Non/Semi-Parametric Estimation of the Failure Time
Distribution in the Presence of Informative Censoring:

Technical Report

Daniel O. Scharfstein and James M. Robins *

November 27, 2000

1. Notation and Data

Let $T^*$ be a (non-negative) failure time random variable which is absolutely continuous with respect to Lebesgue measure. Let $\overline{V}(t)$ denote the history of a covariate process $V(t)$ through time $t-$. For simplicity, we will assume that all subjects have a common, fixed follow-up time $c$. Let $T = \min(T^*, c^*)$. Note that the support of the distribution of $T$ is on the interval $[0, c^*]$. In the absence of a censoring, we think of the “complete” data for an individual as

$$ L \equiv (T, \overline{V}(T)) $$

Because of right censoring, we only observe

$$ O \equiv (X = \min(T, C), \overline{V}(X), \Delta = I(T \leq C)) $$

---

Daniel O. Scharfstein is Assistant Professor of Biostatistics, Johns Hopkins School of Hygiene and Public Health, Baltimore, MD 21205. James M. Robins is Professor of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, MA 02115. This research was partially supported by National Institute of Health grants 1-R29-GM48704-04, 5R01A132475, R01CA74112, 1-R01-MH56639-01A1, and 1-R01-DA10184-01A2.
where $C = \min(C^*, c^*)$ and $C^*$ is a (non-negative) censoring time random variable which is also absolutely continuous with respect to Lebesgue measure.

Our goal is to use $n$ i.i.d. copies of $O$,

$$O = \{O_i = (X_i, \overline{V}_i(X_i), \Delta_i) : i = 1, \ldots, n\}$$

to draw inference about $S_0(u) = P[T \geq u]$. This is of interest because, for $u \in [0, c^*]$, $S_0(u) = P[T^* \geq u]$.

2. General Theory

Let $L$ denote the complete (full) data. Suppose we only observe $(R, L_{(R)})$, where $L_{(R)} = \varphi_R(L)$ and $\varphi_r(L)$ is a known function of $L$ which depends on $r$. Specifically, $R$ indexes the part of $L$ that is actually observed. We assume that there exists a unique value of $R$, $r^*$, such that $\varphi_{r^*}(L) = L$. Let $\Delta = I(R = r^*)$. Furthermore, we assume that (i) $L$ follows an arbitrary semiparametric model, $F_L$, indexed by a $p \times 1$ parameter $\mu$ and an infinite dimensional parameter $\theta$, (ii) $R$ given $L$ follows an arbitrary semiparametric model, $F_{R|L}$, indexed by a $q \times 1$ parameter $\gamma$ and an infinite dimensional parameter $\eta$, and (iii) $P_{\theta} = P[\Delta = 1|L] > \sigma > 0$. We assume that the parameters in model $F_L$ are variation independent of those in the model $F_{R|L}$. We let $\mu_0$, $\gamma_0$, $\theta_0$, and $\eta_0$ denote the true values of $\mu$, $\gamma$, $\theta$, and $\eta$, respectively. We are interested in estimating $\psi_0 = (\mu_0', \gamma_0')'$. We observe $n$ independent identically distributed copies $O_i = (R_i, L_{i(R_i)})$ of $O = (R, L_{(R)})$.

Let $\Lambda_1 = \Lambda(F_L)$ and $\Lambda_2 = \Lambda(F_{R|L})$ denote the (nuisance) tangent spaces for $\theta$ and $\eta$, respectively had we observed $(R, L)$ (see Newey, 1990). Throughout, our spaces are sub-spaces of the Hilbert space of $q + p$ dimensional mean zero random vectors with the covariance inner product. Note that $\Lambda(F_L)$ and $\Lambda(F_{R|L})$ are orthogonal. For the “observed data”, there is an induced semiparametric model which we denote by $O$. In model $O$, the observed data nuisance tangent space is $\Lambda^O = \Lambda^O_1 + \Lambda^O_2$, where $\Lambda^O_1$ is the observed data nuisance tangent space for $\theta$ and $\Lambda^O_2$ is the observed data nuisance tangent space for $\eta$. Specifically, $\Lambda^O_j = \overline{R(\varphi \circ \Pi_j)}$, where $\overline{R(\cdot)}$ is the range of an operator, $g : \Omega^{(R,L)} \to \Omega^{(R,L_{(R)})}$, $g(\cdot) = E[\cdot|R, L_{(R)}]$, $\Omega^{(R,L)}$ and $\Omega^{(R,L_{(R)})}$ are spaces of $p + q$ dimensional mean zero random functions of $(R, L)$ and $(R, L_{(R)})$, $\Pi_j$ is the projection operator from $\Omega^{(R,L)}$ onto $\Lambda_j$ and $\overline{\cdot}$ denotes the close linear span of the set $\mathcal{S}$. (Bickel et al., 1993). A space superscripted by $\perp$ denotes the orthogonal complement of that space. We are interested in finding $\Lambda^{O,\perp}$ because, in sufficiently smooth models including all those studied in this paper, the set of influence functions of all asymptotically linear (RAL)
estimators of $\psi_0$ is the set \( \left\{ E \left[ A S_\psi \right]^{-1} A \right\} \mathcal{A} \rightleftharpoons \mathcal{A} \in \Lambda_0^{O,\perp} \), where $S_\psi$ is the observed data score for $\psi$ evaluated at the truth. Another motivation for our interest in this space is as follows. An element in the $\Lambda^{O,\perp}$ space is a $(p + q)$ dimensional function of the observed data for an individual and the true values of the parameters, $\psi_0$, $\theta_0$, and $\eta_0$. Denote this function by $U \equiv U(\psi_0, \theta_0, \eta_0)$. Suppose we estimate $\psi_0$ by $\hat{\psi}$ solving $\sum_i U_i \left( \psi, \hat{\theta}(\psi), \hat{\eta}(\psi) \right) = 0$ where $\hat{\theta}(\psi_0)$ and $\hat{\eta}(\psi_0)$ converge to $\theta_0$ and $\eta_0$, respectively. Then Bickel et al. (1993) and Newey (1990) show that under suitable regularity conditions $\hat{\psi}$ is a RAL estimator with influence function $\tau^{-1} U$ where $\tau = E \left[ U S_\psi \right] = -E \left[ \partial U (\psi_0, \theta_0, \eta_0) / \partial \psi \right]$. But this is the same influence function as would have been obtained by solving the estimating equation $\sum_i U_i (\psi; \theta_0, \eta_0) = 0$ in which the infinite dimensional components $(\theta_0, \eta_0)$ are known rather than estimated. It is precisely the orthogonality of $U$ to $\Lambda^{O}$ which obviated the need to adjust the asymptotic variance for estimation of the nuisance parameters.

Taking orthogonal complements, $\Lambda^{O,\perp} = \Lambda_1^{O,\perp} \cap \Lambda_2^{O,\perp}$. Let $a(L)$ and $b(R, L_{(R)})$ be $p + q$ dimensional functions of $L$ and $(R, L_{(R)})$, respectively. Rotnitzky and Robins (1997) showed how to compute $\Lambda^{O,\perp}_1$. Specifically,

$$\Lambda^{O,\perp}_1 = \{ \Delta Pr[\Delta = 1|L]^{-1} a(L) + b(R, L_{(R)}) : a(L) \in \Lambda(F_L)^\perp \text{ and } E[b(R, L_{(R)})|L] = 0 \}.$$ 

By the relationship between range and null spaces, we know that $\Lambda^{O,\perp}_2 = N(\Pi_2^T \circ g^T)$, where $N(\cdot)$ is the null space of an operator, and superscript $T$ denotes the adjoint of an operator. As a projection operator $\Pi_2^T = \Pi_2$ and $g^T$ is the identity operator. So,

$$\Lambda^{O,\perp}_2 = \{ b(R, L_{(R)}) : \Pi[b(R, L_{(R)})|\Lambda(F_R|L)] = 0 \} = \{ b(R, L_{(R)}) : b(R, L_{(R)}) \in \Lambda(F_R|L)^\perp \}$$

3. Application of General Theory to Models $A_q$ and $B_q$

We let $R = C$ if $\Delta = 0$ and $R = r^* \equiv \infty$ if $\Delta = 1$. Then, we define

$$\varphi_r(L) = \begin{cases} 
(T > r, V(r)) & r < \infty \\
(T, V(T)) & r = \infty 
\end{cases}$$

November 27, 2000
3.1. Model $A_q$

Let $N(t) = I(X \leq t, \Delta = 0)$ be the counting process for censoring and $M(t) = N(t) - \int_0^t I(X \geq u)\lambda_0(u|\overline{V}(u)) \exp(q(u, \overline{V}(T), T)) du$ be its associated martingale. Define

$$S(t|\overline{V}(T), T) = P[C \geq t|\overline{V}(T), T, T \geq t] = \exp(-\int_0^t \lambda_0(u|\overline{V}(u)) \exp(q(u, \overline{V}(T), T)) du)$$

and $F(t|\overline{V}(T), T) = 1 - S(t|\overline{V}(T), T)$.

$$\Lambda(\mathcal{F}_L)^\perp = \{k(I(T \geq u) - S_0(u)) : k \in R^1\}$$

$$\Lambda(\mathcal{F}_{R|L})^\perp = \{a(\overline{V}(T), T) + \int_0^\infty w(t, \overline{V}(T), T) dM(t) : E[a(\overline{V}(T), T)] = 0, E[w(t, \overline{V}(T), T)|C = t, \overline{V}(t), T \geq t] = 0,$$

$$w(t, \overline{V}(t)) \text{ is an arbitrary function of } t \text{ and } \overline{V}(t)\}$$

Note that any function $c(R, L_{(R)})$ admits the unique representation $\Delta a(\overline{V}(T), T) + (1-\Delta)b(\overline{V}(C), C)$, where $a(\overline{V}(T), T)$ and $b(\overline{V}(C), C)$ are arbitrary functions of $(\overline{V}(T), T)$ and $(\overline{V}(C), C)$, respectively. Thus, $E[c(R, L_{(R)})|L] = 0$ if and only if

$$c(R, L_{(R)}) = -\Delta \frac{E[(1-\Delta)b(\overline{V}(C), C)|\overline{V}(T), T]}{S(T|\overline{V}(T), T)} + (1-\Delta)b(\overline{V}(C), C)$$

With this result, it is easy to express $\Lambda_1^{O,\perp}$ as

$$\Lambda_1^{O,\perp} = \left\{ \frac{\Delta}{S(T|\overline{V}(T), T)}(k(I(T \geq u) - S_0(u)) - E[(1-\Delta)b(\overline{V}(C), C)|\overline{V}(T), T]) + (1-\Delta)b(\overline{V}(C), C) : k \in R^1 \right\}$$

To compute $\Lambda_2^{O,\perp}$, it is useful to note that an observed data random variable, $c(R, L_{(R)}) = \Delta a(\overline{V}(T), T) + (1-\Delta)b(\overline{V}(C), C)$ can be written as
\[ c(R, L_{(R)}) = \Delta a(\nabla(T), T) + (1 - \Delta) b(\nabla(C), C) + \frac{\Delta}{S(T|\nabla(T), T)} E[(1 - \Delta) b(\nabla(C), C)|\nabla(T), T] - \frac{\Delta}{S(T|\nabla(T), T)} E[(1 - \Delta) b(\nabla(C), C)|\nabla(T), T] \]

\[ = \frac{\Delta}{S(T|\nabla(T), T)} m(\nabla(T), T) + (1 - \Delta) b(\nabla(C), C) - \frac{\Delta}{S(T|\nabla(T), T)} E[(1 - \Delta) b(\nabla(C), C)|\nabla(T), T] \]

\[ = m(\nabla(T), T) + \int_0^\infty g(t, \nabla(T), T)dM(t) \]

where

\[ m(\nabla(T), T) = E[\Delta a(\nabla(T), T) + (1 - \Delta) b(\nabla(C), C)|\nabla(T), T] \]

\[ g(t, \nabla(T), T) = b(\nabla(t), t) + \frac{\int_0^t b(\nabla(u), u)dF(\nabla(T), T)}{S(t|\nabla(T), T)} - \frac{m(\nabla(T), T)}{S(t|\nabla(T), T)} \]

Thus, it can be seen that \( c(R, L_{(R)}) \in \Lambda(F_{R|L}) \) if and only if \( E[m(\nabla(T), T)] = 0 \) and

\[ E \left[ b(\nabla(t), t) + \frac{\int_0^t b(\nabla(u), u)dF(\nabla(T), T)}{S(t|\nabla(T), T)} - \frac{m(\nabla(T), T)}{S(t|\nabla(T), T)} | C = t, \nabla(t), T \geq t \right] = 0 \text{ for all } t. \]

This latter restriction implies that \( b(\nabla(t), t) \) is the unique solution to the following Volterra integral equation

\[ b(\nabla(t), t) = J_m(t) - \int_0^t b(\nabla(u), u)f(u, t, \nabla(t))du \]  \hspace{1cm} (3.1)

where

\[ J_m(t) = \frac{E[m(\nabla(T), T) \exp(q(t, \nabla(T), T))I(T \geq t)|\nabla(t)]}{E[S(t|\nabla(T), T) \exp(q(t, \nabla(T), T))I(T \geq t)|\nabla(t)]} \]

\[ f(u, t, \nabla(t)) = \frac{E[\lambda_0(u|\nabla(t))S(u|\nabla(T), T) \exp(q(u, \nabla(T), T) + q(t, \nabla(T), T))I(T \geq t)|\nabla(t)]}{E[S(t|\nabla(T), T) \exp(q(t, \nabla(T), T))I(T \geq t)|\nabla(t)]} \]
Since \( m(\overline{V}(T), T) \) can be any mean 0 function of \( (\overline{V}(T), T) \), we can put these results together to see that

\[
\Lambda_2^{O, \perp} = \{ \frac{\Delta}{S(T|\overline{V}(T), T)} (m(\overline{V}(T), T) - E[(1 - \Delta)b(\overline{V}(C), C)|\overline{V}(T), T]) + (1 - \Delta)b(\overline{V}(C), C) : E[m(\overline{V}(T), T)] = 0, b(\overline{V}(t), T) \text{ solves (2.1)} \}
\]

Intersecting \( \Lambda_1^{O, \perp} \) and \( \Lambda_2^{O, \perp} \), we find that

\[
\Lambda^{O, \perp} = \{ \frac{\Delta}{S(T|\overline{V}(T), T)} (k(I(T \geq u) - S_0(u)) - E[(1 - \Delta)b(\overline{V}(C), C)|\overline{V}(T), T]) + (1 - \Delta)b(\overline{V}(C), C) : k \in R^1, b(\overline{V}(t), T) \text{ solves (2.1)} \}
\]

3.2. Model \( B_q \)

Let \( N(t) = I(X \leq t, \Delta = 0) \) be the counting process for censoring and \( M(t) = N(t) - \int_0^t I(X \geq u)\lambda_0(u) \exp(\gamma_0 r(t, \overline{V}(t)) + q(u, \overline{V}(T), T))du \) be its associated martingale. Define

\[
S(t|\overline{V}(T), T; \gamma_0) = P[C \geq t|\overline{V}(T), T, T \geq t] = \exp(- \int_0^t \lambda_0(u) \exp(\gamma_0 r(T, \overline{V}(t)) + q(u, \overline{V}(T), T))du)
\]

and \( F(t|\overline{V}(T), T; \gamma_0) = 1 - S(t|\overline{V}(T), T; \gamma_0) \).

\[
\Lambda(\mathcal{F}_L)^{\perp} = \{ k(I(T \geq u) - S_0(u)) : k \in R^{q+1} \}
\]

\[
\Lambda(\mathcal{F}_{R|L})^{\perp} = \{ a(\overline{V}(T), T) + \int_0^\infty w(t, \overline{V}(T), T)dM(t) : E[a(\overline{V}(T), T)] = 0, E[w(t, \overline{V}(T), T)|C = t, T \geq t] = 0, w(t, \overline{V}(t)) \text { is an arbitrary } (q + 1) \text { function of } t \text { and } \overline{V}(t) \} \}
\]

Note that any function \( c(R, L_{(R)}) \) admits the unique representation \( \Delta a(\overline{V}(T), T) + (1 - \Delta)b(\overline{V}(C), C) \), where \( a(\overline{V}(T), T) \) and \( b(\overline{V}(C), C) \) are arbitrary functions of \( (\overline{V}(T), T) \) and \( (\overline{V}(C), C) \), respec-
tively. Thus, \( E[c(R, L(R)) | L] = 0 \) if and only if

\[
c(R, L(R)) = -\Delta \frac{E[(1 - \Delta)b(\nabla(C), C)|\nabla(T), T]}{S(T|\nabla(T), T; \gamma_0)} + (1 - \Delta)b(\nabla(C), C)
\]

With this result, it is easy to express \( \Lambda_1^{O,1} \) as

\[
\Lambda_1^{O,1} = \left\{ \frac{\Delta}{S(T|\nabla(T), T; \gamma_0)} (k(I(T \geq u) - S_0(u)) - E[(1 - \Delta)b(\nabla(C), C)|\nabla(T), T]) + (1 - \Delta)b(\nabla(C), C) : k \in R^{q+1} \right\}
\]

To compute \( \Lambda_2^{O,1} \), it is useful to note that an observed data random variable, \( c(R, L(R)) = \Delta a(\nabla(T), T) + (1 - \Delta)b(\nabla(C), C) \) can be written as

\[
c(R, L(R)) = \Delta a(\nabla(T), T) + (1 - \Delta)b(\nabla(C), C) + \frac{\Delta}{S(T|\nabla(T), T; \gamma_0)} E[(1 - \Delta)b(\nabla(C), C)|\nabla(T), T] - \frac{\Delta}{S(T|\nabla(T), T; \gamma_0)} E[(1 - \Delta)b(\nabla(C), C)|\nabla(T), T] = \frac{\Delta}{S(T|\nabla(T), T; \gamma_0)} m(\nabla(T), T) + (1 - \Delta)b(\nabla(C), C) - \frac{\Delta}{S(T|\nabla(T), T; \gamma_0)} E[(1 - \Delta)b(\nabla(C), C)|\nabla(T), T] = m(\nabla(T), T) + \int_0^\infty g(t, \nabla(T), T)dM(t)
\]

where

\[
m(\nabla(T), T) = E[\Delta a(\nabla(T), T) + (1 - \Delta)b(\nabla(C), C)|\nabla(T), T]
\]

\[
g(t, \nabla(T), T) = b(\nabla(t), t) + \frac{\int_0^t b(\nabla(u), u)dF(u|\nabla(T), T)}{S(t|\nabla(T), T; \gamma_0)} - \frac{m(\nabla(T), T)}{S(t|\nabla(T), T; \gamma_0)}
\]

November 27, 2000
Thus, it can be seen that $c(R, L(R)) \in \Lambda(F_{RL})^\perp$ if and only if $E[m(\overline{V}(T), T)] = 0$ and

$$E \left[ b(\overline{V}(t), t) + \int_0^t b(\overline{V}(u), u) dF(u|\overline{V}(T), T) - \frac{m(\overline{V}(T), T)}{S(t|\overline{V}(T), T)} | C = t, T \geq t \right] = 0 \text{ for all } t. $$

This latter restriction implies that

$$E \left[ b(\overline{V}(t), t)S(t|\overline{V}(T), T; \gamma_0) \exp(\gamma_0 r(t, \overline{V}(t)) + q(t, \overline{V}(T), T))I(T \geq t) \right] + $$

$$E \left[ \int_0^t b(\overline{V}(u), u) dF(u|\overline{V}(T), T; \gamma_0) \exp(\gamma_0 r(t, \overline{V}(t)) + q(t, \overline{V}(T), T))I(T \geq t) \right] - $$

$$E \left[ m(\overline{V}(T), T) \exp(\gamma_0 (t, \overline{V}(t)) + q(t, \overline{V}(T), T))I(T \geq t) \right] = 0 \text{ for all } t \quad (3.2)$$

Note that $m(\overline{V}(T), T)$ can be any mean 0 function of $(\overline{V}(T), Y)$. Given any fixed function $m(\overline{V}(T), T)$, there are an infinite number of solutions to Equation (2.2). It is straightforward to check that for any function $\phi(\overline{V}(t), t)$, the solution to the following “Volterra-like” recursive integral equation satisfies (2.2).

$$b(\overline{V}(t), t) = \phi(\overline{V}(t), t) - E \left[ S(t|\overline{V}(T), T; \gamma_0) \exp(\gamma_0 r(t, \overline{V}(t)) + q(t, \overline{V}(T), T))I(T \geq t) \right]^{-1} q_{b, \phi, m}(t)$$

where

$$q_{b, \phi, m}(t) = E \left[ \phi(\overline{V}(t), t)S(t|\overline{V}(T), T; \gamma_0) \exp(\gamma_0 r(t, \overline{V}(t)) + q(t, \overline{V}(T), T))I(T \geq t) \right] + $$

$$E \left[ \int_0^t b(\overline{V}(u), u) dF(u|\overline{V}(T), T; \gamma_0) \exp(\gamma_0 r(t, \overline{V}(t)) + q(t, \overline{V}(T), T))I(T \geq t) \right] - $$

$$E \left[ m(\overline{V}(T), T) \exp(\gamma_0 r(t, \overline{V}(t)) + q(t, \overline{V}(T), T))I(T \geq t) \right]$$

Furthermore, by ranging over all possible functions $\phi(\overline{V}(t), t)$, it is clear that we can generate all solutions to (2.2). The implication of this result is that we can rewrite $\Lambda_2^{0, \perp}$ as
\[ \Lambda_2^{0, \perp} = \left\{ \frac{\Delta}{S(T|\mathbf{V}(T), T; \gamma_0)} (a(\mathbf{V}(T), T) - E[(1 - \Delta)b(\mathbf{V}(C), C)|\mathbf{V}(T), T]) + (1 - \Delta)b(\mathbf{V}(C), C) : \\
\right. \\
a(\mathbf{V}(T), T) \text{ is an arbitrary } q + 1 \text{-dimensional, mean zero, function of } (\mathbf{V}(T), T) \\
\left. b(\mathbf{V}(t), t) = \phi(\mathbf{V}(t), t) - E \left[ \int S(t|\mathbf{V}(T), T; \gamma_0) \exp(\gamma_0 r(t, \mathbf{V}(t)) + q(t, \mathbf{V}(T), T)) I(T \geq t) \right]^{-1} q_{0, \phi, a}(t), \\
\phi(\mathbf{V}(t), t) \text{ is an arbitrary } q + 1 \text{-dimensional function of } (\mathbf{V}(t), t) \right\} \]

Intersecting \( \Lambda_1^{0, \perp} \) and \( \Lambda_2^{0, \perp} \), we find that

\[ \Lambda_1^{0, \perp} = \left\{ \frac{\Delta}{S(T|\mathbf{V}(T), T; \gamma_0)} (a(\mathbf{V}(T), T) - E[(1 - \Delta)b(\mathbf{V}(C), C)|\mathbf{V}(T), T]) + (1 - \Delta)b(\mathbf{V}(C), C) : \\
\right. \\
a(\mathbf{V}(T), T) = k(I(T \geq u) - S_0(u)), k \in R^{q+1} \\
\left. b(\mathbf{V}(t), t) = \phi(\mathbf{V}(t), t) - E \left[ \int S(t|\mathbf{V}(T), T; \gamma_0) \exp(\gamma_0 r(t, \mathbf{V}(t)) + q(t, \mathbf{V}(T), T)) I(T \geq t) \right]^{-1} q_{0, \phi, a}(t), \\
\phi(\mathbf{V}(t), t) \text{ is an arbitrary } q + 1 \text{-dimensional function of } (\mathbf{V}(t), t) \right\} \]