, and using the direct sum notation for vector spaces, $\oplus$, we write
$$\mathcal{T} = \mathcal{T}_\beta \oplus \Lambda.$$
See Tsiatis, pp 42–43.

(e) To identify $\Lambda^{\perp}$, we consider the *projecting* an arbitrary element $h \in \mathcal{H}_q$ (the space of zero-mean random functions of Z identified in lectures) onto Λ and taking the *residual*. The projection of h on Λ, denoted $\Pi(h|\Lambda)$ can be easily computed; we have from the results in lectures that

$$\Pi(h|\Lambda) = \mathbf{B}_0 S_\beta(Z, \theta_0)$$

where $\mathbf{B}_0$ satisfies

$$\mathbb{E}[(h - \mathbf{B}_0 S_\beta(Z, \theta_0))^\top \{\mathbf{B} S_\beta(Z, \theta_0)\}] = 0$$

for any $q \times r$ deterministic matrix $\mathbf{B}$. It follows (Tsiatis pp 16–18) that

$$\mathbf{B}_0 = \mathbb{E}[h S_\beta^\top]\{\mathbb{E}[S_\beta S_\beta^\top]\}^{-1}$$

so that

$$\Pi(h|\Lambda) = \mathbb{E}[h S_\beta^\top]\{\mathbb{E}[S_\beta S_\beta^\top]\}^{-1} S_\beta(Z, \theta).$$

Therefore $\Lambda^{\perp}$ is the space

$$\{h - \Pi(h|\Lambda) : h \in \mathcal{H}_q\} \subset \mathcal{H}_q$$

and we conclude that suitable influence functions must take this form.

(f) The results above characterize the set of all influence functions that satisfy the Key Theorem (see Theorem 3.4, Tsiatis pp 45–46): this set can be written

$$\varphi(Z) + \mathcal{T}^{\perp}$$

where $\varphi(Z)$ is an arbitrary influence function, and $\mathcal{T}^{\perp}$ is the space perpendicular to $\mathcal{T}$.

- **Efficiency:** Once we have identified all influence functions that satisfy the Key Theorem, we can select the optimal or *efficient* influence function, $\varphi^{\text{eff}}(Z)$, in this collection as the influence function that has the *smallest variance*: we know that the asymptotically, the variance of an RAL estimator is given by (2) as

$$\mathbb{E}[\varphi(Z)\{\varphi(Z)\}^\top]$$

so we now look for $\varphi^{\text{eff}}(Z)$ such that

$$\mathbb{E}[\varphi(Z)\{\varphi(Z)\}^\top] - \mathbb{E}[\varphi^{\text{eff}}(Z)\{\varphi^{\text{eff}}(Z)\}^\top]$$

is *non-negative definite* for any other $\varphi(Z)$.

(a) **Efficient Influence Function:** Tsiatis Theorem 3.5 (pp 46–47) demonstrates that the efficient influence function for estimating $\Psi(\theta_0)$ takes the form

$$\varphi^{\text{eff}}(Z) = \varphi(Z) - \Pi(\varphi(Z)|\mathcal{T}^\top) = \Pi(\varphi(Z)|\mathcal{T})$$

where $\varphi(Z)$ is arbitrary and explicitly that

$$\varphi^{\text{eff}}(Z) = \Gamma(\theta_0)\{\mathcal{J}(\theta_0)\}^{-1} S_\theta(Z, \theta_0)$$

where $\Gamma(\theta_0)$ is defined in equation (4) and $\mathcal{J}$ is defined by equation (3).

(b) **Efficient Score Function:** The efficient score function $S_\psi^{\text{eff}}(Z, \theta_0)$ is computed by projecting $S_\psi(Z, \theta_0)$ onto the nuisance tangent space and taking the residual, that is

$$S_\psi^{\text{eff}}(Z, \theta) = S_\psi(Z, \theta) - \Pi(S_\psi(Z, \theta)|\Lambda)$$

specifically

$$S_\psi^{\text{eff}}(Z, \theta) = S_\psi(Z, \theta) - \mathbb{E}[S_\psi S_\beta^\top]\{\mathbb{E}[S_\beta S_\beta^\top]\}^{-1} S_\beta(Z, \theta).$$

Then

$$\varphi^{\text{eff}}(Z) = \{\mathbb{E}[S_\psi^{\text{eff}}\{S_\psi^{\text{eff}}\}^\top\}^{-1} S_\psi^{\text{eff}}(Z, \theta_0)$$

See Corollary 2, p 47.

(c) The **smallest variance** thus is given by

$$\{\mathbb{E}[S_\psi^{\text{eff}}\{S_\psi^{\text{eff}}\}^\top]\}^{-1}$$

where, denoting

$$\mathbf{V}_{\psi\psi} = \mathbb{E}[S_\psi S_\psi^\top] \qquad \mathbf{V}_{\psi\beta} = \mathbb{E}[S_\psi S_\beta^\top] \qquad \mathbf{V}_{\beta\beta} = \mathbb{E}[S_\beta S_\beta^\top]$$

we have

$$\begin{aligned}\mathbb{E}[S_\psi^{\text{eff}}\{S_\psi^{\text{eff}}\}^\top] &= \mathbb{E}\left[\left\{S_\psi - \mathbf{V}_{\psi\beta}\mathbf{V}_{\beta\beta}^{-1}S_\beta\right\}\left\{S_\psi - \mathbf{V}_{\psi\beta}\mathbf{V}_{\beta\beta}^{-1}S_\beta\right\}^\top\right] \\ &= \mathbf{V}_{\psi\psi} - \mathbf{V}_{\psi\beta}\mathbf{V}_{\beta\beta}^{-1}\mathbf{V}_{\psi\beta}^\top\end{aligned}$$

so therefore the smallest variance is

$$\left\{\mathbf{V}_{\psi\psi} - \mathbf{V}_{\psi\beta}\mathbf{V}_{\beta\beta}^{-1}\mathbf{V}_{\psi\beta}^\top\right\}^{-1}.$$

(d) In **likelihood-based** estimation, this smallest variance is **precisely the same variance** obtained when using *joint* estimation of ψ and β, that is, it is the relevant $q \times q$ block of the matrix

$$\{\mathcal{I}(\theta_0)\}^{-1} = \{\mathbb{E}[S_\theta S_\theta^\top]\}^{-1}$$