# Efficient inference for finite-dimensional parameters

Yangjianchen Xu

Department of Biostatistics
University of North Carolina at Chapel Hill

11/04/2021

# Overview

- Score Functions and Estimating Equations
- Example: Semiparametric regression model
- Example: The Cox Model for Right-Censored Data

# Efficient score function

- Semiparametric model $\{P_{\theta,\eta} : \theta \in \Theta, \eta \in H\}$, where $\Theta$ is an open subset of $\mathbb{R}^k$ and $H$ is an arbitrary set that may be infinite dimensional.
- Parameter of interest: $\psi\left(P_{\theta,\eta}\right) = \theta$.
- Tangent sets can be used to develop an efficient estimator for $\psi\left(P_{\theta,\eta}\right) = \theta$ through the formation of an efficient score function.
- Consider submodels of the form $\{P_{\theta+ta,\eta_t}, t \in N_\epsilon\}$ that are differentiable in quadratic mean with score function

$$\partial \log dP_{\theta+ta,\eta_t} / \left(\partial t\right)\Big|_{t=0} = a' \dot{\ell}_{\theta,\eta} + g$$

  - $a \in \mathbb{R}^k$
  - $\dot{\ell}_{\theta,\eta} : \mathcal{X} \mapsto \mathbb{R}^k$ is the ordinary score for $\theta$ when $\eta$ is fixed
  - $g : \mathcal{X} \mapsto \mathbb{R}$ is an element of a tangent set $\dot{\mathcal{P}}^{(\eta)}_{P_{\theta,\eta}}$ for the submodel $\mathcal{P}_\theta = \{P_{\theta,\eta} : \eta \in H\}$ (holding $\theta$ fixed).

# Efficient score function

- The tangent set for the full model is

$$\dot{\mathcal{P}}_{P_{\theta,\eta}} = \left\{ a'\dot{\ell}_{\theta,\eta} + g : a \in \mathbb{R}^k, g \in \dot{\mathcal{P}}^{(\eta)}_{P_{\theta,\eta}} \right\}$$

- $\psi\left(P_{\theta+ta,\eta_t}\right) = \theta + ta$ is clearly differentiable with respect to $t$:

$$\left.\frac{\partial\psi\left(P_{\theta+ta,\eta_t}\right)}{\partial t}\right|_{t=0} = a$$

- The efficient influence function $\tilde{\psi}_{\theta,\eta} : \mathcal{X} \mapsto \mathbb{R}^k$ satisfies:

$$a = P\left[\tilde{\psi}_{\theta,\eta}\left(\dot{\ell}'_{\theta,\eta}a + g\right)\right] \quad (1)$$

- Setting $a = 0$, we see $0 = P\left[\tilde{\psi}_{\theta,\eta}g\right]$, $\tilde{\psi}_{\theta,\eta}$ must be uncorrelated with all of the elements of $\dot{\mathcal{P}}^{(\eta)}_{P_{\theta,n}}$.

# Efficient score function

- Define $\Pi_{\theta,\eta}$ to be the orthogonal projection onto the closed linear span of $\dot{\mathcal{P}}^{(\eta)}_{P_{\theta,\eta}}$ in $L_2^0\left(P_{\theta,\eta}\right)$.
- For any $h \in L_2^0\left(P_{\theta,\eta}\right), h = h - \Pi_{\theta,\eta}h + \Pi_{\theta,\eta}h$, where $\Pi_{\theta,\eta}h \in \dot{\mathcal{P}}^{(\eta)}_{P_{\theta,\eta}}$ but $P\left[\left(h - \Pi_{\theta,\eta}h\right)g\right] = 0$ for all $g \in \dot{\mathcal{P}}^{(\eta)}_{P_{\theta,\eta}}$.
- The efficient score function for $\theta$ is $\tilde{\ell}_{\theta,\eta} = \dot{\ell}_{\theta,\eta} - \Pi_{\theta,\eta}\dot{\ell}_{\theta,\eta}$, while the efficient information matrix for $\theta$ is $\tilde{I}_{\theta,\eta} = P\left[\tilde{\ell}_{\theta,\eta}\tilde{\ell}'_{\theta,\eta}\right]$.

# Efficient influence function

- Provided that $\tilde{I}_{\theta,\eta}$ is nonsingular, the function $\tilde{\psi}_{\theta,\eta} = \tilde{I}_{\theta,\eta}^{-1}\tilde{\ell}_{\theta,\eta}$ satisfies (1) for all $a \in \mathbb{R}^k$ and all $g \in \dot{\mathcal{P}}_{P_{\theta,\eta}}^{(\eta)}$.
- Thus the functional (parameter) $\psi(P_{\theta,\eta}) = \theta$ is differentiable at $P_{\theta,\eta}$ relative to the tangent set $\dot{\mathcal{P}}_{P_{\theta,\eta}}$, with efficient influence function $\tilde{\psi}_{\theta,\eta}$.
- Hence the search for an efficient estimator of $\theta$ is over if one can find an estimator $T_n$ satisfying

$$\sqrt{n}\,(T_n - \theta) = \sqrt{n}\mathbb{P}_n\tilde{\psi}_{\theta,\eta} + o_P(1).$$

- Note that $\tilde{I}_{\theta,\eta} = I_{\theta,\eta} - P\left[\Pi_{\theta,\eta}\dot{\ell}_{\theta,\eta}\left(\Pi_{\theta,\eta}\dot{\ell}_{\theta,\eta}\right)'\right]$, where $I_{\theta,\eta} = P\left[\dot{\ell}_{\theta,\eta}\dot{\ell}'_{\theta,\eta}\right]$.

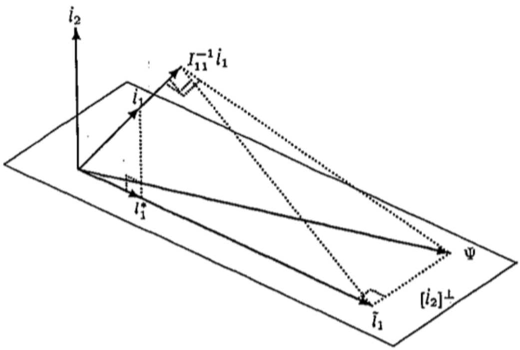


Figure: Score and influence function projections (Bickel et al., 1993)

An intuitive justification for the form of the efficient score is that some information for estimating $\theta$ is lost due to a lack of knowledge about $\eta$. The amount subtracted off of the efficient score, $\Pi_{\theta,\eta}\dot{\ell}_{\theta,\eta}$, is the minimum possible amount for regular estimators when $\eta$ is unknown.

| Name | Notation | Model | |
|---|---|---|---|
| | | $\mathbf{P}$ | $\mathbf{P}_1(\eta_0)$ |
| Efficient score | $\mathbf{l}_1^*(\cdot, P \mid \nu, \cdot)$ | $\mathbf{l}_1^* = \dot{\mathbf{l}}_1 - I_{12}I_{22}^{-1}\dot{\mathbf{l}}_2$ | $\dot{\mathbf{l}}_1$ |
| Information | $I(P \mid \nu, \cdot)$ | $E\mathbf{l}_1^*\mathbf{l}_1^{*T} = I_{11} - I_{12}I_{22}^{-1}I_{21} \equiv I_{11\cdot 2}$ | $I_{11}$ |
| Efficient influence function | $\tilde{\mathbf{l}}_1(\cdot, P \mid \nu, \cdot)$ | $\tilde{\mathbf{l}}_1 = I^{11}\dot{\mathbf{l}}_1 + I^{12}\dot{\mathbf{l}}_2 = I_{11\cdot 2}^{-1}\mathbf{l}_1^* = I_{11}^{-1}\dot{\mathbf{l}}_1 - I_{11}^{-1}I_{12}\tilde{\mathbf{l}}_2$ | $I_{11}^{-1}\dot{\mathbf{l}}_1$ |
| Information bound | $I^{-1}(P \mid \nu, \cdot)$ | $I^{11} = I_{11\cdot 2}^{-1} = I_{11}^{-1} + I_{11}^{-1}I_{12}I_{22\cdot 1}^{-1}I_{21}I_{11}^{-1}$ | $I_{11}^{-1}$ |

Figure: Efficient score functions, efficient influence functions, information, and inverse information for the two models $\mathbf{P}$ and $\mathbf{P}_1(\eta_0)$. (Bickel et al., 1993)

# Example: Semiparametric regression model

- Semiparametric regression model: $Y = \beta' Z + e$
- $E[e \mid Z] = 0$ and $E\left[e^2 \mid Z\right] \leq K < \infty$ almost surely
- We observe $(Y, Z)$, with the joint density $\eta$ of $(e, Z)$ satisfying $\int_{\mathbb{R}} e\eta(e, Z)de = 0$ almost surely.
- Assume $\eta$ has partial derivative with respect to the first argument, $\dot{\eta}_1$, satisfying $\dot{\eta}_1/\eta \in L_2\left(P_{\beta,\eta}\right)$, and hence $\dot{\eta}_1/\eta \in L_2^0\left(P_{\beta,\eta}\right)$, where $P_{\beta,\eta}$ is the joint distribution of $(Y, Z)$.
- The Euclidean parameter of interest is $\theta = \beta$.
- The score for $\beta$, assuming $\eta$ is known, is

$$\dot{\ell}_{\beta,\eta} = -Z\left(\dot{\eta}_1/\eta\right)\left(Y - \beta' Z, Z\right),$$

where we use the shorthand $(f/g)(u, v) = f(u, v)/g(u, v)$ for ratios of functions.

# Example: Semiparametric regression model

- The tangent set $\dot{\mathcal{P}}^{(\eta)}_{P_{\beta,\eta}}$ for $\eta$ is the subset of $L_2^0(P_{\beta,\eta})$ which consists of all functions $g(e,Z) \in L_2^0(P_{\beta,\eta})$ which satisfy

$$E[eg(e,Z) \mid Z] = \frac{\int_{\mathbb{R}} eg(e,Z)\eta(e,Z)de}{\int_{\mathbb{R}} \eta(e,Z)de} = 0$$

  almost surely.
- One can also show that this set is the orthocomplement in $L_2^0(P_{\beta,\eta})$ of all functions of the form $ef(Z)$, where $f$ satisfies $P_{\beta,\eta}f^2(Z) < \infty$.
- Thus $\tilde{\ell}_{\beta,\eta} = (I - \Pi_{\beta,\eta})\,\dot{\ell}_{\beta,\eta}$ is the projection in $L_2^0(P_{\beta,\eta})$ of $-Z(\dot{\eta}_1/\eta)(e,Z)$ onto $\left\{ef(Z) : P_{\beta,\eta}f^2(Z) < \infty\right\}$.

# Example: Semiparametric regression model

- Thus

$$\tilde{\ell}_{\beta,\eta}(Y,Z) = \frac{-Ze\int_{\mathbb{R}} \dot{\eta}_1(e,Z)ede}{P_{\beta,\eta}\left[e^2 \mid Z\right]} = -\frac{Ze(-1)}{P_{\beta,\eta}\left[e^2 \mid Z\right]} = \frac{Z\left(Y-\beta' Z\right)}{P_{\beta,\eta}\left[e^2 \mid Z\right]}$$

- The second-to-last step follows from the identity $\int_{\mathbb{R}} \dot{\eta}_1(e,Z)$ ede $= \partial \int_{\mathbb{R}} \eta(te,Z)de/\left.(\partial t)\right|_{t=1}$
- When the function $z \mapsto P_{\beta,\eta}\left[e^2 \mid Z=z\right]$ is non-constant in $z$,$\tilde{\ell}_{\beta,\eta}(Y,Z)$ is not proportional to $Z\left(Y-\beta' Z\right)$.
- Computation of the efficient estimator requires estimation of the function $z \mapsto \mathrm{E}\left[e^2 \mid Z=z\right]$.

# Score and information operators

- Two very useful tools for computing efficient scores are score and information operators.
- Sometimes it is easier to represent an element $g$ in $\dot{\mathcal{P}}^{(\eta)}_{P_{\theta,\eta}}$, as $B_{\theta,\eta}b$, where $b$ is an element of another set $\mathbb{H}_\eta$ and $B_{\theta,\eta}$ is an operator satisfying $\dot{\mathcal{P}}^{(\eta)}_{P_{\theta,\eta}} = \{B_{\theta,\eta}b : b \in \mathbb{H}_\eta\}$.
- The adjoint of the score operator $B_{\theta,\eta} : \mathbb{H}_\eta \mapsto L_2^0\left(P_{\theta,\eta}\right)$ is another operator $B^*_{\theta,\eta} : L_2^0\left(P_{\theta,\eta}\right) \mapsto \overline{\text{lin}}\mathbb{H}_\eta$ which is similar in spirit to a transpose for matrices and satisfies

$$\langle B_{\theta,\eta}b, h\rangle_{L_2^0(P_{\theta,\eta})} = \langle b, B^*_{\theta,\eta}h\rangle_{\overline{\text{lin}}\mathbb{H}_\eta}$$

  for all $b \in \mathbb{H}_\eta$ and $h \in L_2^0(P_{\theta,\eta})$.
- If $B^*_{\theta,\eta}B_{\theta,\eta}$ has an inverse, then it can be shown that the efficient score for $\theta$ has the form $\tilde{\ell}_{\theta,\eta} = \left(I - B_{\theta,\eta}\left[B^*_{\theta,\eta}B_{\theta,\eta}\right]^{-1}B^*_{\theta,\eta}\right)\dot{\ell}_{\theta,\eta}$.

# Example: Cox model for right-censored data

- We observe a sample of $n$ realizations of $X = (V, d, Z)$, where $V = T \wedge C, d = 1\{V = T\}, Z \in \mathbb{R}^k$ is a covariate vector, $T$ is a failure time, and $C$ is a censoring time.
- $T$ and $C$ are independent given $Z$.
- $T$ given $Z$ has integrated hazard function $e^{\beta' Z}\Lambda(t)$ for $\beta$ in an open subset $B \subset \mathbb{R}^k$ and $\Lambda$ is continuous and monotone increasing with $\Lambda(0) = 0$
- The censoring distribution does not depend on $\beta$ or $\Lambda$ (i.e., censoring is uninformative).
- The counting and at-risk processes: $N(t) = 1\{V \leq t\}d$ and $Y(t) = 1\{V \geq t\}$
- Martingale: $M(t) = N(t) - \int_0^t Y(s)e^{\beta' Z}d\Lambda(s)$.
- For some $0 < \tau < \infty$ with $P\{C \geq \tau\} > 0$, let $H$ be the set of all $\Lambda$ 's satisfying our criteria with $\Lambda(\tau) < \infty$.

# Example: Cox model for right-censored data

- Now the set of models $\mathcal{P}$ is indexed by $\beta \in B$ and $\Lambda \in H$. Let $P_{\beta,\Lambda}$ be the distribution of $(V, d, Z)$ corresponding to the given parameters.
- The likelihood for a single observation is thus proportional to

$$p_{\beta,\Lambda}(X) = \left[e^{\beta' Z}\lambda(V)\right]^d \exp\left[-e^{\beta' Z}\Lambda(V)\right].$$

- Let $L_2(\Lambda)$ be the set of measurable functions $b : [0, \tau] \mapsto \mathbb{R}$ with $\int_0^\tau b^2(s)d\Lambda(s) < \infty$.
- If $b \in L_2(\Lambda)$ is bounded, then $\Lambda_t(s) = \int_0^s e^{tb(u)}d\Lambda(u) \in H$ for all $t$.

# Example: Cox model for right-censored data

We calculate

$$\begin{aligned}\log p_{\beta+ta,\Lambda_t}(X) &= d\left[\log\{e^{(\beta+ta)'Z}\lambda_t(V)\}\right] - e^{(\beta+ta)'Z}\Lambda_t(V) \\ &= d\left[(\beta+ta)'Z + tb(V) + \log\{\lambda(V)\}\right] \\ &\quad - e^{(\beta+ta)'Z}\int_0^V e^{tb(u)}d\Lambda(u)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial t}\log p_{\beta+ta,\Lambda_t}(X) = d\left[a'Z + b(V)\right] \\ - a'Ze^{(\beta+ta)'Z}\int_0^V e^{tb(u)}d\Lambda(u) \\ - e^{(\beta+ta)'Z}\int_0^V b(u)e^{tb(u)}d\Lambda(u)\end{aligned}$$

# Example: Cox model for right-censored data

Thus,

$$
\begin{aligned}
\frac{\partial}{\partial t} \log p_{\beta+ta,\Lambda_t}(X) \mid_{t=0} = {} & d\left[a'Z + b(V)\right] \\
& - a'Ze^{\beta'Z}\Lambda(V) - e^{\beta'Z}\int_0^V b(u)d\Lambda(u) \\
= {} & \int_0^\tau \left[a'Z + b(s)\right] dM(s)
\end{aligned}
$$

- The score function for $\beta$ is therefore $\dot{\ell}_{\beta,\Lambda}(X) = ZM(\tau)$.
- The score function for $\Lambda$ is $\int_0^\tau b(s)dM(s)$.
- In fact, one can show that there exists one-dimensional submodels $\Lambda_t$ such that $\log p_{\beta+ta,\Lambda_t}$ is differentiable with score $a'\dot{\ell}_{\beta,\Lambda}(X) + \int_0^\tau b(s)dM(s)$, for any $b \in L_2(\Lambda)$ and $a \in \mathbb{R}^k$.

# Example: Cox model for right-censored data

- The score operator $B_{\beta,\Lambda} : L_2(\Lambda) \mapsto L_2^0\left(P_{\beta,\Lambda}\right)$ is given by

$$B_{\beta,\Lambda}(b) = \int_0^\tau b(s) dM(s),$$

which generates the tangent set for $\Lambda, \dot{\mathcal{P}}_{P_{\beta,\Lambda}}^{(\Lambda)} \equiv \{B_{\beta,\Lambda} b : b \in L_2(\Lambda)\}$.

- It can be shown that this tangent space spans all square-integrable score functions for $\Lambda$ generated by parametric submodels.
- The adjoint operator $B_{\beta,\Lambda}^* : L_2\left(P_{\beta,\Lambda}\right) \mapsto L_2(\Lambda)$ satisfies

$$\langle B_{\beta,\Lambda} b, h \rangle_{L_2(P_{\beta,\Lambda})} = \langle b, B_{\beta,\Lambda}^* h \rangle_{L_2(\Lambda)}.$$

- $LHS = P_{\beta,\Lambda}\left[h(X) \int_0^\tau b(s) dM(s)\right] = \int_0^\tau b(s) P_{\beta,\Lambda}\left[h(X) dM(s)\right]$
- We have $B_{\beta,\Lambda}^*(h)(t) = P_{\beta,\Lambda}[h(X) dM(t)] / d\Lambda(t)$.

# Example: Cox model for right-censored data

The information operator $B^*_{\beta,\Lambda} B_{\beta,\Lambda} : L_2(\Lambda) \mapsto L_2(\Lambda)$ is thus

$$B^*_{\beta,\Lambda} B_{\beta,\Lambda}(b)(t) = \frac{P_{\beta,\Lambda}\left[\int_0^\tau b(s) dM(s) dM(u)\right]}{d\Lambda(u)} = P_{\beta,\Lambda}\left[Y(t) e^{\beta' Z}\right] b(t)$$

using martingale methods. Similarly, $B^*_{\beta,\Lambda}\left(\dot{\ell}_{\beta,\Lambda}\right)(t) = P_{\beta,\Lambda}\left[ZY(t) e^{\beta' Z}\right]$. Then, the efficient score for $\beta$ is

$$\begin{aligned} \tilde{\ell}_{\beta,\Lambda} &= \left(I - B_{\beta,\Lambda}\left[B^*_{\beta,\Lambda} B_{\beta,\Lambda}\right]^{-1} B^*_{\beta,\Lambda}\right) \dot{\ell}_{\beta,\Lambda} \\ &= \int_0^\tau \left\{ Z - \frac{P_{\beta,\Lambda}\left[ZY(t) e^{\beta' Z}\right]}{P_{\beta,\Lambda}[Y(t) e^{\beta' Z}]} \right\} dM(t). \end{aligned}$$

# Example: Cox model for right-censored data

- When $\tilde{I}_{\beta,\Lambda} \equiv P_{\beta,\Lambda}\left[\tilde{\ell}_{\beta,\Lambda}\tilde{\ell}'_{\beta,\Lambda}\right]$ is positive definite, the resulting efficient influence function is $\tilde{\psi}_{\beta,\Lambda} \equiv \tilde{I}^{-1}_{\beta,\Lambda}\tilde{\ell}_{\beta,\Lambda}$.
- Since the estimator $\hat{\beta}_n$ obtained from maximizing the partial likelihood

$$\tilde{L}_n(\beta) = \prod_{i=1}^{n}\left(\frac{e^{\beta' Z_i}}{\sum_{j=1}^{n} 1\left\{V_j \geq V_i\right\} e^{\beta' Z_j}}\right)^{d_i}$$

can be shown to satisfy $\sqrt{n}\left(\hat{\beta}_n - \beta\right) = \sqrt{n}\mathbb{P}_n\tilde{\psi}_{\beta,\Lambda} + o_P(1)$, this estimator is efficient.

# Score and information operators

- With semiparametric models having score functions of the form $a'\dot{\ell}_{\theta,\eta} + B_{\theta,\eta}b$, for $a \in \mathbb{R}^k$ and $b \in \mathbb{H}_\eta$, we can define a new operator $A_{\beta,\eta}(a, b) = a'\dot{\ell}_{\theta,\eta} + B_{\theta,\eta}b$.
- More generally, we can define the score operator $A_\eta : \text{lin}\,\mathbb{H}_\eta \mapsto L_2(P_\eta)$ for the model $\{P_\eta : \eta \in H\}$, where $H$ indexes the entire model and may include both parametric and nonparametric components, and where $\text{lin}\,\mathbb{H}_\eta$ indexes directions in $H$.
- Let the parameter of interest be $\psi(P_\eta) = \chi(\eta) \in \mathbb{R}^k$. We assume there exists a linear operator $\dot{\chi} : \text{lin}\,\mathbb{H}_\eta \mapsto \mathbb{R}^k$ such that, for every $b \in \text{lin}\,\mathbb{H}_\eta$, there exists a one-dimensional submodel $\{P_{\eta_t} : \eta_t \in H, t \in N_\epsilon\}$ satisfying

$$\int \left[ \frac{(dP_{\eta_t})^{1/2} - (dP_\eta)^{1/2}}{t} - \frac{1}{2} A_\eta b (dP_\eta)^{1/2} \right]^2 \to 0$$

as $t \downarrow 0$, and $\partial\chi(\eta_t) / (\partial t)|_{t=0} = \dot{\chi}(b)$.

# Score and information operators

- If we can find $\tilde{\chi}_\eta \in \mathbb{H}_\eta$ satisfies $\langle \tilde{\chi}_\eta, b \rangle_\eta = \dot{\chi}_\eta(b)$ for all $b \in \mathbb{H}_\eta$, then the efficient influence function is the solution $\tilde{\psi}_{P_\eta} \in \bar{R}(A_\eta) \subset L_2^0(P_\eta)$ of
$$A_\eta^* \tilde{\psi}_{P_n} = \tilde{\chi}_\eta. \tag{2}$$
- $\partial \chi(\eta_t) / (\partial t)|_{t=0} = \dot{\chi}_\eta(b) = \left\langle A_\eta^* \tilde{\psi}_{P_n}, b \right\rangle_\eta = \left\langle \tilde{\psi}_{P_n}, A_\eta b \right\rangle_{L_2(P_\eta)}$.
- When $A_\eta^* A_\eta$ is invertible, then the solution to (2) can be written as $\tilde{\psi}_{P_\eta} = A_\eta \left(A_\eta^* A_\eta\right)^{-1} \tilde{\chi}_\eta$.

# Example: Cox model for right-censored data

- Consider the estimator for $\Lambda$ is $\hat{\Lambda}(t) = \int_0^t \left[\mathbb{P}_n Y(s) e^{\hat{\beta}' Z}\right]^{-1} \mathbb{P}_n dN(s)$.
- We can show that

$$
\begin{aligned}
\sqrt{n}\left[\hat{\Lambda}(t) - \Lambda_0(t)\right] =& \sqrt{n}\left(\mathbb{P}_n - P\right) \int_0^t \frac{dM_{\beta_0}(s)}{P\left[Y(s) e^{\beta_0' Z}\right]} \\
& - \sqrt{n}\left(\hat{\beta} - \beta_0\right)' \int_0^t E\left(s, \beta_0\right) d\Lambda_0(s) + o_P(1)
\end{aligned}
$$

where the remainder term is uniform in t and $E(t, \beta) = \frac{P\left[Z Y(t) e^{\beta' Z}\right]}{P\left[Y(t) e^{\beta' Z}\right]}$.

# Example: Cox model for right-censored data

By previous arguments,

$$\sqrt{n}(\hat{\beta} - \beta_0) = V^{-1}(\beta_0)\,\mathbb{P}_n \int_0^\tau [Z - E(s, \beta_0)]\, dM_{\beta_0}(s) + o_P(1)$$

where

$$V(\beta) = \tilde{I}_{\beta,\Lambda} = P_{\beta,\Lambda}\left[\tilde{\ell}_{\beta,\Lambda}\tilde{\ell}'_{\beta,\Lambda}\right].$$

Thus $\hat{\Lambda}$ is asymptotically linear with influence function

$$\tilde{\psi}(t) = \int_0^t \frac{dM_{\beta_0}(s)}{P\left[Y(s)e^{\beta_0' Z}\right]} - \left\{\int_0^\tau [Z - E(s, \beta_0)]'\, dM_{\beta_0}(s)\right\} V^{-1}(\beta_0) \int_0^t E(s, \beta_0)\, d\Lambda_0(s)$$

# Example: Cox model for right-censored data

- We will now use the previous methods to verify that $\tilde{\psi}(u)$ is the efficient influence function for the parameter $\Lambda(u)$, for each $u \in [0, \tau]$.
- The tangent space for the Cox model (for both parameters together) is $\left\{A(a, b) = \int_0^\tau [Z'a + b(s)]\, dM_\beta(s) : (a, b) \in H\right\}$, where $H = \mathbb{R}^k \times L_2(\Lambda)$.
- The natural inner product for pairs of elements in $H$ is $\langle (a, b), (c, d) \rangle_H = a'b + \int_0^\tau c(s)d(s)d\Lambda(s)$, for $(a, b), (c, d) \in H$.
- Hence the adjoint of $A$, $A^*$, satisfies $\langle A(a, b), g \rangle_{P_{\beta,\Lambda}} = \langle (a, b), A^* g \rangle_H$ for all $(a, b) \in H$ and all $g \in L_2^0(P_{\beta,\Lambda})$.
- It can be shown that $A^* g = \left(P_{\beta,\Lambda}\left[\int_0^\tau Z dM_\beta(s) g\right], P_{\beta,\Lambda}\left[dM_\beta(t) g\right] / d\Lambda_0(t)\right)$ satisfies these equations and is thus the adjoint of $A$.

# Example: Cox model for right-censored data

- We showed that $\Lambda_t(u) = \int_0^u e^{tb(s)} d\Lambda(s) = \int_0^u (1 + tb(s)) d\Lambda(s) + o(t)$.
- Thus, for the parameter $\chi(P_{\beta,\Lambda}) = \Lambda(u)$,

$$\partial\chi(P_t) / (\partial t)|_{t=0} = \int_0^u b(s) d\Lambda(s).$$

- Thus $\langle\tilde{\chi}, (a, b)\rangle_H = \int_0^u b(s) d\Lambda(s)$ and therefore $\tilde{\chi} = (0, 1\{s \leq v\}) \in H$ (s is the function argument here).
- Our proof will be complete if we can show that $A^*\tilde{\psi} = \tilde{\chi}$, since this would imply that $\tilde{\psi}(u)$ is the efficient influence function for $\Lambda(u)$.

# Example: Cox model for right-censored data

- First,

$$
\begin{aligned}
P\left[\int_0^\tau Z dM_\beta(s)\tilde{\psi}(u)\right] &= P_{\beta,\Lambda}\left\{\int_0^\tau Z dM_\beta(s)\int_0^u \frac{dM_\beta(s)}{P_{\beta,\Lambda}\left[Y(s)e^{\beta' Z}\right]}\right\} \\
&- P_{\beta,\Lambda}\left\{\int_0^\tau Z dM_\beta(s)\int_0^\tau [Z - E(s,\beta)]' dM_\beta(s)\right\} \\
&\times V^{-1}(\beta)\int_0^u E(s,\beta)d\Lambda(s) \\
&= \int_0^u E(s,\beta)d\Lambda(s) - V(\beta)V^{-1}(\beta)\int_0^u E(s,\beta)d\Lambda(s) \\
&= 0
\end{aligned}
$$

- Second, it is not difficult to verify that $P_{\beta,\Lambda}\left[dM_\beta(s)\tilde{\psi}(u)\right] = 1\{s \leq u\}d\Lambda(s)$, and thus we obtain the desired result.

# Estimating equation

- Returning to the semiparametric model setting, where $\mathcal{P} = \{P_{\theta,\eta} : \theta \in \Theta, \eta \in H\}$, the efficient score can be used to derive estimating equations for computing efficient estimators of $\theta$.
- An estimating equation is a data dependent function $\Psi_n : \Theta \mapsto \mathbb{R}^k$ for which an approximate zero yields a Z-estimator for $\theta$.
- When $\Psi_n(\tilde{\theta})$ has the form $\mathbb{P}_n \hat{\ell}_{\tilde{\theta},n}$, where $\hat{\ell}_{\tilde{\theta},n}(X \mid X_1, \ldots, X_n)$ is a function for the generic observation $X$ which depends on the value of $\tilde{\theta}$ and the sample data $X_1, \ldots, X_n$, we have the following estimating equation result.

# Estimating equation

## Theorem (3.1)

*Suppose that the model $\{P_{\theta,\eta} : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^k$, is differentiable in quadratic mean with respect to $\theta$ at $(\theta, \eta)$ and let the efficient information matrix $\tilde{I}_{\theta,\eta}$ be nonsingular.* Let $\hat{\theta}_n$ *satisfy* $\sqrt{n}\mathbb{P}_n\hat{\ell}_{\hat{\theta}_n,n} = o_P(1)$ *and be consistent for $\theta$. Also assume that $\hat{\ell}_{\hat{\theta}_n,n}$ is contained in a $P_{\theta,\eta}$-Donsker class with probability tending to 1 and that the following conditions hold:*

$$P_{\hat{\theta}_n,\eta}\hat{\ell}_{\hat{\theta}_n,n} = o_P\left(n^{-1/2} + \|\hat{\theta}_n - \theta\|\right) \tag{3}$$

$$P_{\theta,\eta}\left\|\hat{\ell}_{\hat{\theta}_n,n} - \tilde{\ell}_{\theta,\eta}\right\|^2 \xrightarrow{\mathrm{P}} 0, \quad P_{\hat{\theta}_n,\eta}\left\|\hat{\ell}_{\hat{\theta}_n,n}\right\|^2 = O_P(1) \tag{4}$$

*Then $\hat{\theta}_n$ is asymptotically efficient at $(\theta, \eta)$.*

# Estimating equation

**Proof.** Define $\mathcal{F}$ to be the $P_{\theta,\eta}$-Donsker class of functions that contains both $\hat{\ell}_{\hat{\theta}_n,n}$ and $\tilde{\ell}_{\theta,\eta}$ with probability tending towards 1 . By Condition (4), we now have

$$\mathbb{G}_n \hat{\ell}_{\hat{\theta}_n,n} = \mathbb{G}_n \tilde{\ell}_{\theta,\eta} + o_P(1) = \sqrt{n}\mathbb{P}_n \tilde{\ell}_{\theta,\eta} + o_P(1)$$

Combining this with the "no-bias" Condition (3), we obtain

$$\begin{aligned} \sqrt{n}\left(P_{\hat{\theta}_n,\eta} - P_{\theta,\eta}\right)\hat{\ell}_{\hat{\theta}_n,n} &= o_P\left(1 + \sqrt{n}\left\|\hat{\theta}_n - \theta\right\|\right) + \mathbb{G}_n \hat{\ell}_{\hat{\theta}_n,n} \\ &= \sqrt{n}\mathbb{P}_n \tilde{\ell}_{\theta,\eta} + o_P\left(1 + \sqrt{n}\left\|\hat{\theta}_n - \theta\right\|\right) \end{aligned}$$

If we can show

$$\sqrt{n}\left(P_{\hat{\theta}_n,\eta} - P_{\theta,\eta}\right)\hat{\ell}_{\hat{\theta}_n,n} = \left(\tilde{I}_{\theta,\eta} + o_P(1)\right)\sqrt{n}\left(\hat{\theta}_n - \theta\right), \tag{5}$$

then the proof will be complete.

# Estimating equation

Since $\tilde{I}_{\theta,\eta} = P_{\theta,\eta}\tilde{\ell}_{\theta,\eta}\dot{\ell}'_{\theta,\eta}$,

$$
\begin{aligned}
\sqrt{n}\left(P_{\hat{\theta}_n,\eta} - P_{\theta,\eta}\right)\hat{\ell}_{\hat{\theta}_n,n} - \tilde{I}_{\theta,\eta}\sqrt{n}\left(\hat{\theta}_n - \theta\right)
&= \sqrt{n}\left(P_{\hat{\theta}_n,\eta} - P_{\theta,\eta}\right)\hat{\ell}_{\hat{\theta}_n,n} - P_{\theta,\eta}\left[\tilde{\ell}_{\theta,\eta}\dot{\ell}'_{\theta,\eta}\right]\sqrt{n}\left(\hat{\theta}_n - \theta\right) \\
&= \sqrt{n}\int \hat{\ell}_{\hat{\theta}_n,n}\left(dP^{1/2}_{\hat{\theta}_n,\eta} + dP^{1/2}_{\theta,\eta}\right) \\
&\quad \times \left[\left(dP^{1/2}_{\hat{\theta}_n,n} - dP^{1/2}_{\theta,\eta}\right) - \frac{1}{2}\left(\hat{\theta}_n - \theta\right)'\dot{\ell}_{\theta,\eta}dP^{1/2}_{\theta,\eta}\right] \\
&\quad + \int\left\{\hat{\ell}_{\hat{\theta}_n,n}\left(dP^{1/2}_{\hat{\theta}_n,\eta} - dP^{1/2}_{\theta,\eta}\right)\frac{1}{2}\dot{\ell}'_{\theta,\eta}dP^{1/2}_{\theta,\eta}\right\}\sqrt{n}\left(\hat{\theta}_n - \theta\right) \\
&\quad + \int\left(\hat{\ell}_{\hat{\theta}_n,n} - \tilde{\ell}_{\theta,\eta}\right)\dot{\ell}'_{\theta,\eta}dP_{\theta,\eta}\sqrt{n}\left(\hat{\theta}_n - \theta\right) \\
&\equiv A_n + B_n + C_n
\end{aligned}
\tag{6}
$$

# Estimating equation

Combining the assumed differentiability in quadratic mean with Condition (4), we obtain $A_n = o_P(\sqrt{n}\|\hat{\theta}_n - \theta\|)$. Condition (4) also implies $C_n = o_P(\sqrt{n}\|\hat{\theta}_n - \theta\|)$. Now we consider $B_n$. Note that for any sequence $m_n \to \infty$,

$$\begin{aligned}
&\left\| \int \hat{\ell}_{\hat{\theta}_n,n} \left( dP^{1/2}_{\hat{\theta}_n,\eta} - dP^{1/2}_{\theta,\eta} \right) \frac{1}{2} \dot{\ell}'_{\theta,\eta} dP^{1/2}_{\theta,\eta} \right\| \\
&\quad \leq \quad m_n \int \left\| \hat{\ell}_{\hat{\theta}_n,n} \right\| dP^{1/2}_{\theta,\eta} \left| dP^{1/2}_{\hat{\theta}_n,\eta} - dP^{1/2}_{\theta,\eta} \right| \\
&\quad + \sqrt{\int \left\| \hat{\ell}_{\hat{\theta}_n,n} \right\|^2 \left( dP_{\hat{\theta}_n,\eta} + dP_{\theta,\eta} \right) \int_{\|\dot{\ell}_{\theta,\eta}\| > m_n} \left\| \dot{\ell}_{\theta,\eta} \right\|^2 dP_{\theta,\eta}} \\
&\equiv m_n D_n + E_n
\end{aligned}$$

where $E_n = o_P(1)$ by Condition (4) and the square-integrability of $\dot{\ell}_{\theta,n}$.

Now

$$\begin{aligned} D_n^2 &\leq \int \left\| \hat{\ell}_{\hat{\theta}_n,n} \right\|^2 dP_{\theta,\eta} \times \int \left( dP_{\hat{\theta}_n,\eta}^{1/2} - dP_{\theta,\eta}^{1/2} \right)^2 \\ &\equiv F_n \times G_n \end{aligned}$$

where $F_n = O_P(1)$ by reapplication of Condition (4) and $G_n = o_P(1)$ by differentiability in quadratic mean combined with the consistency of $\hat{\theta}_n$. Thus there exists some sequence $m_n \to \infty$ slowly enough so that $m_n^2 G_n = o_P(1)$. Hence, for this choice of $m_n, m_n D_n = o_P(1)$. Thus $B_n = o_P\left(\sqrt{n}\left\|\hat{\theta}_n - \theta\right\|\right)$, and we obtain (5). The desired result now follows.