Estimation from Incomplete Longitudinal Data – What We Learn from Event History Data Analysis

Xiaoqiong Joan Hu

Department of Statistics and Actuarial Science

Simon Fraser University

Vancouver, Canada

A Presentation at the Fields Institute on October 13, 2005

Outline

1. Introduction
2. Estimation in Nonparametric Models
3. Estimation in Semiparametric Models
4. Situations with Non-Random Missing
5. Final Remarks

1. Introduction

1.1. Motivating Example

ACTG 359 prospective, randomized, 2×3 factorial, multicentered (Gulick et al, 2000 and 2002)

- study population: HIV-infected with indinavir experience, HIV-RNA $\geq 2,000$ copies/ml
- study regimens (“salvage therapies"): 6 combinations of SQV with RTV or NFV together with DLV, ADV, or both
- response of primary interest: viral load (HIV-RNA) overtime

ACTG359 used its *Observed Sample Means* at different time points to study the trend of HIV-RNA overtime

1. Introduction

1.1. Motivating Example

ACTG 359 prospective, randomized, 2×3 factorial, multicentered (Gulick et al, 2000 and 2002)

- study population: HIV-infected with indinavir experience, HIV-RNA $\geq 2,000$ copies/ml
- study regimens (“salvage therapies"): 6 combinations of SQV with RTV or NFV together with DLV, ADV, or both
- response of primary interest: viral load (HIV-RNA) overtime

ACTG359 used its *Observed Sample Means* at different time points to study the trend of HIV-RNA overtime

In recent AIDS treatment clinical trials,

- primary response – a marker overtime:
  e.g. HIV-RNA copies or CD4 counts (virologic/immunologic measures);
  e.g. weight, height, or IQ (age-adjusted) for children
- missing data
- robust analysis methods are desirable:
  a rapidly evolving area

Similar situations in many other medical studies.

In recent AIDS clinical trials,

- primary response – a marker overtime:
  e.g. HIV-RNA copies or CD4 counts (virologic/immunologic measures);
  e.g. weight, height, or IQ (age-adjusted) for children
- missing data
- robust analysis methods are desirable:
  a rapidly evolving area

Similar situations in many other medical studies.

For repeated measures with missing in general,

Observed Sample Mean is commonly used in practice in a descriptive way.

- how does *Observed Sample Mean* perform?
- any alternatives?
- what can we learn from survival analysis?

For repeated measures with missing in general,

Observed Sample Mean is commonly used in practice in a descriptive way.

- how does *Observed Sample Mean* perform?
- any alternatives?
- what can we learn from survival analysis?

Recall

- Marginal analysis in counting process setting: Lawless (1995), Lawless and Nadeau (1995)

  followed by e.g. Cook, Lawless and Nadeau (1996), Lin, Wei, Yang and Ying (2000), Hu, Sun and Wei (2003)

- Longitudinal analysis: GEE

  recent work, e.g. Robins and Rotnitzky (1995), Lin and Carroll (2000, 2001), Wang (2003).

1.2. Framework

- Response: $X(t), t \in \mathcal{T}$
- Observation Indicator: $\delta(t), t \in \mathcal{T}$,
  with $\delta(t) = 1$ if $X(t)$ observed; $= 0$ if not.
- Covariate: $Z(t), t \in \mathcal{T}$

Goals:

- to estimate $\mu(t) = \mathsf{E}\{X(t)\}, t \in \mathcal{T}$.
- to estimate $\mu_Z(t) = \mathsf{E}\{X(t) \big| Z(s) : s \leq t\}, t \in \mathcal{T}$

1.2. Framework

- Response: $X(t), t \in \mathcal{T}$
- Observation Indicator: $\delta(t), t \in \mathcal{T}$,
  with $\delta(t) = 1$ if $X(t)$ observed; $= 0$ if not.
- Covariate: $Z(t), t \in \mathcal{T}$

Goals:

- to estimate $\mu(t) = \mathsf{E}\{X(t)\}, t \in \mathcal{T}$.
- to estimate $\mu_Z(t) = \mathsf{E}\{X(t) \big| Z(s) : s \leq t\}, t \in \mathcal{T}$

Illustrative Examples for the Framework:

- *Repeated Measures with Missing*
  $X(t)$: the measure of an quantity at time t
  $\delta(t) = \mathsf{I}\big(t = \xi_1, \ldots, \xi_K\big)$, ξ_j and K rvs
  $\mu(t)$: the average over time of the quantity in the population
- *Right-censored Survival times*
  $X(t) = \mathsf{I}(T \leq t)$: the indicator process of death
  $\delta(t) = \mathsf{I}\big(t \leq C\big)$, C a censoring time
  $\mu(t)$: the cdf of T
- *Panel Counts*
  $X(t)$: a counting process
  $\delta(t) = \mathsf{I}\big(t = \xi_1, \ldots, \xi_K\big)$, ξ_j and K rvs
  $\mu(t)$: the cumulative intensity of X if X is Poisson

Illustrative Examples for the Framework:

- *Repeated Measures with Missing*
  $X(t)$: the measure of an quantity at time t
  $\delta(t) = \mathrm{I}\big(t = \xi_1, \ldots, \xi_K\big)$, ξ_j and K rvs
  $\mu(t)$: the average over time of the quantity in the population
- *Right-censored Survival times*
  $X(t) = \mathrm{I}(T \leq t)$: the indicator process of death
  $\delta(t) = \mathrm{I}\big(t \leq C\big)$, C a censoring time
  $\mu(t)$: the cdf of T
- *Panel Counts*
  $X(t)$: a counting process
  $\delta(t) = \mathrm{I}\big(t = \xi_1, \ldots, \xi_K\big)$, ξ_j and K rvs
  $\mu(t)$: the cumulative intensity of X if X is Poisson

Illustrative Examples for the Framework:

- *Repeated Measures with Missing*
  $X(t)$: the measure of an quantity at time t
  $\delta(t) = \mathsf{I}\big(t = \xi_1, \ldots, \xi_K\big)$, ξ_j and K rvs
  $\mu(t)$: the average over time of the quantity in the population
- *Right-censored Survival times*
  $X(t) = \mathsf{I}(T \leq t)$: the indicator process of death
  $\delta(t) = \mathsf{I}\big(t \leq C\big)$, C a censoring time
  $\mu(t)$: the cdf of T
- *Panel Counts*
  $X(t)$: a counting process
  $\delta(t) = \mathsf{I}\big(t = \xi_1, \ldots, \xi_K\big)$, ξ_j and K rvs
  $\mu(t)$: the cumulative intensity of X if X is Poisson

2. Estimation in Nonparametric Models

(Hu and Lagakos, 2004; Hu, Lagakos, and Lockhart, 2005)

Goal:
To estimate $\mu(t) = \mathrm{E}\{X(t)\}$ from iid $\{X_i, \delta_i : i = 1, \ldots n\}$ nonparametrically

Assumptions:

- $X(\cdot)$ and $\delta(\cdot)$ independent
- *Periodic Observations*: all times of interest $\mathcal{T} = \{t_1, t_2, \ldots, t_M\}, 0 < M < \infty$; $\mathrm{E}\{\delta(t)\} > 0$ for $t \in \mathcal{T}$

the Assumptions ?

2.1. Estimation Procedures

2.1.1. Observed sample mean (OSM)

For $t \in \mathcal{T}$, a natural estimator and commonly used in a descriptive way:

$$\bar{\mu}(t) = \frac{\sum_{i=1}^{n} X_i(t)\delta_i(t)}{\sum_{i=1}^{n} \delta_i(t)}.$$

- Unbiased
- Consistent and Asymptotically Gaussian

- A weighted least squares estimator: it minimizes

$$\sum_{i=1}^{n} \sum_{t \in \mathcal{T}} \delta_i(t) \Big\{ X_i(t) - \mu(t) \Big\}^2,$$

i.e., it's the solution of

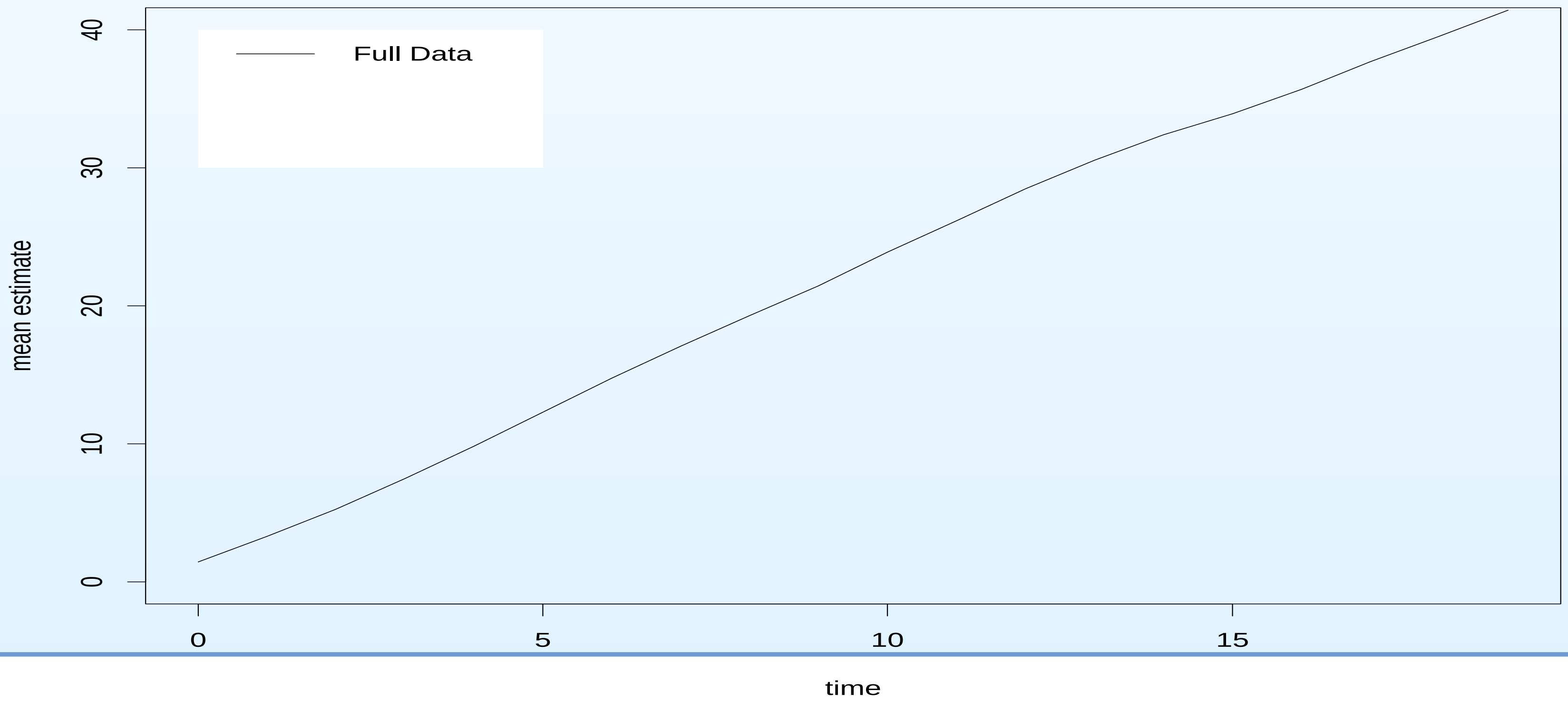
$$\sum_{i=1}^{n} \Phi_i \Big\{ \underline{X}_i - \underline{\mu} \Big\} = 0$$

with $\underline{X}_i = (X_i(t_1), \ldots, X_i(t_M))'$, $\underline{\mu} = (\mu(t_1), \ldots, \mu(t_M))'$, and $\Phi_i = diag(\delta_i(t) : t \in \mathcal{T})$.

How does it perform numerically?

Simulation

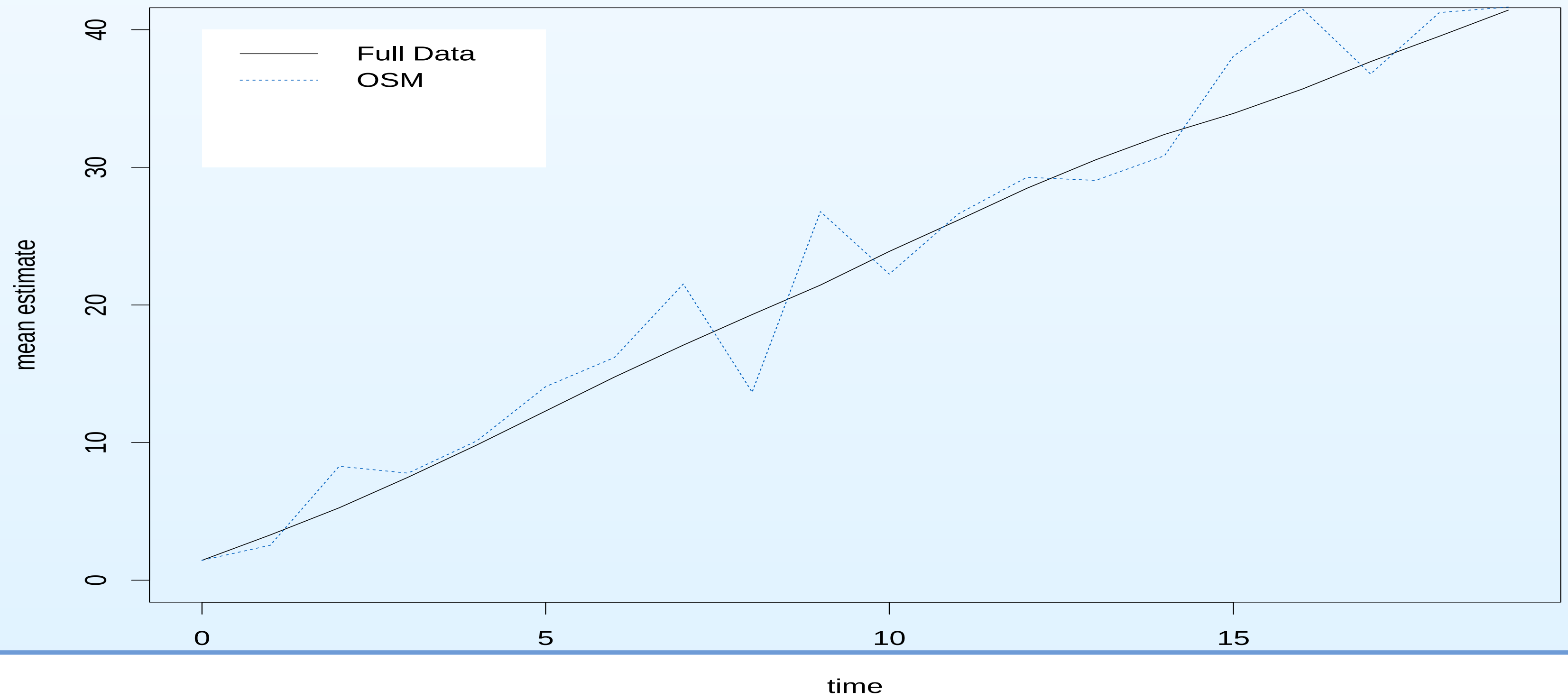
Generate independent $\{X_i(t) : t \in \mathcal{T} = \{0, 1, \ldots, 19\}\}$, $i = 1, \ldots, 100$: $X_i(t) = e^{Q_i(t)} + e^{Q_i(t-1)}$, $\underline{Q} \sim MN(\underline{\nu}, \Sigma)$, AR with $\rho = 0.8$;
Generate random missing with obs rate of 20% for $t \in \mathcal{T}$.



Simulation

Generate independent $\{X_i(t) : t \in \mathcal{T} = \{0, 1, \ldots, 19\}\}$, $i = 1, \ldots, 100$: $X_i(t) = e^{Q_i(t)} + e^{Q_i(t-1)}$, $\underline{Q} \sim MN(\underline{\nu}, \Sigma)$, AR with $\rho = 0.8$;

Generate random missing with obs rate of 20% for $t \in \mathcal{T}$.



Recall *“Reduced Sample Estimator”* from right-censored survival times (Kaplan and Meier, 1958):

$$\bar{\mu}(t) = \frac{\sum_{i=1}^{n} X_i(t)\delta_i(t)}{\sum_{i=1}^{n} \delta_i(t)}, \quad t \in \mathcal{T}$$

Compared to Kaplan-Meier estimator for $S(t)$?

How about to consider

$$\mu(t) = [\mu(t_1) - \mu(t_0)] + [\mu(t_2) - \mu(t_1)] + \ldots + [\mu(t) - \mu(t_l)],$$

and have $\tilde{\mu}(t) = \sum_{t_j \leq t} \tilde{\nu}_j$???

Recall *“Reduced Sample Estimator”* from right-censored survival times (Kaplan and Meier, 1958):

$$\bar{\mu}(t) = \frac{\sum_{i=1}^{n} X_i(t)\delta_i(t)}{\sum_{i=1}^{n} \delta_i(t)}, \quad t \in \mathcal{T}$$

Compared to Kaplan-Meier estimator for $S(t)$?

How about to consider

$$\mu(t) = [\mu(t_1) - \mu(t_0)] + [\mu(t_2) - \mu(t_1)] + \ldots + [\mu(t) - \mu(t_l)],$$

and have $\tilde{\mu}(t) = \sum_{t_j \leq t} \tilde{\nu}_j$???

2.1.2. Cumulative observed increments (COI)

Consider to minimize, wrt $\nu_j = \mu(t_j) - \mu(t_{j-1})$,

$$\sum_{i=1}^{n} \sum_{t \in \mathcal{T}} \delta_i(t) \Big\{ \Delta X_i(t) - \Delta \mu_i(t) \Big\}^2,$$

$\Delta X_i(t) = X_i(t) - X_i(s_i(t))$, $\Delta \mu_i(t) = \sum_{s_i(t) < t_j \leq t} \nu_j$. The weighted least squares estimator:

$$\tilde{\mu}(t) = \sum_{t_j \in \mathcal{T} : t_j \leq t} \tilde{\nu}_j, \quad t \in \mathcal{T}.$$

- Unbiased
- Consistent, Asymptotically Gaussian

2.1.2. Cumulative observed increments (COI)

Consider to minimize, wrt $\nu_j = \mu(t_j) - \mu(t_{j-1})$,

$$\sum_{i=1}^{n} \sum_{t \in \mathfrak{T}} \delta_i(t) \Big\{ \Delta X_i(t) - \Delta\mu_i(t) \Big\}^2,$$

$\Delta X_i(t) = X_i(t) - X_i(s_i(t))$, $\Delta\mu_i(t) = \sum_{s_i(t) < t_j \leq t} \nu_j$. The weighted least squares estimator:

$$\tilde{\mu}(t) = \sum_{t_j \in \mathfrak{T} : t_j \leq t} \tilde{\nu}_j, \quad t \in \mathfrak{T}.$$

- Unbiased
- Consistent, Asymptotically Gaussian

- Nelson-Aalen estimator from right-censored Poisson counts, Lawless-Nadeau for the mean of a counting process:

$$\tilde{\mu}(t) = \sum_{i=1}^{n} \int_0^t \frac{\delta_i(u)}{\sum_{j=1}^{n} \delta_j(u)} dX_i(u), \quad t > 0.$$

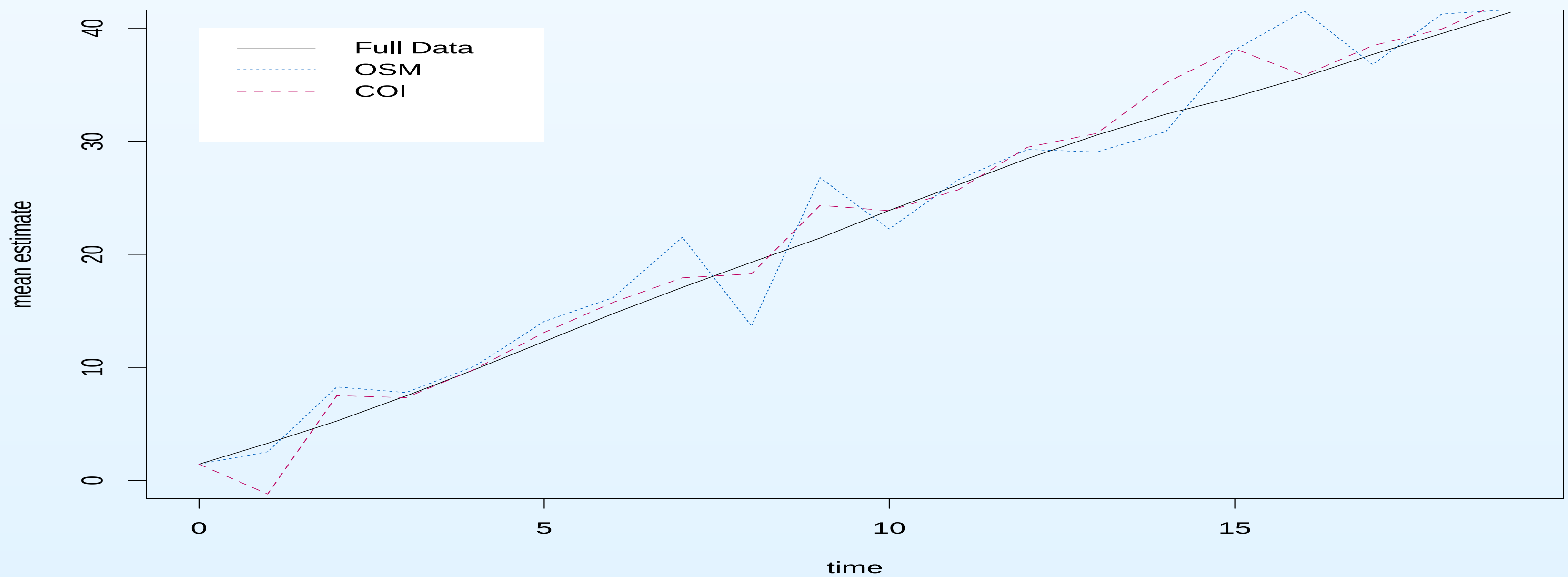
How does it perform numerically?

Simulation (cont'd)

- Nelson-Aalen estimator from right-censored Poisson counts, Lawless-Nadeau for the mean of a counting process:

$$\tilde{\mu}(t) = \sum_{i=1}^{n} \int_0^t \frac{\delta_i(u)}{\sum_{j=1}^{n} \delta_j(u)} dX_i(u), \quad t > 0.$$

How does it perform numerically? **Simulation** (cont'd)



Recall

- $\bar{\mu}(\cdot)$ (OSM) minimizes

$$\sum_{i=1}^{n}\left\{\Phi_i \underline{X}_i - \Phi_i \underline{\mu}\right\}'\left\{\Phi_i \underline{X}_i - \Phi_i \underline{\mu}\right\} = \sum_{i=1}^{n}\left\{\underline{X}_i - \underline{\mu}\right\}'\Phi_i'\Phi_i\left\{\underline{X}_i - \underline{\mu}\right\};$$

- $\tilde{\mu}(\cdot)$ (COI) minimizes

$$\sum_{i=1}^{n}\left\{\underline{X}_i - \underline{\mu}\right\}'\Phi_i'\Omega_i'\Omega_i\Phi_i\left\{\underline{X}_i - \underline{\mu}\right\},$$

$$\Omega_i \underline{X}_i = (\delta_i(t)\Delta X_i(t), t \in \mathcal{T})'.$$

How about to minimize (W_i symmetric weight)

$$\sum_{i=1}^{n} \left\{ \underline{X}_i - \underline{\mu} \right\}' \Phi_i' W_i \Phi_i \left\{ \underline{X}_i - \underline{\mu} \right\}?$$

or, to consider the estimation equation (GEE type)

$$\sum_{i=1}^{n} W_i \Phi_i \left\{ \underline{X}_i - \underline{\mu} \right\} = 0?$$

What W_i to use?

- the inverse of $Var(\Phi_i \underline{X}_i)$?
- COI: $W_i = \Omega_i' \Omega_i$.
- What else?

How about to minimize (W_i symmetric weight)

$$\sum_{i=1}^{n} \left\{ \underline{X}_i - \underline{\mu} \right\}^{'} \Phi_i^{'} W_i \Phi_i \left\{ \underline{X}_i - \underline{\mu} \right\}?$$

or, to consider the estimation equation (GEE type)

$$\sum_{i=1}^{n} W_i \Phi_i \left\{ \underline{X}_i - \underline{\mu} \right\} = 0?$$

What W_i to use?

- the inverse of $Var(\Phi_i \underline{X}_i)$?
- COI: $W_i = \Omega_i^{'} \Omega_i$.
- What else?

2.1.3. Projection of Nelson-Aalen estimator (PNA)

Recall Nelson-Aalen estimator, the solution of the EE based on right-censored data:

$$\sum_{i=1}^{n} Y_i(t)\Big\{\big[X_i(t) - X_i(s(t))\big] - \Delta\mu(t)\Big\} = 0, t \in \mathcal{T},$$

$\Delta\mu(t) = \mu(t) - \mu(s(t)) = \nu(t)$ and $Y_i(t) = \mathsf{I}(t \leq C_i)$.

For the current situation, to consider for $t \in \mathcal{T}$

$$\sum_{i=1}^{n} Y_i(t)\Big\{\mathsf{E}\big[X_i(t) - X_i(s(t))\big| X_i(u) : \delta_i(u) = 1, u \in \mathcal{T}\big] - \Delta\mu(t)\Big\} = 0.$$

2.1.3. Projection of Nelson-Aalen estimator (PNA)

Recall Nelson-Aalen estimator, the solution of the EE based on right-censored data:

$$\sum_{i=1}^{n} Y_i(t)\Big\{\big[X_i(t) - X_i(s(t))\big] - \Delta\mu(t)\Big\} = 0, t \in \mathcal{T},$$

$\Delta\mu(t) = \mu(t) - \mu(s(t) = \nu(t)$ and $Y_i(t) = \mathsf{I}(t \leq C_i)$.

For the current situation, to consider for $t \in \mathcal{T}$

$$\sum_{i=1}^{n} Y_i(t)\Big\{\mathsf{E}\big[X_i(t) - X_i(s(t))\big| X_i(u) : \delta_i(u) = 1, u \in \mathcal{T}\big] - \Delta\mu(t)\Big\} = 0.$$

The EE gives, for $t \in \mathcal{T}$,

$$\mu(t) = \sum_{v \in \mathcal{T}: v \leq t} \sum_{i=1}^{n} \frac{Y_i(v)}{\sum_{j=1}^{n} Y_j(v)} \mathsf{E}\left[X_i(v) - X_i(s(v)) \middle| X_i(u) : \begin{array}{c} \delta_i(u) = 1, \\ u \in \mathcal{T} \end{array}\right].$$

How to get $\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathcal{T}\right]$?

Denote $\Delta^* X_i(t) = X_i(s_i^*(t)) - X_i(s_i(t))$.

- In survival setting,

$$\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathcal{T}\right]$$
$$= \begin{cases} 0 & \text{if } \Delta^* X_i(t) = 0 \\ \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)} & \text{if } \Delta^* X_i(t) = 1 \end{cases} = \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)} \Delta^* X_i(t)$$

- For Poisson counts,

$$\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathcal{T}\right] = \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)} \Delta^* X_i(t).$$

- In general, $\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathcal{T}\right]$ and $\Delta^* X_i(t) \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)}$ have the same expectation $\Delta\mu(t)$, conditional on the observation.

Denote $\Delta^* X_i(t) = X_i(s_i^*(t)) - X_i(s_i(t))$.

- In survival setting,

$$\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathcal{T}\right]$$

$$= \begin{cases} 0 & \text{if } \Delta^* X_i(t) = 0 \\ \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)} & \text{if } \Delta^* X_i(t) = 1 \end{cases} = \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)} \Delta^* X_i(t)$$

- For Poisson counts,

$$\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathcal{T}\right] = \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)} \Delta^* X_i(t).$$

- In general, $\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathcal{T}\right]$ and $\Delta^* X_i(t) \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)}$ have the same expectation $\Delta\mu(t)$, conditional on the observation.

Denote $\Delta^* X_i(t) = X_i(s_i^*(t)) - X_i(s_i(t))$.

- In survival setting,

$$\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathfrak{T}\right]$$

$$= \begin{cases} 0 & \text{if } \Delta^* X_i(t) = 0 \\ \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)} & \text{if } \Delta^* X_i(t) = 1 \end{cases} = \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)} \Delta^* X_i(t)$$

- For Poisson counts,

$$\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathfrak{T}\right] = \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)} \Delta^* X_i(t).$$

- In general, $\mathsf{E}\left[X_i(t) - X_i(s(t)) \middle| X_i(u) : \delta_i(u) = 1, u \in \mathfrak{T}\right]$ and $\Delta^* X_i(t) \frac{\Delta\mu(t)}{\Delta^*\mu_i(t)}$ have the same expectation $\Delta\mu(t)$, conditional on the observation.

Recall the projection of NA estimator: for $t \in \mathcal{T}$,

$$\mu(t) = \sum_{v \in \mathcal{T}: v \le t} \sum_{i=1}^{n} \frac{Y_i(v)}{\sum_{j=1}^{n} Y_j(v)} \mathsf{E}\left[X_i(v) - X_i(s(v)) \middle| X_i(u) : \begin{array}{c} \delta_i(u) = 1, \\ u \in \mathcal{T} \end{array} \right].$$

Thus

$$\mu(t) = \sum_{v \in \mathcal{T}: v \le t} \sum_{i=1}^{n} \frac{Y_i(v)}{\sum_{j=1}^{n} Y_j(v)} \Delta^* X_i(v) \frac{\Delta \mu(v)}{\Delta^* \mu_i(v)}, t \in \mathcal{T}.$$

Its solution (by an iterative algorithm) is $\hat{\mu}(\cdot)$.

Recall the projection of NA estimator: for $t \in \mathcal{T}$,

$$\mu(t) = \sum_{v \in \mathcal{T}: v \leq t} \sum_{i=1}^{n} \frac{Y_i(v)}{\sum_{j=1}^{n} Y_j(v)} \mathsf{E}\Big[X_i(v) - X_i(s(v)) \Big| X_i(u) : \begin{array}{c} \delta_i(u) = 1, \\ u \in \mathcal{T} \end{array} \Big].$$

Thus

$$\mu(t) = \sum_{v \in \mathcal{T}: v \leq t} \sum_{i=1}^{n} \frac{Y_i(v)}{\sum_{j=1}^{n} Y_j(v)} \Delta^* X_i(v) \frac{\Delta\mu(v)}{\Delta^* \mu_i(v)}, t \in \mathcal{T}.$$

Its solution (by an iterative algorithm) is $\hat{\mu}(\cdot)$.

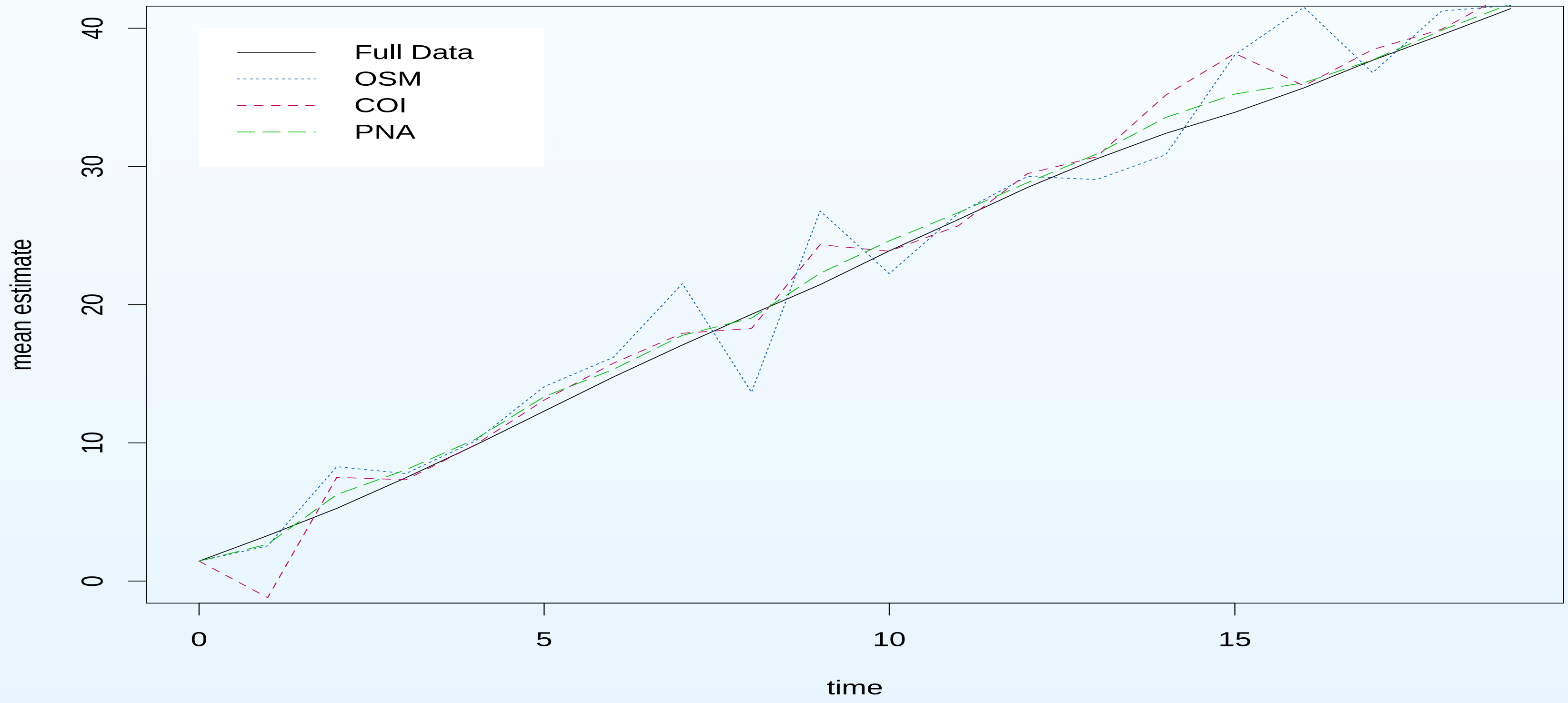
Estimator $\hat{\mu}(\cdot)$,

- self-consistent
- it uses $W_i = \Omega_i' \Sigma_i^{-1} \Omega_i$ in the GEE type EE:
  $\Sigma_i = diag(\Omega_i \underline{\mu} : t \in \mathcal{T})$, it's $Var(\Omega_i \underline{X}_i)$ when $X(\cdot)$ is Poisson.
- the same as NMLE of $\mu(\cdot)$ from panel counts under Poisson assumption, given by Wellner and Zhang (2000)
- consistent and asymptotically Gaussian

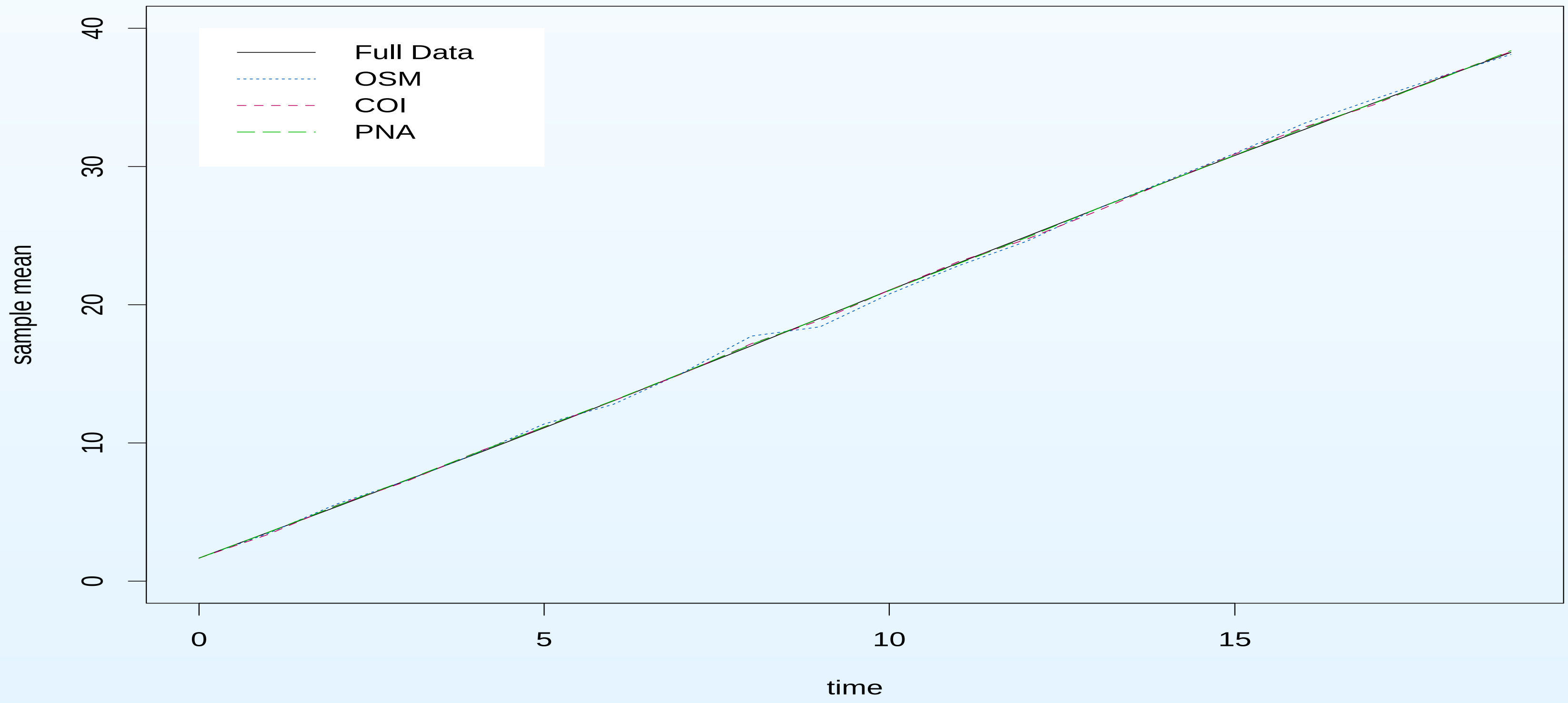
Estimator $\hat{\mu}(\cdot)$,

- self-consistent
- it uses $W_i = \Omega_i' \Sigma_i^{-1} \Omega_i$ in the GEE type EE:
  $\Sigma_i = diag(\Omega_i \underline{\mu} : t \in \mathcal{T})$, it's $Var(\Omega_i \underline{X}_i)$ when $X(\cdot)$ is Poisson.
- the same as NMLE of $\mu(\cdot)$ from panel counts under Poisson assumption, given by Wellner and Zhang (2000)
- consistent and asymptotically Gaussian

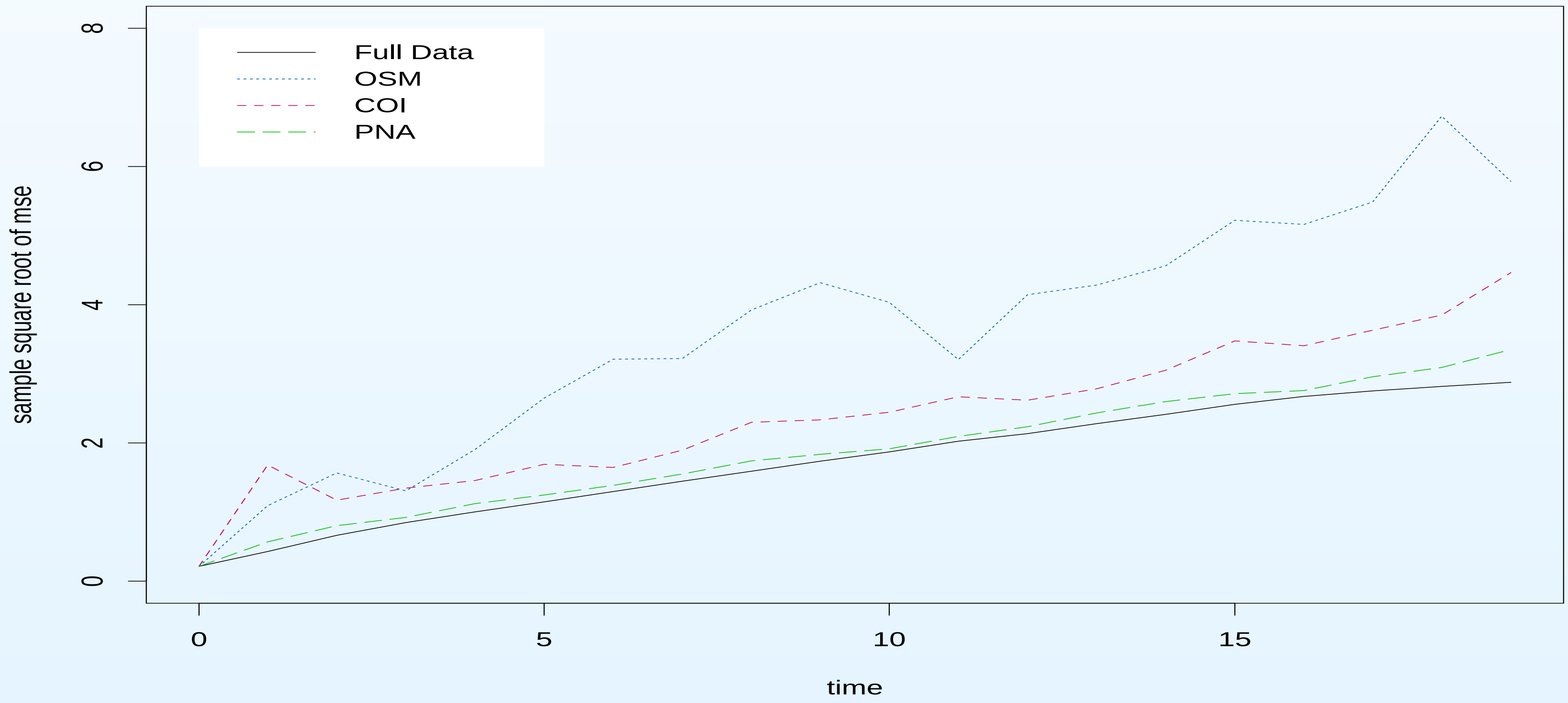
How does it perform numerically? **Simulation** (cont'd)



Based on 100 repetitions: the sample means?



Based on 100 repetitions: the sample mean square erros?



2.2. Estimation for Monotone Mean

When $\mu(\cdot)$ is monotone?

- In survival setting, $\mu(\cdot) = F(\cdot)$
- In counting process setting, $\mu(\cdot) = \Lambda(\cdot)$
- $X_i(\cdot)$ as height overtime, or IQ (age adjusted) overtime of HIV children

Note

- $\bar{\mu}(\cdot)$ and $\tilde{\mu}(\cdot)$ not necessarily monotone
- $\hat{\mu}(\cdot)$ is monotone, when $X(\cdot)$ is monotone

Consider to minimize wrt $\mu(\cdot)$

$$\sum_{i=1}^{n} \left\{ \underline{X}_i - \underline{\mu} \right\}' \Phi_i' W_i \Phi_i \left\{ \underline{X}_i - \underline{\mu} \right\}$$

under the monotone constraint.

- if use OSM weight, $\Longrightarrow \bar{\mu}^*(\cdot)$, the isotonic regression of $\bar{\mu}(\cdot)$ with weights $\{M(t):$ the num of obs at $t \in \mathcal{T}\}$ the same as the estimator given by Sun and Kalbfleisch (1995), called the NPMLE by Wellner and Zhang (2000)

Consider to minimize wrt $\mu(\cdot)$

$$\sum_{i=1}^{n} \left\{ \underline{X}_i - \underline{\mu} \right\}' \Phi_i' W_i \Phi_i \left\{ \underline{X}_i - \underline{\mu} \right\}$$

under the monotone constraint.

- if use OSM weight, $\Longrightarrow \bar{\mu}^*(\cdot)$, the isotonic regression of $\bar{\mu}(\cdot)$ with weights $\{M(t):$ the num of obs at $t \in \mathcal{T}\}$ the same as the estimator given by Sun and Kalbfleisch (1995), called the NPMLE by Wellner and Zhang (2000)

Consider to minimize wrt $\mu(\cdot)$

$$\sum_{i=1}^{n} \left\{ \underline{X}_i - \underline{\mu} \right\}' \Phi_i' W_i \Phi_i \left\{ \underline{X}_i - \underline{\mu} \right\}$$

under the monotone constraint.

- if use COI weight, $\Longrightarrow \tilde{\mu}^*(\cdot)$, obtained by the iterative convex minorant (ICM) algorithm
  slightly different from the isotonic regression of $\tilde{\mu}(\cdot)$ with weights $\{M(\cdot)\}$

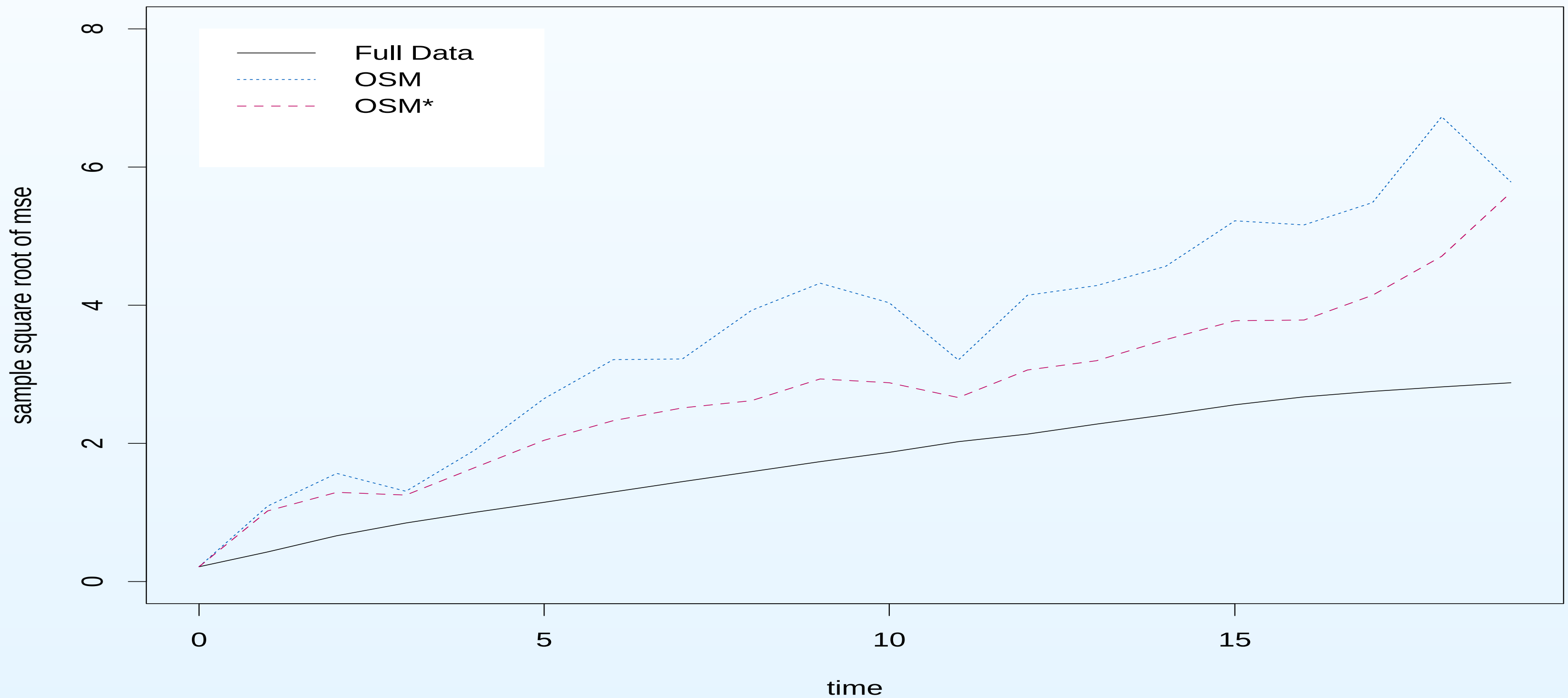
Consider to minimize wrt $\mu(\cdot)$

$$\sum_{i=1}^{n} \left\{ \underline{X}_i - \underline{\mu} \right\}' \Phi_i' W_i \Phi_i \left\{ \underline{X}_i - \underline{\mu} \right\}$$

under the monotone constraint.

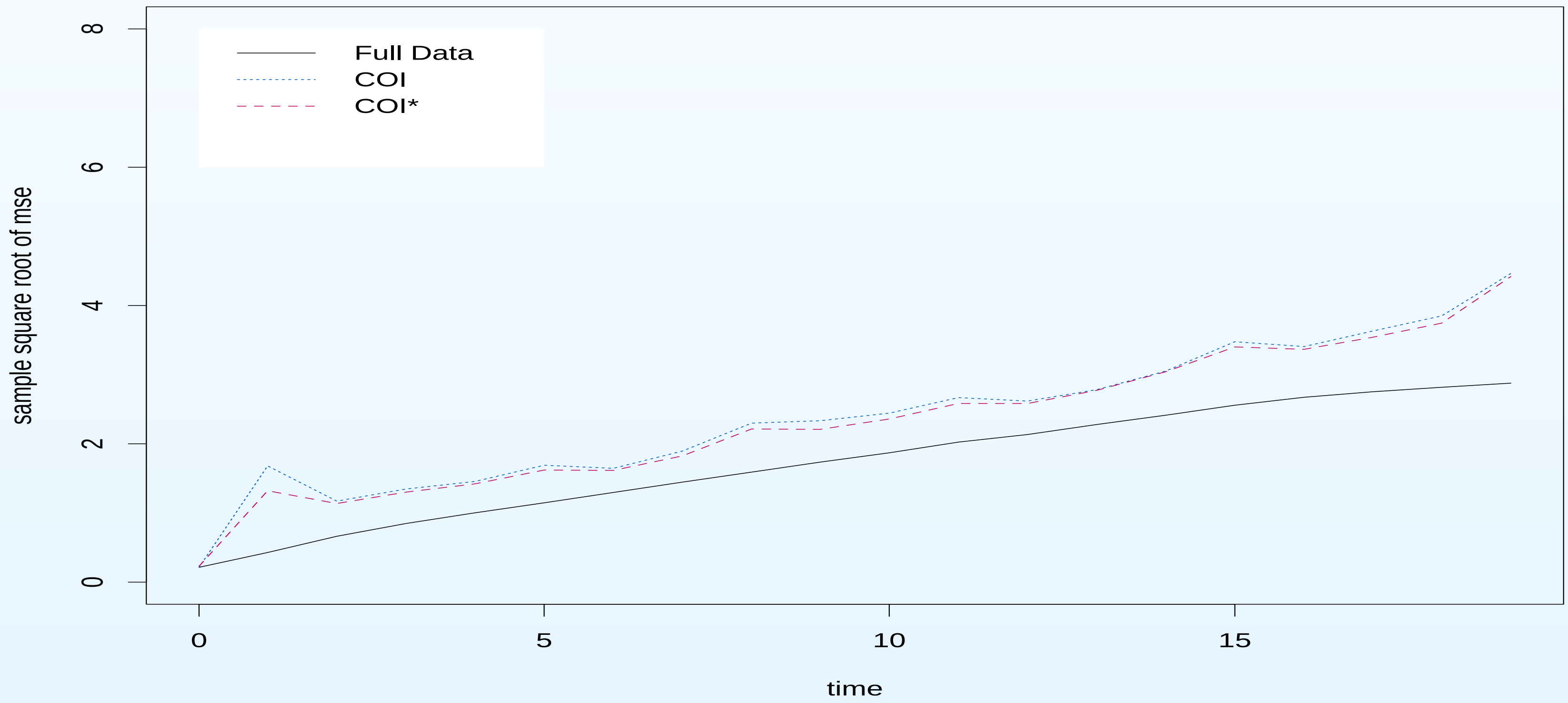
- if use PNA weight, $\Longrightarrow \hat{\mu}^*(\cdot)$, obtained by the iterative convex minorant (ICM) algorithm
  slightly different from the PNA $\hat{\mu}(\cdot)$, likely more efficient than $\hat{\mu}(\cdot)$
  to improve $\hat{\mu}(\cdot)$ in counting process setting?

Simulation (cont'd)
Based on 100 repetitions: the OSM estimators $\bar{\mu}$ and $\bar{\mu}^*$



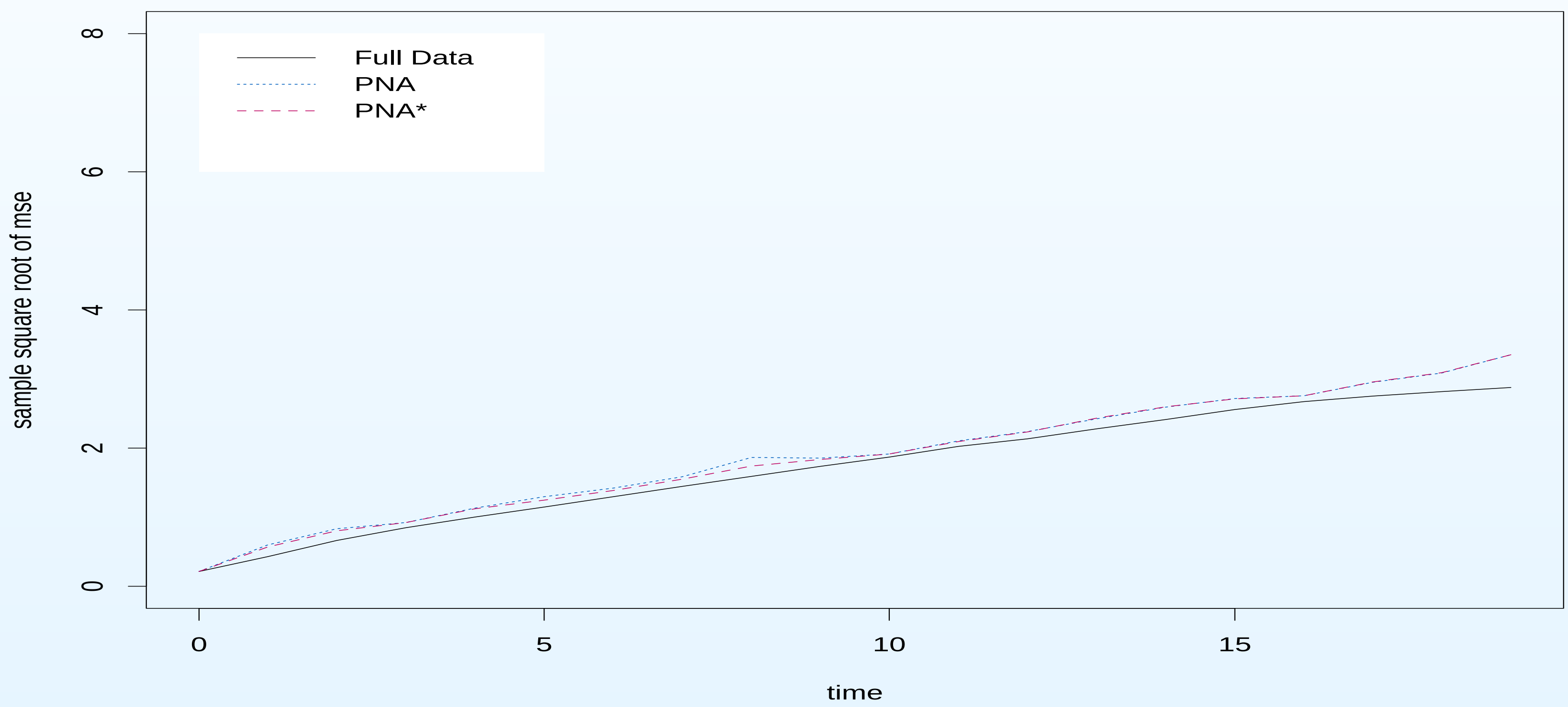
Simulation (cont'd)

Based on 100 repetitions: the COI estimators $\tilde{\mu}$ and $\tilde{\mu}^*$



Simulation (cont'd)

Based on 100 repetitions: the PNA estimators $\hat{\mu}$ and $\hat{\mu}^*$



3. Estimation in Semiparametric Models

(Hu, Jin and Lagakos, 2005)

Goal:
To estimate $\mu_Z(t) = \mathsf{E}\{X(t)\big|Z(s), s \le t\}$ from iid $\{X_i, \delta_i : i = 1, \ldots n\}$,

$$\mu_Z(t) = G(h(\cdot), \beta; Z(\cdot))$$

Assumptions:

- $X(\cdot)$ and $\delta(\cdot)$ independent
- *Periodic Observation*: all times of interest $\mathcal{T} = \{t_1, t_2, \ldots, t_M\}, 0 < M < \infty$; $\mathsf{E}\{\delta(t)\} > 0$ for $t \in \mathcal{T}$

3. Estimation in Semiparametric Models

(Hu, Jin and Lagakos, 2005)

Goal:
To estimate $\mu_Z(t) = \mathsf{E}\{X(t) \big| Z(s), s \le t\}$ from iid $\{X_i, \delta_i : i = 1, \dots n\}$,

$$\mu_Z(t) = G(h(\cdot), \beta; Z(\cdot))$$

Assumptions:

- $X(\cdot)$ and $\delta(\cdot)$ independent
- *Periodic Observation*: all times of interest $\mathcal{T} = \{t_1, t_2, \dots, t_M\}, 0 < M < \infty$; $\mathsf{E}\{\delta(t)\} > 0$ for $t \in \mathcal{T}$

3.1. Semiparametric Transformation Models

Suppose $\mu_Z(t)$ follows, with either $h(\cdot)$ or $g(\cdot)$ unknown,

$$g(\mu_Z(t)) = h(t) + \beta Z(t).$$

Illustrative Examples for the Models:

- *Cox's regression models*:
  - in survival setting, Cox's proportional hazards model (Cox, 1972)

    $$\mu_Z(t) = S_0(t)^{\exp\{\beta Z(t)\}};$$

  - in counting process setting, conditional cumulative intensity (c.f. Andersen, Borgan, Gill and Keiding, 1991)

    $$\mu_Z(t) = \Lambda_0(t) \exp\{\beta Z(t)\};$$

3.1. Semiparametric Transformation Models

Suppose $\mu_Z(t)$ follows, with either $h(\cdot)$ or $g(\cdot)$ unknown,

$$g(\mu_Z(t)) = h(t) + \beta Z(t).$$

Illustrative Examples for the Models:

- *Cox's regression models*:
  - in survival setting, Cox's proportional hazards model (Cox, 1972)

    $$\mu_Z(t) = S_0(t)^{\exp\{\beta Z(t)\}};$$

  - in counting process setting, conditional cumulative intensity (c.f. Andersen, Borgan, Gill and Keiding, 1991)

    $$\mu_Z(t) = \Lambda_0(t) \exp\{\beta Z(t)\};$$

3.1. Semiparametric Transformation Models

Suppose $\mu_Z(t)$ follows, with either $h(\cdot)$ or $g(\cdot)$ unknown,

$$g(\mu_Z(t)) = h(t) + \beta Z(t).$$

Illustrative Examples for the Models:

- *Cox's regression models*:

Application: a generalization of the classical model for the surplus process of an insurance company, where

$$X(t) = u + c(t) - \sum_{k=1}^{N(t)} U_k,$$

$u =$ the initial surplus, $c(t) =$ the cumulative premiums upto t, $N(t) =$ the cumulative counts of claims, U_k the size of kth claim

3.1. Semiparametric Transformation Models

Suppose $\mu_Z(t)$ follows, with either $h(\cdot)$ or $g(\cdot)$ unknown,

$$g(\mu_Z(t)) = h(t) + \beta Z(t).$$

Illustrative Examples for the Models:

- *Proportional odds models*: for binary $X(\cdot)$,

$$logit(\mu_Z(t)) = \beta Z(t) + h(t).$$

- *Accelerated failure time models*:

$$\mu_Z(t) = \mu_0(te^{\beta Z(t)})$$

(cfs, Wei, 1992; Lin, Wei and Ying, 1998)

3.1. Semiparametric Transformation Models

Suppose $\mu_Z(t)$ follows, with either $h(\cdot)$ or $g(\cdot)$ unknown,

$$g(\mu_Z(t)) = h(t) + \beta Z(t).$$

Illustrative Examples for the Models:

- *Generalized linear models for repeated measures*: (c.f., Zeger and Diggle, 1994)

$$\mu_Z(t) = \beta Z(t) + \alpha(t)$$

Lin and Carroll (2001) consider the general model. So is it mentioned in Lin and Ying (2001).

3.1. Semiparametric Transformation Models

Suppose $\mu_Z(t)$ follows, with either $h(\cdot)$ or $g(\cdot)$ unknown,

$$g(\mu_Z(t)) = h(t) + \beta Z(t).$$

Illustrative Examples for the Models:

- *Generalized autoregressive models*:
  eg, the AR(1) Poisson model (McKenzie, 1988)

$$X(t) = \beta * X(t-1) + W(t),$$

  $\beta * X(t)$ defined as $\sum_{k=1}^{X(t)} B_k(\beta)$ with $\{B_k(\beta) : k = 1, 2, \ldots\}$ iid binary rvs and $p = \beta$, $W(t)$ a mean $h(t)$ Poisson process independent of $X(t-1)$.

3.2. Esimtation Procedures

3.2.1. With $G(\cdot)$ known and $h(\cdot)$ unknown

Generalized least squares estimation:

Consider to minimize wrt $h(\cdot)$ and β

$$\sum_{i=1}^{n} \left\{ \underline{X}_i - \underline{\mu}_{Z_i} \right\}' \Phi_i' W_i \Phi_i \left\{ \underline{X}_i - \underline{\mu}_{Z_i} \right\}$$

i.e. To jointly solve the GEE type EEs, with $\mathbf{Z}_i$ the $p \times M$ matrix with columns $Z_i(t_1), \ldots, Z_i(t_M)$ and
$\dot{\mathbf{G}}_i = diag\Big(\dot{G}(\beta Z_i(t_l) + h(t_l)) : l = 1, \ldots, M\Big)$,

$$\begin{cases} \sum_{i=1}^{n} \mathbf{Z}_i \dot{\mathbf{G}}_i \Phi_i' W_i \Phi_i \Big\{ \underline{X}_i - \underline{\mu}_{Z_i} \Big\} = 0 \\ \sum_{i=1}^{n} \dot{\mathbf{G}}_i \Phi_i' W_i \Phi_i \Big\{ \underline{X}_i - \underline{\mu}_{Z_i} \Big\} = 0 \end{cases}$$

Some approaches in situations for counting process data use

$$\begin{cases} \sum_{i=1}^{n} \mathbf{Z}_i \dot{\mathbf{G}}_i \Phi_i' W_i \Phi_i \Big\{ \underline{X}_i - \underline{\mu}_{Z_i} \Big\} = 0 \\ \sum_{i=1}^{n} \dot{\mathbf{G}}_i \Phi_i' W_i \Phi_i \Big\{ \underline{X}_i - \underline{\mu}_{Z_i} \Big\} = 0 \end{cases}$$

i.e. To jointly solve the GEE type EEs, with $\mathbf{Z}_i$ the $p \times M$ matrix with columns $Z_i(t_1), \ldots, Z_i(t_M)$ and $\dot{\mathbf{G}}_i = diag\Big(\dot{G}(\beta Z_i(t_l) + h(t_l)) : l = 1, \ldots, M\Big)$,

$$\begin{cases} \sum_{i=1}^{n} \mathbf{Z}_i \dot{\mathbf{G}}_i \Phi_i' W_i \Phi_i \Big\{ \underline{X}_i - \underline{\mu}_{Z_i} \Big\} = 0 \\ \sum_{i=1}^{n} \dot{\mathbf{G}}_i \Phi_i' W_i \Phi_i \Big\{ \underline{X}_i - \underline{\mu}_{Z_i} \Big\} = 0 \end{cases}$$

Some approaches in situations for counting process data use

$$\begin{cases} \sum_{i=1}^{n} \mathbf{Z}_i \cancel{\dot{\mathbf{G}}_i} \cancel{\Phi_i'} \cancel{W_i} \Phi_i \Big\{ \underline{X}_i - \underline{\mu}_{Z_i} \Big\} = 0 \\ \sum_{i=1}^{n} \cancel{\dot{\mathbf{G}}_i} \cancel{\Phi_i'} \cancel{W_i} \Phi_i \Big\{ \underline{X}_i - \underline{\mu}_{Z_i} \Big\} = 0 \end{cases}$$

3.2. Esimtation Procedures

3.2.2. With $G(\cdot)$ unknown and $h(\cdot)$ known

Consider for situations with time-independent Z

$$\tilde{X}(t;\beta) = X(h^{-1}(t - \beta' Z)).$$

Thus

$$\mathsf{E}\{\tilde{X}(t;\beta)\big|Z\} = G(t)$$

– For fixed β, use $\tilde{X}_i(\cdot;\beta)$ to estimate $G(\cdot)$, using the EEs in *3.2.1*.
– Use the estimated $G(\cdot)$, to estimate β, using the EEs in *3.2.1*

4. Situations with Non-Random Missing

In many situations, $X_i(\cdot) \not\perp \delta_i(\cdot)$. Two special cases are considered.

4.1. Longitudinal Data with Informative Censoring Time:

Motivating Example: (Jin et al, 2004) Quality of life score collected over time, censored at either the time that the study ends or the death time.

Consider the conditional independent model:

$$X_i(\cdot) \perp \delta_i(\cdot) \Big| Z_i(\cdot).$$

- Procedures in 3. may be used with some modification
- How to check for the model?

4.2. Incomplete Longitudinal Data due to Quantification Limit of the Assay

Motivating Example: The assay used in ACTG359 to quantify HIV-RNA was Amplicor bioassay, with lower detection limit 500 copies/ml. (LIKE MANY LAB DATA IN PRACTICE.)

- impute 500 for all censored HIV-RNA
- impute all censored HIV-RNA by HIV-RNA copies obtained using the Ultrasensitive bioassay

Schroeder (2004) studies the relationship between HIV-RNA obtained by Amplicor assay and obtained by the Ultrasensitive assay at one time point:

$$X \big| X^* \sim f(x \big| X^*).$$

How about to use it to obtain

$$\mathsf{E}\{X_i(t) \big| X_i^*(t)\},$$

and substitute the unobserved $X_i(t)$ with the conditional expectation?

- strong assumption
- not fully utilize the information from the neighborhood

5. Final Remarks

- the Approaches
  - intuitive: “adaptive GEE"
  - easy to implement
- their Extensions
  - more general observation settings
  - spatial data, clustered data

5. Final Remarks

- the Approaches
  - intuitive: “adaptive GEE"
  - easy to implement
- their Extensions
  - more general observation settings
  - spatial data, clustered data