Chapter 6

April 22, 2022

- We will now discuss the extension of a well known result in parametric inference to semiparametric inference.
- Consider a parametric model with "factorized likelihood", i.e.

$$\mathcal{F}_{par} = \{f(z;\theta) = g_1(z;\theta_1)\, g_2(z;\theta_2) : (\theta_1,\theta_2) \in \Theta_1 \times \Theta_2 \subseteq \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}\}$$

where g_1 and g_2 are some known functions.

- Note that in model $\mathcal{F}_{par}$ the parameters θ_1 and θ_2 are variation independent, since they vary over a Cartesian product set $\Theta_1 \times \Theta_2$.

- Suppose enough regularity conditions hold so that the score is the derivative of the log-likelihood and the information for θ is minus the expectation of the second derivative of the log-likelihood. Then

$$\underbrace{s_\theta(Z;\theta)}_{(p_1+p_2)\times 1} \equiv \begin{bmatrix} s_{\theta_1}(Z;\theta) \\ s_{\theta_2}(Z;\theta) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial\theta_1}\log f(z;\theta) \\ \frac{\partial}{\partial\theta_2}\log f(z;\theta) \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial\theta_1}\log g_1(z;\theta_1) \\ \frac{\partial}{\partial\theta_2}\log g_2(z;\theta_2) \end{bmatrix}$$

and

$$\underbrace{I(\theta)}_{(p_1+p_2)\times(p_1+p_2)} = -E_\theta\left[\frac{\partial^2}{\partial\theta\partial\theta^T}\log f(z;\theta)\right]$$

- Because $s_{\theta_2}(Z;\theta)$ depends on θ ONLY through θ_2, it holds that

$$\frac{\partial^2}{\partial\theta_1\partial\theta_2^T}\log f(z;\theta) = \frac{\partial}{\partial\theta_1} s_{\theta_2}(Z;\theta)^T = 0$$

- Consequently, $I(\theta)$ is a block- diagonal matrix. So,

$$E_\theta\left[s_{\theta_1}(Z;\theta)\, s_{\theta_2}(Z;\theta)^T\right] = 0$$

- We thus conclude that the CR- bound for estimation of θ_1 in model $\mathcal{F}_{par}$ is the same as the CR-bound for estimation of θ_1 in the parametric model in which θ_2 is known, say equal to θ_2^*

$$\mathcal{F}_{par,1,\theta_2^*} = \{f(z;\theta) = g_1(z;\theta_1)\, g_2(z;\theta_2^*) : \theta_1 \in \Theta_1 \subseteq \mathbb{R}^{p_1}\}$$

for any θ_2^* (recall the discussion on parameter orthogonality in slide 190 of ch. 3).

- Also, if $\beta(\theta)$ depends on θ only through θ_1, then the CR-bound for $\beta(\theta)$ at any $F^* = F_{\theta^*}$ is the same in models $\mathcal{F}_{par,1,\theta_2^*}$ and $\mathcal{F}_{par}$.
- In plain words:
  in a factorized likelihood model, the efficiency with which we can estimate with large samples a parameter that depends only on one of the factors is the same regardless of whether the remaining factor is known or unknown.

- The preceding result implies also the following geometric result.
- The tangent space in model $\mathcal{F}_{par,1,\theta_2^*}$ at $F^* = F_{\theta^*}$ is equal to

  $$\Lambda_{\mathcal{F}_{par,1,\theta_2^*}}(F^*) = \left\{ a_1^T s_{\theta_1}(Z;\theta^*) : a_1 \in \mathbb{R}^{p_1} \right\}$$

  and because $s_{\theta_1}(Z;\theta^*)$ is the same for all θ_2^*, then the space $\Lambda_{\mathcal{F}_{par,1,\theta_2^*}}(F^*)$ does not change with θ_2^*.
- Likewise $\Lambda_{\mathcal{F}_{par,2,\theta_1^*}}(F^*)$, the tangent space of $\mathcal{F}_{par,2,\theta_1^*}$ at F^*, does not depend on θ_1^*.

- Now, because the tangent space $\Lambda_{\mathcal{F}_{par}}(F)$ at F^* is comprised of linear combinations of the components of $s_{\theta_1}(Z;\theta^*)$ and $s_{\theta_2}(Z;\theta^*)$ then, the orthogonality of $s_{\theta_1}(Z;\theta^*)$ and $s_{\theta_2}(Z;\theta^*)$ implies that

  $$\Lambda_{\mathcal{F}_{par}}(F^*) = \mathcal{F}_{par,1,\theta_2^*}(F^*) \oplus \mathcal{F}_{par,2,\theta_1^*}(F^*)$$

- Thus we conclude that in factorized likelihood parametric models with variation independent parameters, the tangent space is the sum of two orthogonal tangent spaces: each tangent space being the tangent space of the model where one of the factors in the likelihood is known.

- We now extend the preceding result to semiparametric models.
- A "factorized likelihood" semiparametric model is a model of the form

  $$\mathcal{F} = \{ f(z) = g_1(z)\, g_2(z) : g_1 \in G_1, g_2 \in G_2 \}$$

  where G_1 and G_2 are some sets of functions.

- **Example 6.1:** An example of a factorized likelihood model is one in which $Z = (Z_1, Z_2)^T$, and

$$g_1(z) \equiv f_{Z_1}(z_1) \text{ and } g_2(z) \equiv f_{Z_2|Z_1}(z_2|z_1)$$

  - if $\mathcal{F}$ is the non-parametric model, then G_1 is the set of all possible densities of Z_1 and G_2 is the set of all possible conditional densities of Z_2 given Z_1.
  - if instead, for instance, G_2 is the set of all densities $f_{Z_2|Z_1}$ of $Z_2|Z_1$ such that there exists β verifying $E_{F_{Z_2|Z_1}}(Z_2|Z_1) = \beta^T Z_1$, then $\mathcal{F}$ is a strictly semiparametric model.

- For given $g_1^*(z)$ and $g_2^*(z)$ we define:
  - the semiparametric model

$$\mathcal{F}_1 = \{f(z) = g_1(z)\, g_2^*(z) : g_1 \in G_1\}$$

    in which g_2 is known and equal to g_2^*,
  - the semiparametric model

$$\mathcal{F}_2 = \{f(z) = g_1^*(z)\, g_2(z) : g_2 \in G_2\}$$

    in which g_1 is known and equal to g_1^*.
- Note that both $\mathcal{F}_1$ and $\mathcal{F}_2$ are submodels of $\mathcal{F}$.

- To state the Lemma which extends the orthogonality of tangent spaces of factorized likelihood models to semiparametric models we define a class $\mathcal{A}$ of regular parametric submodels

$$\mathcal{F}_{sub} = \{f(z;\theta) = g_1(z;\theta)\, g_2(z;\theta) : \theta \in \Theta \subseteq \mathbb{R}^p, g_j(z;\theta) \in G_j, j = 1, 2\}$$

  through F^*, i.e. such that $g_1^*(z) = g_1(z;\theta^*)$ and $g_2^*(z) = g_2(z;\theta^*)$.
- We also define for each $\mathcal{F}_{sub}$ in the class $\mathcal{A}$ and each $\theta' \in \Theta$, the submodels

$$\mathcal{F}_{sub,1,\theta'} = \{f(z;\theta) = g_1(z;\theta)\, g_2(z;\theta') : \theta \in \Theta \subseteq \mathbb{R}^p\}$$

  and

$$\mathcal{F}_{sub,2,\theta'} = \{f(z;\theta) = g_1(z;\theta')\, g_2(z;\theta) : \theta \in \Theta \subseteq \mathbb{R}^p\}$$

- Note that $\mathcal{F}_{sub,1,\theta'}$ is a parametric submodel of $\mathcal{F}_1$ in which $g_2(z)$ is known and equal to $g_2(z;\theta')$ (and likewise for $\mathcal{F}_{sub,2,\theta'}$ with roles of the subscripts 1 and 2 interchanged).

- For a given θ^* we also define

$$\mathcal{A}_1 \equiv \{\mathcal{F}_{sub,1,\theta^*} : \mathcal{F}_{sub} \in \mathcal{A}\}$$

  and

$$\mathcal{A}_2 \equiv \{\mathcal{F}_{sub,2,\theta^*} : \mathcal{F}_{sub} \in \mathcal{A}\}$$

- If the models in the class $\mathcal{A}_1$ are regular at $F^* = F_{\theta^*}$, we then let $\Lambda_{\mathcal{F}_1}(F^*)$ denote the tangent space at F^* wrt the class $\mathcal{A}_1$ in model $\mathcal{F}_1$
- We define $\Lambda_{\mathcal{F}_2}(F^*)$ likewise.
- We are now ready to state and prove two lemmas that extend the results discussed earlier for parametric models to semiparametric models.
- In the following Lemmas, $\mathcal{F}, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_{sub}, \mathcal{F}_{sub,1,\theta^*}, \mathcal{F}_{sub,2,\theta^*}, \mathcal{A}_1, \mathcal{A}_2$, $\Lambda_{\mathcal{F}_1}(F^*)$ and $\Lambda_{\mathcal{F}_2}(F^*)$ are defined as above.

- **Lemma 6.1:** Suppose that
  - For each $\mathcal{F}_{sub}$ in the class $\mathcal{A}$,
    - $\mathcal{F}_{sub,1,\theta'}$ and $\mathcal{F}_{sub,2,\theta'}$ are regular parametric models for any $\theta' \in \Theta$, with scores at θ^* that do not depend on θ', and are denoted as $s_1(Z;\theta^*)$ and $s_2(Z;\theta^*)$.
    - $\mathcal{F}_{sub,1,\theta^*}$ and $\mathcal{F}_{sub,2,\theta^*}$ are submodels in the class $\mathcal{A}$.
  - For each $\mathcal{F}_{sub}$ in the class $\mathcal{A}$, the score $s_\theta(Z;\theta^*)$ for θ at θ^* in model $\mathcal{F}_{sub}$ can be decomposed as

    $$s_\theta(Z;\theta^*) = s_1(Z;\theta^*) \ + \ s_2(Z;\theta^*)$$

  - The following map is continuous at $\theta = \theta^*$:

    $$\theta \to E_\theta\left[\{s_1(Z;\theta^*)\}^2\right]$$

    where $E_\theta(\cdot)$ is expectation under $f(z;\theta)$ in model $\mathcal{F}_{sub,2,\theta^*}$, i.e.

    $$E_\theta\left[\{s_1(Z;\theta^*)\}^2\right] = \int s_1(z;\theta^*)^2\, g_1(z;\theta^*)\, g_2(z;\theta)\, dz$$

  - Then

    $$\boxed{\Lambda_{\mathcal{F}}(F^*) = \Lambda_{\mathcal{F}_1}(F^*) \oplus \Lambda_{\mathcal{F}_2}(F^*)}$$

- **Lemma 6.2:** Suppose the assumptions of lemma 6.1 hold. Suppose that a parameter $\beta(F)$ of interest is pathwise differentiable wrt $\mathcal{A}$ at F^* and depends on F only through g_1, that is

  $$f = g_1 g_2 \text{ and } f' = g_1 g_2' \Rightarrow \beta(F) = \beta(F')$$

  then,
  - for any gradient $\psi_{F^*}(Z)$ of $\beta(F)$ wrt $\mathcal{A}$ in model $\mathcal{F}$ at F^* it holds that

    $$\psi_{F^*}(Z) \perp \Lambda_{\mathcal{F}_2}(F^*)$$

  - The efficient influence function $\psi_{F^*,eff}(Z)$ of $\beta(F)$ wrt $\mathcal{A}$ in model $\mathcal{F}$ at F^* satisfies

    $$\psi_{F^*,eff}(Z) \in \Lambda_{\mathcal{F}_1}(F^*)$$

  - If $\psi_{1,F^*,eff}(Z)$ denotes the efficient influence function of $\beta(F)$ wrt $\mathcal{A}_1$ in model $\mathcal{F}_1$ at F^* then

    $$\psi_{F^*,eff}(Z) = \psi_{1,F^*,eff}(Z)$$

    Consequently, the CR-bound for $\beta(F)$ at F^* wrt $\mathcal{A}_1$ in model $\mathcal{F}_1$ is the same the CR-bound wrt to $\mathcal{A}$ in model $\mathcal{F}$.

- Before we prove the Lemmas let us examine the assumptions made on the class $\mathcal{A}$ in an example.
- **Example 6.1 (continuation):** recall in this example $Z = (Z_1, Z_2)^T$,

  $$g_1(z) \equiv f_{Z_1}(z_1) \text{ and } g_2(z) \equiv f_{Z_2|Z_1}(z_2|z_1)$$

  The assumptions of Lemma 6.1 boil down to assumptions about a class $\mathcal{A}$ of regular parametric submodels

  $$\mathcal{F}_{sub} = \left\{f(z_1, z_2;\theta) : f_{Z_1}(z_1;\theta)\, f_{Z_2|Z_2}(z_2|z_1\theta) : \theta \in \mathbb{R}^p\right\}$$

  The assumptions are not at all restrictive. Specifically,

- The first assumption requires that for each $\mathcal{F}_{sub}$ in the class $\mathcal{A}$,
  -

    $$\mathcal{F}_{sub,1,\theta'} = \left\{f(z_1, z_2;\theta) : f_{Z_1}(z_1;\theta)\, f_{Z_2|Z_2}(z_2|z_1\theta') : \theta \in \mathbb{R}^p\right\}$$

    $$\mathcal{F}_{sub,1,\theta'} = \left\{f(z_1, z_2;\theta) : f_{Z_1}(z_1;\theta')\, f_{Z_2|Z_2}(z_2|z_1\theta) : \theta \in \mathbb{R}^p\right\}$$

    are regular parametric models for any $\theta' \in \Theta$, with scores at θ^* that do not depend on θ', and are denoted as $s_1(Z;\theta^*)$ and $s_2(Z;\theta^*)$.

    [Note that this is not at all restrictive, most "well behaved" submodels F_{sub} will satisfy this condition]

  - $\mathcal{F}_{sub,1,\theta^*}$ and $\mathcal{F}_{sub,2,\theta^*}$ are submodels in the class $\mathcal{A}$.

    [Note that this indeed can always be achieved by enlarging the class $\mathcal{A}$ to include the submodels $F_{sub,1,\theta^*}$ and $F_{sub,2,\theta^*}$ in the class, so it is not in essence a restriction on $\mathcal{A}$]

- The second assumption requires that for each $\mathcal{F}_{sub}$ in the class $\mathcal{A}$, the score $s_\theta(Z;\theta^*)$ for θ at θ^* in model $\mathcal{F}_{sub}$ can be decomposed as

$$s_\theta(Z;\theta^*) = s_1(Z;\theta^*) \ + \ s_2(Z;\theta^*)$$

Note this assumption holds trivially if scores are derivatives of the log-likelihood, i.e. if

$$\begin{aligned} s_\theta(Z;\theta^*) &= \frac{\partial}{\partial\theta^T}\log\left\{f_{Z_1}(z_1;\theta)\, f_{Z_2|Z_2}(z_2|z_1\theta)\right\}\Big|_{\theta=\theta^*} \\ s_1(Z;\theta^*) &= \frac{\partial}{\partial\theta^T}\log\left\{f_{Z_1}(z_1;\theta)\, f_{Z_2|Z_2}(z_2|z_1\theta')\right\}\Big|_{\theta=\theta^*} \\ s_2(Z;\theta^*) &= \frac{\partial}{\partial\theta^T}\log\left\{f_{Z_1}(z_1;\theta')\, f_{Z_2|Z_2}(z_2|z_1\theta)\right\}\Big|_{\theta=\theta^*} \end{aligned}$$

- The third assumption of Lemma 6.1 which requires the continuity of the map

$$\theta \to E_\theta\left[\{s_1(Z;\theta^*)\}^2\right] = \int s_1(z;\theta^*)^2\, g_1(z;\theta^*)\, g_2(z;\theta)\, dz$$

holds trivially because the map is constant, since

$$\begin{aligned} & \int s_1(z;\theta^*)^2\, g_1(z;\theta^*)\, g_2(z;\theta)\, dz \\ = & \int\int s_1(z_1;\theta^*)^2\, f_{Z_1}(z_1;\theta^*)\, f_{Z_2|Z_2}(z_2|z_1\theta)\, dz_1 dz_2 \\ = & \int s_1(z_1;\theta^*)^2\, f_{Z_1}(z_1;\theta^*)\, dz_1 \underbrace{\int f_{Z_2|Z_2}(z_2|z_1\theta)\, dz_2}_{=1} \\ = & \int s_1(z_1;\theta^*)^2\, f_{Z_1}(z_1;\theta^*)\, dz_1 \end{aligned}$$

- **Proof of Lemma 6.1:** Because $s_1(Z;\theta^*)$ is a score at $\theta=\theta^*$ in model $\mathcal{F}_{sub,1,\theta'}$ for all $\theta' \in \Theta$, then

$$\int s_1(z;\theta^*)\, g_1(z;\theta^*)\, g_2\left(z;\theta'\right) dz = 0 \text{ for all } \theta' \in \Theta$$

then

$$0 = \frac{\partial}{\partial\theta^T}\left[\int s_1(z;\theta^*)\, g_1(z;\theta^*)\, g_2(z;\theta)\, dz\right]\Big|_{\theta=\theta^*}$$

Now, since, by assumption, the model

$$\mathcal{F}_{sub,2,\theta^*} = \left\{f(z;\theta) = g_1(z;\theta^*)\, g_2(z;\theta) : \theta \in \Theta \subseteq \mathbb{R}^p\right\}$$

is regular and the map

$$\theta \to E_\theta\left[s_1(Z;\theta^*)^2\right] = \int s_1(z;\theta^*)^2\, g_1(z;\theta^*)\, g_2(z;\theta)\, dz$$

is continuous at $\theta=\theta^*$, then by Lemma 3.3,

$$\frac{\partial}{\partial\theta^T}\left[\int s_1(z;\theta^*)\, g_1(z;\theta^*)\, g_2(z;\theta)\, dz\right]\Big|_{\theta=\theta^*} = \int s_1(z;\theta^*)\, s_2(z;\theta^*)^T f(z;\theta^*)\, dz$$

which then shows that $\Lambda_{\mathcal{F}_1}(F^*)$ and $\Lambda_{\mathcal{F}_2}(F^*)$ are orthogonal.

- Now, by assumption (1.b), $s_1(Z;\theta^*)$ is a score at F^* in a submodel in class $\mathcal{A}$. This implies that $\Lambda_{\mathcal{F}_1}(F^*) \subset \Lambda_{\mathcal{F}}(F^*)$. Likewise $\Lambda_{\mathcal{F}_2}(F^*) \subset \Lambda_{\mathcal{F}}(F^*)$. Consequently,

$$\Lambda_{\mathcal{F}_1}(F^*) \oplus \Lambda_{\mathcal{F}_2}(F^*) \subset \Lambda_{\mathcal{F}}(F^*)$$

- On the other hand, by assumption (2), for any $a \in \mathbb{R}^p$ and any submodel $\mathcal{F}_{sub}$ in $\mathcal{A}$ with score $s_\theta(Z;\theta^*)$ it holds that

$$a^T s_\theta(Z;\theta^*) \in \Lambda_{\mathcal{F}_1}(F^*) \oplus \Lambda_{\mathcal{F}_2}(F^*)$$

Therefore

$$\Lambda_{\mathcal{F}}(F^*) \subset \overline{\left[\Lambda_{\mathcal{F}_1}(F^*) \oplus \Lambda_{\mathcal{F}_2}(F^*)\right]}$$

- But

$$\overline{\left[\Lambda_{\mathcal{F}_1}(F^*) \oplus \Lambda_{\mathcal{F}_2}(F^*)\right]} = \Lambda_{\mathcal{F}_1}(F^*) \oplus \Lambda_{\mathcal{F}_2}(F^*)$$

because $\Lambda_{\mathcal{F}_j}(F^*) = \overline{\left[\Lambda_{\mathcal{F}_j}(F^*)\right]}, (j=1,2)$ and $\Lambda_{\mathcal{F}_1}(F^*) \perp \Lambda_{\mathcal{F}_2}(F^*)$.

- Thus,

$$\Lambda_{\mathcal{F}}(F^*) \subset \Lambda_{\mathcal{F}_1}(F^*) \oplus \Lambda_{\mathcal{F}_2}(F^*)$$

which then concludes the proof.

- **Proof of Lemma 6.2:** By assumption (1.b),

$$\mathcal{F}_{sub,2,\theta^*} = \{f(z;\theta) = g_1(z;\theta^*) g_2(z;\theta) : \theta \in \Theta \subseteq \mathbb{R}^p\}$$

is a submodel in the class $\mathcal{A}$. Since $\beta(F)$ does not depend on g_2 then

$$\beta(F_\theta) \text{ is constant over all } f(z;\theta) \text{ in } \mathcal{F}_{sub,2,\theta^*}$$

- Therefore, since $\beta(F)$ is a pathwise differentiable parameter with respect to $\mathcal{A}$ at F^* we have that with F_θ varying over model $\mathcal{F}_{sub,2,\theta^*}$,

$$\begin{aligned} 0 &= \left.\frac{\partial \beta(F_\theta)}{\partial \theta^T}\right|_{\theta^*} \\ &= E_{\theta^*}\left[\psi_{F^*}(Z) s_2(Z;\theta^*)^T\right] \end{aligned}$$

where $\psi_{F^*}(Z)$ is any gradient of $\beta(F)$. This shows part (1) of Lemma 6.2 that $\psi_{F^*}(Z) \perp \Lambda_{\mathcal{F}_2}(F^*)$.

- The proof of part (2) follows from

$$\begin{aligned} \psi_{F^*,eff}(Z) &= \Pi\left[\psi_{F^*}(Z) | \Lambda_{\mathcal{F}}(F^*)\right] \\ &= \Pi\left[\psi_{F^*}(Z) | \Lambda_{\mathcal{F}_1}(F^*) \oplus \Lambda_{\mathcal{F}_2}(F^*)\right] \\ &= \Pi\left[\psi_{F^*}(Z) | \Lambda_{\mathcal{F}_1}(F^*)\right] + \underbrace{\Pi\left[\psi_{F^*}(Z) | \Lambda_{\mathcal{F}_2}(F^*)\right]}_{=0 \text{ bc } \psi_{F^*}(Z) \perp \Lambda_{\mathcal{F}_2}(F^*)} \\ &= \Pi\left[\psi_{F^*}(Z) | \Lambda_{\mathcal{F}_1}(F^*)\right] \end{aligned}$$

which shows that $\psi_{F^*,eff}(Z) \in \Lambda_{\mathcal{F}_1}(F^*)$.

- Finally, to prove part (3), we first note that $\psi_{F^*}(Z)$ is a gradient of $\beta(F)$ wrt $\mathcal{A}_1$ at F^* in model $\mathcal{F}_1$ because by assumption (1.b) $\mathcal{A}_1 \subset \mathcal{A}$. Then, the efficient influence function of $\beta(F)$ wrt $\mathcal{A}_1$ at F^* in model $\mathcal{F}_1$ satisfies

$$\psi_{1,F^*,eff}(Z) = \Pi\left[\psi_{F^*}(Z) | \Lambda_{\mathcal{F}_1}(F^*)\right]$$

But we have just showed that $\Pi\left[\psi_{F^*}(Z) | \Lambda_{\mathcal{F}_1}(F^*)\right] = \psi_{F^*,eff}(Z)$. So $\psi_{F^*,eff}(Z) = \psi_{1,F^*,eff}(Z)$. This concludes the proof of Lemma 6.2.

- **Example 6.1. (continuation).** Suppose that model $\mathcal{F}$ for $Z = (Z_1, Z_2)$ is defined by the sole restriction that for each $F \in \mathcal{F}$ there exists $\beta(F) \in \mathbb{R}^k$ such that

$$E_F(Z_2|Z_1) = \beta(F)^T Z_1$$

- Then, Lemma 6.2 implies that the CR-bound for $\beta(F)$ in a model that does not impose restrictions on the marginal law of the covariate Z_1 is the same as in a model in which the law of Z_1 is known (provided the class $\mathcal{A}$ satisfies the restrictions of Lemma 6.1 which as we have seen are not at all restrictive).

- Let F be the cdf of a random vector $Z = \left(Y, R, X^T\right)^T$ where R is a binary r.v., Y is a scalar r.v. (discrete or continous) and X is a random vector with discrete and/or continuous components.
- We will study inference about the parameter

$$\beta(F) \equiv E_F\left[E_F(Y|R=1, X)\right]$$

under three different models, namely:

  - $$\mathcal{F}_{np} = \left\{F : E_F\left[\{E_F(Y|R=1,X)\}^2\right] < \infty, P_F(R=1|X) > \sigma_F > 0\right\},$$
  - $$\mathcal{F}_{sem,fixed} = \{F \in \mathcal{F}_{np} : P_F(R=1|X) = \pi^*(X)\}$$

    where $\pi^*(x)$ is specified, i.e. known , and
  - $$\mathcal{F}_{sem,par} = \{F \in \mathcal{F}_{np} : P_F(R=1|X) = \pi(X;\alpha),\ \alpha \in \Xi \subseteq \mathbb{R}^r\}$$

    where $\pi(x;\alpha)$ is a specified, i.e. known, function of x and α which is differentiable wrt α at every x.

- The motivation for the parameter $\beta(F)$ is as follows.
- Suppose that on a random sample of units from a population of interest, we always measure the random vector X, but we measure the vector Y^f only on a subsample, that is, Y^f is missing in some study units.
- Suppose that R is the missing data indicator, i.e.

$$R = \begin{cases} 1 \text{ if } Y^f \text{ is observed} \\ 0 \text{ if } Y^f \text{ is missing} \end{cases}$$

and

$$Y = \begin{cases} Y^f & \text{if } Y^f \text{ is observed} \\ NA & \text{if } Y^f \text{ is missing} \end{cases}$$

- Then, on each unit of our random sample we actually observe $\left(Y, R, X^T\right)^T$.

- The outcome Y^f is said to be *missing at random iff*

$$P\left(R = 1|Y^f, X\right) = P\left(R = 1|X\right)$$

  - The missing at random assumption essentially postulates that X contains all the predictors of the outcome Y^f that are associated with non-response.
  - The MAR assumption is untestable, i.e. it does not impose any restriction on the law of the observed data $\left(Y, R, X^T\right)$ and thus cannot be rejected by any test.
  - MAR holds in two-stage study designs, where at the first stage, a cheap surrogate X for the outcome Y^f is measured on all study units and at the second stage, an expensive outcome Y^f is measured in a subsample selected with known probability that may depend on X.

- Under the missing at random assumption, the mean of Y^f happens to be equal to $\beta(F)$.
- To see this, let F^f denote the law of the "full" data vector $\left(Y^f, R, X^T\right)^T$ and let F be the law of $\left(Y, R, X^T\right)^T$ implied by F^f.
- Then

$$\begin{aligned} E_{F^f}\left(Y^f\right) &= E_{F^f}\left[E_{F^f}\left(Y^f|X\right)\right] \\ &= E_{F^f}\left[E_{F^f}\left(Y^f|R=1, X\right)\right] \text{ (by MAR)} \\ &= E_F\left[E_F\left(Y|R=1, X\right)\right] \text{ (by the definition of } Y) \\ &= \beta(F) \end{aligned}$$

- Let's now introduce the class of regular parametric submodels $\mathcal{A}_{np}$ of the non-parametric model $\mathcal{F}_{np}$ with respect to which we will compute the efficient influence function and the CR-bound.

- To understand the conditions that define the model class $\mathcal{A}_{np}$, consider an arbitrary parametric submodel, say,

$$\mathcal{F}_{sub} = \left\{f_{Y,R,X}\left(y, r, x; \theta\right) : \theta \in \Theta \subset \mathbb{R}^p\right\}$$

through F^* at θ^*, i.e. $F^* = F_{\theta^*}$. Writing

$$f_{Y,R,X}\left(y, r, x; \theta\right) = f_X\left(x; \theta\right) f_{R|X}\left(r|x; \theta\right) f_{Y|R,X}\left(y|r, x; \theta\right)$$

we see that $\mathcal{F}_{sub}$ implies a parametric submodel for each of the following:

  - the marginal law of X,

$$\mathcal{F}_{X,sub} = \left\{f_X\left(x; \theta\right) : \theta \in \Theta \subset \mathbb{R}^p\right\}$$

  - for each x, the conditional probability of R given $X = x$,

$$\mathcal{F}_{R|X=x,sub} = \left\{P_{R|X}\left(r|x; \theta\right) : \theta \in \Theta \subset \mathbb{R}^p\right\}$$

  - for each $R = r$ and $X = x$, the law of Y given $(R = r, X = x)$,

$$\mathcal{F}_{Y|R,X,sub} = \left\{f_{Y|R,X}\left(y|r, x; \theta\right) : \theta \in \Theta \subset \mathbb{R}^p\right\}$$

- Suppose that the parametric submodel $\mathcal{F}_{sub}$ is regular, it goes through F^* at θ^*, i.e. $F^* = F_{\theta^*}$, and its score at θ^* is denoted as $s_\theta(Y,R,X;\theta^*) = S_\theta(\theta^*)$.
- The class $\mathcal{A}_{np}$ that we will consider will be comprised of submodels $\mathcal{F}_{sub}$ such that the implied submodels $\mathcal{F}_{X,sub}$, $\mathcal{F}_{R|X=x,sub}$ for all x and $\mathcal{F}_{Y|R=r,X=x,sub}$ for all (r,x) are *all regular*.
- We will denote the scores at θ^* in each submodel $\mathcal{F}_{X,sub}$, $\mathcal{F}_{R|X=x,sub}$ and $\mathcal{F}_{Y|R=r,X=x,sub}$ respectively with

$$s_{X,\theta}(X;\theta^*), s_{R|X,\theta}(R,x;\theta^*) \text{ and } s_{Y|R,X,\theta}(Y,r,x;\theta^*)$$

- Note that by the mean zero property of scores we have that

$$\begin{aligned} 0 &= E_{F_X^*}[s_{X,\theta}(X;\theta^*)], \\ 0 &= E_{F_{R|X}^*}\left[s_{R|X,\theta}(R,X;\theta^*)|X=x\right], \\ 0 &= E_{F_{Y|R,X}^*}\left[s_{Y|R,X,\theta}(Y,R,X;\theta^*)|R=r,X=x\right] \end{aligned}$$

- In addition, each parametric submodel $\mathcal{F}_{sub}=\{F_\theta:\theta\in\Theta\}$ in class $\mathcal{A}_{np}$ of $\mathcal{F}_{np}$ goes through F^* at θ^*, i.e. $F^*=F_{\theta^*}$, and satisfies the following conditions:
  - the following map is continuous at θ^*

$$\theta \to E_\theta\left[E_{F^*}(Y|R=1,X)^2\right]$$

  - for each x, the following map is continuous at θ^*

$$\theta \to E_\theta\left(Y^2|R=1,X=x\right)$$

  - the derivative in the left hand side below exists and equals the expression in the right hand side

$$\frac{\partial}{\partial\theta^T}E_{\theta^*}\{E_\theta(Y|R=1,X)\}\Big|_{\theta^*} = E_{\theta^*}\left\{\frac{\partial}{\partial\theta^T}E_\theta(Y|R=1,X)\Big|_{\theta^*}\right\}$$

  - the scores $s_\theta(Y,R,X;\theta^*), s_{X,\theta}(X;\theta^*), s_{R|X,\theta}(R,x;\theta^*)$ and $s_{Y|R,X,\theta}(Y,r,x;\theta^*)$ satisfy

$$s_\theta(Y,R,X;\theta^*) = s_{X,\theta}(X;\theta^*)+s_{R|X,\theta}(R,X;\theta^*)+s_{Y|R,X,\theta}(Y,R,X;\theta^*)$$

  (note that this last condition holds if scores are computed as derivatives of the log-likelihood).

- We shall now derive the gradients of $\beta(F)$ wrt to the class $\mathcal{A}_{np}$ at a given F^*. Let $\mathcal{F}_{sub}=\{F_\theta:\theta\in\Theta\}$ be in $\mathcal{A}_{np}$.

$$\begin{aligned} &\frac{\partial}{\partial\theta^T}\beta(\theta)\Big|_{\theta^*} = \frac{\partial}{\partial\theta^T}E_\theta\{E_\theta(Y|R=1,X)\}\Big|_{\theta^*} \\ =\ & \frac{\partial}{\partial\theta^T}E_\theta\{E_{\theta^*}(Y|R=1,X)\}\Big|_{\theta^*} + \frac{\partial}{\partial\theta^T}E_{\theta^*}\{E_\theta(Y|R=1,X)\}\Big|_{\theta^*} \\ =\ & E_{\theta^*}\left\{E_{\theta^*}(Y|R=1,X)\,s_{X,\theta}(X;\theta^*)^T\right\} + E_{\theta^*}\left\{\frac{\partial}{\partial\theta^T}E_\theta(Y|R=1,X)\right\}\Big|_{\theta^*} \\ =\ & E_{\theta^*}\left\{E_{\theta^*}(Y|R=1,X)\,s_{X,\theta}(X;\theta^*)^T\right\} \\ &+ E_{\theta^*}\left[E_{\theta^*}\left\{Y\ s_{Y|R=1,X;\theta}(Y,R=1,X;\theta^*)^T|R=1,X\right\}\right] \end{aligned}$$

- Now, letting

$$s_{Y,R|X,\theta}(Y,R,X;\theta^*) = s_{Y|R,X,\theta}(Y,R,X;\theta^*) + s_{R|X,\theta}(R,X;\theta^*),$$

we have

$$\begin{aligned} & E_{\theta^*}\left\{E_{\theta^*}(Y|R=1,X)\,s_{Y,R|X,\theta}(Y,R,X;\theta^*)^T|X\right\} \\ =\ & E_{\theta^*}(Y|R=1,X)\underbrace{E_{\theta^*}\left\{s_{Y,R|X,\theta}(Y,R,X;\theta^*)^T|X\right\}}_{=0} \end{aligned}$$

- Therefore

$$E_{\theta^*}\left\{E_{\theta^*}(Y|R=1,X)\,s_{X,\theta}(X;\theta^*)^T\right\} = E_{\theta^*}\left\{E_{\theta^*}(Y|R=1,X)\,S_\theta(\theta^*)^T\right\}$$

where $S_\theta(\theta^*) = s_{X,\theta}(X;\theta^*) + s_{Y,R|X,\theta}(Y,R,X;\theta^*)$ is the score at θ^* in the submodel $\mathcal{F}_{sub}$.

- On the other hand

$$
\begin{aligned}
& E_{\theta^*}\left[E_{\theta^*}\left\{Y\ s_{Y|R=1,X;\theta}\left(Y,R=1,X;\theta^*\right)^T|R=1,X\right\}\right] \\
& \overset{(1)}{=} E_{\theta^*}\left[E_{\theta^*}\left\{\frac{RY}{P_{\theta^*}\left(R=1|X\right)}s_{Y|R=1,X;\theta}\left(Y,R=1,X;\theta^*\right)^T\middle| X\right\}\right] \\
& \overset{(2)}{=} E_{\theta^*}\left[\frac{RY}{P_{\theta^*}\left(R=1|X\right)}s_{Y|R,X;\theta}\left(Y,R,X;\theta^*\right)^T\right] \\
& \overset{(3)}{=} E_{\theta^*}\left[\frac{R\left\{Y-E_{\theta^*}\left(Y|R,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}s_{Y|R,X;\theta}\left(Y,R,X;\theta^*\right)^T\right] \\
& \overset{(4)}{=} E_{\theta^*}\left[\frac{R\left\{Y-E_{\theta^*}\left(Y|R,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}\left[s_{Y|R,X;\theta}\left(Y,R,X;\theta^*\right)^T+s_{R,X,\theta}\left(R,X;\theta^*\right)^T\right]\right] \\
= \quad & E_{\theta^*}\left[\frac{R\left\{Y-E_{\theta^*}\left(Y|R=1,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}S_\theta\left(\theta^*\right)^T\right]
\end{aligned}
$$

- (1) holds because for any W, $E\left[W|R=1,X\right]=E\left[RW|X\right]/P\left(R=1|X\right)$
- (2) holds because $Rs_{Y|R=1,X;\theta}\left(Y,R=1,X;\theta^*\right)=Rs_{Y|R,X;\theta}\left(Y,R,X;\theta^*\right)$
- (3) holds because $E_{\theta^*}\left[\frac{RE(Y|R,X)}{P_{\theta^*}(R=1|X)}s_{Y|R,X;\theta}\left(Y,R,X;\theta^*\right)^T\middle| R,X\right]$ $=\frac{RE(Y|R,X)}{P_{\theta^*}(R=1|X)}\underbrace{E_{\theta^*}\left[s_{Y|R,X;\theta}\left(Y,R,X;\theta^*\right)^T\middle| R,X\right]}_{=0}=0.$
- (4) holds because $E_{\theta^*}\left[q\left(R,X\right)\left\{Y-E\left(Y|R,X\right)\right\}\right]=0$ for any $q\left(R,X\right)$, in particular, for $q\left(R,X\right)=\frac{R}{P_{\theta^*}(R=1|X)}s_{R,X,\theta}\left(R,X;\theta^*\right)^T$ where $s_{R,X,\theta}\left(R,X;\theta^*\right)\equiv s_{X,\theta}\left(X;\theta^*\right)+s_{R|X,\theta}\left(R|X;\theta^*\right)$.

So finally we obtain

$$
\begin{aligned}
& \frac{\partial}{\partial\theta^T}\beta\left(\theta\right)\Big|_{\theta^*}= \\
= \quad & E_{\theta^*}\left\{E_{\theta^*}\left(Y|R=1,X\right)s_{X,\theta}\left(X;\theta^*\right)^T\right\}\ + \\
& E_{\theta^*}\left[E_{\theta^*}\left\{Y\ s_{Y|R=1,X;\theta}\left(Y,R=1,X;\theta^*\right)^T|R=1,X\right\}\right] \\
= \quad & E_{\theta^*}\left\{E_{\theta^*}\left(Y|R=1,X\right)S_\theta\left(\theta^*\right)^T\right\}\ +\ E_{\theta^*}\left[\frac{R\left\{Y-E_{\theta^*}\left(Y|R=1,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}S_\theta\left(\theta^*\right)^T\right] \\
= \quad & E_{\theta^*}\left\{\left[E_{\theta^*}\left(Y|R=1,X\right)+\frac{R\left\{Y-E_{\theta^*}\left(Y|R=1,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}\right]S_\theta\left(\theta^*\right)^T\right\}
\end{aligned}
$$

- We therefore conclude that the following is a gradient of $\beta\left(F\right)$ at F^* w.r.t $\mathcal{A}_{np}$ in the model $\mathcal{F}_{np}$

$$
E_{F^*}\left(Y|R=1,X\right)+\frac{R\left\{Y-E_{F^*}\left(Y|R=1,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}
$$

- Furthermore because

$$
\begin{aligned}
& E_{F^*}\left[\frac{R\left\{Y-E_{F^*}\left(Y|R=1,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}\right] \\
= \quad & E_{F^*}\left\{E_{F^*}\left[\frac{R\left\{Y-E_{F^*}\left(Y|R=1,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}\middle| R,X\right]\right\} \\
= \quad & E_{F^*}\left\{\frac{R\left\{E_{F^*}\left(Y|R,X\right)-E_{F^*}\left(Y|R=1,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}\right\} \\
= \quad & E_{F^*}\left\{\frac{R\left\{E_{F^*}\left(Y|R=1,X\right)-E_{F^*}\left(Y|R=1,X\right)\right\}}{P_{\theta^*}\left(R=1|X\right)}\right\} \\
= \quad & 0
\end{aligned}
$$

- then the mean of the preceding gradient is $E_{F^*}\left[E_{F^*}\left(Y|R=1,X\right)\right]=\beta\left(F\right)$ from where we conclude that a mean zero gradient of $\beta\left(F\right)$ at F^* w.r.t $\mathcal{A}_{np}$ in the model $\mathcal{F}_{np}$ is

$$
\psi_{F^*}\left(Y,R,X\right)=E_{F^*}\left(Y|R=1,X\right)+\frac{R\left\{Y-E_{F^*}\left(Y|R=1,X\right)\right\}}{P_{F^*}\left(R=1|X\right)}-\beta\left(F^*\right)
$$

- To find the efficient influence function we must find the tangent space at F^* wrt $\mathcal{A}_{np}$ in the semiparametric model $\mathcal{F}_{np}$.
- The tangent space of $\mathcal{A}_{np}$ is indeed $\mathcal{L}_2^0\left(F^*\right)$.
- To see this, given any $s\left(Y,R,X\right)\in\mathcal{L}_2^0\left(F^*\right)$ we decompose it as

$$
s\left(Y,R,X\right)=s_1\left(X\right)+s_2\left(R,X\right)+s_3\left(Y,R,X\right)
$$

where

$$
\begin{aligned}
s_1\left(X\right) &= E_{F^*}\left[s\left(Y,R,X\right)|X\right] \\
s_2\left(R,X\right) &= E_{F^*}\left[s\left(Y,R,X\right)|R,X\right]-E_{F^*}\left[s\left(Y,R,X\right)|X\right] \\
s_3\left(Y,R,X\right) &= s\left(Y,R,X\right)-E_{F^*}\left[s\left(Y,R,X\right)|R,X\right]
\end{aligned}
$$

- Notice that by construction,

$$
\begin{aligned}
s_1\left(X\right) &\in \mathcal{L}_2^0\left(F_X^*\right), \\
s_2\left(R,x\right) &\in \mathcal{L}_2^0\left(F_{R|X=x}^*\right) \\
s_3\left(Y,r,x\right) &\in \mathcal{L}_2^0\left(F_{Y|R=rX=x}^*\right)
\end{aligned}
$$

- Consider the parametric submodel

$$\mathcal{F}_{sub} = \{f_{Y,R,X}(y,r,x;\theta) = f_X(x;\theta_1) f_{R|X}(r|x;\theta_2) f_{Y|R,X}(y|r,x;\theta_3) : \theta_j \in \mathbb{R}\}$$

where

$$f_X(x;\theta_1) = f_X^*(x)\, k(\theta_1 s_1(x))\, c_1(\theta_1)$$
$$f_{R|X}(r|x;\theta_2) = f_{R|X}^*(r|x)\, k(\theta_2 s_2(r,x))\, c_2(x;\theta_2)$$
$$f_{Y|R,X}(y|r,x;\theta_3) = f_{R|X}^*(y|r,x)\, k(\theta_3 s_3(y,r,x))\, c_3(r,x;\theta_3)$$
$$k(u) = 2[1+\exp(-2u)]^{-1}$$

and

$$c_1(\theta_1) = \left[\int f_X^*(x)\, k(\theta_1 s_1(x))\, dx\right]^{-1}$$
$$c_2(x;\theta_2) = \left[f_{R|X}^*(1|x)\, k(\theta_2 s_2(1,x)) + f_{R|X}^*(0|x)\, k(\theta_2 s_2(0,x))\right]^{-1}$$
$$c_3(r,x;\theta_3) = \left[\int f_{R|X}^*(y|r,x)\, k(\theta_3 s_3(y,r,x))\, dy\right]^{-1}$$

- This model is in class $\mathcal{A}_{np}$ (check it in the privacy of your own room) and its score at $\theta^* = (0,0,0)$ is

$$\left.\frac{\partial \log}{\partial \theta^T} f_{Y,R,X}(Y,R,X;\theta)\right|_{\theta^*} = [s_1(X), s_2(R,X), s_3(Y,R,X)]$$

so,

$$\begin{aligned}\left.\frac{\partial \log}{\partial \theta} f_{Y,R,X}(Y,R,X;\theta)\right|_{\theta^*} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} &= s_1(X) + s_2(R,X) + s_3(Y,R,X) \\ &= s(Y,R,X)\end{aligned}$$

which proves that $s(Y,R,X)$ is in the tangent space of model $\mathcal{F}_{np}$ wrt $\mathcal{A}_{np}$ at F^*.

- Since $s(Y,R,X)$ is an arbitrary element of $\mathcal{L}_2^0(F^*)$, this proves that $\mathcal{L}_2^0(F^*)$ is included in the tangent space wrt to $\mathcal{A}_{np}$ at F^*.

- Since the reverse inclussion always holds, then this shows that the tangent spce of model $\mathcal{F}_{np}$ wrt to $\mathcal{A}_{np}$ at F^* is $\mathcal{L}_2^0(F^*)$

- Because the tangent space of model $\mathcal{F}_{np}$ wrt $\mathcal{A}_{np}$ at F^* is $\mathcal{L}_2^0(F^*)$, then $\psi_{F^*}(X)$ is the *unique* mean zero gradient and therefore it is equal to the efficient influence function.

- Thus, if $\psi_{F^*,eff,np}(Y,R,X)$ denotes the efficient influence function for the parameter $\beta(F)$ at F^* w.r.t $\mathcal{A}_{np}$ in the model $\mathcal{F}_{np}$, we have

$$\boxed{\psi_{F^*,eff,np}(Y,R,X) = E_{F^*}(Y|R=1,X) + \frac{R\{Y - E_{F^*}(Y|R=1,X)\}}{P_{F^*}(R=1|X)} - \beta(F^*)}$$

- Note, the condition $P_{F^*}(R=1|X) > \sigma_{F^*} > 0$ that holds for laws F^* in $\mathcal{F}_{np}$, ensures that $\psi_{F^*,eff,np}(Y,R,X)$ has finite variance under F^*, i.e. that $\psi_{F^*,eff,np}(Y,R,X) \in \mathcal{L}_2^0(F^*)$.

- Note also that for any $s_2(R,X)$ satisfying $E_{F^*}[s_2(R,X)|X] = 0$ it holds that

$$E_{F^*}\left[\psi_{F^*,eff,np}(Y,R,X)\, s_2(R,X)\right] = 0$$

- This can be checked by directly calculating the expectation on the RHS of the display, or otherwise by noticing that model $\mathcal{F}_{np}$ is a factorized likelihood model as we argue in the next slide.

- We now turn to the calculation of the gradients of $\beta(F)$ and the semiparametric efficient score in the semiparametric model $\mathcal{F}_{sem,fixed}$.

- Recall, this model is restricted only by the condition that the conditional probability $P(R=1|X=x)$ if fixed and known and equal to a given $\pi^*(x)$.

- Writing

$$\begin{aligned} g_1(y,r,x) &\equiv f_X(x)\, f_{Y|R,X}(y|r,x) \\ g_2(y,r,x) &\equiv f_{R|X}(r|x) \end{aligned}$$

we recognize that

$$f_{Y,R,X}(y,r,x) = g_1(y,r,x)\, g_2(y,r,x)$$

- Consequently, model $\mathcal{F}_{np}$ is a factorized likelihood model and model $\mathcal{F}_{sem,fixed}$ is like model $\mathcal{F}_1$ in chapter 6, in which g_2 is fixed and known and equal to

$$g_2^*(y,r,x) \equiv \pi^*(x)^r \{1 - \pi^*(x)\}^{1-r}$$

- Furthermore,

$$\beta(F) = E_F[E_F(Y|R=1,X)] = \int f_X(x) \int y f_{Y|R=1,X}(y|r=1,x)\, dy dx$$

depends only on $g_1(y,r,x)$.

- Now, suppose that in addition to the conditions (1)-(4) in slide 26, we also require from the submodels in class $\mathcal{A}_{np}$ that they also satisfy the assumptions (1)-(3) of Lemma 6.1 (see slide 9).
- It can be checked that the tangent space wrt this newly defined $\mathcal{A}_{np}$ remains equal to $\mathcal{L}_2^0(F^*)$.
- Thus, we conclude that

$$\boxed{\psi_{F^*,eff,np}(Y,R,X) = E_{F^*}(Y|R=1,X) + \frac{R\{Y - E_{F^*}(Y|R=1,X)\}}{P_{F^*}(R=1|X)} - \beta(F^*)}$$

is also the efficient score of $\beta(F)$ at F^* wrt the newly defined $\mathcal{A}_{np}$.

- Next, for each submodel

$$\mathcal{F}_{sub} = \left\{f_{Y,R,X}(y,r,x;\theta) = f_X(x;\theta) f_{R|X}(r|x;\theta) f_{Y|R,X}(y|r,x;\theta) : \theta \in \mathbb{R}^p\right\}$$

in class $\mathcal{A}_{np}$ define the submodel

$$\mathcal{F}_{sub,fixed} = \left\{f_{Y,R,X}(y,r,x;\theta) = f_X(x;\theta) f^*_{R|X}(r|x) f_{Y|R,X}(y|r,x;\theta) : \theta \in \mathbb{R}^p\right\}$$

where $f^*_{R|X}(r|x) = \pi^*(x)^r \{1 - \pi^*(x)\}^{1-r}$ is fixed and known.

- Define the class

$$\mathcal{A}_{sem,fixed} = \left\{\mathcal{F}_{sub,fixed} : \mathcal{F}_{sub} \text{ is in } \mathcal{A}_{np}\right\}$$

- From Lemma 6.1 we now conclude that $\psi_{F^*,eff,np}(Y,R,X)$ is also the efficient influence function of $\beta(F)$ at F^* wrt $\mathcal{A}_{sem,fixed}$ in the semiparametric model $\mathcal{F}_{sem,fixed}$. (to see this let $\mathcal{A}_{sem,fixed}$ play the role of $\mathcal{A}_1$ in that Lemma).
- Thus, the CR- bound of $\beta(F)$ at F^* is the same in model $\mathcal{F}_{np}$ wrt $\mathcal{A}_{np}$ as in model $\mathcal{F}_{sem,fixed}$ wrt $\mathcal{A}_{sem,fixed}$.

- Finally, consider the semiparametric model

$$\mathcal{F}_{sem,par} = \{F \in \mathcal{F}_{np} : P_F(R=1|X) = \pi(X;\alpha) \text{ for some } \alpha \in \mathbb{R}^r\}$$

where $\pi(x;\alpha)$ is a specified, i.e. known, function of x and α which is differentiable wrt α at every x.

- Model $\mathcal{F}_{sem,par}$ is also a semiparametric factorized likelihood model. Each $f(y,r,x)$ in $\mathcal{F}_{sem,par}$ factorizes as

$$f(y,r,x) = g_1(y,r,x)\, g_2(y,r,x;\alpha)$$

where

$$\begin{aligned} g_1(y,r,x) &\equiv f_X(x;\theta) f_{Y|R,X}(y|r,x;\theta) \\ g_2(y,r,x;\alpha) &\equiv \pi(x;\alpha)^r \{1 - \pi(x;\alpha)\}^{1-r} \end{aligned}$$

- Next, for each

$$\mathcal{F}_{sub} = \left\{f_{Y,R,X}(y,r,x;\theta) = f_X(x;\theta) f_{R|X}(r|x;\theta) f_{Y|R,X}(y|r,x;\theta) : \theta \in \Theta \subseteq \mathbb{R}^p\right\}$$

in class $\mathcal{A}_{np}$ define the submodel

$$\begin{aligned} \mathcal{F}_{sub,par} = & \{f_{Y,R,X}(y,r,x;\theta) = f_X(x;\theta) f_{R|X}(r|x;\alpha) f_{Y|R,X}(y|r,x;\theta) : \\ & \theta \in \Theta \subseteq \mathbb{R}^p, \alpha \in \Xi \subseteq \mathbb{R}^r\} \end{aligned}$$

- Define the class

$$\mathcal{A}_{sem,par} = \left\{\mathcal{F}_{sub,par} : \mathcal{F}_{sub} \in \mathcal{A}_{np}\right\}$$

- From Lemma 6.1 with $\mathcal{A}_{sem,par}$ and $\mathcal{F}_{sem,par}$ playing the role of $\mathcal{A}$ and $\mathcal{F}$ in that lemma, and $\mathcal{A}_{np,fixed}$ and $\mathcal{F}_{sem,fixed}$ playing the role of $\mathcal{A}_1$ and $\mathcal{F}_1$ in that Lemma, we conclude that the efficient influence function for $\beta(F)$ in both models is the same. Since we have just shown that the efficient influence function wrt $\mathcal{A}_{np,fixed}$ is $\psi_{F^*,eff,np}(Y,R,X)$ we thus conclude that

$$\boxed{\psi_{F^*,eff,np}(Y,R,X) = E_{F^*}(Y|R=1,X) + \frac{R\{Y - E_{F^*}(Y|R=1,X)\}}{P_{F^*}(R=1|X)} - \beta(F^*)}$$

is also the efficient influence function of $\beta(F)$ at F^* wrt $\mathcal{A}_{sem,par}$ in the semiparametric model $\mathcal{F}_{sem,par}$.

- Recall we had already shown that the CR-bounds of $\beta(F)$ in model $\mathcal{F}_{sem,fixed}$ wrt $\mathcal{A}_{sem,fixed}$ and in model $\mathcal{F}_{np}$ wrt $\mathcal{A}_{np}$ are the same.
- We thus conclude that: the semiparametric CR-bound of $\beta(F)$ is the same regardless of whether the missingness probabilities are known, modeled, or fully unspecified .

- An interesting application of the preceding results is that examination of the CR-bound allows us to quantify the information loss for estimating $E_{F^f}\left(Y^f\right)$ due to Y^f missing in some study subjects and the value of measuring the predictor X for reducing the information loss as will now see.
- First we re-write the efficient influence function as

$$\begin{aligned}
&\psi_{F^*,eff,np}(Y,R,X) = \\
= \;& E_{F^*}(Y|R=1,X) + \frac{R\{Y - E_{F^*}(Y|R=1,X)\}}{P_{F^*}(R=1|X)} - \beta(F^*) \\
= \;& \frac{R}{P_{F^*}(R=1|X)}Y - \left\{\frac{R}{P_{F^*}(R=1|X)} - 1\right\} E_{F^*}(Y|R=1,X) - \beta(F^*)
\end{aligned}$$

- Next, we notice that since by definition $Y = Y^f$ if $R=1$ then

$$\begin{aligned}
&\psi_{F^*,eff,np}(Y,R,X) = \\
= \;& \frac{R}{P_{F^*}(R=1|X)}Y^f - \left\{\frac{R}{P_{F^*}(R=1|X)} - 1\right\} E_{F^*}(Y|R=1,X) - \beta(F^*) \\
= \;& \left\{Y^f - \beta(F^*)\right\} + \left\{\frac{R}{P_{F^*}(R=1|X)} - 1\right\}\left\{Y^f - E_{F^*}(Y|R=1,X)\right\}
\end{aligned}$$

- Thus,

$$\begin{aligned}
&var_{F^*}\left[\psi_{F^*,eff,np}(Y,R,X)\right] \\
= \;& var_{F^{*f}}\left[\left\{Y^f - \beta(F^*)\right\} + \left\{\frac{R}{P_{F^{*f}}(R=1|X)} - 1\right\}\left\{Y^f - E_{F^{*f}}(Y|R=1,X)\right\}\right] \\
= \;& var_{F^{*f}}\left(Y^f\right) \; + \; var_{F^{*f}}\left[\left\{\frac{R}{P_{F^{*f}}(R=1|X)} - 1\right\}\left\{Y^f - E_{F^{*f}}(Y|R=1,X)\right\}\right]
\end{aligned}$$

because

$$\begin{aligned}
& E_{F^{*f}}\left[\left\{\frac{R}{P_{F^{*f}}(R=1|X)} - 1\right\}\left\{Y^f - E_{F^{*f}}(Y|R=1,X)\right\}\left\{Y^f - \beta(F^*)\right\}\right] \\
= \;& E_{F^{*f}}\left[\underbrace{\left\{\frac{E_{F^{*f}}\left[R|Y^f,X\right]}{P_{F^{*f}}(R=1|X)} - 1\right\}}_{=0 \text{ by MAR}}\left\{Y^f - E_{F^{*f}}(Y|R=1,X)\right\}\left\{Y^f - \beta(F^*)\right\}\right] \\
= \;& 0
\end{aligned}$$

- Notice that
$$var_{F^f}\left(Y^f\right)$$
is the CR-bound for estimating $E_{F^f}\left(Y^f\right)$ in the non-parametric model when Y^f is fully observed (i.e. not missing in any study unit)
- Whereas
$$var_{F^{*f}}\left(Y^f\right) + var_{F^{*f}}\left[\left\{\frac{R}{P_{F^{*f}}(R=1|X)} - 1\right\}\left\{Y^f - E_{F^{*f}}(Y|R=1,X)\right\}\right]$$
is the CR-bound for estimating $E_{F^f}\left(Y^f\right)$ when Y^f is missing in some study units and we (correctly) assume that the missingness process is MAR but assume nothing else.
- Then, the second term in the last display is the penalty for not observing Y^f in all study units when the data are MAR and we observe a predictor X in all study units.

Notice that

$$
\begin{aligned}
& var_{F^*f}\left[\left\{\frac{R}{P_{F^*f}(R=1|X)}-1\right\}\left\{Y^f-E_{F^*f}(Y|R=1,X)\right\}\right] \\
= \ & E_{F^*f}\left[\left\{\frac{R}{P_{F^*f}(R=1|X)}-1\right\}^2\left\{Y^f-E_{F^*f}(Y|R=1,X)\right\}^2\right] \\
= \ & E_{F^*f}\left[E_{F^*f}\left[\left\{\frac{R}{P_{F^*f}(R=1|X)}-1\right\}^2\middle| Y^f,X\right]\left\{Y^f-E_{F^*f}(Y|R=1,X)\right\}^2\right] \\
= \ & E_{F^*f}\left[\frac{E_{F^*f}\left[\{R-P_{F^*f}(R=1|X)\}^2|Y^f,X\right]}{P_{F^*f}(R=1|X)^2}\left\{Y^f-E_{F^*f}(Y|R=1,X)\right\}^2\right] \\
= \ & E_{F^*f}\left[\frac{\{1-P_{F^*f}(R=1|X)\}}{P_{F^*f}(R=1|X)}\left\{Y^f-E_{F^*f}(Y|R=1,X)\right\}^2\right] \\
= \ & E_{F^*f}\left[\frac{\{1-P_{F^*f}(R=1|X)\}}{P_{F^*f}(R=1|X)}\left\{Y^f-E_{F^*f}\left(Y^f|X\right)\right\}^2\right] \\
= \ & E_{F^*f}\left[\frac{\{1-P_{F^*f}(R=1|X)\}}{P_{F^*f}(R=1|X)}E_{F^*f}\left[\left\{Y^f-E_{F^*f}\left(Y^f|X\right)\right\}^2|X\right]\right] \\
= \ & E_{F^*f}\left[\frac{\{1-P_{F^*f}(R=1|X)\}}{P_{F^*f}(R=1|X)}var_{F^*f}\left(Y^f|X\right)\right]
\end{aligned}
$$

- The penalty, i.e. the increase, in the CR bound for not fully observing Y^f is therefore equal to

$$\boxed{E_{F^*f}\left[\frac{\{1-P_{F^*f}(R=1|X)\}}{P_{F^*f}(R=1|X)}var_{F^*f}\left(Y^f|X\right)\right]}$$

- Note that if X is a perfect predictor of Y^f, $var_{F^*f}\left(Y^f|X\right)=0$ and consequently the penalty is 0.
- Also, note that if X is not observed, then the penalty is $\frac{\{1-P_{F^*f}(R=1)\}}{P_{F^*f}(R=1)}var_{F^*f}\left(Y^f\right)$ (this formula is obtained by replacing X by a constant in the formula for the penalty).
- The expression for the CR-bound allows us to compute the fraction of the information loss due to missing Y^f that is recovered by measuring the predictor X of the missing outcome on all study subjects.

- Specifically, if we let CR_f, CR_X, CR_N denote respectively the C-R bounds for estimating $E_{F^*f}\left(Y^f\right)$ when Y^f is fully observed, when Y^f is missing in a subsample but X is observed in the entire sample and when Y^f is missing and X is not measured.
- The information lost due to Y^f missing when X is not measured is

$$(CR_f)^{-1}-(CR_N)^{-1}$$

- The information lost due to Y^f missing when X is measured is

$$(CR_f)^{-1}-(CR_X)^{-1}$$

- Then, the information recovered by measuring X is

$$\left[(CR_f)^{-1}-(CR_N)^{-1}\right]-\left[(CR_f)^{-1}-(CR_X)^{-1}\right]=(CR_X)^{-1}-(CR_N)^{-1}$$

- The fraction of the information lost due to Y^f missing when X is not measured that is recovered by measuring X is therefore

$$\frac{(CR_X)^{-1}-(CR_N)^{-1}}{(CR_f)^{-1}-(CR_N)^{-1}}=\frac{CR_f}{CR_X}\times\frac{CR_N-CR_X}{CR_N-CR_f}$$

- The expression for the fraction of information recovered can be simplified in the special case in which $P_{F^*f}(R=1|X)=\pi^*$ does not depend on X. In such case, after some algebra it can be checked that

$$
\begin{aligned}
& \frac{CR_f}{CR_X}\times\frac{CR_N-CR_X}{CR_N-CR_f}= \\
= \ & \frac{var_{F^*f}(Y)}{var_{F^*f}(Y)+\left(\frac{1}{\pi^*}-1\right)E_{F^*f}\left[var_{F^*f}(Y|X)\right]}\frac{var_{F^*f}(Y)-E_{F^*f}\left[var_{F^*f}(Y|X)\right]}{var_{F^*f}(Y)} \\
= \ & \frac{(1-\lambda)}{(1-\lambda)+(\lambda/\pi^*)}
\end{aligned}
$$

where $\lambda=E_{F^*f}\left[var_{F^*f}(Y|X)\right]/var_{F^*f}(Y)$.

- Note that when X is a perfect predictor of Y, $\lambda=0$ and the fraction recovered is equal to 1
- In contrast, when X has no predictive value, i.e. when $var_{F^*f}(Y|X)=var_{F^*f}(Y)$, $\lambda=1$ and the fraction recovered is 0.

- **Recap:** We have derived the efficient influence function and the CR-bound for

$$\beta(F) \equiv E_F\left[E_F(Y|R=1,X)\right]$$

in the three semiparametric models,

$$\mathcal{F}_{np} = \left\{F : E_F\left[\{E_F(Y|R=1,X)\}^2\right] < \infty, P_F(R=1|X) > \sigma_F > 0\right\},$$

$$\mathcal{F}_{sem,fixed} = \{F \in \mathcal{F}_{np} : P_F(R=1|X) = \pi^*(X)\}$$

where $\pi^*(x)$ is specified, i.e. known, and

$$\mathcal{F}_{sem,par} = \{F \in \mathcal{F}_{np} : P_F(R=1|X) = \pi(X;\alpha) \text{ for some } \alpha \in \Xi \subseteq \mathbb{R}^r\}$$

with respect to adequate classes of parametric models, $\mathcal{A}_{np}, \mathcal{A}_{sem,fixed}$ and $\mathcal{A}_{sem,par}$ respectively.

- We have seen that the efficient influence function for $\beta(F)$ is the same in the three models.

- We have seen that the tangent space $\Lambda_{\mathcal{F}_{np}}(F^*)$ at a given F^* wrt $\mathcal{A}_{np}$ of model $\mathcal{F}_{np}$ is $\mathcal{L}_2^0(F^*)$ and consequently, the efficient influence function is the UNIQUE mean zero gradient of $\beta(F)$ in model $\mathcal{F}_{np}$.

- We are now going to derive the set of mean zero gradients of $\beta(F)$ wrt $\mathcal{A}_{sem,fixed}$ in model $\mathcal{F}_{sem,fixed}$ and wrt $\mathcal{A}_{sem,par}$ in model $\mathcal{F}_{sem,par}$

- To derive the set of all influence functions in each model we must derive the tangent spaces of each model, characterize the orthogonal complement of the tangent space and finally construct each mean zero gradient as the sum of the efficient influence function with an element of the orthogonal complement of the tangent space.

- We thus start with the characterization of the tangent space $\Lambda_{\mathcal{F}_{sem,fixed}}(F^*)$ wrt $\mathcal{A}_{sem,fixed}$ of model $\mathcal{F}_{sem,fixed}$ and its orthogonal complement in $\mathcal{L}_2^0(F^*)$.

- We will use the following useful lemma (prove it in the privacy of your own room).

- **Lemma 7.1:** Suppose that H_1, H_2 and H_3 are three subspaces of a Hilbert space such that

$$H_1 = H_2 \oplus H_3$$

then

$$H_2^{\perp} = H_1^{\perp} \oplus H_3$$

In particular, if

$$\mathcal{L}_2^0(F^*) = H_2 \oplus H_3$$

then

$$H_2^{\perp} = H_3$$

- To compute $\Lambda_{\mathcal{F}_{sem,fixed}}(F^*)$, we recall that since $\mathcal{F}_{np}$ is a factorized likelihood model, see slide 17 of this chapter, the class $\mathcal{A}_{np}$ verifies the assumptions of Lemma 6.1 and $\mathcal{A}_{sem,fixed}$ is defined as $\mathcal{A}_1$ in that Lemma, then we can apply the conclusions of Lemma 6.1.

- To apply lemma 6.1, we must first define the analogous of the semip. model $\mathcal{F}_2$ and the class $\mathcal{A}_2$ in that Lemma and compute its tangent space.

- The analogous of $\mathcal{F}_2$ in Lemma 6.1 corresponds in our problem to the model

$$\mathcal{F}_{sem,miss} = \left\{f_{Y,R,X}(y,r,x) = f_X^*(x)\, f_{R|X}(r|x)\, f_{Y|R,X}^*(y|r,x) : \ f_{R|X}(1|x) > \sigma_f > 0\right\}$$

- To define the analogous of class $\mathcal{A}_2$ we consider for every

$$\mathcal{F}_{sub} = \left\{f_{Y,R,X}(y,r,x;\theta) = f_X(x;\theta)\, f_{R|X}(r|x;\theta)\, f_{Y|R,X}(y|r,x;\theta) : \theta \in \Theta \subseteq \mathbb{R}^p\right\}$$

in class $\mathcal{A}_{np}$, the submodel

$$\mathcal{F}_{sub,miss} = \left\{f_{Y,R,X}(y,r,x;\theta) = f_X^*(x)\, f_{R|X}(r|x;\theta)\, f_{Y|R,X}^*(y|r,x) : \theta \in \Theta \subseteq \mathbb{R}^p\right\}$$

- Then, the analogous of class $\mathcal{A}_2$ in Lemma 6.1 is the class

$$\mathcal{A}_{sem,miss} = \left\{\mathcal{F}_{sub,miss} : \mathcal{F}_{sub} \in \mathcal{A}_{np}\right\}$$

- Note that $\mathcal{F}_{sub,miss}$ is a parametric model for the law of (Y,R,X) in which f_X and $f_{Y|R,X}$ are fixed and known and the law of $R|X$ is unrestricted.

- Because the class $\mathcal{A}_{sem,miss}$ is essentially restricted only by smoothness conditions, it is natural to conjecture that the tangent space $\Lambda_{\mathcal{F}_{sem,miss}}(F^*)$ at F^* wrt $\mathcal{A}_{sem,miss}$ in model $\mathcal{F}_{sem,miss}$ is

$$\Lambda_{\mathcal{F}_{sem,miss}}(F^*) = \{s(R,X) \in \mathcal{L}_2(F^*) : E_{F^*}[s(R,X)|X] = 0\}$$

- This conjecture is indeed correct.

- The conjecture can be checked, once again, by considering for each $s(R,X)$ in the set in the RHS of the last display, the submodel

$$\left\{f_{Y,R,X}(y,r,x;\theta) = f_X^*(x)\, f_{R|X}(r|x;\theta)\, f_{Y|R,X}^*(y|r,x) : \theta \in \mathbb{R}\right\}$$

where

$$f_{R|X}(r|x;\theta_2) = f_{R|X}^*(r|x)\, k(\theta s(r,x))\, c(x;\theta)$$

$$k(u) = 2\left[1 + \exp(-2u)\right]^{-1}$$

and

$$c(x;\theta) = \left[f_{R|X}^*(1|x)\, k(\theta s(1,x)) + f_{R|X}^*(0|x)\, k(\theta s(0,x))\right]^{-1}$$

This submodel is regular, goes through F^* at $\theta = 0$ and its score is equal to $s(R,X)$. This shows that the conjectured set is included in $\Lambda_{\mathcal{F}_{sem,fixed}}(F^*)$. The inclusion in the other direction is straightforward as any score in model $\mathcal{F}_{sub,miss}$ must be a score for the law of R given X and, as such, it must be a function of just (R,X) and have mean zero given X.

- Now, letting

$\Lambda_{\mathcal{F}_{sub,fixed}}(F^*)$ denote the tangent space at F^* wrt $\mathcal{A}_{sub,fixed}$ in model $\mathcal{F}_{sub,fixed}$

we conclude by Lemma 6.1 that

$$\Lambda_{\mathcal{F}_{np}}(F^*) = \Lambda_{\mathcal{F}_{sub,fixed}}(F^*) \oplus \Lambda_{\mathcal{F}_{sub,miss}}(F^*)$$

- Since we know that $\Lambda_{\mathcal{F}_{np}}(F^*) = \mathcal{L}_2^0(F^*)$, we then conclude that

$$\Lambda_{\mathcal{F}_{sub,fixed}}(F^*)^{\perp} = \Lambda_{\mathcal{F}_{sub,miss}}(F^*)$$

- So finally, we conclude that

$$\Lambda_{\mathcal{F}_{sub,fixed}}(F^*)^{\perp} = \{s(R,X) \in \mathcal{L}_2(F^*) : E_{F^*}[s(R,X)|X] = 0\}$$

- Thus, $\psi_{F^*}(Y,R,X)$ is a mean zero gradient of $\beta(F)$ wrt at F^* wrt $\mathcal{A}_{sub,fixed}$ in model $\mathcal{F}_{sub,fixed}$ iff it is of the form

$$\psi_{F^*}(Y,R,X) = \psi_{F^*,eff,np}(Y,R,X) + s(R,X)$$

for some $s(R,X)$ satisfying $E_{F^*}[s(R,X)|X] = 0$.

- We can write any $\psi_{F^*,eff,np}(Y,R,X) + s(R,X)$ more explicitly.
- First, we note that $s(R,X)$ satisfies $E_{F^*}[s(R,X)|X] = 0$ iff

$$s(R,X) = d(X)[R - P_{F^*}(R=1|X)]$$

for some $d(X)$ (in fact, for $d(X) = s(1,X) - s(0,X)$).

- Next, we write

$$\psi_{F^*,eff,np}(Y,R,X) = E_{F^*}(Y|R=1,X) + \frac{R\{Y - E_{F^*}(Y|R=1,X)\}}{P_{F^*}(R=1|X)} - \beta(F^*)$$

$$= \frac{RY}{P_{F^*}(R=1|X)} - \left\{\frac{R}{P_{F^*}(R=1|X)} - 1\right\} E_{F^*}(Y|R=1,X) - \beta(F^*)$$

$$= \frac{R\{Y - \beta(F^*)\}}{P_{F^*}(R=1|X)} - \{R - P_{F^*}(R=1|X)\} \underbrace{\frac{\{E_{F^*}(Y|R=1,X) - \beta(F^*)\}}{P_{F^*}(R=1|X)}}_{\text{a fcn of } X \text{ only}}$$

- We thus conclude that $\psi_{F^*}(Y,R,X)$ is a mean zero gradient of $\beta(F)$ wrt at F^* wrt $\mathcal{A}_{sub,fixed}$ in model $\mathcal{F}_{sub,fixed}$ iff is of the form

$$\boxed{\psi_{F^*}(Y,R,X) = \frac{R}{P_{F^*}(R=1|X)}\{Y - \beta(F^*)\} + d(X)\{R - P_{F^*}(R=1|X)\}}$$

for some $d(X)$.

- We now derive the set of mean zero gradients of $\beta(F)$ at F^* wrt $\mathcal{A}_{sem,par}$ in model

$$\mathcal{F}_{sem,par} = \{F \in \mathcal{F}_{np} : P_F(R=1|X) = \pi(X;\alpha) \text{ for some } \alpha \in \Xi \subseteq \mathbb{R}^r\}$$

- Since the only restriction imposed by $\mathcal{F}_{sem,par}$ is a parametric model for the law of $R|X$, and since the class $\mathcal{A}_{sem,par}$ only imposes smoothness conditions, it is natural to conjecture that the tangent space $\Lambda_{\mathcal{F}_{sub,par}}(F^*)$ wrt to $\mathcal{A}_{sem,par}$ of model $\mathcal{F}_{sub,par}$ is equal to

$$\Omega_1(F^*) \oplus \Omega_3(F^*) \oplus \Omega_{2,par}(F^*)$$

where

$$\Omega_1(F^*) \equiv \{s_1(X) \in \mathcal{L}_2(F^*) : E_{F^*}[s_1(X)] = 0\}$$

$$\Omega_3(F^*) \equiv \{s_3(Y,R,X) \in \mathcal{L}_2(F^*) : E_{F^*}[s_3(Y,R,X)|R,X] = 0\}$$

$$\Omega_{2,par}(F^*) \equiv \{a^T s_\alpha(R,X;\alpha^*), a \in \mathbb{R}^r\}$$

and

$$s_\alpha(R,X;\alpha^*) = \frac{d}{d\alpha}\log\left[\pi(X;\alpha)^R\{1-\pi(X;\alpha)\}^{1-R}\right]\Big|_{\alpha=\alpha^*}$$

- This conjucture is indeed true.

- The conjecture can be checked, once again, by exhibiting for any given $s_1(X) \in \Omega_1(F^*)$, $s_3(Y,R,X) \in \Omega_3(F^*)$ and $a \in \mathbb{R}^r$, a submodel whose score is such that a linear combination of its components is equal to

$$s_1(X) + s_3(Y,R,X) + a^T s_\alpha(R,X)$$

- Such submodel is (check it in the privacy of your own room)

$$\{f_{Y,R,X}(y,r,x;\theta,\alpha) = f_X(x;\theta_1) f_{R|X}(r|x;\alpha) f_{Y|R,X}(y|r,x;\theta_3) : \theta = (\theta_1,\theta_3) \in \mathbb{R}^2, \alpha \in \mathbb{R}^r\}$$

where

$$f_X(x;\theta_1) = f_X^*(x) k(\theta_1 s_1(x)) c_1(\theta_1)$$

$$f_{Y|R,X}(y|r,x;\theta_3) = f_{R|X}^*(y|r,x) k(\theta_3 s_3(y,r,x)) c_3(r,x;\theta_3)$$

$$k(u) = 2[1 + \exp(-2u)]^{-1}$$

and

$$c_1(\theta_1) = \left[\int f_X^*(x) k(\theta_1 s_1(x)) dx\right]^{-1}$$

$$c_3(r,x;\theta_3) = \left[\int f_{R|X}^*(y|r,x) k(\theta_3 s_3(y,r,x)) dy\right]^{-1}$$

- Note that the sets $\Omega_1(F^*), \Omega_2(F^*)$ and $\Omega_{2,par}(F^*)$ are mutually orthogonal in $\mathcal{L}_2(F^*)$. This can be seen by directly checking this condition or otherwise by applying Lemma 6.1, after noticing that model $\mathcal{F}_{sub,par}$ is a factorized likelihood semiparametric model.

- To compute the set of mean zero gradients of $\beta(F)$ at F^* wrt $\mathcal{A}_{sem,par}$ we must therefore compute

$$\Lambda_{\mathcal{F}_{sub,par}}(F^*)^{\perp} = [\Omega_1(F^*) \oplus \Omega_3(F^*) \oplus \Omega_{2,par}(F^*)]^{\perp}$$

- We carry out this calculation as follows.
- First we notice that

$$\mathcal{L}_2^0(F^*) = [\Omega_1(F^*) \oplus \Omega_3(F^*) \oplus \Omega_2(F^*)]$$

where

$$\Omega_2(F^*) = \{s_2(R,X) \in \mathcal{L}_2(F^*) : E_{F^*}[s_2(R,X)|X] = 0\}$$

- To check this, simply decompose any $s(Y,R,X)$ in $\mathcal{L}_2^0(F^*)$ as

$$\begin{aligned} s(Y,R,X) &= \underbrace{\{s(Y,R,X) - E_{F^*}[s(Y,R,X)|R,X]\}}_{\in\,\Omega_3(F^*)} \\ &\quad + \underbrace{\{E_{F^*}[s(Y,R,X)|R,X] - E_{F^*}[s(Y,R,X)|X]\}}_{\in\,\Omega_2(F^*)} \\ &\quad + \underbrace{\{E_{F^*}[s(Y,R,X)|X] - E_{F^*}[s(Y,R,X)]\}}_{\in\,\Omega_1(F^*)} \end{aligned}$$

- Next we note that

$$\Omega_{2,par}(F^*) \subset \Omega_2(F^*)$$

and that the set $\Omega_{2,par}(F^*)$ can be re-written as

$$\Omega_{2,par}(F^*) = \{\Pi[s_2(R,X)|\Omega_{2,par}(F^*)] : s_2(R,X) \in \Omega_2(F^*)\}$$

- Furthermore, we also note that any $s_2(R,X) \in \Omega_2(F^*)$ can be decomposed as

$$s_2(R,X) = \{s_2(R,X) - \Pi[s_2(R,X)|\Omega_{2,par}(F^*)]\} + \Pi[s_2(R,X)|\Omega_{2,par}(F^*)]$$

- So, we can write

$$\Omega_2(F^*) = \Omega_{2,resid}(F^*) \oplus \Omega_{2,par}(F^*)$$

where

$$\Omega_{2,resid}(F^*) \equiv \{s_2(R,X) - \Pi[s_2(R,X)|\Omega_{2,par}(F^*)] : s_2(R,X) \in \Omega_2(F^*)\}$$

- Thus, we conclude that can write,

$$\mathcal{L}_2^0(F^*) = \left[\Omega_1(F^*) \oplus \Omega_3(F^*) \oplus \Omega_{2,resid}(F^*)\right] \oplus \Omega_{2,par}(F^*)$$

- Consequently, we conclude that

$$\Lambda_{\mathcal{F}_{sub,par}}(F^*)^{\perp} = [\Omega_1(F^*) \oplus \Omega_3(F^*) \oplus \Omega_{2,par}(F^*)]^{\perp} = \Omega_{2,resid}(F^*)$$

- From $\Lambda_{\mathcal{F}_{sub,par}}(F^*)^{\perp} = \Omega_{2,resid}(F^*)$ we conclude that $\psi_{F^*}(Y,R,X)$ is a mean zero gradient of $\beta(F)$ at F^* wrt $\mathcal{A}_{sub,par}$ in model $\mathcal{F}_{sub,par}$ if and only if

$$\boxed{\psi_{F^*}(Y,R,X) = \psi_{F^*,eff,np}(Y,R,X) + s_2(R,X) - \Pi[s_2(R,X)|\Omega_{2,par}(F^*)]}$$

for some $s_2(R,X)$ satisfying $E_{F^*}[s_2(R,X)|X] = 0$.

- In the next slide we rewrite the expression for $\psi_{F^*}(Y,R,X)$ more explicitly.

- First we note that

$$\Pi\left[\psi_{F^*,eff,np}(Y,R,X)|\Omega_{2,par}\left(F^*\right)\right] = 0$$

because, recall, for any $s_2(R,X) \in \Omega_2(F^*)$ it holds that, $E_{F^*}\left[\psi_{F^*,eff,np}(Y,R,X)\, s_2(R,X)\right] = 0$, and $\Omega_{2,par}(F^*) \subseteq \Omega_2(F^*)$.

- Therefore,

$$\begin{aligned} & \psi_{F^*,eff,np}(Y,R,X) + s_2(R,X) - \Pi\left[s_2(R,X)|\Omega_{2,par}\left(F^*\right)\right] \\ = \; & \psi_{F^*,eff,np}(Y,R,X) + s_2(R,X) - \Pi\left[\psi_{F^*,eff,np}(Y,R,X) + s_2(R,X)|\Omega_{2,par}\left(F^*\right)\right] \end{aligned}$$

- Furthermore, recalling that

$$\psi_{F^*,eff,np}(Y,R,X) + s_2(R,X) = \frac{R[Y - \beta(F^*)]}{P_{F^*}(R=1|X)} + d(X)[R - P_{F^*}(R=1|X)]$$

for some $d(X)$,

- we conclude that any mean zero gradient of a $\beta(F)$ at F^* wrt class $\mathcal{A}_{sem,par}$ in model $\mathcal{F}_{sem,par}$ is of the form

$$\begin{aligned} \psi_{F^*}(X) &= \frac{R[Y - \beta(F^*)]}{P_{F^*}(R=1|X)} + d(X)[R - P_{F^*}(R=1|X)] \\ &\quad - \Pi\left[\frac{R[Y - \beta(F^*)]}{P_{F^*}(R=1|X)} + d(X)[R - P_{F^*}(R=1|X)]|\Omega_{2,par}\left(F^*\right)\right] \end{aligned}$$

for some $d(X)$.