In order to stress that $dM$ depends on $\mu, \Lambda_0$ and that $g(h)$ depends on $\mu$, we will now and then denote these quantities with $dM_{\mu, \Lambda_0}$ and $g_{\mu}$, respectively. Here $\int g(h)(t) dM(t)$ equals the projection of $\int h(t, \tilde{Z}(t^-)) dM(t)$ onto the tangent space $\{ \int g(t) dM(t) : g \}$ of $\Lambda_0$. In other words, the full data estimating functions are of the form $\int h(t, \tilde{Z}(t^-)) dM(t)$ for $h$ chosen so that it is orthogonal to the tangent space of $\Lambda_0$. Thus, each $h$ yields an estimating function for $\mu = \beta$ indexed by a nuisance parameter $\Lambda_0$. The optimal estimating function in the full data model in which one observes $n$ i.i.d. observations of $X$ is obtained by selecting $h(t, \tilde{Z}(t^-)) = W(t)$, which corresponds with the full data efficient score $S_{\text{eff}}^F(X | F_X)$. Our general choice (3.6) $IC_0(Y | G, D_h)$ is given by

$$IC_0(Y | G, D_h(\cdot | \mu, \Lambda_0)) = D_h(X | \mu, \Lambda_0) \Delta \frac{\Delta}{G(T | X)},$$

where $\Delta = I(C \geq T)$ and $\tilde{G}(t | X) = P(C \geq t | X)$. Therefore, $D_h$ is an integral (sum) of unbiased estimating functions, an alternative choice of $IC_0(Y | G, D_h)$ is given by

$$IC_{02}(Y | G, D_h(\cdot | \mu, \Lambda_0)) = \int h(t, \tilde{Z}(t^-)) dM(t) \frac{IC(C \geq t)}{G(t | X)} dM_{\mu, \Lambda_0}(t).$$

We have the following lemma (as in Robins, 1993a).

**Lemma 3.1** If $D(X)I(\tilde{G}(T | X) > 0) = D(X) F_X$-a.e., then

$$E(\int h(t, \tilde{Z}(t^-)) dM(t) = D(X) F_X$$

$$E(\int h(t, \tilde{Z}(t^-)) dM(t) = D(X) F_X$$

If $\int h(t, \tilde{Z}(t^-)) I(\tilde{G}(t | X) > 0) dM(t) = \int h(t, \tilde{Z}(t^-)) dM(t)$. Then

$$E(\int h(t, \tilde{Z}(t^-)) I(\tilde{G}(t | X) > 0) dM(t).$$

**Proof.** We have

$$E\left( \frac{D(X)I(T \leq C)}{G(T | X)} \right) = D(X) \frac{I(\tilde{G}(T | X) > 0)D(X)}{I(\tilde{G}(T | X) > 0)}.$$

$$E(\int h(t, \tilde{Z}(t^-)) dM(t) = \int h(t, \tilde{Z}(t^-)) dM(t).$$

Let $IC_0(Y | G, D_h(\cdot | \mu, \Lambda_0))$ denote one of these two choices of estimating functions for $\mu$ with nuisance parameters $\Lambda_0, G$. Given estimators $G_n, \Lambda_{0,n}$ of $G, \Lambda_0$, each of these observed data estimating functions yields an estimating equation for $\mu = \beta$:

$$0 = \sum_{i=1}^{n} IC_0(Y_i | G_n, D_h(\cdot | \mu, \Lambda_{0,n}).$$

As discussed in Section 3.1, if we assume a multiplicative intensity model for $A(t) = I(C \leq t)$ w.r.t. $F(t) = (\tilde{A}(t), \tilde{X}(\min(t, C)))$, then Coxph() can be used to obtain an estimate of $G$. We also need a reasonable estimator of $\Lambda_0$. Since our estimating function is orthogonal to the nuisance tangent space in the observed data model $\mathcal{M}(G)$ with $G$ known, which thus includes the tangent space generated by $\Lambda_0$, the influence curve of $\mu_n$ is not affected by the first-order behavior of $\Lambda_{0,n}$ (except that it needs to be consistent at an appropriate rate). Therefore, it suffices to construct an ad hoc estimator of $\Lambda_0$. Since $E(dN(t)) = EE(dN(t) | \tilde{Z}(t^-)) = \lambda_0(t)E(Y(t) \exp(\beta W(t)))$ it follows that

$$\lambda_0(t) = \lambda_0(t | \beta) = \int_0^t \frac{E(dN(t))}{E(Y(t) \exp(\beta W(t)))}.$$

For general $\beta$, we denoted the right-hand side of the last equation with $\lambda_0(t | \beta)$, while at the true $\beta$ it equals $\lambda_0(t)$. Now, note that

$$E(dN(t)) = E(\frac{dN(t)I(C > t)}{G(t | X)}),$$

$$E(Y(t) \exp(\beta W(t))) = E(Y(t) \exp(\beta W(t)) I(C > t)) \frac{I(C > t)}{G(t | X)}.$$

This suggests the following estimator of $\lambda_0(t | \beta)$:

$$\lambda_{0,n}(t | \beta) = \frac{1}{n} \sum_{i=1}^{n} \int \frac{1}{n} \sum_{i=1}^{n} \frac{dN_i(t)I(C_i > t)}{G_n(t | X_i)}.$$

Substitution of $\lambda_{0,n}(t | \beta)$ for $\lambda_0$ in our estimating function yields the following estimating equation for $\beta$:

$$0 = \sum_{i=1}^{n} [IC_0(Y_i | G_n, D_h(\cdot | \beta, \lambda_{0,n}(\cdot | \beta))).$$

(3.10)

We now reparametrize the full data estimating function so that it has a variation-independent nuisance parameter

$$D_h^*(X | \beta, \rho) = D_h(X | \beta, \lambda_0(\beta)),$$

where $\rho$ denotes the additional parameters beyond $\beta$ identifying $\lambda_0(\cdot | \beta)$. We denote this reparametrized class of full data structure estimating functions with $D_h(\cdot | \beta, \rho)$ again.

We will now provide a sensible data-adaptive choice for the full data index $h$. If $\lambda_C(t | X) = \lambda_C(t | Z)$ so that censoring is explained by the covariates in our multiplicative intensity model, then a sensible estimating function is the one corresponding with the score of the partial likelihood for $\beta$ and $\lambda_0$ ignoring $V_2(t)$ which is given by (Andersen, Borgan, Gill, and Keiding, 1993)

$$\int \left \{ W(t) - \frac{E(W(t)Y(t)I(C > t) \exp(\beta W(t)))}{E(Y(t)I(C > t) \exp(\beta W(t)))} \right \} I(C > t) dM(t).$$

(3.11)