Web-based Supporting Materials for "Choosing profile double-sampling designs with application to PEPFAR evaluation" by An, Frangakis, and Yiannoutsos

Ming-Wen An [1], Constantine E. Frangakis [2], Constantin T. Yiannoutsos [3]

July 19, 2013

Appendix

A.1. Description of EM algorithm.

Complete data log likelihood. We consider the complete data log likelihood defined as the likelihood assuming R known:

$$\begin{aligned}\log P(\mathcal{D}_i, R_i \mid Z_i, \theta) &= R_i^{obs}\Delta_i\{\log P(T_i \mid R_i = 1, Z_i, \theta) + \log P(R_i = 1|Z_i, \theta)\} \\ &+ R_i^{obs}(1-\Delta_i)\left\{I(1 = R_i)\log \mathrm{Surv}_{T,R=1}(C_i; Z_i, \theta) P(R_i = 1 \mid Z_i, \theta)\right. \\ &\qquad \left. + I(0 = R_i)\log \mathrm{Surv}_{L,R=0}(C_i; Z_i, \theta) P(R_i = 0 \mid Z_i, \theta)\right\} \\ &+ (1 - R_i^{obs})\log P(L_i \mid R_i = 0, Z_i, \theta) P(R_i = 0 \mid Z_i, \theta) \\ &+ (1 - R_i^{obs}) S_i \Delta_i \log \mathrm{Surv}_{T,R=0}(C_i; \mathrm{Profile}_i, \theta) \\ &+ (1 - R_i^{obs}) S_i (1-\Delta_i) \log \mathrm{Surv}_{T,R=0}(C_i; \mathrm{Profile}_i, \theta),\end{aligned}$$

[1] Department of Mathematics, Vassar College, Poughkeepsie, NY 12604, USA; mian@vassar.edu
[2] Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA; cfrangak@jhsph.edu
[3] Division of Biostatistics, Indiana University School of Medicine, Indianapolis, IN 46202, USA; cyiannou@iupui.edu

with $\mathrm{Surv}_{T,R=1}(c;\ldots) = P(T_i > c \mid R_i = 1, \ldots)$, $\quad \mathrm{Surv}_{L,R=0}(c;\ldots) = P(L_i > c \mid R_i = 0, \ldots)$,

and $\mathrm{Surv}_{T,R=0}(c;\ldots) = P(T_i > c \mid R_i = 0, \ldots)$.

<u>E-Step.</u> Let $\theta_{(k)}$ be the current estimate of θ, then

$$
\begin{aligned}
Q(\theta;\theta_{(k)}) &= E\{\log P(\mathcal{D}_i, R_i \mid Z_i, \theta) \mid \mathcal{D}_i, \theta_{(k)}\} \\
&= R_i^{obs}\{\Delta_i \log P(T_i \mid R_i = 1, Z_i, \theta) + \log P(R_i = 1 | Z_i, \theta)\} \\
&\quad + R_i^{obs}(1-\Delta_i)\big\{P(R_i = 1 \mid T_i > C_i, Z_i, \theta_{(k)}) \log \mathrm{Surv}_{T,R=1}(C_i; Z_i, \theta) P(R_i = 1 \mid Z_i, \theta) \\
&\qquad + P(R_i = 0 \mid L_i > C_i, Z_i, \theta_{(k)}) \log \mathrm{Surv}_{L,R=0}(C_i; Z_i, \theta) P(R_i = 0 \mid Z_i, \theta)\big\} \\
&\quad + (1 - R_i^{obs}) \log\{P(L_i \mid R_i = 0, Z_i, \theta) P(R_i = 0 \mid Z_i, \theta)\} \\
&\quad + (1 - R_i^{obs}) S_i \Delta_i \log \mathrm{Surv}_{T,R=0}(C_i; \mathrm{Profile}_i, \theta) \\
&\quad + (1 - R_i^{obs}) S_i (1-\Delta_i) \log \mathrm{Surv}_{T,R=0}(C_i; \mathrm{Profile}_i, \theta)
\end{aligned}
$$

where

$$P(R_i = 1 \mid T_i > C_i, Z_i, \theta_{(k)}) = \frac{p_{i,R1;\theta_{(k)}}}{p_{i,R1;\theta_{(k)}} + p_{i,R0;\theta_{(k)}}}$$

and

$$P(R_i = 0 \mid L_i > C_i, Z_i, \theta_{(k)}) = \frac{p_{i,R0;\theta_{(k)}}}{p_{i,R1;\theta_{(k)}} + p_{i,R0;\theta_{(k)}}};$$

and

$$p_{i,R1;\theta_{(k)}} = \mathrm{Surv}_{T,R=1}(C_i; Z_i, \theta_{(k)}) P(R_i = 1 \mid Z_i, \theta_{(k)})$$

and

$$p_{i,R0;\theta_{(k)}} = \mathrm{Surv}_{L,R=1}(C_i; Z_i, \theta_{(k)}) P(R_i = 0 | Z_i, \theta_{(k)}).$$

M-Step. The log likelihood can be separated into distinct terms involving the partition $(\theta_C, \theta_{T1}, \theta_{T0}, \theta_L)$ of the parameter space θ as

$$\log P(\mathcal{D}_i, R_i \mid Z_i, \theta) = L_{\theta_{T1}} + L_{\theta_{T0}} + L_{\theta_L} + L_{\theta_R}$$

where

$$\begin{aligned}
L_{\theta_{T1}} &= \sum_{i \in \{(a)\}} \log P(T_i \mid Z_i, R_i = 1, \theta_{T1}) + \sum_{i \in \{(b)\}} \omega_{i1} \log \mathrm{Surv}_{T,R=1}(C_i; Z_i, \theta_{T1}) \\
&\quad \text{with } w_{i1} = \frac{p_{R1;;\theta_{(k)}}}{p_{R1;\theta_{(k)}} + p_{R0;\theta_{(k)}}} \\
L_{\theta_{T0}} &= \sum_{i \in \{(c)\}} \log P(T_i \mid L_i, Z_i, R_i = 0, \theta_T) + \sum_{i \in \{(d)\}} \log \mathrm{Surv}_{T,R=0}(C_i; Z_i, L_i, \theta_{T0}) \\
L_{\theta_L} &= \sum_{i \in \{(b)\}} \omega_{i0} \log \mathrm{Surv}_{L,R=0}(C_i; Z_i, \theta_L) + \sum_{i \in \{(c),(d),(e)\}} \log P(L_i | R_i = 0, Z_i, \theta_L) \\
&\quad \text{with } w_{i0} = \frac{p_{R0;;\theta_{(k)}}}{p_{R1;\theta_{(k)}} + p_{R0;\theta_{(k)}}} \\
L_{\theta_R} &= \sum_{i \in \{(a),(b),(c),(d),(e)\}} \omega_{ir} \log P(R_i \mid Z_i, \theta_R) \\
&\quad \text{with } w_{ir} = \begin{cases} \frac{p_{Rr;;\theta_{(k)}}}{p_{R1;\theta_{(k)}} + p_{R0;\theta_{(k)}}} & \text{for } i \in (b); r = 0,1 \text{ (on ``expanded'' dataset)} \\ 1 & \text{otherwise} \end{cases}
\end{aligned}$$

where the groups (a), (b), (c), (d), and (e) correspond to those in Section 3.

Therefore the M-Step is equivalent to maximizing the terms separately. $L_{\theta_{T1}}$ and $L_{\theta L}$ can be maximized for any model for which survival analysis maximization can be performed with weights, and L_{θ_R} can be maximized for any model for which weighted logistic regression can be performed. $L_{\theta_{T0}}$ can be maximized using numerical optimization methods. We used a Guass-Seidel algorithm.

A.2. Proof of Result 1. By taking the second derivative of (4; as defined in the paper) and defining

$$I(e(\cdot),\theta) = -E\left\{S_i \frac{\partial^2}{\partial\theta^2}\ell^{\text{dble}}_{\text{sample}}\Big|^{\text{first}}_{\text{phase}}(\mathcal{D}_i;\theta) \mid \theta\right\},$$

$$\text{the latter becomes } = -E\left\{S_i \frac{\partial^2}{\partial\theta^2}\ell^{\text{dble}}_{\text{sample}}\Big|^{\text{first}}_{\text{phase}}(\mathcal{D}_i;\theta) \mid R_i^{obs}=0,\theta\right\}\cdot P(R_i^{obs}=0\mid\theta)$$

because a person is double sampled ($S_i = 1$) only if (but not necessarily if) they are observed dropouts ($R_i^{obs} = 0$). By iterating the expectation in the latter expression over the variables Profile_i that determine the double-sampling rule, the last expression becomes

$$= -E\left[E\left\{S_i \frac{\partial^2}{\partial\theta^2}\ell^{\text{dble}}_{\text{sample}}\Big|^{\text{first}}_{\text{phase}}(\mathcal{D}_i;\theta) \mid R_i^{obs}=0,\text{Profile}_i,\theta\right\} \mid R_i^{obs}=0,\theta\right]$$

which, by design condition 2, becomes

$$= E\left[E\left\{S_i \mid R_i^{obs}=0,\text{Profile}_i,\theta\right\}\cdot E\left\{-\frac{\partial^2}{\partial\theta^2}\ell^{\text{dble}}_{\text{sample}}\Big|^{\text{first}}_{\text{phase}}(\mathcal{D}_i;\theta) \mid R_i^{obs}=0,\text{Profile}_i,\theta\right\} \mid R_i^{obs}=0,\theta\right]$$

$$= E\left\{e(\text{Profile}_i)\,\delta(\text{Profile}_i,\theta) \mid R^{obs}=0,\theta\right\},$$

where the last expression follows from (6; as defined in the paper) since, inside the conditioning on $R^{obs} = 0$, the factor $(1 - R^{obs})$ is 1.

A.3. Proof of Result 2. Using notation defined in the paper, in particular, $S(t \mid Z, \theta) := P(T > t \mid Z, \theta)$ and $\tilde{S}(t \mid \theta) := \frac{1}{n} \sum_i S(t \mid Z_i, \theta)$,

$$
\begin{aligned}
& \sqrt{n}\left[\tilde{S}(t \mid \hat{\theta}) - S(t \mid \theta)\right] \\
= \quad & \sqrt{n}\left[\tilde{S}(t \mid \hat{\theta}) - \tilde{S}(t \mid \theta)\right] + \sqrt{n}\left[\tilde{S}(t \mid \theta) - S(t \mid \theta)\right] \\
= \quad & \sqrt{n}\left[\frac{\partial}{\partial \theta}\tilde{S}(t \mid \theta)\Big|_{\theta=\hat{\theta}}(\hat{\theta} - \theta)\right] && \text{(A.1)} \\
& -\sqrt{n}\left[\frac{\partial^2}{\partial \theta^2}\tilde{S}(t \mid \theta)\Big|_{\theta=\xi}(\hat{\theta} - \theta)'(\hat{\theta} - \theta)/2!\right] && \text{(A.2)} \\
& +\sqrt{n}\left[\tilde{S}(t \mid \theta) - S(t \mid \theta)\right] && \text{(A.3)}
\end{aligned}
$$

with the last equality following from a Taylor expansion of $\tilde{S}(t \mid \theta)$ around $\theta = \hat{\theta}$, and where ξ is some vector whose elements ξ_j are between θ_j and $\hat{\theta}_j$, for elements θ_j and $\hat{\theta}_j$ of θ and $\hat{\theta}$, respectively.

Result 2 is obtained by applying Slutzky's Theorem to the following two facts:

FACT 1. $\sqrt{n}\left[\frac{\partial^2}{\partial \theta^2}\tilde{S}(t \mid \theta)\big|_{\theta=\xi}(\hat{\theta} - \theta)'(\hat{\theta} - \theta)/2!\right] \xrightarrow{p} 0.$

FACT 2. $\sqrt{n}\left[\frac{\partial}{\partial \theta}\tilde{S}(t \mid \theta)\big|_{\theta=\hat{\theta}}(\hat{\theta} - \theta)\right] + \sqrt{n}\left[\tilde{S}(t \mid \theta) - S(t \mid \theta)\right] \xrightarrow{d} N(0, V(e(\cdot), \theta)),$
where $V(e(\cdot), \theta) = E_\theta\{S_\theta(t \mid \theta)'\} \;\; I(e(\cdot), \theta)^{-1} \;\; E_\theta\{S_\theta(t \mid \theta)\} \; + \mathrm{var}_\theta\{S(t \mid Z_i, \theta)\},$
and $S_\theta(t \mid \theta) = \partial S(t \mid \theta)/\partial \theta$.

We now verify the two facts.

Proof of Fact 1. Under regularity conditions, we have the following standard result from maximum likelihood estimation theory: $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I(e(\cdot), \theta)^{-1})$. Then $n(\hat{\theta} - \theta)'(\hat{\theta} - \theta)$

converges in distribution to some chi-squared distribution, and so

$$\frac{(\hat{\theta}-\theta)'(\hat{\theta}-\theta)}{n^{\delta-1}} \xrightarrow{p} 0$$
$$\Rightarrow \quad (\hat{\theta}-\theta)'(\hat{\theta}-\theta) = o_p(n^{\delta-1})$$

for any $\delta > 0$. Letting $\delta = 1/2$, we have $(\hat{\theta}-\theta)'(\hat{\theta}-\theta) = o_p(1/\sqrt{n})$, or equivalently,

$$\sqrt{n}(\hat{\theta}-\theta)'(\hat{\theta}-\theta) \xrightarrow{p} 0. \tag{A.4}$$

Moreover, by the law of large numbers, we have:

$$\frac{\partial^2}{\partial\theta^2}\tilde{S}(t \mid \theta)\Big|_{\theta=\hat{\theta}} \xrightarrow{p} E\left[\frac{\partial^2}{\partial\theta^2}S(t \mid Z_i, \theta)\right]. \tag{A.5}$$

Therefore applying Slutzky's Theorem to (A.4) and (A.5), we have that

$$\sqrt{n}\left[\frac{\partial^2}{\partial\theta^2}\tilde{S}(t \mid \theta)\Big|_{\theta=\xi}(\hat{\theta}-\theta)'(\hat{\theta}-\theta)/2!\right] \xrightarrow{p} 0.$$

Proof of Fact 2. We will show that $\varphi(\mathcal{D}_i)$ and $S(t \mid Z_i, \theta) - S(t \mid \theta)$ are asymptotically independently normal:

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, I((e(\cdot), \theta)^{-1})$$

and

$$\sqrt{n}\left[\tilde{S}(t \mid \theta) - S(t;\theta)\right] \xrightarrow{d} N(0, \mathrm{var}_\theta\{S(t \mid Z_i, \theta)\}) \tag{A.6}$$

are asymptotically independent. Further, by the strong law of large numbers, we have:

$$\frac{\partial}{\partial\theta}\tilde{S}(t \mid \theta)\Big|_{\theta=\hat{\theta}} \xrightarrow{p} E_\theta\left[\frac{\partial}{\partial\theta}S(t \mid \theta)\right].$$

Then we obtain Fact 2 by application of Slutzky's Theorem:

$$\sqrt{n}\frac{\partial}{\partial\theta}\tilde{S}(t \mid \theta)\Big|_{\theta=\hat{\theta}}(\hat{\theta}-\theta) + \sqrt{n}\left[\tilde{S}(t \mid \theta) - S(t;\theta)\right] \xrightarrow{d} N(0, V(e(\cdot),\theta)),$$

where $V(e(\cdot),\theta) = E_\theta\{S_\theta(t \mid \theta)'\} \quad I(e(\cdot),\theta)^{-1} \quad E_\theta\{S_\theta(t \mid \theta)\} \ + \mathrm{var}_\theta\{S(t \mid Z_i,\theta)\}$, and $S_\theta(t \mid \theta) = \partial S(t \mid \theta)/\partial\theta$.

We now verify (A.6) by showing that $\varphi(\mathcal{D}_i)$ and $S(t \mid Z_i,\theta) - S(t \mid \theta)$ are jointly asymptotically normal with zero covariance. Using the theory of influence functions, we can express $\sqrt{n}(\hat{\theta}-\theta)$ from (A.1) in a linearized form as follows:

$$\sqrt{n}(\hat{\theta}-\theta) \quad = \quad \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}\varphi(\mathcal{D}_i|Z_i),$$

where $\varphi(x) = I^{-1}(\theta)\frac{\partial}{\partial\theta}\log f(x|\theta)$. Since $\varphi(\mathcal{D}_i)$ and $S(t;\theta,Z_i) - S(t;\theta)$ are data from individuals, we have

$$\sqrt{n}\sum_{i=1}^{n}\begin{pmatrix} \varphi(\mathcal{D}_i) \\ S(t|Z_i,\theta) - S(t|\theta) \end{pmatrix} \xrightarrow{d} N\left(\begin{matrix} 0 \\ 0 \end{matrix}, V_1\right), \tag{A.7}$$

where

$$V_1 = \begin{pmatrix} E_\theta[\varphi(\mathcal{D}_i)'\varphi(\mathcal{D}_i)] & \mathrm{cov}[\varphi(\mathcal{D}_i), S(t|Z_i,\theta) - S(t|\theta)] \\ \mathrm{cov}[\varphi(\mathcal{D}_i), S(t|Z_i,\theta) - S(t|\theta))] & \mathrm{var}_\theta\{S(t \mid Z_i,\theta)) \end{pmatrix}.$$

We now show the covariance term is 0:

$$\begin{aligned}
\mathrm{cov}[\varphi(\mathcal{D}_i), S(t|Z_i,\theta) - S(t|\theta)] &= E_\theta\{\varphi(\mathcal{D}_i)\cdot[S(t|Z_i,\theta) - S(t|\theta)]\} \\
&= E_Z\big\{E_\theta\{\varphi(\mathcal{D}_i)\cdot[S(t;\theta,Z_i) - S(t;\theta)]\ |Z\}\big\} \\
&= E_Z\big\{E_\theta\{I^{-1}(\theta)\frac{\partial}{\partial\theta}\log f(\mathcal{D}_i \mid \theta)\cdot[S(t|Z_i,\theta) - S(t\mid\theta)]\ |Z\}\big\} \\
&= 0,
\end{aligned}$$

where the last equality follows since the score function, $\frac{\partial}{\partial\theta}\log f(\mathcal{D}_i \mid \theta)$, has zero expectation with respect to θ given Z, and the remaining product terms are constant given Z. Therefore

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, I(e(\cdot),\theta)^{-1})$$

and

$$\sqrt{n}\left[\tilde{S}(t\mid\theta) - S(t;\theta)\right] \xrightarrow{d} N(0, \mathrm{var}_\theta\{S(t\mid Z_i,\theta)\})$$

are asymptotically independent, where $I(e(\cdot),\theta)^{-1} = E_\theta[\varphi(\mathcal{D}_i)\varphi(\mathcal{D}_i)']$.