Chapter 6

Semiparametric Approach

6.1 Introduction and notation

The problem we study arises in the epidemiological context. Our goal is to make inferences about the time to an event ϵ according to the values of some covariates of interest. All the variables have been measured at the beginning of the survival time, but in some individuals some part of the covariates of interest are missing. For all the subjects we have the censored or uncensored survival time and other variables possibly correlated with the covariates of interest.

Let T be the survival (or time to ϵ) time and C the censoring time, with cumulative distribution functions F and G, respectively. Denote by $Y = T \wedge C$ the observed survival time and by $\delta = I(T \leq C) = I(Y = T)$ the censoring indicator. For each individual we have one vector $\boldsymbol{X} = (X_1, \ldots, X_p)'$ of p discrete covariates of interest and confounding variables, and another vector $\boldsymbol{V} = (V_1, \ldots, V_s)'$ of s other variables, possibly surrogates of $\boldsymbol{X}$, and variables in the causal pathway from $\boldsymbol{X}$ to Y. As usual in this kind of studies, we suppose that C is independent of T, given $\boldsymbol{X}$ and $\boldsymbol{V}$.

We are interested in making inferences about some functional of the distribution of T given $\boldsymbol{X}$, when part of the covariates $\boldsymbol{X}$ have not been observed (for example, we may want to estimate $E(T|\boldsymbol{X})$ for each category in $\boldsymbol{X}$, the median of $T|\boldsymbol{X}$, some percentiles of $T|\boldsymbol{X}$, the Kaplan–Meier estimator of $T|\boldsymbol{X}$, the differences between $T|\boldsymbol{X} = \boldsymbol{a}$ and $T|\boldsymbol{X} = \boldsymbol{b}$,...).

We define the vector of potential data as $\boldsymbol{L}_i = (Y_i, \delta_i, \boldsymbol{X}'_i, \boldsymbol{V}'_i)'$ for $i = 1, \ldots, n$. $\boldsymbol{L}_i$ is a vector of $2 + p + s$ components. For the j-th component of $\boldsymbol{X}$, we define R_j as the binary variable which takes value 1 if this component of $\boldsymbol{X}$ has been observed, and 0 otherwise. $\boldsymbol{R} = (R_1, \ldots, R_p)'$ is the indicator vector of response in the covariates $\boldsymbol{X}$. Denote by $\Omega_{\boldsymbol{R}} = \{0, 1\}^p$ the sample space of the random vector $\boldsymbol{R}$.

For the i-th individual, $i = 1, \ldots, n$, we consider the realization $\boldsymbol{R}_i$ of the variable $\boldsymbol{R}$, and we denote by $\boldsymbol{L}_{(\boldsymbol{R}_i)i}$ the subvector of $\boldsymbol{L}_i$ formed by the observed components and by $\boldsymbol{L}_{\left(\overline{\boldsymbol{R}_i}\right)i}$ the subvector of $\boldsymbol{L}_i$ formed by the non-observed ones. So, the observed data are $\left\{\boldsymbol{R}_i, \boldsymbol{L}_{(\boldsymbol{R}_i)i}\right\}_{i=1,\ldots,n}$.

As we already mentioned our goal is to derive an estimator for the survival function in each category of $\boldsymbol{X}$ when some of the values in the covariate $\boldsymbol{X}$ have not been observed. In order to do that we develop first, in Section 6.2, a grouped Kaplan–Meier estimator and its stratified version assuming that the values of $\boldsymbol{X}$ have been observed. We use this grouped estimator as a platform to build in Section 6.3 a Kaplan–Meier semiparametric estimator when missing data is present. At the end of the chapter we apply the proposed methodology to the HIV+PTB cohort.

6.2 Grouped Kaplan–Meier (GKM) estimator

6.2.1 Definition

In many situations the inferences are only needed at a finite number of calendar times (*e.g.,* years, months, weeks ...). If data were complete, the Kaplan–Meier estimator estimates the survival function at each uncensored time and, from this, the survival curve could be estimated at each time of interest. However, the corresponding survival function would be estimated at a number of times that increases with the sample size. In this section we develop a methodology to estimate the survival function at each time of interest when our sample has missing data. The advantage of the proposed methodology is that it will avoid the above mentioned dimensionality issue.

Suppose that the data are collected in a observation window $(0, T_{max}]$. Therefore, in order to deal with and to estimate a finite dimensional vector, in what follows we assume that we are interested in the inference of the survival function at a finite number of points $\tau_0 = 0 < \tau_1 < \tau_2 < \ldots < \tau_K < T_{max} = \tau_{K+1}$. We define the $K+1$ intervals $I_k = (\tau_{k-1}, \tau_k]$, $k = 1, \ldots, K+1$ that form a partition of the window $(0, T_{max}]$.

For each time of interest τ_k, $k = 1, \ldots, K$, we define the survival at time τ_k, $S_k = S(\tau_k) = P(T > \tau_k)$. We can rewrite the survival function in the usual product limit form

$$\begin{aligned} S_k = P(T > \tau_k) &= P(T > \tau_1)P(T > \tau_2|T > \tau_1) \cdot \ldots \cdot P(T > \tau_k|T > \tau_{k-1}) \\ &= \prod_{\ell=1}^{k} P(T > \tau_\ell|T > \tau_{\ell-1}), \end{aligned} \tag{6.1}$$

and each of these conditional probabilities as

$$\begin{aligned} P(T > \tau_\ell|T > \tau_{\ell-1}) &= \frac{P(T > \tau_\ell)}{P(T > \tau_{\ell-1})} = \frac{1 - F(\tau_{\ell-1}) + F(\tau_{\ell-1}) - F(\tau_\ell)}{1 - F(\tau_{\ell-1})} \\ &= 1 - \frac{F(\tau_\ell) - F(\tau_{\ell-1})}{1 - F(\tau_{\ell-1})} = 1 - q_\ell \end{aligned} \tag{6.2}$$

where q_ℓ is the probability that the event ϵ happens in the ℓ-th interval, I_ℓ, given that it has not happened before $\tau_{\ell-1}$.

For $k = 1, \ldots, K$, we denote by r_k the number of individuals at risk at the beginning of I_k, by e_k the number of events in I_k, by c_k the number of censored individuals in I_k and by $r_{K+1} = n - \sum_{k=1}^{K} (e_k + c_k)$ the number of individuals in the sample with observed survival time strictly greater than τ_K. Following the Kaplan–Meier's idea, we propose to estimate the conditional probabilities q_k by the naive estimator

$$\widehat{q_k} = \frac{e_k}{r_k} = \frac{e_k}{\sum_{\ell=k}^{K} (e_\ell + c_\ell) + r_{K+1}}. \tag{6.3}$$

We define the *Grouped Kaplan–Meier (GKM) estimator* as the vector of estimators $\widehat{\boldsymbol{S}} = \left(\widehat{S_1}, \ldots, \widehat{S_K}\right)'$, where for every $k = 1, \ldots, K$

$$\widehat{S_k} = \prod_{\ell=1}^{k} (1 - \widehat{q_\ell}). \tag{6.4}$$

In what follows, we introduce some vectorial notation that will enable us to characterize the Grouped Kaplan–Meier estimator as a $(2K+1)$-dimensional vector. For each $k = 1, \ldots, K$ we define $\widehat{\beta_{k1}} = e_k$ and $\widehat{\beta_{k0}} = c_k$, $\widehat{\beta_{K+1}} = r_{K+1}$ and

$$\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta_{11}}, \widehat{\beta_{10}}, \widehat{\beta_{21}}, \widehat{\beta_{20}}, \ldots, \widehat{\beta_{K1}}, \widehat{\beta_{K0}}, \widehat{\beta_{K+1}}\right)'$$

and it follows that the GKM estimator $\widehat{\boldsymbol{S}}$ can be uniquely determined through $\widehat{\boldsymbol{\beta}}$. Let $\mathcal{J}$ be the ordered set of $2K+1$ indexes, that is, $\mathcal{J} = \{11, 10, 21, 20, \ldots, K1, K0, K+1\}$, and rewrite the vector $\widehat{\boldsymbol{\beta}}$ as $\left(\widehat{\beta_j}\right)'_{j\in\mathcal{J}}$. If for each $k = 1, \ldots, K$ denote by $\mathcal{J}_k$ the subset of indexes $\{k1, k0, \ldots, K1, K0, K+1\}$, thus, expression (6.4) can be written as

$$\widehat{S_k} = \prod_{\ell=1}^{k} \left(1 - \frac{\widehat{\beta_{\ell 1}}}{\sum\limits_{j\in\mathcal{J}_\ell} \widehat{\beta_j}} \right). \tag{6.5}$$

6.2.2 Asymptotic behavior

Breslow and Crowley (1974) described the large sample properties of the life table and product limit estimators (Kaplan and Meier, 1958) under random censorship. For the estimation of the life table they considered that the withdrawal mechanism takes place just in the middle of the intervals I_k, in order to adjust for the fact that the r_k individuals are not at risk for the entire interval. This estimator is in general inconsistent and therefore they use the Kaplan–Meier estimator to derive the asymptotic properties.

Before to proceed to derive the asymptotic properties of the GKM estimator we make the following consideration.

Remark 6.2.1 .

a) The KM estimator is the limit of the GKM estimator of nested sequences when the norm of the partition, defined by $\rho = \max_{k=1,\ldots,K+1}(\tau_k - \tau_{k-1})$*, converges to 0.*

b) In practice, with discrete time data, the GKM estimator and the KM estimator yield the same estimates if we take the partition $\{\tau_k\}_{1,\ldots,K}$ *equal to the grid for* *the discrete values of Y. In fact, we are considering all the information in the same way that it was only observable in the right endpoint of each interval I_k.*

We define the events Ω_j, $j \in \mathcal{J}$:

$$\Omega_j = \begin{cases} Y \in I_k, \delta = 1 & \text{if } j = k1 \quad k = 1, \ldots, K \\ Y \in I_k, \delta = 0 & \text{if } j = k0 \quad k = 1, \ldots, K \\ Y > \tau_K & \text{if } j = K + 1 \end{cases} \tag{6.6}$$

According to these definitions the probabilities of the events Ω_j, $j \in \mathcal{J}$, are respectively:

$$P(\Omega_j) = \begin{cases} P(T \in I_k, C \geq T) = \int_{\tau_{k-1}}^{\tau_k} (1 - G) dF & j = k1 \quad k = 1, \ldots, K \\ P(C \in I_k, T \geq C) = \int_{\tau_{k-1}}^{\tau_k} (1 - F) dG & j = k0 \quad k = 1, \ldots, K \\ P(T > \tau_K, C > \tau_K) = (1 - F(\tau_K))(1 - G(\tau_K)) & j = K + 1 \end{cases}$$

The vector $\widehat{\boldsymbol{\beta}}$ is integrated by the counts of the events Ω_j, $j \in \mathcal{J}$, that is, $\widehat{\beta_{k1}} = e_k = \sum_{i=1}^n I(\Omega_{k1i})$, $\widehat{\beta_{k0}} = c_k = \sum_{i=1}^n I(\Omega_{k0i})$ and $\widehat{\beta_{K+1}} = r_{K+1} = \sum_{i=1}^n I(\Omega_{(K+1)i})$, and follows a multinomial distribution with vector of probabilities $\boldsymbol{p} = (P(\Omega_j))'_{j \in \mathcal{J}}$. Denote by $\boldsymbol{p^*}$ the true vector of probabilities and by $\boldsymbol{\beta^*} = \left(\beta_j^*\right)'_{j \in \mathcal{J}} = n \cdot \boldsymbol{p^*}$ the true vector of expected counts when the sample size is n. So, by computing the moment generating functions we can derive the following lemmas and theorems.

Lemma 6.2.1 *The standardized random vector* $\dfrac{\widehat{\boldsymbol{\beta}} - n\boldsymbol{p^*}}{\sqrt{n}}$ *converges in distribution to a multivariate normal* $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ *with* $\boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{p^*}) - \boldsymbol{p^*} \cdot \boldsymbol{p^{*\prime}}$*, or, in other words,*

$$\sqrt{n} \left(\frac{\widehat{\boldsymbol{\beta}}}{n} - \boldsymbol{p^*} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}).$$

Proof: See, for instance, page 470 in Bishop *et al.* (1975). □

In order to derive the asymptotic distribution of the GKM estimator we will use twice the multivariate version of the δ-method.

Theorem 6.2.1 Multivariate δ-method *(Bishop et al., 1975) Let $\boldsymbol{f}$ be a R-valued function defined in the T-dimensional parameter space $\boldsymbol{\Theta}$. Let $\boldsymbol{\theta}, \boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}$ be vectors of $\boldsymbol{\Theta}$. If*

a) $\boldsymbol{f}$ is differentiable in one open subset of $\boldsymbol{\Theta}$ containing $\boldsymbol{\theta}^*$ and

b) $\widehat{\boldsymbol{\theta}}$ has an asymptotic normal distribution in the sense that

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Delta}),$$

then the asymptotic distribution of $\boldsymbol{f}\left(\widehat{\boldsymbol{\theta}}\right)$ is given by

$$\sqrt{n}\left(\boldsymbol{f}\left(\widehat{\boldsymbol{\theta}}\right) - \boldsymbol{f}\left(\boldsymbol{\theta}^*\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}}\right)\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \boldsymbol{\Delta} \left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}}\right)'\Bigg|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right).$$

In other words, for large n, if $\widehat{\boldsymbol{\theta}}$ has an approximate $\mathcal{N}(\boldsymbol{\theta}^*, n^{-1}\boldsymbol{\Delta})$ distribution, then $\boldsymbol{f}\left(\widehat{\boldsymbol{\theta}}\right)$ has an approximate $\mathcal{N}\left(\boldsymbol{f}(\boldsymbol{\theta}^*), n^{-1}\left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}}\right)\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*} \boldsymbol{\Delta} \left(\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{\theta}}\right)'\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}\right)$ distribution.

Define the Kaplan–Meier map, $\mathcal{KM}$, as the function $\mathcal{KM} : [0,\infty)^{2K+1} \longrightarrow [0,1]^K$ that assigns the GKM estimator $\widehat{\boldsymbol{S}}$ where each component is defined by (6.5) to each realization of the vector $\widehat{\boldsymbol{\beta}}$ (*i.e.,* $\widehat{\boldsymbol{S}} = \mathcal{KM}(\widehat{\boldsymbol{\beta}})$). Note that the $\mathcal{KM}$ map is homogeneous, that is $\mathcal{KM}(\widehat{\boldsymbol{\beta}}/n) = \mathcal{KM}(\widehat{\boldsymbol{\beta}})$. Define as well the log-Kaplan–Meier map, $\ell\mathcal{KM}$, by

$$\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_j\right)_{j\in\mathcal{J}} \xrightarrow{\ell\mathcal{KM}} \log \widehat{\boldsymbol{S}} = \left(\sum_{\ell\leq k} \log\left(1 - \frac{\widehat{\beta_{\ell 1}}}{\sum_{j\in\mathcal{J}_\ell}\widehat{\beta}_j}\right)\right)_{\{k=1,\ldots,K\}}. \tag{6.7}$$

Note that $1 - \widehat{\beta_{\ell 1}}/\sum_{j\in\mathcal{J}_\ell}\widehat{\beta}_j = \sum_{j\in\mathcal{J}_\ell\setminus\{\ell 1\}}\widehat{\beta}_j/\sum_{j\in\mathcal{J}_\ell}\widehat{\beta}_j$.

In order to prove that the Kaplan–Meier map satisfies the conditions of 6.2.1, it is sufficient to verify it for the log-function $\ell\mathcal{KM}$.

Lemma 6.2.2 *The function $\ell\mathcal{KM} : [0,\infty)^{2K+1} \longrightarrow (-\infty, 0]^K$ defined by (6.7) is differentiable in all the points $\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta_{11}}, \widehat{\beta_{10}}, \widehat{\beta_{21}}, \widehat{\beta_{20}}, \ldots, \widehat{\beta_{K1}}, \widehat{\beta_{K0}}, \widehat{\beta_{K+1}}\right)' \in [0,\infty)^{2K+1}$ such as $\widehat{\beta_{K+1}} > 0$*

Proof: Indeed, due to $\widehat{\beta_{K+1}} > 0$ all the quantities $\sum_{j\in\mathcal{J}_\ell}\widehat{\beta}_j$ are positive and the ratios $\frac{\widehat{\beta_{\ell 1}}}{\sum_{j\in\mathcal{J}_\ell}\widehat{\beta}_j}$ are strictly less than 1. So, each one of the k-th components of $\log\widehat{\boldsymbol{S}}$ is differentiable with continuous partial derivatives in a neighbourhood of $\widehat{\boldsymbol{\beta}}$. □

Note that the condition $\widehat{\beta_{K+1}} > 0$ means that $S(\tau_K) > 0$ (*i.e.*, there are alive individuals in the last estimation point), or, analogously, that the distribution of F is not concentrated in the interval $(0, \tau_K]$. In practice, if $r_{K+1} = 0$ we can shift K to $K-1$.

Let us calculate the differential matrix $\partial\ell\mathcal{KM}/\partial\widehat{\boldsymbol{\beta}}$:

$$\frac{\partial\ell\mathcal{KM}_k}{\partial\widehat{\beta}_j} = \sum_{\ell\le k}\frac{\partial\log\left(1-\dfrac{\widehat{\beta_{\ell 1}}}{\sum\limits_{j\in\mathcal{J}_\ell}\widehat{\beta}_j}\right)}{\partial\widehat{\beta}_j} = \sum_{\ell\le k}\frac{\partial}{\partial\widehat{\beta}_j}\left\{\log\left(\sum_{j\in\mathcal{J}_\ell\setminus\{\ell 1\}}\widehat{\beta}_j\right) - \log\left(\sum_{j\in\mathcal{J}_\ell}\widehat{\beta}_j\right)\right\}$$

and, after straightforward calculations, we obtain

$$\frac{\partial\ell\mathcal{KM}_k}{\partial\widehat{\beta}_j} = \sum_{\ell\le k}\begin{cases}\dfrac{1}{\sum\limits_{j\in\mathcal{J}_\ell\setminus\{\ell 1\}}\widehat{\beta}_j} - \dfrac{1}{\sum\limits_{j\in\mathcal{J}_\ell}\widehat{\beta}_j} & j\in\mathcal{J}_\ell\setminus\{\ell 1\}\\ \dfrac{-1}{\sum\limits_{j\in\mathcal{J}_\ell}\widehat{\beta}_j} & j=\ell 1\\ 0 & j\notin\mathcal{J}_\ell\end{cases} \tag{6.8}$$

If we retake the notation e_k and r_k for the number of events and individuals at risk in the interval I_k, and we define $g_k = \dfrac{e_k}{r_k(r_k-e_k)}$, expression (6.8) can be rewritten in matricial form as

$$\frac{\partial\ell\mathcal{KM}}{\partial\widehat{\boldsymbol{\beta}}} = \begin{pmatrix}1 & 0 & 0 & \cdots & 0\\ 1 & 1 & 0 & \cdots & 0\\ 1 & 1 & 1 & \cdots & 0\\ \cdots & \cdots & \cdots & \cdots & \cdots\\ 1 & 1 & 1 & \cdots & 1\end{pmatrix}\left(\begin{array}{cc|cc|cc|c|cc|c}\frac{-1}{r_1} & g_1 & g_1 & g_1 & g_1 & g_1 & \cdots & g_1 & g_1 & g_1\\ 0 & 0 & \frac{-1}{r_2} & g_2 & g_2 & g_2 & \cdots & g_2 & g_2 & g_2\\ 0 & 0 & 0 & 0 & \frac{-1}{r_3} & g_3 & \cdots & g_3 & g_3 & g_3\\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \frac{-1}{r_K} & g_K & g_K\end{array}\right). \tag{6.9}$$

Let us denote by $\mathcal{T}$ and $\mathcal{G}$ the respective $K \times K$ and $K \times (2K+1)$ matrices in (6.9). So, after composing with the exponential function we have the following theorem that summarizes the asymptotic properties of the *Grouped Kaplan–Meier* estimator.

Theorem 6.2.2 *The Grouped Kaplan–Meier estimator defined in (6.5) asymptotically follows a* $\mathcal{N}\left(\mathcal{KM}(\boldsymbol{p}^*), n^{-1}(\mathcal{STG})\boldsymbol{\Sigma}(\mathcal{STG})'\right)$ *distribution being* $\mathcal{S}$ *the* $K \times K$ *diagonal matrix* $\operatorname{diag}(\mathcal{KM}(\boldsymbol{p}^*))$*,* $\mathcal{T}$ *and* $\mathcal{G}$ *as in (6.9) and evaluated in* $\boldsymbol{p}^*$ *and* $\boldsymbol{\Sigma} = \operatorname{diag}(\boldsymbol{p}^*) - \boldsymbol{p}^* \cdot \boldsymbol{p}^{*\prime}$ *as in Lemma 6.2.1.*

Proof: Due to the homogeneity of the $\mathcal{KM}$ map, if we apply the δ-method to the function

$$(x_1, \ldots, x_K)' \rightarrow (\exp(x_1), \ldots, \exp(x_K))'$$

acting on the estimator $\ell\mathcal{KM}(\widehat{\boldsymbol{\beta}})$, we obtain

$$\sqrt{n}\left(\mathcal{KM}\left(\widehat{\boldsymbol{\beta}}\right) - \mathcal{KM}(\boldsymbol{p}^*)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, (\mathcal{STG})\boldsymbol{\Sigma}(\mathcal{STG})'\right)$$

that yields the proposed result. □

6.2.3 Asymptotic bias

It is important to note that the Grouped Kaplan–Meier estimator defined in (6.4), $\widehat{S_k}$, has a positive bias with respect to the true survival $S(\tau_k)$. In fact, due to the "shift to the right", the conditional probabilities $\widehat{q_k}$ in (6.3) underestimate the true values q_k in (6.2). Consequently, $\widehat{S_k}$ overestimates S_k, for $k = 1, \ldots, K$. However, the upper bound for the asymptotic relative bias at τ_k decreases when we add cut points previous to τ_k to the former partition. The following two lemmas summarize these facts.

Lemma 6.2.3 *Let* F *and* G *be the cumulative distribution functions of* T *and* C*, respectively. Then, asymptotically, we have that*

a) the inequality

$$\frac{1 - G(\tau_k)}{1 - G(\tau_{k-1})} q_k \leq \widehat{q_k} \leq q_k \tag{6.10}$$

holds a.s., for every $k = 1, \ldots, K$*,* *b) the ratio* $\dfrac{G(\tau_k)-G(\tau_{k-1})}{1-G(\tau_{k-1})}$ *is an upper bound for the relative bias in the approximation of* q_k *by* $\widehat{q_k}$*, and*

c) for every $k=1,\ldots,K$*,*

$$\frac{\widehat{S_k}-S(\tau_k)}{S(\tau_k)} \leq \sum_{\ell=1}^{k}\Delta_\ell^F\nabla_\ell^G + \sum_{\eta=2}^{k}\sum_{\ell_1<\ell_2<\ldots<\ell_\eta\leq k}\prod_{i=1}^{\eta}\Delta_{\ell_i}^F\nabla_{\ell_i}^G, \tag{6.11}$$

where $\Delta_\ell^F = \dfrac{F(\tau_\ell)-F(\tau_{\ell-1})}{1-F(\tau_\ell)}$ *and* $\nabla_\ell^G = \dfrac{G(\tau_\ell)-G(\tau_{\ell-1})}{1-G(\tau_{\ell-1})}$ *for* $\ell=1,\ldots,K$*.*

Proof: According to the definitions

a)

$$\widehat{q_k} = \frac{\widehat{\beta_{k1}}}{\sum_{j\in\mathcal{J}_k}\widehat{\beta_j}} = \frac{\widehat{\beta_{k1}}/n}{\sum_{j\in\mathcal{J}_k}\widehat{\beta_j}/n} \text{ converges, almost surely, when } n\to\infty \text{ to}$$

$$\frac{\int_{\tau_{k-1}}^{\tau_k}(1-G)dF}{\sum_{\ell=k}^{K}\left\{\int_{\tau_{\ell-1}}^{\tau_\ell}(1-G)dF+\int_{\tau_{\ell-1}}^{\tau_\ell}(1-F)dG\right\}+(1-F(\tau_K))(1-G(\tau_K))}$$

$$= \frac{\int_{\tau_{k-1}}^{\tau_k}(1-G)dF}{\int_{\tau_{k-1}}^{\tau_K}\{(1-G)dF+(1-F)dG\}+(1-F(\tau_K))(1-G(\tau_K))}$$

$$= \frac{\int_{\tau_{k-1}}^{\tau_k}(1-G)dF}{-((1-F)(1-G))|_{\tau_{k-1}}^{\tau_K}+(1-F(\tau_K))(1-G(\tau_K))}$$

$$= \frac{\int_{\tau_{k-1}}^{\tau_k}(1-G)dF}{(1-F(\tau_{k-1}))(1-G(\tau_{k-1}))} = \frac{\int_{\tau_{k-1}}^{\tau_k}\dfrac{1-G(t)}{1-G(\tau_{k-1})}dF(t)}{1-F(\tau_{k-1})}. \tag{6.12}$$

Since G is increasing, $1-G(t)<1-G(\tau_{k-1}), \forall t\in I_k=(\tau_{k-1},\tau_k]$ and therefore asymptotically

$$\widehat{q_k} \leq \frac{\int_{\tau_{k-1}}^{\tau_k}dF(t)}{1-F(\tau_{k-1})} = \frac{F(\tau_k)-F(\tau_{k-1})}{1-F(\tau_{k-1})} = q_k.$$

On the other hand, $1-G(\tau_k) < 1-G(t), \forall t \in (\tau_{k-1}, \tau_k)$ and, consequently the left inequality in (6.10) holds.

b) From part a)

$$\frac{|\widehat{q_k}-q_k|}{q_k} \leq \frac{q_k - \dfrac{1-G(\tau_k)}{1-G(\tau_{k-1})}q_k}{q_k} = 1 - \frac{1-G(\tau_k)}{1-G(\tau_{k-1})} = \frac{G(\tau_k)-G(\tau_{k-1})}{1-G(\tau_{k-1})}.$$

c) If we denote by $\boldsymbol{q} = (q_1, \ldots, q_K)'$ and $\widehat{\boldsymbol{q}} = (\widehat{q_1}, \ldots, \widehat{q_K})'$ the vector of probabilities and its estimators, and f_k the function $f_k(x_1, \ldots, x_K) = \prod_{\ell \leq k}(1-x_\ell)$ we can write

$$\widehat{S_k} - S(\tau_k) = f_k(\widehat{\boldsymbol{q}}) - f_k(\boldsymbol{q}).$$

After observing that

$$\begin{aligned} \frac{\partial f_k}{\partial x_m} &= \begin{cases} \dfrac{-f_k}{1-x_m} = \displaystyle\prod_{\ell \leq k, \ell \neq m}(1-x_\ell) & \text{if } m \leq k \\ 0 & \text{if } m > k \end{cases} \quad \text{and} \\ \frac{\partial^2 f_k}{\partial x_m^2} &= 0 \ \forall m, \end{aligned}$$

we expand f_k by Taylor in a neighbourhood of $\boldsymbol{q}$ and we obtain

$$\begin{aligned} f_k(\widehat{\boldsymbol{q}}) &= f_k(\boldsymbol{q}) + \sum_{\ell=1}^{k} \left.\frac{-f_k}{1-x_\ell}\right|_{\boldsymbol{q}} (\widehat{q_\ell} - q_\ell) + \\ &+ \sum_{\ell_1<\ell_2\leq k} \left.\frac{f_k}{(1-x_{\ell_1})(1-x_{\ell_2})}\right|_{\boldsymbol{q}} (\widehat{q_{\ell_1}} - q_{\ell_1})(\widehat{q_{\ell_2}} - q_{\ell_2}) + \\ &+ \sum_{\ell_1<\ell_2<\ell_3\leq k} \left.\frac{-f_k}{(1-x_{\ell_1})(1-x_{\ell_2})(1-x_{\ell_3})}\right|_{\boldsymbol{q}} (\widehat{q_{\ell_1}} - q_{\ell_1})(\widehat{q_{\ell_2}} - q_{\ell_2})(\widehat{q_{\ell_3}} - q_{\ell_3}) + \\ &+ \ldots + \left.\frac{(-1)^k f_k}{\prod_{\ell\leq k}(1-x_\ell)}\right|_{\boldsymbol{q}} \prod_{\ell \leq k}(\widehat{q_\ell} - q_\ell) \end{aligned}$$

and hence

$$\frac{f_k(\widehat{\boldsymbol{q}}) - f_k(\boldsymbol{q})}{f_k(\boldsymbol{q})} = \sum_{\ell=1}^{k} \frac{q_\ell - \widehat{q_\ell}}{1-q_\ell} + \sum_{\ell_1<\ell_2\leq k} \frac{q_{\ell_1} - \widehat{q_{\ell_1}}}{1-q_{\ell_1}} \frac{q_{\ell_2} - \widehat{q_{\ell_2}}}{1-q_{\ell_2}} + \ldots + \prod_{\ell\leq k} \frac{q_\ell - \widehat{q_\ell}}{1-q_\ell}. \tag{6.13}$$

But using expression (6.12) for $\widehat{q}_\ell$ we derive that, asymptotically,

$$\begin{aligned}\frac{q_\ell - \widehat{q}_\ell}{1-q_\ell} &= \frac{\dfrac{F(\tau_\ell)-F(\tau_{\ell-1})}{1-F(\tau_{\ell-1})} - \dfrac{\int_{\tau_{\ell-1}}^{\tau_\ell} \frac{1-G(t)}{1-G(\tau_{\ell-1})} dF(t)}{1-F(\tau_{\ell-1})}}{1-\dfrac{F(\tau_\ell)-F(\tau_{\ell-1})}{1-F(\tau_{\ell-1})}} \\ &= \frac{\{F(\tau_\ell)-F(\tau_{\ell-1})\} - \int_{\tau_{\ell-1}}^{\tau_\ell} \frac{1-G(t)}{1-G(\tau_{\ell-1})} dF(t)}{1-F(\tau_\ell)},\end{aligned}$$

and since G is increasing with t we have

$$\begin{aligned}\frac{q_\ell - \widehat{q}_\ell}{1-q_\ell} &\leq \frac{\{F(\tau_\ell)-F(\tau_{\ell-1})\} - \frac{1-G(\tau_\ell)}{1-G(\tau_{\ell-1})}\{F(\tau_\ell)-F(\tau_{\ell-1})\}}{1-F(\tau_\ell)} \\ &= \left(\frac{F(\tau_\ell)-F(\tau_{\ell-1})}{1-F(\tau_\ell)}\right)\left(\frac{G(\tau_\ell)-G(\tau_{\ell-1})}{1-G(\tau_{\ell-1})}\right) \\ &= \Delta_\ell^F \nabla_\ell^G.\end{aligned}$$

From (6.13) and the previous inequality we derive (6.11). □

Note that the upper bound for the relative bias in the estimation of the conditional probabilities q_k only depends on the censoring distribution. For absolute continuous distributions, a) the bound increases if we move the endpoint τ_k to the right and, b) the bound converges to zero if τ_{k-1} approaches τ_k.

On the other hand, the cumulative sum of the expression (6.11) could increase if we add points before a fixed cut point τ_k. The next lemma studies this aspect and proves that the asymptotic relative bias in the estimation of $S(\tau_k)$ decreases when the partition gets thinner.

Lemma 6.2.4 *Let F and G be the cumulative distribution functions of T and C, respectively, and $H(x,y)$ $(0 \leq x \leq y)$ defined by*

$$H(x,y) = \left(\frac{F(y)-F(x)}{1-F(y)}\right)\left(\frac{G(y)-G(x)}{1-G(x)}\right).$$

Then, for every $0 \leq a < c < b$ we have the following inequality

$$H(a,b) > H(a,c) + H(c,b).$$

Proof: We start proving that this "split" property holds for the F-term of the function H. Indeed, since $F(b) > F(c)$

$$\frac{F(b)-F(a)}{1-F(b)} = \frac{F(c)-F(a)}{1-F(b)} + \frac{F(b)-F(c)}{1-F(b)} > \frac{F(c)-F(a)}{1-F(c)} + \frac{F(b)-F(c)}{1-F(b)}.$$

On the other hand, $G(b) > G(c)$ and the function $\dfrac{G(b)-G(x)}{1-G(x)} = 1 - \dfrac{1-G(b)}{1-G(x)}$ is strictly decreasing with x, therefore

$$\begin{aligned}
H(a,b) &> \left(\frac{F(c)-F(a)}{1-F(c)} + \frac{F(b)-F(c)}{1-F(b)}\right)\frac{G(b)-G(a)}{1-G(a)} \\
&= \left(\frac{F(c)-F(a)}{1-F(c)}\right)\left(\frac{G(b)-G(a)}{1-G(a)}\right) + \left(\frac{F(b)-F(c)}{1-F(b)}\right)\left(\frac{G(b)-G(a)}{1-G(a)}\right) \\
&> \left(\frac{F(c)-F(a)}{1-F(c)}\right)\left(\frac{G(c)-G(a)}{1-G(a)}\right) + \left(\frac{F(b)-F(c)}{1-F(b)}\right)\left(\frac{G(b)-G(c)}{1-G(c)}\right) \\
&= H(a,c) + H(c,b)
\end{aligned}$$

□

6.2.4 Stratified Grouped Kaplan–Meier estimator $\widehat{\boldsymbol{S}_{\boldsymbol{x}}}$

For each category $\boldsymbol{x} = (x_1, \dots, x_p)'$ of the covariates vector $\boldsymbol{X}$ if the above subsections are rewritten conditioning on $\boldsymbol{X} = \boldsymbol{x}$, we obtain the stratified version of the Grouped Kaplan–Meier estimator for the category $\boldsymbol{X} = \boldsymbol{x}$. Without loss of generality we will assume that $0 < P(\boldsymbol{X} = \boldsymbol{x}) < 1$. Our vector of interest will be $\boldsymbol{S}_{\boldsymbol{x}} = (S_{\boldsymbol{x}1}, \dots, S_{\boldsymbol{x}K})'$ where $S_{\boldsymbol{x}k} = P(T > \tau_k | \boldsymbol{X} = \boldsymbol{x})$ for each time τ_k, $k = 1, \dots, K$.

In a similar way, we can derive equivalent expressions to (6.1) and (6.2), with

$$\begin{aligned}
P(T > \tau_\ell | T > \tau_{\ell-1}, \boldsymbol{X} = \boldsymbol{x}) &= \frac{P(T > \tau_\ell | \boldsymbol{X} = \boldsymbol{x})}{P(T > \tau_{\ell-1} | \boldsymbol{X} = \boldsymbol{x})} \\
&= \frac{1 - F_{\boldsymbol{x}}(\tau_{\ell-1}) + F_{\boldsymbol{x}}(\tau_{\ell-1}) - F_{\boldsymbol{x}}(\tau_\ell)}{1 - F_{\boldsymbol{x}}(\tau_{\ell-1})} \\
&= 1 - \frac{F_{\boldsymbol{x}}(\tau_\ell) - F_{\boldsymbol{x}}(\tau_{\ell-1})}{1 - F_{\boldsymbol{x}}(\tau_{\ell-1})} = 1 - q_{\boldsymbol{x}\ell} \qquad (6.14)
\end{aligned}$$

where $F_{\boldsymbol{x}}$ is the conditional distribution function of T given $\boldsymbol{X} = \boldsymbol{x}$.

Analogously, if in the category $\boldsymbol{X} = \boldsymbol{x}$ we denote by $r_{\boldsymbol{x}k}$ the number of individuals at risk at the beginning of I_k, by $e_{\boldsymbol{x}k}$ the number of events in I_k, by

$c_{\boldsymbol{x}k}$ the number of censored individuals in I_k for $k = 1, \dots, K$ and by $r_{\boldsymbol{x}K+1}$ the number of individuals with observed survival time strictly greater than τ_K, we can estimate the conditional probabilities $q_{\boldsymbol{x}k}$ by the ratio $\widehat{q_{\boldsymbol{x}k}} = e_{\boldsymbol{x}k}/r_{\boldsymbol{x}k}$ and compute the *Stratified Grouped Kaplan–Meier estimator* as the vector of estimators $\widehat{\boldsymbol{S_x}} = \left(\widehat{S_{\boldsymbol{x}1}}, \dots, \widehat{S_{\boldsymbol{x}K}}\right)'$, where for every $k = 1, \dots, K$

$$\widehat{S_{\boldsymbol{x}k}} = \prod_{\ell=1}^{k} (1 - \widehat{q_{\boldsymbol{x}\ell}}). \tag{6.15}$$

Also the conditional version of the equation (6.5) can be obtained after defining, for each $k = 1, \dots, K$, $\widehat{\beta_{\boldsymbol{x},k1}} = e_{\boldsymbol{x}k}$, $\widehat{\beta_{\boldsymbol{x},k0}} = c_{\boldsymbol{x}k}$, $\widehat{\beta_{\boldsymbol{x},K+1}} = r_{\boldsymbol{x}K+1}$, $\widehat{\boldsymbol{\beta_x}} = \left(\widehat{\beta_{\boldsymbol{x},j}}\right)'_{j\in\mathcal{J}}$ as

$$\widehat{\boldsymbol{\beta_x}} = (\widehat{\beta_{\boldsymbol{x},11}}, \widehat{\beta_{\boldsymbol{x},10}}, \widehat{\beta_{\boldsymbol{x},21}}, \widehat{\beta_{\boldsymbol{x},20}}, \dots, \widehat{\beta_{\boldsymbol{x},K1}}, \widehat{\beta_{\boldsymbol{x},K0}}, \widehat{\beta_{\boldsymbol{x},K+1}})'$$

and, consequently,

$$\widehat{S_{\boldsymbol{x}k}} = \prod_{\ell=1}^{k} \left(1 - \frac{\widehat{\beta_{\boldsymbol{x},\ell 1}}}{\sum_{j\in\mathcal{J}_\ell} \widehat{\beta_{\boldsymbol{x},j}}} \right). \tag{6.16}$$

In order to derive the asymptotic behavior of the stratified Grouped Kaplan–Meier estimator, we denote by $\boldsymbol{p_x} = (P(\Omega_j | \boldsymbol{X} = \boldsymbol{x}))'_{j\in\mathcal{J}}$ the vector of conditional probabilities for the events Ω_j defined in (6.6), by $\boldsymbol{p}^*_{\boldsymbol{x}}$ the true values vector for $\boldsymbol{p_x}$ and by $n_{\boldsymbol{x}} = n \cdot P(\boldsymbol{X} = \boldsymbol{x})$ the expected number of individuals with $\boldsymbol{X} = \boldsymbol{x}$ when the sample size is n. Then, we can obtain the stratified version of Lemma 6.2.1 and Theorem 6.2.2.

Lemma 6.2.5 *The standardized random vector* $\dfrac{\widehat{\boldsymbol{\beta_x}} - n_{\boldsymbol{x}} \boldsymbol{p}^*_{\boldsymbol{x}}}{\sqrt{n_{\boldsymbol{x}}}}$ *converges in distribution to a multivariate normal* $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_x})$ *with* $\boldsymbol{\Sigma_x} = \mathrm{diag}(\boldsymbol{p}^*_{\boldsymbol{x}}) - \boldsymbol{p}^*_{\boldsymbol{x}} \cdot \boldsymbol{p}^{*\prime}_{\boldsymbol{x}}$, *or, in other words,*

$$\sqrt{n_{\boldsymbol{x}}} \left(\frac{\widehat{\boldsymbol{\beta_x}}}{n_{\boldsymbol{x}}} - \boldsymbol{p}^*_{\boldsymbol{x}} \right) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma_x}).$$

Proof: It is enough to restrict us to the conditional probability space given $\boldsymbol{X} = \boldsymbol{x}$ and to apply Lemma 6.2.1 after replacing $\widehat{\boldsymbol{\beta}}$ by $\widehat{\boldsymbol{\beta_x}}$, n by $n_{\boldsymbol{x}}$ and $\boldsymbol{p}^*$ by $\boldsymbol{p}^*_{\boldsymbol{x}}$. □

Theorem 6.2.3 *The stratified Grouped Kaplan–Meier estimator defined in (6.16) asymptotically follows a* $\mathcal{N}\left(\mathcal{KM}(\boldsymbol{p}^*_{\boldsymbol{x}}), n^{-1}_{\boldsymbol{x}}(\mathcal{S}_{\boldsymbol{x}}\mathcal{T}\mathcal{G}_{\boldsymbol{x}})\boldsymbol{\Sigma}_{\boldsymbol{x}}(\mathcal{S}_{\boldsymbol{x}}\mathcal{T}\mathcal{G}_{\boldsymbol{x}})'\right)$ *distribution being* $\mathcal{S}_{\boldsymbol{x}}$ *the* $K \times K$ *diagonal matrix* $\mathrm{diag}(\mathcal{KM}(\boldsymbol{p}^*_{\boldsymbol{x}}))$, $\mathcal{T}$ *and* $\mathcal{G}_{\boldsymbol{x}}$ *as in (6.9) and evaluated in* $\boldsymbol{p}^*_{\boldsymbol{x}}$ *and* $\boldsymbol{\Sigma}_{\boldsymbol{x}} = \mathrm{diag}(\boldsymbol{p}^*_{\boldsymbol{x}}) - \boldsymbol{p}^*_{\boldsymbol{x}} \cdot \boldsymbol{p}^{*\prime}_{\boldsymbol{x}}$ *as in Lemma 6.2.5.*

Proof: Analogously to the proof of Theorem 6.2.2, the result follows from the application of the δ-method to the $\mathcal{KM}$ map acting on the distributions of the above Lemma 6.2.5. □

Concerning the asymptotic bias of the stratified Grouped Kaplan–Meier estimator, previous Subsection 6.2.3 holds by replacing cumulative distributions F and G by their respective conditional on $\boldsymbol{X} = \boldsymbol{x}$ cumulative distributions, $F_{\boldsymbol{x}}$ and $G_{\boldsymbol{x}}$.

6.3 Estimated (stratified) Grouped Kaplan–Meier (EGKM) estimator $\widetilde{\boldsymbol{S}_{\boldsymbol{x}}}$

When the vector $\boldsymbol{X}$ is not totally observed, we cannot obtain the quantities $\widehat{\beta_{\boldsymbol{x},j}}$, $j \in \mathcal{J}$, directly from the sample. We propose to estimate $\boldsymbol{p}^*_{\boldsymbol{x}}$, using all the available data in the sample and the non-response mechanism, by means of the semiparametric estimates $\widetilde{\widetilde{p_j}}$, $j \in \mathcal{J}$, to be defined below. From these, we define the *Estimated Grouped Kaplan–Meier* (EGKM) estimator as

$$\widetilde{S_{\boldsymbol{x}k}} = \prod_{\ell=1}^{k} \left(1 - \frac{\widetilde{\widetilde{p_{\boldsymbol{x},\ell 1}}}}{\sum\limits_{j\in\mathcal{J}_\ell} \widetilde{\widetilde{p_{\boldsymbol{x},j}}}} \right). \tag{6.17}$$

Although these values could be estimated independently, we choose to estimate them jointly so that we can derive an estimate for the variance of the new estimator $\widetilde{\boldsymbol{S}_{\boldsymbol{x}}} = \left(\widetilde{S_{\boldsymbol{x}1}}, \ldots, \widetilde{S_{\boldsymbol{x}K}}\right)'$. Taking into account that the category $\boldsymbol{X} = \boldsymbol{x}$ is fixed through all the section, we will omit, unless it would be necessary, the subindex $\boldsymbol{x}$ in all the estimators and parameters.

6.3.1 An introductory example

To motivate the methodology that we will develop in this section we start with an illustrative example based on the fictitious data in Table 6.1. The data correspond to the observed survival time (Y, δ) for $n = 10$ individuals for whom a binary covariate X could be also recorded. Table 6.1 presents the data for these 10 individuals ordered by the observed value of Y. We can observe that for 6 individuals the covariate X is available, while is missing for the other 4. Let the variable R be the binary response indicator to the covariate X. Let τ_1 and τ_2 be two times of interest such that $t_7 \leq \tau_1 < t_8$ and $t_{10} \leq \tau_2$. Suppose that the partition $\{\tau_k\}_{k=1,\ldots,K}$ is $\{\tau_1, \tau_2\}$.

Y	δ	X	R
t_1	1	0	1
t_2	1	1	1
t_3	1	1	1
t_4	0	0	1
t_5	0	1	1
t_6	1	NA	0
t_7	0	NA	0
t_8	1	0	1
t_9	1	NA	0
t_{10}	0	NA	0

Table 6.1: *Data example to illustrate the Estimated Grouped Kaplan–Meier estimator.* $n = 10$, $\{\tau_1, \tau_2\}$ *such that* $t_i \leq \tau_1$ *for* $i = 1, \ldots, 7$ *and* $\tau_1 < t_i \leq \tau_2$ *for* $i = 8, 9, 10$

Complete data from Table 6.1 can be equivalently summarized as in Table 6.2. Note that to build the life table for the stratified Grouped Kaplan–Meier estimator at times τ_1 and τ_2 we only use 6 individuals (3 belonging to $X = 0$ and 3 to $X = 1$).

On the other hand if, for only pedagogical reasons, we assume that the non-

	$X=0$				$X=1$			
Y	r	e	c	S_{cc}	r	e	c	S_{cc}
τ_1	3	1	1	0.667	3	2	1	0.333
τ_2	1	1	0	0	0	0	0	0.333
Total		2	1			2	1	
		$n_0=3$				$n_1=3$		
				$n_{ef}=6$				

Table 6.2: *Complete case life table and stratified Grouped Kaplan–Meier estimator for categories $X=0$ and $X=1$ for the data in Table 6.1. $n_x, x=0,1$, number of individuals belonging to the category $X=x$. n_{ef}, effective sample size*

response mechanism is MCAR or MAR, we can apply the following strategy. If missing data are MAR, *i.e.,* $P(R=1|Y,\delta,X)=P(R=1|Y,\delta)$, then

$$P(X,R|Y,\delta)=P(X|Y,\delta)\cdot P(R|Y,\delta,X)=P(X|Y,\delta)\cdot P(R|Y,\delta).$$

So, X and R are conditional independent given (Y,δ) or, in other words,

$$P(X=x|Y,\delta,R=1)=P(X=x|Y,\delta)=P(X=x|Y,\delta,R=0),$$

that is, the fully observed subsample is a good representation of the partially observed subsample.

Using this idea we can reproduce the empirical distribution of the fully observed data on the partially observed subsample. This allows us to estimate, for each $j\in\mathcal{J}$, the values of the expected counts of events or censored individuals in $X=x$ given $I(\Omega_j)$. For example, when $X=0$ if we are interested in the number of events in $I_1=(0,\tau_1]$, *i.e.,* $j=11$, looking at Table 6.1 we observe only the first individual with $X=0$, $\delta=1$ and $Y\le\tau_1$. But, we also observe that the empirical distribution of $X=0$ given $I(\Omega_{11})$ is 1/3. So, 1/3 of the individuals for those $I(\Omega_{11})=1$ and X is missing (in our data example, the sixth individual) can be distributed to the category $X=0$. Then, the estimated number of events in I_1 with $X=0$ will be $1+1/3$ as it is shown in Table 6.3. Analogously, the estimated number of events in I_1 with $X=1$ will be $2+2/3$.

	$X=0$				$X=1$			
Y	r	e	c	$\tilde{S}$	r	e	c	$\tilde{S}$
τ_1	$\frac{29}{6}$	$1+\frac{1}{3}$	$1+\frac{1}{2}$	0.724	$\frac{25}{6}$	$2+\frac{2}{3}$	$1+\frac{1}{2}$	0.360
τ_2	2	$1+1$	$0+0$	0	0	$0+0$	$0+0$	0.360
Total		$\frac{10}{3}$	$\frac{3}{2}$			$\frac{8}{3}$	$\frac{3}{2}$	
	$n_0=\frac{29}{6}$				$n_1=\frac{25}{6}$			
	$n_{ef}=9$							

Table 6.3: *Estimated life table and stratified Grouped Kaplan–Meier estimator for categories $X=0$ and $X=1$ for the data in Table 6.1, under the MCAR and MAR hypotheses. $n_x, x=0,1$, estimated number of individuals belonging to the category $X=x$. n_{ef}, effective sample size*

We can observe that the effective sample size has increased to 9 and that the estimated number of individuals in $X=0$ is slightly higher than in $X=1$. This is a consequence of the small size of the sample. As illustration note that there is no individuals in the sample with $Y \in I_2$, $\delta=1$ and $X=1$. So, the ninth individual is entirely assigned to $X=0$. It is also important to observe that the last individual cannot be distributed in the categories in X because we do not have information about the covariate in I_2 for censored individuals. These two problems disappear when the sample size is large enough.

In both categories, survival estimates are also different from those in the complete analysis due to the gain in the sample size (*e.g.,* at time τ_1, 0.724 vs 0.667 in $X=0$ and 0.360 vs 0.333 in $X=1$).

6.3.2 Semiparametric estimation of $\boldsymbol{p}_{\boldsymbol{x}}^*$

In this subsection we first define an estimator $\widetilde{\boldsymbol{p}}$ for $\boldsymbol{p}^*$ when both $P(\Omega_j), j \in \mathcal{J}$ and $P(\boldsymbol{X}=\boldsymbol{x})$ are known, and based on this we derive a second estimator $\widetilde{\widetilde{\boldsymbol{p}}}$ when none of those probabilities are known. We remark that this second estimator $\widetilde{\widetilde{\boldsymbol{p}}}$ is what we were looking for in order to estimate $\widetilde{\boldsymbol{S}}$.

Case I: $P(\Omega_j), j \in \mathcal{J}$ **and** $P(\boldsymbol{X} = \boldsymbol{x})$ **known**

Assume that we have fixed a category $\boldsymbol{X} = \boldsymbol{x}$ and a value $j \in \mathcal{J}$. If the data are complete and $P(\Omega_j)$ and $P(\boldsymbol{X} = \boldsymbol{x})$ are known, by Theorem 5.4.1 a consistent estimator for p_j^* can be found solving the estimating equation

$$\sum_{i=1}^{n} \left(I\left(\Omega_{ji}\right)\left(P(\Omega_j)I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x})p_j\right)\right) = 0. \qquad (6.18)$$

The crucial requirement in Theorem 5.4.1 is that the equation (6.18) has to be unbiased only for the true value p_j^*. Next lemma proves that this hypothesis is fulfilled.

Lemma 6.3.1 $E\left\{I\left(\Omega_{ji}\right)\left(P(\Omega_j)I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x})p_j\right)\right\} = 0$ *if and only if* $p_j = p_j^*$.

Proof:

$$\begin{aligned}
&E\left\{I\left(\Omega_{ji}\right)\left(P(\Omega_j)I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x})p_j\right)\right\} = \\
&= E\left\{I\left(\Omega_{ji}\right)P(\Omega_j)I(\boldsymbol{X}_i = \boldsymbol{x})\right\} - E\left\{I\left(\Omega_{ji}\right)P(\boldsymbol{X} = \boldsymbol{x})p_j\right) \\
&= P(\Omega_j)E\left\{I\left(\Omega_{ji}, \boldsymbol{X}_i = \boldsymbol{x}\right)\right\} - p_j P(\boldsymbol{X} = \boldsymbol{x})E\left\{I\left(\Omega_{ji}\right)\right\} \\
&= P(\Omega_j)P(\Omega_j, \boldsymbol{X} = \boldsymbol{x}) - p_j P(\boldsymbol{X} = \boldsymbol{x})P(\Omega_j)
\end{aligned}$$

that is equal to 0 if and only if $p_j = \dfrac{P(\Omega_j, \boldsymbol{X} = \boldsymbol{x})}{P(\boldsymbol{X} = \boldsymbol{x})} = P(\Omega_j | \boldsymbol{X} = \boldsymbol{x}) = p_j^*$. □

It is important to note that terms in (6.18) can be weighted by a real number d_j, depending only on j, and the equation remains unbiased.

When missing data is present, we already noted in Chapter 5 that equation (6.18) can only be used when the non-response mechanism is MCAR, because otherwise provides biased estimates.

However, if the probabilities of non-response, $\pi(\boldsymbol{r}; \boldsymbol{\alpha}^*) = P(\boldsymbol{R} = \boldsymbol{r} \mid \boldsymbol{L}; \boldsymbol{\alpha}^*)$, are known and the non-response mechanism is MAR, p_j^* can be consistently estimated by the solution of the inverse probability of being observed weighted equations (Liang and Zeger, 1986; Newey and McFadden, 1994; Robins et al., 1994)

$$\sum_{i=1}^{n} \left(\frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \boldsymbol{\alpha}^*\right)} I\left(\Omega_{ji}\right)\left(P(\Omega_j) I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x}) p_j\right) \right) = 0, \tag{6.19}$$

and the resulting estimator is asymptotically normal. The key for this result is again that (6.19) is unbiased.

Lemma 6.3.2 $E\left\{ \frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \boldsymbol{\alpha}^*\right)} I\left(\Omega_{ji}\right)\left(P(\Omega_j) I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x}) p_j\right) \right\} = 0$ *if and only if* $p_j = p_j^*$.

Proof: Conditioning on the data $\boldsymbol{L}$ we obtain

$$\begin{aligned}
& E\left\{ \frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \boldsymbol{\alpha}^*\right)} I\left(\Omega_{ji}\right)\left(P(\Omega_j) I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x}) p_j\right) \right\} = \\
= \; & E\left\{ E\left\{ \frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \boldsymbol{\alpha}^*\right)} I\left(\Omega_{ji}\right)\left(P(\Omega_j) I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x}) p_j\right) \mid \boldsymbol{L} \right\} \right\} \\
= \; & E\left\{ \frac{P(\boldsymbol{R}_i = \mathbf{1} \mid \boldsymbol{L})}{\pi_i(\mathbf{1}; \boldsymbol{\alpha}^*)} I\left(\Omega_{ji}\right)\left(P(\Omega_j) I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x}) p_j\right) \right\} \\
= \; & E\left\{ I\left(\Omega_{ji}\right)\left(P(\Omega_j) I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x}) p_j\right) \right\}.
\end{aligned}$$

The result follows after applying Lemma 6.3.1. □

If we do not know the true value of the parameter $\boldsymbol{\alpha}$, but a $\sqrt{n}$-consistent estimator $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}^*$ can be found, then we can replace $\boldsymbol{\alpha}^*$ by $\widehat{\boldsymbol{\alpha}}$ in (6.19) and, under regularity conditions, a consistent and asymptotically normal estimator of the true value of the parameter $\boldsymbol{p}_{\boldsymbol{x}}^*$ can be derived by solving the resulting equations.

Next lemma provides a general estimating equation to obtain $\sqrt{n}$-consistent estimators for $\boldsymbol{\alpha}^*$.

Lemma 6.3.3 *If* $\boldsymbol{\alpha}^*$ *is the true value for the q-dimensional vector of the non-response probabilities* $\pi_i(r; \boldsymbol{\alpha})$*, then for each set* $\{\phi_{\boldsymbol{r}}\}_{\boldsymbol{r} \in \{0,1\}^p}$ *of q-valued functions the solution* $\widehat{\boldsymbol{\alpha}}$ *to*

$$\sum_{i=1}^{n} \boldsymbol{A}_i(\phi) = \sum_{i=1}^{n} \sum_{\boldsymbol{r} \neq \mathbf{1}} \left(\left\{ I(\boldsymbol{R}_i = \boldsymbol{r}) - \frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i(\mathbf{1}; \boldsymbol{\alpha})} \pi_i(\boldsymbol{r}; \boldsymbol{\alpha}) \right\} \phi_{\boldsymbol{r}}(\boldsymbol{L}_{(\boldsymbol{r})i}) \right) = 0 \tag{6.20}$$

is a $\sqrt{n}$-consistent estimator for $\boldsymbol{\alpha}^*$.

Proof: By Theorem 5.4.2 it is enough to prove that $E\{\boldsymbol{A}_i(\phi)\} = 0$ for $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$. This result straightforwardly follows after conditioning on the data $\boldsymbol{L}$, and applying the definition of $\pi_i(\boldsymbol{r};\boldsymbol{\alpha}^*)$ as the conditional probabilities $P(\boldsymbol{R}=\boldsymbol{r}|\boldsymbol{L})$. □

Now, in a similar way than in Theorem 5.5.1, if we define $\boldsymbol{\gamma} = (\boldsymbol{p}', \boldsymbol{\alpha}')'$ and $\boldsymbol{\gamma}^* = (\boldsymbol{p}^{*\prime}, \boldsymbol{\alpha}^{*\prime})'$ as the parameter of interest and its true value, respectively, then we can build a class of $\sqrt{n}$-consistent estimators for $\boldsymbol{\gamma}^*$ (in particular for $\boldsymbol{p}^*$ and $\boldsymbol{\alpha}^*$).

For ease of notation, for each $j \in \mathcal{J}$ and $i = 1, \ldots, n$ we will denote by ϵ_{ji} the potential contribution of the i-th individual in the j-th equation in (6.18), that is,

$$\epsilon_{ji} = I\left(\Omega_{ji}\right)\left(P(\Omega_j)I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x})p_j\right). \tag{6.21}$$

Theorem 6.3.1 *Assuming that a model $\pi(\boldsymbol{r};\boldsymbol{\alpha})$ is correctly specified and $\pi(\boldsymbol{1};\boldsymbol{\alpha})$ is bounded away from 0 with probability 1, then the solution $\widehat{\boldsymbol{\gamma}} = \left(\widetilde{\boldsymbol{p}}', \widehat{\boldsymbol{\alpha}}'\right)'$ to the equations*

$$\boldsymbol{U}(\boldsymbol{p},\boldsymbol{\alpha};\boldsymbol{D},\phi^{(1)},\phi^{(2)}) = \sum_{i=1}^{n} \boldsymbol{U}_i(\boldsymbol{p},\boldsymbol{\alpha};\boldsymbol{D},\phi^{(1)},\phi^{(2)}) = \boldsymbol{0}_{2K+1+q} \tag{6.22}$$

defined by

$$\left.\begin{array}{rcl} \displaystyle\sum_{i=1}^{n}\left\{\frac{I(\boldsymbol{R}_i=\boldsymbol{1})}{\pi_i\left(\boldsymbol{1};\boldsymbol{\alpha}\right)}\boldsymbol{D}\begin{pmatrix}\epsilon_{11i}\\ \epsilon_{10i}\\ \vdots\\ \epsilon_{(K+1)i}\end{pmatrix} + \boldsymbol{A}_i(\phi^{(1)})\right\} & = & \boldsymbol{0}_{2K+1} \\ \displaystyle\sum_{i=1}^{n}\boldsymbol{A}_i(\phi^{(2)}) & = & \boldsymbol{0}_q \end{array}\right\} \tag{6.23}$$

where $\boldsymbol{D}$ is a $(2K+1)$-squared nonsingular constant matrix, $\left\{\phi^{(1)}_{\boldsymbol{r}}\right\}_{\boldsymbol{r}\in\{0,1\}^p}$ a set of $(2K+1)$-valued functions and $\left\{\phi^{(2)}_{\boldsymbol{r}}\right\}_{\boldsymbol{r}\in\{0,1\}^p}$ a set of q-valued functions, is a consistent estimator of the parameter $\boldsymbol{\gamma}^* = (\boldsymbol{p}^{*\prime}, \boldsymbol{\alpha}^{*\prime})'$.

Proof: We will verify the hypotheses in Theorem 5.4.1 about consistency of GMM estimators.

1. The matrix $\boldsymbol{W}$ that we are considering is $\boldsymbol{\Omega}^{-1} = (\text{Var}(\boldsymbol{U}_i(\boldsymbol{\gamma}^*)))^{-1}$ that is definite positive.

2. Since $\boldsymbol{W}$ is definite positive, it is only necessary to proof that $E(\boldsymbol{U}_i(\boldsymbol{\gamma})) = \mathbf{0}$ only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$. Let $\boldsymbol{\gamma}_0 = (\boldsymbol{p}_0', \boldsymbol{\alpha}_0')'$ be a potential vector for which $E(\boldsymbol{U}_i(\boldsymbol{\gamma}_0)) = \mathbf{0}$. By Lemma 6.3.3 the last q equations in (6.23), that is $\sum_{i=1}^n \boldsymbol{A}_i(\phi^{(2)}) = \mathbf{0}_q$, provide consistent estimators for $\boldsymbol{\alpha}^*$, and therefore $\boldsymbol{\alpha}_0 = \boldsymbol{\alpha}^*$.

   On the other hand, in the first $2K+1$ equations, $E\left\{\boldsymbol{A}_i(\phi^{(1)})\right\}|_{\left(\boldsymbol{p}_0', \boldsymbol{\alpha}^{*\prime}\right)'} = \mathbf{0}_{2K+1}$ and we obtain

$$E\left\{\frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \boldsymbol{\alpha}^*\right)}\boldsymbol{D}\begin{pmatrix}\epsilon_{11i}\\ \epsilon_{10i}\\ \vdots\\ \epsilon_{(K+1)i}\end{pmatrix}\right\} = \boldsymbol{D}E\left\{\frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \boldsymbol{\alpha}^*\right)}\begin{pmatrix}\epsilon_{11i}\\ \epsilon_{10i}\\ \vdots\\ \epsilon_{(K+1)i}\end{pmatrix}\right\} = \mathbf{0}_{2K+1}$$

   and, as $\boldsymbol{D}$ is nonsingular, then $E\left\{\dfrac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \boldsymbol{\alpha}^*\right)}\epsilon_{ji}\right\} = 0$ for each $j \in \mathcal{J}$. This implies by Lemma 6.3.2 that $p_{0j} = p_j^*$ for each $j \in \mathcal{J}$, thus $\boldsymbol{\gamma}_0 = \boldsymbol{\gamma}^*$.

3. We can warranty that the joint parameter $\boldsymbol{\gamma}^*$ is in the interior of a compact set of parameters. Indeed, the first $2K+1$ components correspond to the parameter $\boldsymbol{p}^*$ that obviously lies in the interior of a bounded set in R^{2K+1}, (note that $0 < p_j^* < 1$, for $j \in \mathcal{J}$). On the other hand, if the conditional probability $\pi_i(\boldsymbol{R} = \mathbf{1}|\boldsymbol{L})$ is bounded away from 0, it admits some parameterization (*e.g.,* as logit model) such that the corresponding space parameter for $\boldsymbol{\alpha}$ would also be bounded in R^q.

4. $\boldsymbol{U}_i(\boldsymbol{\gamma})$ is continuous for each $\boldsymbol{\gamma}$ because $\pi(\mathbf{1}; \boldsymbol{\alpha})$ is bounded away from 0.

5. For each $\boldsymbol{\gamma}$, $\boldsymbol{U}_i(\boldsymbol{\gamma})$ is uniformly bounded because $\pi(\mathbf{1}; \boldsymbol{\alpha})$ is bounded away from 0, the matrix $\boldsymbol{D}$ is constant and all the other quantities are uniformly bounded, hence $E(\sup_{\boldsymbol{\gamma}} \|\boldsymbol{U}_i(\boldsymbol{\gamma})\|) < \infty$. □

Next two lemmas consider the particular case of the estimating equations in (6.23) when $\phi_{\boldsymbol{r}}^{(1)}(.) = \mathbf{0}, \forall \boldsymbol{r} \neq \mathbf{1}$ and therefore $\boldsymbol{A}_i(\phi^{(1)}) = \mathbf{0}$. The resulting estimating equations are very useful, on one hand, in what follows in this section and, on the other, from an applied point of view, in the illustration with the HIV+PTB cohort in Section 6.4.

Lemma 6.3.4 *If in the expression (6.23) we setup* $\phi_{\boldsymbol{r}}^{(1)}(.) = \mathbf{0}, \forall \boldsymbol{r} \neq \mathbf{1}$*, then each one of the first* $2K+1$ *equations is equivalent to*

$$\sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \boldsymbol{\alpha}\right)} \epsilon_{ji} = 0,\ \ j \in \mathcal{J} \tag{6.24}$$

and then

$$\widetilde{p}_j = \frac{P(\Omega_j)}{P(\boldsymbol{X} = \boldsymbol{x})} \frac{\displaystyle\sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \widehat{\boldsymbol{\alpha}}\right)} I(\Omega_{ji}) I(\boldsymbol{X}_i = \boldsymbol{x})}{\displaystyle\sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1}; \widehat{\boldsymbol{\alpha}}\right)} I(\Omega_{ji})},\ \ j \in \mathcal{J}. \tag{6.25}$$

Proof: After substituting $\phi_{\boldsymbol{r}}^{(1)}(.) = \mathbf{0}, \forall \boldsymbol{r} \neq \mathbf{1}$, in (6.23), the equivalency follows as a direct consequence of the nonsingularity of the matrix $\boldsymbol{D}$. The explicit expression for $\widetilde{p}_j$ is derived by basic algebraic computations. □

Note that with the setup $\phi_{\boldsymbol{r}}^{(1)}(.) = \mathbf{0}, \forall \boldsymbol{r} \neq \mathbf{1}$ the nonsingular matrix $\boldsymbol{D}$ becomes irrelevant. Observe that in (6.24) $\boldsymbol{D}$ is equal to the identity.

Lemma 6.3.5 *If* $\phi_{\boldsymbol{r}}^{(1)}(.) = \mathbf{0}\ \forall \boldsymbol{r} \neq \mathbf{1}$*, then for every value* $j = ku \in \mathcal{J}$ *we have the following two facts:*

a) if there are no individuals in I_k *with* $\delta = u$ *that belong to category* $\boldsymbol{x}$ *in the subsample with fully observed covariate vector, then* $\widetilde{p}_j = 0$*, and*

b) if all the individuals in I_k *with* $\delta = u$ *and fully observed covariate vector* $\boldsymbol{X}$ *belong to the category* $\boldsymbol{x}$*, then* $\widetilde{p}_j = P(\Omega_j)/P(\boldsymbol{X} = \boldsymbol{x})$*.*

Proof: By the previous lemma, we can substitute in equation (6.25) $I(\boldsymbol{X}_i = \boldsymbol{x})$ by 0 or by 1, respectively, and the result follows straightforwardly. □

As a consequence of Lemma 6.3.5 we have that if $\phi_{\boldsymbol{r}}^{(1)}(.) = 0,\ \forall \boldsymbol{r} \neq \mathbf{1}$, then

Remark 6.3.1 .

1. *if we are in case a) of the previous lemma, we do not have information for the category $\boldsymbol{x}$ and the estimating equations do not allow to make inferences about the p_j^* corresponding to category $\boldsymbol{x}$, and*

2. *on the other hand, if the empirical distribution of $\boldsymbol{X}$ in the subsample with fully observed covariate vector is concentrated in one category $\boldsymbol{x}$, all the subjects with missing data in $\boldsymbol{X}$ will be estimated to belong to this category $\boldsymbol{x}$.*

It is important to note, as we illustrated in the introductory example, that if we are in the unidimensional case (*i.e.,* $p = 1$), the equations (6.23) distribute the individuals with missing covariate between the observed values of the covariate X. However, in the multivariate case, if we would setup $\phi_{\boldsymbol{r}}^{(1)}(.) = 0$, $\forall \boldsymbol{r} \neq \mathbf{1}$ and, if jointly with individuals belonging to category $\boldsymbol{x}$ we would observe partially individuals not belonging to $\boldsymbol{x}$, then the resulting estimates from the estimating equations would not be supported by the data. In these situations it will be necessary to use other configurations for $\phi_{\boldsymbol{r}}^{(1)}(.)$ to take the available information into account in the subjects with partially observed covariates. Obviously, this issue disappears when the sample size is large enough.

We illustrate this drawback through the following example. Assume that $\boldsymbol{X}$ is a bivariate binary covariate vector, and that for a fixed $j = ku \in \mathcal{J}$ we have four individuals with covariates (1,1), (NA,1), (0,NA) and (NA,NA). Note that based on this data the number of individuals belonging to categories $(0, 0)$ or $(0, 1)$ is at least 1, and the number of individuals that could be assigned to the category $(1, 1)$ is at most 3.

If we would try to estimate S_k for categories (0,0) and (0,1), using $\phi_{\boldsymbol{r}}^{(1)}(.) = 0$, $\forall \boldsymbol{r} \neq \mathbf{1}$, then the respective p_j^* would be wrongly estimated as $\widetilde{p}_j = 0$, and for the category (1,1) we would estimate $\widetilde{p}_j = P(\Omega_j)/P(\boldsymbol{X} = \boldsymbol{x})$.

The concerning issue is that the resulting estimated number of individuals in the category (1,1) does not have to be, necessarily, lower than or equal to 3 (take, for instance, $P(\Omega_j) = 1/4$ and $P(\boldsymbol{X} = \boldsymbol{x}) = 1/4$, thus $\widetilde{p}_j = 1$ and we would obtain that 4 individuals would be in category (1,1)).

Case II: $P(\Omega_j), j \in \mathcal{J}$ **and** $P(\boldsymbol{X} = \boldsymbol{x})$ **unknown**

In many practical situations we do not know the values of $P(\Omega_j)$, for $j \in \mathcal{J}$ and $P(\boldsymbol{X} = \boldsymbol{x})$. So, we cannot neither evaluate the residuals ϵ_{ji} in (6.21) nor compute the explicit expression for $\widetilde{p_j}$ in (6.25). However, last q equations in (6.23), that is $\sum_{i=1}^{n} \boldsymbol{A}_i(\phi^{(2)}) = \boldsymbol{0}_q$, still provide a $\sqrt{n}$-consistent estimator $\widehat{\boldsymbol{\alpha}}$ for $\boldsymbol{\alpha}^*$.

Henceforth, in this section we consider that the estimating equations in (6.22) and (6.23) use $\phi_{\boldsymbol{r}}^{(1)}(.) = 0$, $\forall \boldsymbol{r} \neq \boldsymbol{1}$. Therefore, if we were in the case I, for each estimate $\widehat{\boldsymbol{\alpha}}$ we could derive $\widetilde{p_j}$ as the solution to

$$\sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i\left(\boldsymbol{1}; \widehat{\boldsymbol{\alpha}}\right)} I(\Omega_{ji}) \left(P(\Omega_j) I(\boldsymbol{X}_i = \boldsymbol{x}) - P(\boldsymbol{X} = \boldsymbol{x}) p_j\right) = 0. \tag{6.26}$$

However note that this equation is equivalent to

$$\sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i\left(\boldsymbol{1}; \widehat{\boldsymbol{\alpha}}\right)} I(\Omega_{ji}) \left(n \cdot P(\Omega_j) I(\boldsymbol{X}_i = \boldsymbol{x}) - n \cdot P(\boldsymbol{X} = \boldsymbol{x}) p_j\right) = 0. \tag{6.27}$$

We can observe that $n \cdot P(\Omega_j)$ corresponds to the expected number of individuals in the sample for which $I(\Omega_{ji}) = 1$. Denote by $m_j = \sum_{i=1}^{n} I(\Omega_{ji})$ the number of individuals in the sample with $I(\Omega_{ji}) = 1$, and by $\boldsymbol{m} = (m_j)_{j \in \mathcal{J}}$ the corresponding vector of observed counts.

Analogously let $n_{\boldsymbol{x}} = n \cdot P(\boldsymbol{X} = \boldsymbol{x})$ be the expected number of individuals in the sample with $\boldsymbol{X}_i = \boldsymbol{x}$. Since $p_j = P(\Omega_j | \boldsymbol{X} = \boldsymbol{x})$, then $n \cdot P(\boldsymbol{X} = \boldsymbol{x}) p_j = n_{\boldsymbol{x}} \cdot p_j$ is the expected number of individuals in $\boldsymbol{X} = \boldsymbol{x}$ such that $I(\Omega_{ji}) = 1$. Define $\beta_j = n_{\boldsymbol{x}} \cdot p_j$ the above expectation (note that β_j, as well as p_j, depends on $\boldsymbol{x}$).

If we use this notation in (6.27) we obtain

$$\sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i\left(\boldsymbol{1}; \widehat{\boldsymbol{\alpha}}\right)} I(\Omega_{ji}) (m_j I(\boldsymbol{X}_i = \boldsymbol{x}) - \beta_j) = 0 \tag{6.28}$$

and we can derive the solution for β_j

$$\widetilde{\beta}_j = m_j \frac{\displaystyle\sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i\left(\boldsymbol{1}; \widehat{\boldsymbol{\alpha}}\right)} I(\Omega_{ji}) I(\boldsymbol{X}_i = \boldsymbol{x})}{\displaystyle\sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i\left(\boldsymbol{1}; \widehat{\boldsymbol{\alpha}}\right)} I(\Omega_{ji})}, \; j \in \mathcal{J}. \tag{6.29}$$

From these values we can estimate $n_{\boldsymbol{x}}$ by $\widetilde{n_{\boldsymbol{x}}} = \sum_{k\in\mathcal{J}} \widetilde{\beta}_k$ and we can define a new estimator $\widetilde{\widetilde{\boldsymbol{p}}} = (\widetilde{\widetilde{p_j}})_{j\in\mathcal{J}}$ for $\boldsymbol{p}$ where

$$\widetilde{\widetilde{p_j}} = \frac{\widetilde{\beta_j}}{\widetilde{n_{\boldsymbol{x}}}} = \frac{\widetilde{\beta_j}}{\sum\limits_{k\in\mathcal{J}} \widetilde{\beta_k}} = \frac{m_j \dfrac{\sum\limits_{i=1}^{n} \dfrac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} I(\Omega_{ji}) I(\boldsymbol{X}_i = \boldsymbol{x})}{\sum\limits_{i=1}^{n} \dfrac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} I(\Omega_{ji})}}{\sum\limits_{k\in\mathcal{J}} m_k \dfrac{\sum\limits_{i=1}^{n} \dfrac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} I(\Omega_{ki}) I(\boldsymbol{X}_i = \boldsymbol{x})}{\sum\limits_{i=1}^{n} \dfrac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} I(\Omega_{ki})}}. \tag{6.30}$$

For ease of notation in expressions (6.29) and (6.30) we denote by ω_j the weights $\sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} I(\Omega_{ji}) I(\boldsymbol{X}_i = \boldsymbol{x}) / \sum_{i=1}^{n} \frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} I(\Omega_{ji})$ and thus $\widetilde{\beta_j} = m_j\omega_j$ and $\widetilde{\widetilde{p_j}} = m_j\omega_j / \sum_{k\in\mathcal{J}} m_k\omega_k$. Next lemma shows that $\widetilde{\widetilde{p_j}}$ is a consistent estimator for the true value p_j^*.

Lemma 6.3.6 *According to the definitions in (6.29) and (6.30), for each $j \in \mathcal{J}$, when $n \to \infty$*

a) $\omega_j \xrightarrow{a.s.} P(\boldsymbol{X} = \boldsymbol{x}|\Omega_j)$ *and*

b) $\widetilde{\widetilde{p_j}} \xrightarrow{a.s.} P(\Omega_j|\boldsymbol{X} = \boldsymbol{x}) = p_j^*$.

Proof:

a) Rewrite ω_j as follows

$$\omega_j = \frac{\sum\limits_{i=1}^{n} \dfrac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} I(\Omega_{ji}) I(\boldsymbol{X}_i = \boldsymbol{x})}{\sum\limits_{i=1}^{n} \dfrac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} I(\Omega_{ji})} = \frac{\sum\limits_{i=1}^{n} \dfrac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} \dfrac{I(\Omega_{ji}, \boldsymbol{X}_i = \boldsymbol{x})}{n}}{\sum\limits_{i=1}^{n} \dfrac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})} \dfrac{I(\Omega_{ji})}{n}}.$$

By the strong law of large numbers (see, for instance, Schervish (1995)) it converges almost surely to $P(\Omega_j, \boldsymbol{X} = \boldsymbol{x})/P(\Omega_j) = P(\boldsymbol{X} = \boldsymbol{x}|\Omega_j)$, when $n \to \infty$.

b) If we note that

$$\widetilde{\widetilde{p}}_j = \frac{m_j \omega_j}{\displaystyle\sum_{k \in \mathcal{J}} m_k \omega_k} = \frac{\dfrac{\sum_{i=1}^{n} I(\Omega_{ji})}{n} \omega_j}{\displaystyle\sum_{k \in \mathcal{J}} \dfrac{\sum_{i=1}^{n} I(\Omega_{ki})}{n} \omega_k}$$

therefore, by the strong law of large numbers and the previous part a), $\widetilde{\widetilde{p}}_j$ converges almost surely to

$$\frac{P(\Omega_j)P(\boldsymbol{X} = \boldsymbol{x}|\Omega_j)}{\displaystyle\sum_{k \in \mathcal{J}} P(\Omega_k)P(\boldsymbol{X} = \boldsymbol{x}|\Omega_k)} = \frac{P(\Omega_j, \boldsymbol{X} = \boldsymbol{x})}{\displaystyle\sum_{k \in \mathcal{J}} P(\Omega_k, \boldsymbol{X} = \boldsymbol{x})} =$$
$$= \frac{P(\Omega_j, \boldsymbol{X} = \boldsymbol{x})}{P(\boldsymbol{X} = \boldsymbol{x})} = P(\Omega_j | \boldsymbol{X} = \boldsymbol{x}) = p_j^*$$

when $n \to \infty$. □

In the next section we will prove the asymptotic properties of $\widetilde{\boldsymbol{p}}$ and $\widetilde{\widetilde{\boldsymbol{p}}}$. In particular, we will prove that $\widetilde{\widetilde{\boldsymbol{p}}}$ is also a $\sqrt{n}$-consistent estimator of $\boldsymbol{p}^*$.

6.3.3 Asymptotic properties of $\widetilde{\boldsymbol{p}_{\boldsymbol{x}}}$ and $\widetilde{\widetilde{\boldsymbol{p}_{\boldsymbol{x}}}}$

The asymptotic behavior of the vector $\widehat{\boldsymbol{\gamma}} = (\widetilde{\boldsymbol{p}}', \widehat{\boldsymbol{\alpha}}')'$ follows straightforwardly from Theorem 5.4.2. We need to derive the expression of the expected score matrix for the estimating equations (6.23) in order to verify the hypothesis 5 in Theorem 5.4.2. After that, the asymptotic properties of $\widetilde{\boldsymbol{p}}$ are obtained.

Lemma 6.3.7 *The estimating equations (6.22) and (6.23) have a score matrix*

$$\frac{\partial \boldsymbol{U}_i}{\partial(\boldsymbol{p}', \boldsymbol{\alpha}')'} = \begin{pmatrix} \boldsymbol{D}_i & \boldsymbol{B}_i \\ \boldsymbol{0} & \boldsymbol{C}_i \end{pmatrix}$$

where

$$\begin{aligned}
\boldsymbol{D}_i &= -\frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1};\boldsymbol{\alpha}\right)} P(\boldsymbol{X} = \boldsymbol{x})\boldsymbol{D}\ \mathrm{diag}\left(I(\Omega_{ji})_{j\in\mathcal{J}}\right), \\
\boldsymbol{B}_i &= I(\boldsymbol{R}_i = \mathbf{1})\boldsymbol{D}\begin{pmatrix} \epsilon_{11i} \\ \epsilon_{10i} \\ \vdots \\ \epsilon_{(K+1)i} \end{pmatrix}\frac{\partial \pi_i^{-1}(\mathbf{1};\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} + \frac{\partial \boldsymbol{A}_i(\phi^{(1)})}{\partial \boldsymbol{\alpha}}, \text{ and} \\
\boldsymbol{C}_i &= \frac{\partial \boldsymbol{A}_i(\phi^{(2)})}{\partial \boldsymbol{\alpha}}
\end{aligned}$$

are matrices of dimension equal to $(2K+1)\times(2K+1), (2K+1)\times q$ *and* $q\times q$*, respectively.*

Proof: The result follows straightforwardly, after taking the derivative of each term in the sum of the left-hand part of equations (6.23) and using the notation introduced in (6.21). □

Lemma 6.3.8 *For large* n *and for all those choices of functions* $\phi^{(2)}_{\boldsymbol{r}}$ *such that* $E\{\boldsymbol{C}_i\}$ *is nonsingular then* $\boldsymbol{\Gamma} = E\left\{(\partial \boldsymbol{U}_i/\partial \boldsymbol{\gamma})\,|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^*}\right\}$ *is nonsingular.*

Proof: By the previous lemma it is only necessary to prove that the expectation of the matrix $\boldsymbol{D}_i$ is nonsingular for the true value $\boldsymbol{\gamma}^*$. But

$$E\left\{\boldsymbol{D}_i\right\} = -P(\boldsymbol{X} = \boldsymbol{x})\boldsymbol{D}\ \mathrm{diag}\left(E\left\{\frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1};\boldsymbol{\alpha}^*\right)} I(\Omega_{ji})\right\}_{j\in\mathcal{J}}\right)$$

and therefore to derive the nonsingularity it is enough to prove that each one of the expectations in the diagonal matrix is different from 0.

Without loss of generality in the family of the usual distributions for Y and δ, we will assume that the joint density function of (Y,δ) assigns positive mass to all neighbourhoods of every observed point in the sample. We have to take into account the fact that $\boldsymbol{X}$ is discrete and that for each value $Y \in I_k$ and $\delta = u$ the probabilities $\pi_i(\mathbf{1};\boldsymbol{\alpha}^*)$ depend only on $\boldsymbol{X}$. We shall distinguish two cases: $j = ku$ (with $k = 1,\dots,K$ and $u = 1, 0$) and $j = K+1$.

If $j = ku$, we can rewrite

$$\begin{aligned}
& E\left\{\frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1};\boldsymbol{\alpha}^*\right)} I(\Omega_{ji})\right\} = \\
&= \sum_{\text{all } \boldsymbol{x}} \frac{1}{\pi_i(\mathbf{1};\boldsymbol{\alpha}^*)} P(\boldsymbol{R}_i = \mathbf{1}, Y_i \in I_k, \delta_i = u, \boldsymbol{X}_i = \boldsymbol{x}) \\
&= \sum_{\text{all } \boldsymbol{x}} P(Y_i \in I_k, \delta_i = u, \boldsymbol{X}_i = \boldsymbol{x}) \frac{P(\boldsymbol{R}_i = \mathbf{1} | Y_i \in I_k, \delta_i = u, \boldsymbol{X}_i = \boldsymbol{x})}{\pi_i(\mathbf{1};\boldsymbol{\alpha}^*)} \\
&= \sum_{\text{all } \boldsymbol{x}} P(Y_i \in I_k, \delta_i = u, \boldsymbol{X}_i = \boldsymbol{x}) = P(Y_i \in I_k, \delta_i = u) = P(\Omega_j) > 0,
\end{aligned}$$

and if $j = K + 1$ then

$$\begin{aligned}
& E\left\{\frac{I(\boldsymbol{R}_i = \mathbf{1})}{\pi_i\left(\mathbf{1};\boldsymbol{\alpha}^*\right)} I(\Omega_{ji})\right\} = \\
&= \sum_{\text{all } \boldsymbol{x}} \frac{1}{\pi_i(\mathbf{1};\boldsymbol{\alpha}^*)} P(\boldsymbol{R}_i = \mathbf{1}, Y_i > \tau_K, \boldsymbol{X}_i = \boldsymbol{x}) \\
&= \sum_{\text{all } \boldsymbol{x}} P(Y_i > \tau_K, \boldsymbol{X}_i = \boldsymbol{x}) \frac{P(\boldsymbol{R}_i = \mathbf{1} | Y_i > \tau_K, \boldsymbol{X}_i = \boldsymbol{x})}{\pi_i(\mathbf{1};\boldsymbol{\alpha}^*)} \\
&= \sum_{\text{all } \boldsymbol{x}} P(Y_i > \tau_K, \boldsymbol{X}_i = \boldsymbol{x}) = P(Y_i > \tau_K) = P(\Omega_{K+1}) > 0.
\end{aligned}$$

□

Theorem 6.3.2 *In the estimating equations (6.22) and (6.23), assuming that the model $\pi(\boldsymbol{r};\boldsymbol{\alpha})$ is correctly specified, under the regularity conditions*

1. *$\boldsymbol{\gamma}$ lies in the interior of a compact set,*
2. *$(\boldsymbol{L}_i, \boldsymbol{R}_i)$, $i = 1, \ldots, n$ are independently and identically distributed,*
3. *for some c, $\pi(\mathbf{1};\boldsymbol{\alpha}) > c > 0$ for all $\boldsymbol{\alpha}$,*
4. *$E(\boldsymbol{U}_i(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})) \neq \mathbf{0}$ if $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}^*$,*
5. *$\boldsymbol{\Omega} = Var(\boldsymbol{U}_i(\boldsymbol{\gamma}^*;\boldsymbol{D},\phi^{(1)},\phi^{(2)}))$ is finite and positive definite,*
6. *$\boldsymbol{\Gamma} = E\left(\partial \boldsymbol{U}_i(\boldsymbol{\gamma}^*;\boldsymbol{D},\phi^{(1)},\phi^{(2)})/\partial\boldsymbol{\gamma}\right)$ exists and is invertible,*

7. *there exist a neighborhood* N *of* $\boldsymbol{\gamma}^*$ *such that* $E(\sup_{\boldsymbol{\gamma}\in N}\|\boldsymbol{U}_i(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})\|)$, $E(\sup_{\boldsymbol{\gamma}\in N}\|\partial\boldsymbol{U}_i(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})/\partial\boldsymbol{\gamma}\|)$, *and* $E(\sup_{\boldsymbol{\gamma}\in N}\|\boldsymbol{U}_i(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})\cdot \boldsymbol{U}_i'(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})\|)$ *are all finite, where* $\|\boldsymbol{A}\| = (\sum_{ij} a_{ij}^2)^{1/2}$ *for any matrix* $\boldsymbol{A} = (a_{ij})$,

8. *for all* $\bar{\boldsymbol{\gamma}}$ *in a neighborhood* N *of* $\boldsymbol{\gamma}^*$, $E_{\bar{\boldsymbol{\gamma}}}(\boldsymbol{U}_i(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)}))$ *and* $E_{\bar{\boldsymbol{\gamma}}}(\sup_{\boldsymbol{\gamma}\in N}\|\boldsymbol{U}_i(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})\boldsymbol{U}_i'(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})\|)$ *are bounded, where* $E_{\bar{\boldsymbol{\gamma}}}$ *refers to expectation with respect to the density* $f(\boldsymbol{L},\boldsymbol{R};\bar{\boldsymbol{\gamma}})$,

then

a) *with probability approaching 1, there is a unique solution* $\widehat{\boldsymbol{\gamma}}$ *to (6.23) and* $\widehat{\boldsymbol{\gamma}} \xrightarrow{\mathcal{P}} \boldsymbol{\gamma}^*$,

b) *the random vector* $\sqrt{n}\,(\widehat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^*) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0},\boldsymbol{\Upsilon})$, *with* $\boldsymbol{\Upsilon} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Omega}\boldsymbol{\Gamma}^{-1'}$,

c) *the asymptotic variance-covariance matrix* $\boldsymbol{\Upsilon}$ *can be consistently estimated by* $\widehat{\boldsymbol{\Upsilon}} = \widehat{\boldsymbol{\Gamma}}^{-1}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{\Gamma}}^{-1'}$ *where* $\widehat{\boldsymbol{\Gamma}} = n^{-1}\sum\partial\boldsymbol{U}_i(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})/\partial\boldsymbol{\gamma}$ *and* $\widehat{\boldsymbol{\Omega}} = n^{-1}\sum\boldsymbol{U}_i(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})\boldsymbol{U}_i'(\boldsymbol{\gamma};\boldsymbol{D},\phi^{(1)},\phi^{(2)})$ *are evaluated in* $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}$.

Proof: Observe that we are considering $\boldsymbol{W} = \boldsymbol{\Omega}^{-1}$ and $\boldsymbol{\Omega}$ is definite positive. Regularity conditions 5, 4, 1 and 7 correspond to the hypothesis in Theorem 5.4.1. Part a) is then straightforwardly derived. Regularity conditions 1, 4, 6 and 7 imply, respectively, hypothesis 1, 3, 5, 2 and 4 in Theorem 5.4.2 (hypothesis 5 in Theorem 5.4.2 is replaced by the nonsingularity of $\boldsymbol{\Gamma}$, which is proved in Lemma 6.3.8). Therefore, we conclude the asymptotic normality and the asymptotic variance-covariance matrix (part b)). Conditions 7 and 8 allow to apply Theorem 5.4.3 in order to replace the true value $\boldsymbol{\gamma}^*$ by its estimator $\widehat{\boldsymbol{\gamma}}$ and get the consistent variance-covariance estimator (part c)). □

Next corollary establishes the asymptotic distribution of $\widetilde{\boldsymbol{p}}$ and a consistent estimator for its asymptotic variance-covariance matrix.

Corollary 6.3.1 *If $\widetilde{\boldsymbol{p}}$ denotes the estimator of $\boldsymbol{p}^*$ in (6.23), under the same hypotheses of Theorem 6.3.2 we have*

a) $\sqrt{n}\,(\widetilde{\boldsymbol{p}} - \boldsymbol{p}^*) \xrightarrow{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_1)$, *where $\boldsymbol{\Lambda}_1$ is the $(2K+1) \times (2K+1)$ upper-left squared matrix of $\boldsymbol{\Upsilon}$ in part b) of Theorem 6.3.2.*

b) the asymptotic variance-covariance matrix $\boldsymbol{\Lambda}_1$ can be consistently estimated by the $(2K+1) \times (2K+1)$ upper-left squared matrix of $\widehat{\boldsymbol{\Upsilon}} = \widehat{\boldsymbol{\Gamma}}^{-1}\widehat{\boldsymbol{\Omega}}\widehat{\boldsymbol{\Gamma}}^{-1'}$ where $\widehat{\boldsymbol{\Gamma}} = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} \boldsymbol{D}_i & \boldsymbol{B}_i \\ \mathbf{0} & \boldsymbol{C}_i \end{pmatrix}$ and $\widehat{\boldsymbol{\Omega}} = \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{U}_i(\boldsymbol{\gamma}; \boldsymbol{D}, \phi^{(1)}, \phi^{(2)})\boldsymbol{U}_i'(\boldsymbol{\gamma}; \boldsymbol{D}, \phi^{(1)}, \phi^{(2)})$ are evaluated in $\boldsymbol{\gamma} = \widehat{\boldsymbol{\gamma}}$.

Next lemmas and Theorem 6.3.3 establish the asymptotic behavior of the semiparametric estimator $\widetilde{\widetilde{\boldsymbol{p}}}$ introduced in (6.30) in relation to the estimator $\widetilde{\boldsymbol{p}}$.

Lemma 6.3.9 *Let $\{X_n\}_{n\geq 1}$ and $\{Y_n\}_{n\geq 1}$ two sequences of random variables such that X_n and Y_n are independent for all n. If*

1. $X_n \xrightarrow{\mathcal{P}} a, Y_n \xrightarrow{\mathcal{P}} b$ *and*
2. $\sqrt{n}(X_n - a) \xrightarrow{\mathcal{D}} X \sim \mathcal{N}(0, \sigma_X^2)$ *and* $\sqrt{n}(Y_n - b) \xrightarrow{\mathcal{D}} Y \sim \mathcal{N}(0, \sigma_Y^2)$,

then $\sqrt{n}(X_nY_n - ab) \xrightarrow{\mathcal{D}} \mathcal{N}(0, b^2\sigma_X^2 + a^2\sigma_Y^2)$.

Proof: By Theorem 7.20 at page 400 in Shervish (1995),

$$(\sqrt{n}(X_n - a), \sqrt{n}(Y_n - b)) \xrightarrow{\mathcal{D}} (X, Y)$$

and X and Y are independent.

On the other hand, we can write $\sqrt{n}(X_nY_n - ab)$ as follows

$$\sqrt{n}(X_nY_n - ab) = \sqrt{n}(X_nY_n - aY_n + aY_n - ab) = \sqrt{n}\;Y_n(X_n - a) + \sqrt{n}\;a(Y_n - b).$$

Since $\sqrt{n}(X_n - a) \xrightarrow{\mathcal{D}} X$ and $\sqrt{n}(Y_n - b) \xrightarrow{\mathcal{D}} Y$, after applying the Slutzky's Theorem $\sqrt{n}\;Y_n(X_n - a) \xrightarrow{\mathcal{D}} bX$ and $\sqrt{n}\;a(Y_n - b) \xrightarrow{\mathcal{D}} aY$ and, by the normality and the independence of the random variables X and Y, the lemma is derived. □

Lemma 6.3.10 *According to the definitions in (6.29) and (6.30), for each $j \in \mathcal{J}$, when $n \to \infty$*

a) $\frac{m_j}{n} \xrightarrow{a.s.} P(\Omega_j)$ *and* $\sqrt{n}\left(\frac{m_j}{n} - P(\Omega_j)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P(\Omega_j)(1 - P(\Omega_j)))$

b) $\sqrt{n}(\omega_j - P(\boldsymbol{X} = \boldsymbol{x}|\Omega_j)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_j^2)$ *where* σ_j^2 *can be consistently estimated by*

$$\widehat{\sigma_j^2} = \left(\frac{-1}{n}\sum_{i=1}^{n}\frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i\left(\boldsymbol{1};\widehat{\boldsymbol{\alpha}}\right)}I(\Omega_{ji})\right)^{-2}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})^2}I(\Omega_{ji})(I(\boldsymbol{X}_i = \boldsymbol{x}) - \omega_j)^2\right).$$

Proof:

a) For each $j \in \mathcal{J}$, $m_j/n = \sum_{i=1}^{n} I(\Omega_{ji})/n$ and it converges almost surely to $P(\Omega_j)$ by the strong law of large nunmbers, and, by the central limit theorem $\sqrt{n}(\sum_{i=1}^{n} I(\Omega_{ji})/n - P(\Omega_j)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P(\Omega_j)(1 - P(\Omega_j)))$.

b) By Lemma 6.3.6, $\omega_j \xrightarrow{\text{a.s.}} P(\boldsymbol{X} = \boldsymbol{x}|\Omega_j)$. On the other hand, $\boldsymbol{\omega} = (\omega_k)_{k\in\mathcal{J}}$ by construction is the solution to the diagonal estimating equations

$$\sum_{i=1}^{n}\boldsymbol{V}_i = \sum_{i=1}^{n}\frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i\left(\boldsymbol{1};\widehat{\boldsymbol{\alpha}}\right)}\text{diag}(I(\Omega_{ki})_{k\in\mathcal{J}})(I(\boldsymbol{X}_i = \boldsymbol{x}) - \boldsymbol{\omega}) = \boldsymbol{0}_{2K+1}. \tag{6.31}$$

Therefore, after following similar steps to those in the Theorem 6.3.2 for the vector $(\boldsymbol{\omega}', \boldsymbol{\alpha}')'$ instead of the vector $(\boldsymbol{p}', \boldsymbol{\alpha}')'$, we obtain a similar result to the Corollary 6.3.1 for the vector $\boldsymbol{\omega}$,

$$\sqrt{n}(\omega_j - P(\boldsymbol{X} = \boldsymbol{x}|\Omega_j)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_j^2),$$

where σ_j^2 can be consistently estimated by

$$\begin{aligned}\widehat{\sigma_j^2} &= \left(\left(\widehat{\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\omega}}}\right)^{-1}\widehat{\text{Var}(\boldsymbol{V}_i)}\left(\widehat{\frac{\partial \boldsymbol{V}_i}{\partial \boldsymbol{\omega}}}\right)^{-1}\right)_{jj} \\ &= \left(\frac{-1}{n}\sum_{i=1}^{n}\frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\pi_i\left(\boldsymbol{1};\widehat{\boldsymbol{\alpha}}\right)}I(\Omega_{ji})\right)^{-2}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{I(\boldsymbol{R}_i = \boldsymbol{1})}{\boldsymbol{\pi}_i(\boldsymbol{1};\widehat{\boldsymbol{\alpha}})^2}I(\Omega_{ji})(I(\boldsymbol{X}_i = \boldsymbol{x}) - \omega_k)^2\right).\end{aligned}$$

□

Remark 6.3.2 *In order to derive the asymptotic normality of the resulting estimators we have to conjecture the following independence conditions:*

C1. *For each $k \in \mathcal{J}$, m_k and ω_k are independent.*

C2. *For each $k_1, k_2 \in \mathcal{J}, k_1 \neq k_2$, $\widetilde{\beta_{k_1}} = m_{k_1}\omega_{k_1}$ and $\widetilde{\beta_{k_2}} = m_{k_2}\omega_{k_2}$ are independent.*

C3. *For each $j \in \mathcal{J}$, m_j and $\widetilde{n_{\boldsymbol{x}}} = \sum_{k\in\mathcal{J}} m_k\omega_k$ are independent.*

C4. *The random vector $(\widetilde{\boldsymbol{p}} - \boldsymbol{p}^*)$ and $\left(\widetilde{\widetilde{\boldsymbol{p}}} - \widetilde{\boldsymbol{p}}\right)$ are independent.*

Conjecture C1 is based on the fact that m_k counts the event Ω_k while ω_k is a weight from the complete observed subsample that depends on the non-response pattern and the distribution of $\boldsymbol{X}$. The main idea about conjecture C2 is that the vector of counts $\widetilde{\beta_{k_1}}$ and $\widetilde{\beta_{k_2}}$ refer to two different pairs of time-delta indexes in $\mathcal{J}$. In a similar way than in conjecture C1, in conjecture C3, $\widetilde{n_{\boldsymbol{x}}}$ refers to the estimated number of individuals belonging to category $\boldsymbol{X} = \boldsymbol{x}$. Finally, conjecture C4 will be supported by the asymptotically uncorrelation between both residuals (see Lemma 6.3.12).

Lemma 6.3.11 *If X_n denotes the random variable $\dfrac{P(\boldsymbol{X}=\boldsymbol{x})}{\sum\limits_{k\in\mathcal{J}} \frac{m_k}{n}\omega_k}$ then*

a) $X_n \xrightarrow{a.s.} 1$ *and*

b) $\sqrt{n}(1 - X_n) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \nu^2)$ *with*

$$\nu^2 = \frac{\sum_{k\in\mathcal{J}} \left(P(\boldsymbol{X}=\boldsymbol{x}|\Omega_k)^2 P(\Omega_k)(1-P(\Omega_k)) + P(\Omega_k)^2\sigma_k^2\right)}{P(\boldsymbol{X}=\boldsymbol{x})^2}$$

where σ_k^2 is the same as the previous lemma.

Proof:

a) By the strong law of large numbers,

$$X_n = \frac{P(\boldsymbol{X}=\boldsymbol{x})}{\sum\limits_{k\in\mathcal{J}} \frac{m_k}{n}\omega_k} \xrightarrow{\text{a.s.}} \frac{P(\boldsymbol{X}=\boldsymbol{x})}{\sum\limits_{k\in\mathcal{J}} P(\Omega_k)P(\boldsymbol{X}=\boldsymbol{x}|\Omega_k)} = \frac{P(\boldsymbol{X}=\boldsymbol{x})}{\sum\limits_{k\in\mathcal{J}} P(\Omega_k, \boldsymbol{X}=\boldsymbol{x})} = 1.$$

b) Rewrite $\sqrt{n}(1 - X_n)$ as follows

$$\begin{aligned}
\sqrt{n}(1 - X_n) &= \sqrt{n}\left(1 - \frac{P(\boldsymbol{X} = \boldsymbol{x})}{\sum_{k\in\mathcal{J}} \frac{m_k}{n}\omega_k}\right) = \sqrt{n}\frac{\sum_{k\in\mathcal{J}} \frac{m_k}{n}\omega_k - P(\boldsymbol{X} = \boldsymbol{x})}{\sum_{k\in\mathcal{J}} \frac{m_k}{n}\omega_k} \\
&= \sqrt{n}\frac{\sum_{k\in\mathcal{J}} \left(\frac{m_k}{n}\omega_k - P(\Omega_k, \boldsymbol{X} = \boldsymbol{x})\right)}{\sum_{k\in\mathcal{J}} \frac{m_k}{n}\omega_k} \\
&= \frac{1}{\sum_{k\in\mathcal{J}} \frac{m_k}{n}\omega_k} \cdot \sum_{k\in\mathcal{J}} \sqrt{n}\left(\frac{m_k}{n}\omega_k - P(\Omega_k)P(\boldsymbol{X} = \boldsymbol{x}|\Omega_k)\right)
\end{aligned} \tag{6.32}$$

On one hand, the factor $\frac{1}{\sum_{k\in\mathcal{J}} \frac{m_k}{n}\omega_k}$ in (6.32), as we have previously seen, converges almost surely to $P(\boldsymbol{X} = \boldsymbol{x})^{-1}$. On the other, for each $k \in \mathcal{J}$, using conjecture C1, Lemma 6.3.10 and Lemma 6.3.9, we derive the asymptotic distribution of $\sqrt{n}\left(\frac{m_k}{n}\omega_k - P(\Omega_k)P(\boldsymbol{X} = \boldsymbol{x}|\Omega_k)\right)$. So, since the terms in the sum are independent (conjecture C2), the result is derived after summing independent normal distributed random variables and applying the Slutzky's Theorem. □

Theorem 6.3.3 *If $\widetilde{\widetilde{\boldsymbol{p}}}$ denotes the estimator of $\boldsymbol{p}^*$ in (6.30), under the same hypotheses of Theorem 6.3.2, then*

a) $\sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}} - \widetilde{\boldsymbol{p}}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}_2)$, *where $\boldsymbol{\Lambda}_2$ is the $(2K+1)\times(2K+1)$ squared matrix defined as*

$$\text{diag}((p_j^*/P_j)_{j\in\mathcal{J}})\,(\text{diag}(\boldsymbol{P}) - \boldsymbol{P}\cdot\boldsymbol{P}')\,\text{diag}((p_j^*/P_j)_{j\in\mathcal{J}}) + \text{diag}((p_j^{*2})_{j\in\mathcal{J}})\nu^2$$

with $\boldsymbol{P} = (P_j)_{j\in\mathcal{J}} = (P(\Omega_j))_{j\in\mathcal{J}}$ and ν^2 the same as in Lemma 6.3.11 and

b) the asymptotic variance-covariance matrix $\boldsymbol{\Lambda}_2$ can be consistently estimated by replacing $P(\Omega_j)$ by m_j/n, $\boldsymbol{p}^*$ by $\widetilde{\widetilde{\boldsymbol{p}}}$ and $P(\boldsymbol{X} = \boldsymbol{x})$ by $\widetilde{n\boldsymbol{x}}/n$.

Proof:

a) For each $j \in \mathcal{J}$, from the expressions (6.30) and (6.25) we have

$$\begin{aligned}\sqrt{n}(\widetilde{\widetilde{p}}_j - \widetilde{p}_j) &= \sqrt{n}\left(\frac{m_j\omega_j}{\sum_{k\in\mathcal{J}} m_k\omega_k} - \frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}\,\omega_j\right) \\ &= \sqrt{n}\frac{\omega_j}{P(\boldsymbol{X}=\boldsymbol{x})}\left(\frac{P(\boldsymbol{X}=\boldsymbol{x})}{\sum_{k\in\mathcal{J}} m_k\omega_k}\,m_j - P(\Omega_j)\right) \\ &= \frac{\omega_j}{P(\boldsymbol{X}=\boldsymbol{x})}\cdot\sqrt{n}\left(\frac{P(\boldsymbol{X}=\boldsymbol{x})}{\sum_{k\in\mathcal{J}} \frac{m_k}{n}\omega_k}\,\frac{m_j}{n} - P(\Omega_j)\right) \qquad (6.33)\end{aligned}$$

Note that, by Lemma 6.3.6, $\omega_j \xrightarrow{\text{a.s.}} P(\boldsymbol{X}=\boldsymbol{x}|\Omega_j)$ and therefore the first term in the product in (6.33) $\omega_j/P(\boldsymbol{X}=\boldsymbol{x}) \xrightarrow{\text{a.s.}} P(\boldsymbol{X}=\boldsymbol{x}|\Omega_j)/P(\boldsymbol{X}=\boldsymbol{x}) = P(\Omega_j|\boldsymbol{X}=\boldsymbol{x})/P(\Omega_j) = p_j^*/P_j$.

On the other hand, by using the expression of X_n introduced in Lemma 6.3.11, second term in (6.33) can be decomposed as follows:

$$\sqrt{n}\left(X_n\frac{m_j}{n} - P(\Omega_j)\right) = X_n\sqrt{n}\left(\frac{m_j}{n} - P(\Omega_j)\right) - \sqrt{n}(1-X_n)P(\Omega_j). \qquad (6.34)$$

We study now the asymptotic distribution of each term in the right-hand part of (6.34). Firstly, since $X_n \xrightarrow{\text{a.s.}} 1$ and the quantities $(m_j)_{j\in\mathcal{J}}$ follow a multinomial distribution with vector of probabilities $\boldsymbol{P}$, therefore, by the Slutzky's Theorem, first term follows the asymptotic distribution $\mathcal{N}(0, P_j(1-P_j))$. Secondly, since $P(\Omega_j)$ is a constant, by Lemma 6.3.11, part b), and the Slutzky's Theorem, second term converges in distribution to a $\mathcal{N}(0, P(\Omega_j)^2\cdot\nu^2)$ distribution. Since terms in (6.34) are independent (conjecture C3),

$$\sqrt{n}\left(X_n\frac{m_j}{n} - P(\Omega_j)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, P_j(1-P_j) + P_j^2\nu^2).$$

Finally, to derive the asymptotic distribution of (6.33), since $\dfrac{\omega_j}{P(\boldsymbol{X}=\boldsymbol{x})} \xrightarrow{\mathcal{P}} \dfrac{p_j^*}{P_j}$ for each $j \in \mathcal{J}$, we apply again the Slutzky's Theorem and $\sqrt{n}(\widetilde{\widetilde{\boldsymbol{p}}} - \widetilde{\boldsymbol{p}}) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}_2)$ as we wanted to prove.

b) The result follows by replacing all the unknown values in $\mathbf{\Lambda}_2$ by their respective consistent estimators. □

Next lemma proves that the random vectors $\sqrt{n}\left(\widetilde{\boldsymbol{p}}-\boldsymbol{p}^*\right)$ and $\sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}}-\widetilde{\boldsymbol{p}}\right)$ are asymptotically uncorrelated. This result is useful in order to support the conjecture C4 and combine both estimators to derive the asymptotic behavior of $\sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}}-\boldsymbol{p}^*\right)$ in Theorem 6.3.4.

Lemma 6.3.12 *The random vectors $\sqrt{n}\left(\widetilde{\boldsymbol{p}}-\boldsymbol{p}^*\right)$ and $\sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}}-\widetilde{\boldsymbol{p}}\right)$ are asymptotically uncorrelated.*

Proof:

We will prove that $Cov\left(\sqrt{n}\left(\widetilde{p}_j-p_j^*\right),\sqrt{n}\left(\widetilde{\widetilde{p}}_j-\widetilde{p}_j\right)\right)\to 0$ when $n\to\infty$, for each $j\in\mathcal{J}$.

According to the definitions

$$
\begin{aligned}
&Cov\left(\sqrt{n}\left(\widetilde{p}_j-p_j^*\right),\sqrt{n}\left(\widetilde{\widetilde{p}}_j-\widetilde{p}_j\right)\right)=\\
&= E\left\{\sqrt{n}\left(\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}\omega_j-p_j^*\right)\sqrt{n}\left(\widetilde{\widetilde{p}}_j-\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}\omega_j\right)\right\}-\\
&- E\left\{\sqrt{n}\left(\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}\omega_j-p_j^*\right)\right\}E\left\{\sqrt{n}\left(\widetilde{\widetilde{p}}_j-\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}\omega_j\right)\right\}=\\
&= nE\left\{\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}\widetilde{\widetilde{p}}_j\omega_j-\frac{P(\Omega_j)^2}{P(\boldsymbol{X}=\boldsymbol{x})^2}\omega_j^2-\widetilde{\widetilde{p}}_jp_j^*+\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}p_j^*\omega_j\right\}-\\
&- n\left(\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}E(\omega_j)-p_j^*\right)\left(E(\widetilde{\widetilde{p}}_j)-\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}E(\omega_j)\right)=\\
&= n\left(\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}E(\widetilde{\widetilde{p}}_j\omega_j)-\frac{P(\Omega_j)^2}{P(\boldsymbol{X}=\boldsymbol{x})^2}E(\omega_j^2)-p_j^*E(\widetilde{\widetilde{p}}_j)+\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}p_j^*E(\omega_j)\right)-\\
&- n\left(\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}E(\omega_j)E(\widetilde{\widetilde{p}}_j)-\left(\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}E(\omega_j)\right)^2-p_j^*E(\widetilde{\widetilde{p}}_j)+\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}p_j^*E(\omega_j)\right)=\\
&= n\left(\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}Cov(\widetilde{\widetilde{p}}_j,\omega_j)-\frac{P(\Omega_j)^2}{P(\boldsymbol{X}=\boldsymbol{x})^2}Var(\omega_j)\right)\\
&= \frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}\left(nCov(\widetilde{\widetilde{p}}_j,\omega_j)-\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}nVar(\omega_j)\right).
\end{aligned}
$$

By Lemma 6.3.6, the proportion ω_j converges almost surely to the probability $P(\boldsymbol{X} = \boldsymbol{x}|\Omega_j)$ and then, by the central limit theorem, $nVar(\omega_j)$ converges to the finite quantity $P(\boldsymbol{X} = \boldsymbol{x}|\Omega_j)(1 - P(\boldsymbol{X} = \boldsymbol{x}|\Omega_j))$.

On the other hand, by the strong law of large numbers, $\frac{mj}{\sum_{k\in\mathcal{J}} m_k\omega_k}$ converges almost surely to $\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}$ and therefore $nCov(\frac{mj}{\sum_{k\in\mathcal{J}} m_k\omega_k}\omega_j, \omega_j)$ has the same asymptotic behavior as $\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}nCov(\omega_j, \omega_j)$, that is $\frac{P(\Omega_j)}{P(\boldsymbol{X}=\boldsymbol{x})}nVar(\omega_j)$, and the result holds. □

Theorem 6.3.4 *If $\widetilde{\widetilde{\boldsymbol{p}}}$ denotes the estimator of $\boldsymbol{p}^*$ in (6.30), under the same hypotheses of Theorem 6.3.2,*

a) $\sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}} - \boldsymbol{p}^*\right) \xrightarrow{\mathcal{D}} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda})$, *where $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2$ is a $(2K+1) \times (2K+1)$ squared matrix with $\boldsymbol{\Lambda}_1$ the same as in Corollary 6.3.1 and $\boldsymbol{\Lambda}_2$ the same as in Theorem 6.3.3.*

b) the asymptotic variance-covariance matrix $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2$ can be consistently estimated by replacing $P(\Omega_j)$ by m_j/n, $P(\boldsymbol{X} = \boldsymbol{x})$ by $\widetilde{n_{\boldsymbol{x}}}/n$, $\boldsymbol{p}^*$ by $\widetilde{\widetilde{\boldsymbol{p}}}$, $\boldsymbol{\alpha}^*$ by $\widehat{\boldsymbol{\alpha}}$, $\boldsymbol{D}$ by the identity matrix, $\phi^{(1)}$ by $\boldsymbol{0}$ in the evaluations of $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ according to parts b) in Corollary 6.3.1 and Theorem 6.3.3.

Proof:

a) First of all, note that $\sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}} - \boldsymbol{p}^*\right)$ can be written as

$$\sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}} - \boldsymbol{p}^*\right) = \sqrt{n}\left(\widetilde{\boldsymbol{p}} - \boldsymbol{p}^*\right) + \sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}} - \widetilde{\boldsymbol{p}}\right).$$

By Corollary 6.3.1, first term in the sum converges in distribution to a random vector, let denote $\boldsymbol{X}_1$, $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}_1)$ distributed and, by Theorem 6.3.3, the last term converges in distribution to a random vector, let denote $\boldsymbol{X}_2$, $\mathcal{N}(\boldsymbol{0}, \boldsymbol{\Lambda}_2)$ distributed.

Since $\sqrt{n}\left(\widetilde{\boldsymbol{p}} - \boldsymbol{p}^*\right)$ and $\sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}} - \widetilde{\boldsymbol{p}}\right)$ are supposed to be independent (conjecture C4), the vector $\left(\sqrt{n}\left(\widetilde{\boldsymbol{p}} - \boldsymbol{p}^*\right), \sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}} - \widetilde{\boldsymbol{p}}\right)\right)$, by Theorem 7.20 at page 400 in Shervish (1995), converges in distribution to $(\boldsymbol{X}_1, \boldsymbol{X}_2)$ and $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent. So, by the continuity theorem applied to the sum operator,

$$\sqrt{n}\left(\widetilde{\boldsymbol{p}} - \boldsymbol{p}^*\right) + \sqrt{n}\left(\widetilde{\widetilde{\boldsymbol{p}}} - \widetilde{\boldsymbol{p}}\right) \xrightarrow{\mathcal{D}} \boldsymbol{X}_1 + \boldsymbol{X}_2.$$

And, since $\boldsymbol{X}_1 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_1)$ and $\boldsymbol{X}_2 \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_2)$ and they are independent, the limit random variable $\boldsymbol{X}_1 + \boldsymbol{X}_2$ follows a $\mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2)$ distribution.

b) It straightforwardly follows from parts b) in Corollary 6.3.1 and Theorem 6.3.3.

□

6.3.4 Asymptotic properties of $\widetilde{\boldsymbol{S}_{\boldsymbol{x}}}$

From the asymptotic properties of $\widetilde{\widetilde{\boldsymbol{p}}}$ described just above, we can obtain the asymptotic behavior of the estimator $\widetilde{\boldsymbol{S}}$ defined in (6.17) from the $\widetilde{\widetilde{\boldsymbol{p}}}$ estimator. In the next theorem we derive the asymptotic distribution of $\widetilde{\boldsymbol{S}}$ respect to the true vector $\mathcal{KM}(\boldsymbol{p}^*)$.

Theorem 6.3.5 *If $\widetilde{\widetilde{\mathbf{p}}}$ is the estimator proposed in (6.30) for the parameter $\mathbf{p}$, and the true value is $\mathbf{p}^*$, under the same hypotheses of Theorem 6.3.4 then*

a)

$$\sqrt{n}\left(\widetilde{\boldsymbol{S}} - \mathcal{KM}(\boldsymbol{p}^*)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, (\mathcal{STG})\boldsymbol{\Lambda}(\mathcal{STG})'\right), \tag{6.35}$$

being $\mathcal{S}$ the $K \times K$ diagonal matrix $\operatorname{diag}(\mathcal{KM}(\boldsymbol{p}^*))$, $\mathcal{T}$ and $\mathcal{G}$ as in (6.9), $\mathcal{G}$ is evaluated in $\boldsymbol{p} = \boldsymbol{p}^*$ and $\boldsymbol{\Lambda}$ is the same as in Theorem 6.3.3.

b) the asymptotic variance-covariance matrix $(\mathcal{STG})\boldsymbol{\Lambda}(\mathcal{STG})'$ can be consistently estimated by replacing $P(\Omega_j)$ by m_j/n, $P(\boldsymbol{X} = \boldsymbol{x})$ by $\widetilde{n_{\boldsymbol{x}}}/n$, $\mathbf{p}^*$ by $\widetilde{\widetilde{\mathbf{p}}}$, $\boldsymbol{\alpha}^*$ by $\widehat{\boldsymbol{\alpha}}$, $\boldsymbol{D}$ by the identity matrix, $\phi^{(1)}$ by $\mathbf{0}$ in the evaluation of the matrices $\mathcal{S}$, $\mathcal{G}$ and $\boldsymbol{\Lambda}$.

Proof:

a) Without loss of generality we can assume that $p^*_{K+1} > 0$. In Theorem 6.3.4 we proved that $\widetilde{\widetilde{\boldsymbol{p}}}$ is a $\sqrt{n}$-consistent and asymptotically normal estimator of $\boldsymbol{p}^*$ with asymptotic variance-covariance matrix $\boldsymbol{\Lambda}$. So, since by definition $\widetilde{\boldsymbol{S}} = \mathcal{KM}(\widetilde{\widetilde{\boldsymbol{p}}})$, if we apply the δ-method to the $\mathcal{KM}$ map, analogously to Theorem 6.2.2, we conclude the proposed result.

b) The result follows by replacing all the parameters in (6.35) by their respective consistent estimators. □

6.4 Returning to the HIV+PTB cohort example

In this section we apply the semiparametric approach to the estimation of the univariate stratified survival for the HIV+PTB cohort introduced in Chapter 2, according to our covariates of interest $CD4$ and PPD. In particular, we consider that our data are $(Y, \delta, CD4, PPD)$ where Y is the survival time between the beginning of TB treatment and death, δ is the censoring indicator, $CD4$ is the immunosuppression level (0 = high, 1 = low) and PPD is the result to the tuberculin skin test (0 = negative, 1 = positive). Suppose we are interested in comparing, between categories, the survivorship after 1 year from the beginning of the TB treatment .

6.4.1 Design of the estimation

As we saw in Chapter 2, the observation window is 3 years, the sample size is $n = 494$, the proportion of censoring is 63.8% and the proportion of missing in the covariates $CD4$ and PPD is 38.9% and 50.4%, respectively.

In order to choose an appropriated τ_k partition, we show in Figure 6.1 the corresponding histograms for the survival times of individuals with observed covariate and unobserved covariate for the natural partitions in months and in weeks. The figure is for the $CD4$ covariate, but similar graphics can be obtained for the covariate PPD. As we can see, when the grid is in months the effective sample size is quite all the sample size while, when the grid is in weeks, there is a small reduction because there are few weeks where we only observe survival times with unobserved covariate. Indeed, the effective sample size is 490 and 434, respectively, for the covariate $CD4$, and 479 and 420 for the covariate PPD. On the other hand, in Chapter 3 we have shown that the $CD4$ covariate (in its continuous form as well as categorized) is the best prognosis factor for survival. In other words, there is dependence between the $CD4$ changes and survival time. So, in order to get less biased and more precise estimates, we decide to perform the analysis using a grid in weeks.

We specify a model for the non-response probabilities depending on the survival time, Y, the censoring indicator, δ, and the covariate of interest, X, as

$$\text{logit}\ (P(R = 1|Y, \delta, X)) = \alpha_0 + \alpha_1 \cdot Y + \alpha_2 \cdot \delta + \tau \cdot X \qquad (6.36)$$

where the covariate X is $CD4$ or PPD.

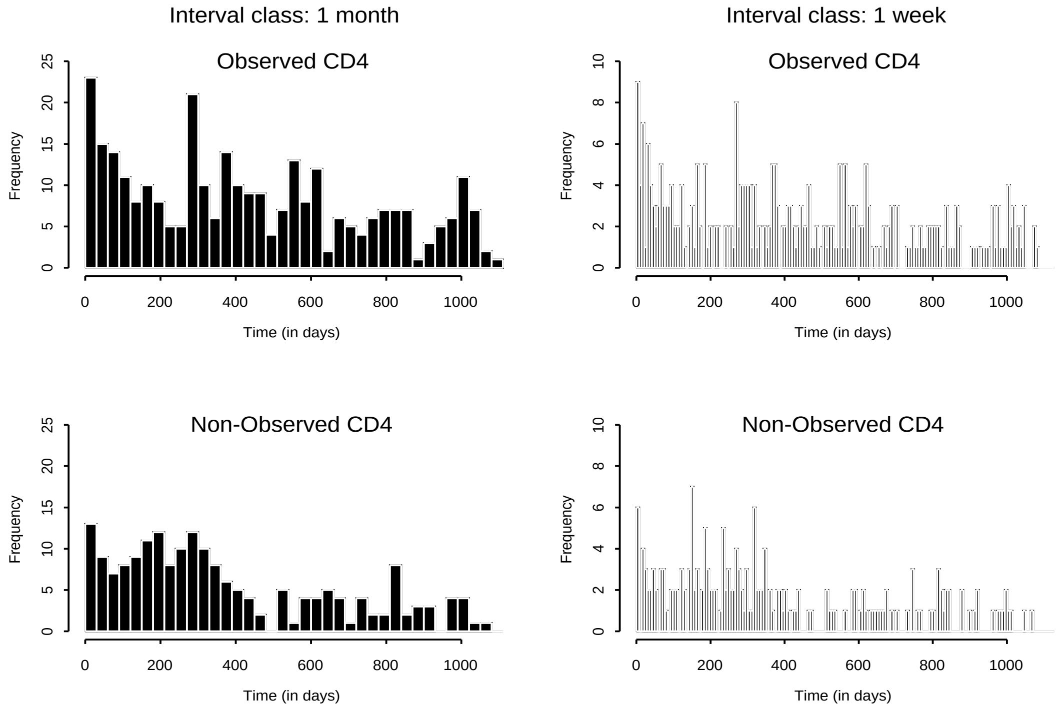


Figure 6.1: *Histograms of the survival times whether the CD4 covariate has been observed or not, and the interval class is in months or weeks*

The model (6.36) allows us to setup the non-response probabilities as MCAR (if $\alpha_1 = \alpha_2 = \tau = 0$), as MAR (if $\tau = 0$) and, otherwise, as NI. The non-ignorability parameter τ is the log odds-ratio of being observed in category $X = 1$ versus being observed in category $X = 0$, *i.e.*,

$$\tau = \log \frac{P(R = 1|X = 1)/P(R = 0|X = 1)}{P(R = 1|X = 0)/P(R = 0|X = 0)}.$$

If we denote by p_i the probability of being observed in category i ($i = 0, 1$), $p_i = P(R = 1|X = i)$, we derive

$$p_1 = \frac{\exp(\tau) \cdot p_0}{1 - (1 - \exp(\tau)) \cdot p_0}. \tag{6.37}$$

Figure 6.2 shows the contour lines for p_1 as a function of p_0 and τ. Each curve corresponds to a fixed value for the probability p_1 and illustrates all the pairs (p_0, τ) that verify (6.37). This graph helps us to determine a range of plausible values for

τ, based on the information on the p_i's probabilities. For example, as it is illustrated in the figure, if $0.4 \leq p_0 \leq 0.6$ and $0.7 \leq p_1 \leq 0.9$ then an interval for τ values is $[0.44, 2.60]$. In our analysis, in order to cover a big range of possibilities we consider $\tau \in [-6, 6]$ (in other words, we are allowing an odds-ratio between 0.002 and 403).

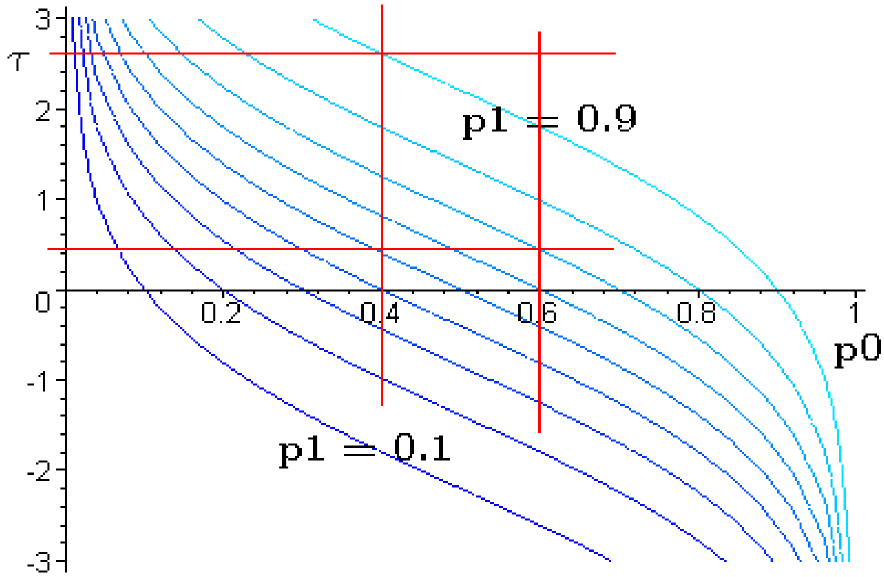


Figure 6.2: *Contour lines for* $p_1 = 0.1, \dots, 0.9$ *as a function of* p_0 *and* τ

To estimate semiparametrically the non-response parameter $\boldsymbol{\alpha}^*$, using the estimating equations (6.23), we setup $\phi^{(2)}_{r=0}(Y, \delta) = (1, Y, Y^2)'$.

6.4.2 Sensitivity analysis

Following the semiparametric methodology introduced in Section 6.3, and the considerations in Section 5.6 about the convenience of performing a sensitivity analysis, for each $\tau \in [-6, 6]$ we can estimate the vector $(\boldsymbol{p}^{*\prime}, \boldsymbol{\alpha}^{*\prime})'$ and the corresponding estimates for the survival at 1 year and its standard error. Table 6.4 displays these estimates for the integer values of τ in $[-6, 6]$, and for the two covariates of interest. Results corresponding to the complete case analysis are also displayed. A similar table can be obtained for grid in months.

In the scope of the sensitivity analysis we plot in Figures 6.3 and 6.4 the estimates and the corresponding confidence bands for the survival at 1 year for each category

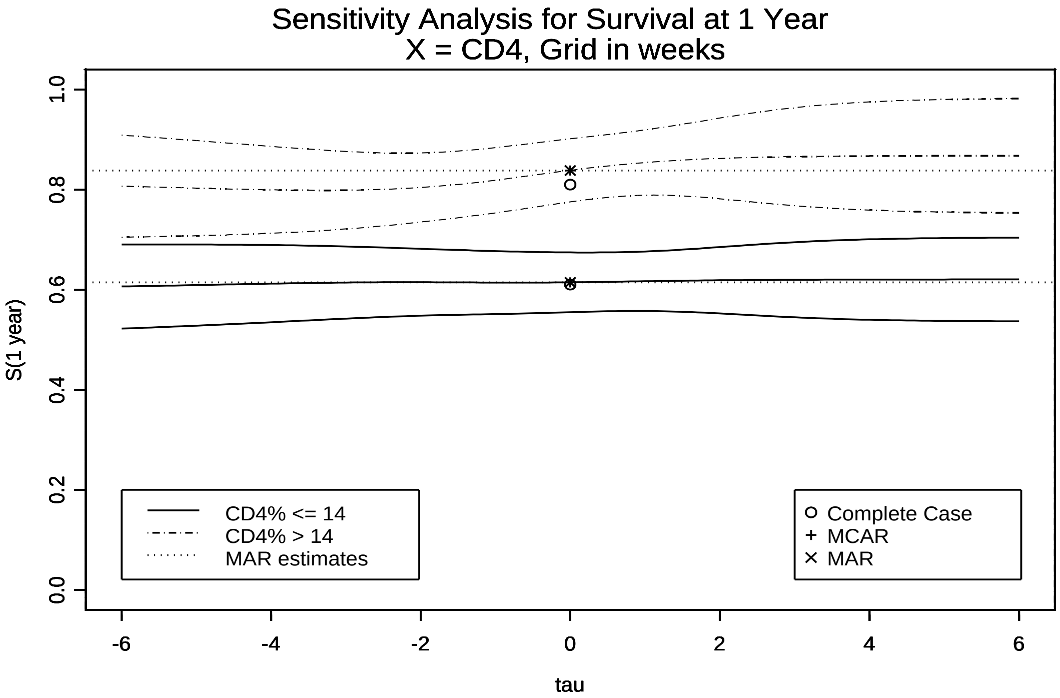
	$X = CD4$		$X = PPD$	
Analysis	$\widetilde{S_{X=0}}$	$\widetilde{S_{X=1}}$	$\widetilde{S_{X=0}}$	$\widetilde{S_{X=1}}$
CC	0.610 (0.037)	0.810 (0.040)	0.657 (0.042)	0.877 (0.037)
MCAR	0.615 (0.030)	0.838 (0.032)	0.602 (0.031)	0.873 (0.034)
$\tau = -6$	0.606 (0.043)	0.807 (0.052)	0.591 (0.531)	0.845 (0.995)
$\tau = -5$	0.609 (0.041)	0.803 (0.048)	0.593 (0.424)	0.842 (0.779)
$\tau = -4$	0.612 (0.039)	0.799 (0.044)	0.595 (0.273)	0.839 (0.493)
$\tau = -3$	0.614 (0.037)	0.799 (0.040)	0.597 (0.133)	0.839 (0.233)
$\tau = -2$	0.615 (0.034)	0.804 (0.035)	0.599 (0.055)	0.842 (0.086)
$\tau = -1$	0.614 (0.032)	0.818 (0.033)	0.601 (0.034)	0.853 (0.039)
MAR $\tau = 0$	0.615 (0.030)	0.838 (0.032)	0.602 (0.031)	0.873 (0.034)
$\tau = 1$	0.617 (0.030)	0.854 (0.033)	0.604 (0.030)	0.891 (0.034)
$\tau = 2$	0.619 (0.034)	0.862 (0.041)	0.606 (0.031)	0.903 (0.035)
$\tau = 3$	0.620 (0.038)	0.866 (0.050)	0.607 (0.031)	0.908 (0.036)
$\tau = 4$	0.620 (0.041)	0.867 (0.055)	0.607 (0.031)	0.910 (0.037)
$\tau = 5$	0.620 (0.042)	0.867 (0.057)	0.607 (0.031)	0.911 (0.037)
$\tau = 6$	0.620 (0.043)	0.868 (0.058)	0.607 (0.031)	0.912 (0.037)

Table 6.4: *Estimates for the survival at 1 year for categories in $CD4$ and PPD covariates (standard error, in parentheses) resulting from the complete case analysis and the semiparametric methodology for different values of τ and grid in weeks*

in the covariates $CD4$ and PPD. In Figure 6.3 we use a τ_k partition in weeks, and in Figure 6.4 the grid is in months.

Looking at the top figure in Figure 6.3 we can see that, no matter which type of assumptions we make about τ in the non-response mechanism specification in (6.36) for the $CD4$ covariate,

a) the resulting estimates in everyone of the categories are essentially the same (mainly in group $CD4\% \leq 14$) and we can see that the semiparametric analysis



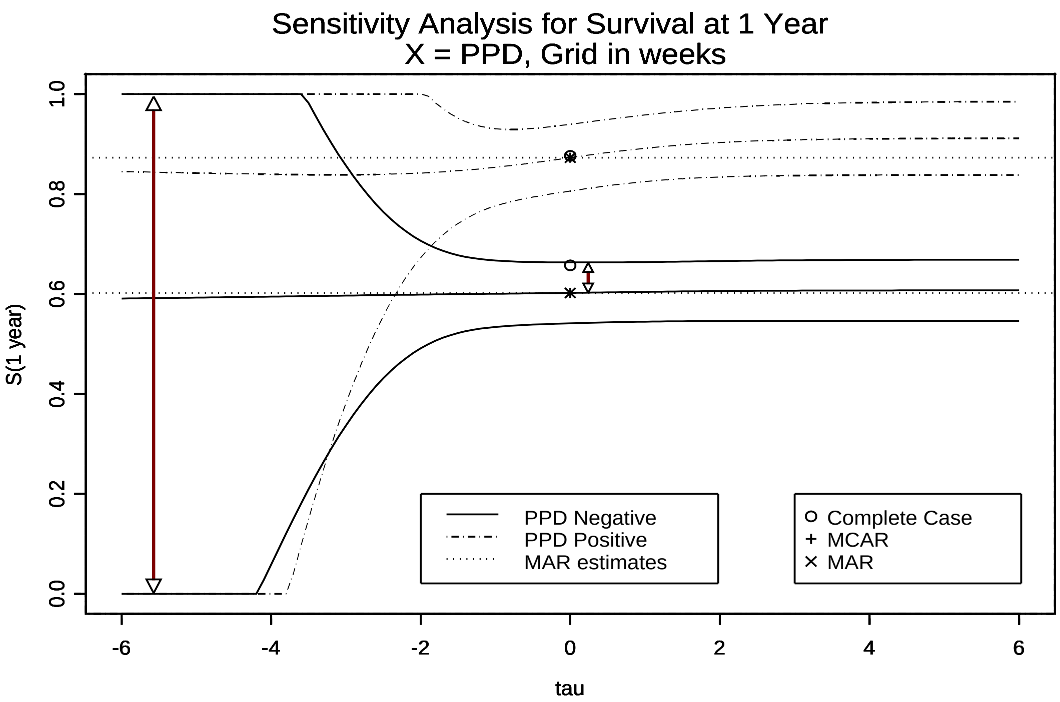


Figure 6.3: *Estimates and 95% confidence bands for the stratified survival at 1 year, for the covariates CD4 and PPD, as a function of the non-ignorability parameter τ and when the grid is in weeks*

do not differ too much from the CC analysis. This fact suggests that the missingness in the $CD4$ covariate could be ignorable,

b) the survival for the less immunosuppression group (*i.e.,* $CD4 > 14$) is better, in general. However, if the less immunosuppression group would be heavily underrepresented in the subsample with observed $CD4$ covariate (*e.g.,* $\tau \ll -6$), we could not infer differences between both survivals.

About the PPD covariate (Figure 6.3, bottom) we observe

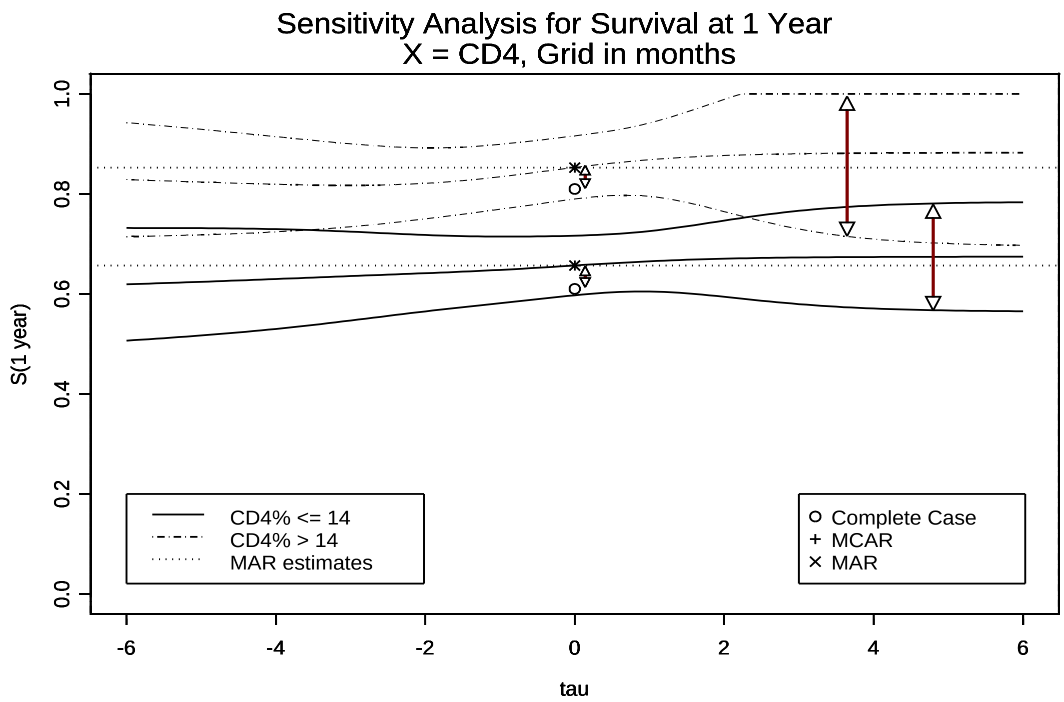
a) for individuals with negative PPD, the complete case methodology provides a positive biased estimate. Looking at Table 6.4, all the semiparametric estimates are closed to 60%, while the CC analysis provides a 65.7% estimate,

b) when τ is lower than -2 the estimates for the estandard error quickly increase and the confidence bands become $[0, 1]$.

c) with this dataset, we can only infer that the positivity in the tuberculin skin test implies a better survival if and only if we can suppose that individuals with positive PPD are represented enough in the subsample with observed PPD covariate (*e.g.,* $\tau > -1$).

As we said at the beginning of this section, the choice of the τ_k partition is actually decisive. In Figure 6.4 where the grid is setup in months we can observe that all the estimates have a bigger positive bias and a bigger standard error. In essence, in such situation the differences between groups are no so evident.

On the other hand, if we compare Figure 6.3 and Figure 6.4 for the $CD4$ covariate, we can see another side effect of taking the grid in months: the increase of the standard error estimates when τ is far away from 0. This is due to the fact that when the grid is in months we are mixing in the same estimating equation (6.23) individuals potentially belonging to different categories, because the $CD4$ percentage could be actually different from the beginning to the end of the month. In some sense, we are not using the information coming from the relationship between $CD4$ and (Y, δ) and, consequently, the standard error will be bigger. In particular because the weights $\pi_i(1; \boldsymbol{\alpha})^{-1}$ are bigger (for the category $CD4 = 1$ when $\tau \ll 0$ and for the category $CD4 = 0$ when $\tau \gg 0$) and then $\boldsymbol{\Lambda}$ in Theorem 6.3.4 is also bigger. This



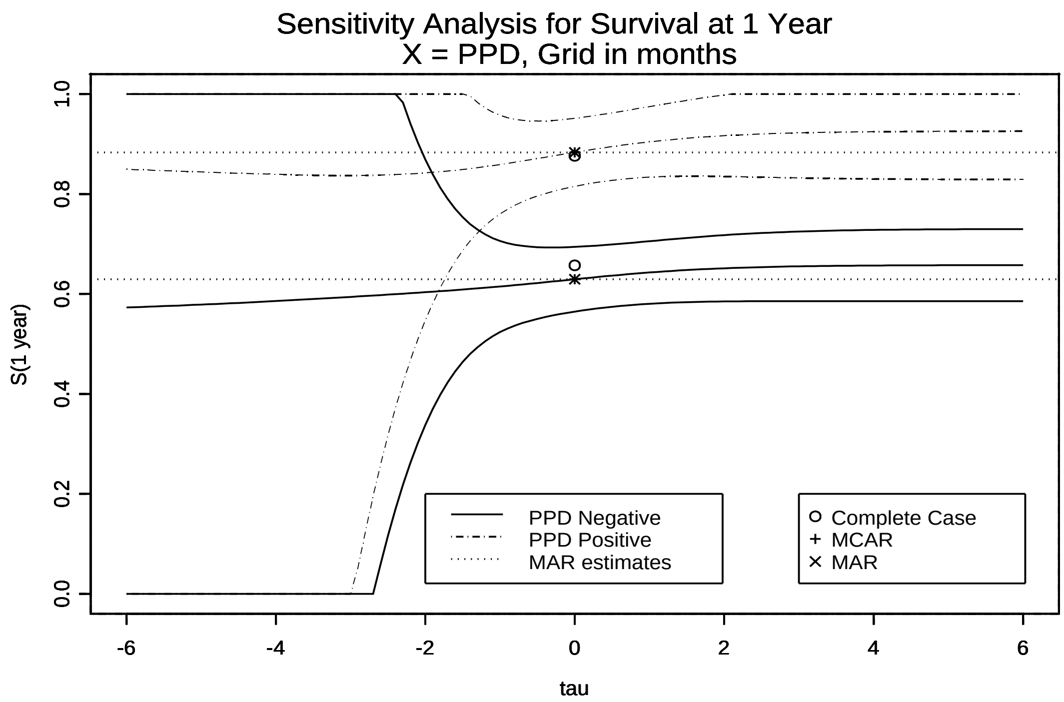


Figure 6.4: *Estimates and 95% confidence bands for the stratified survival at 1 year, for the covariates CD4 and PPD, as a function of the non-ignorability parameter* τ *and when the grid is in months*

remark is specially important to be taken into consideration when the covariate of interest, even being measured at the beginning of the study, is time dependent.

Another strategy for conducting a global sensitivity analysis, in a graphical way, is to plot in the same graph different estimated survival functions for different values of the non-ignorability parameter. Figure 6.5 shows this type of sensitivity analysis for different strategies of analyzing the data, when the grid considered is in weeks. The complete case analysis is also shown in the picture for the purpose of comparison. As in Figure 6.3 and Figure 6.4, the upper part of the figure corresponds to the $CD4$ covariate and the bottom part to the PPD covariate. The proposed semiparametric estimator is illustrated when we consider the MAR non-response pattern and the NI non-response pattern with $\tau = -2$ or $\tau = 2$.

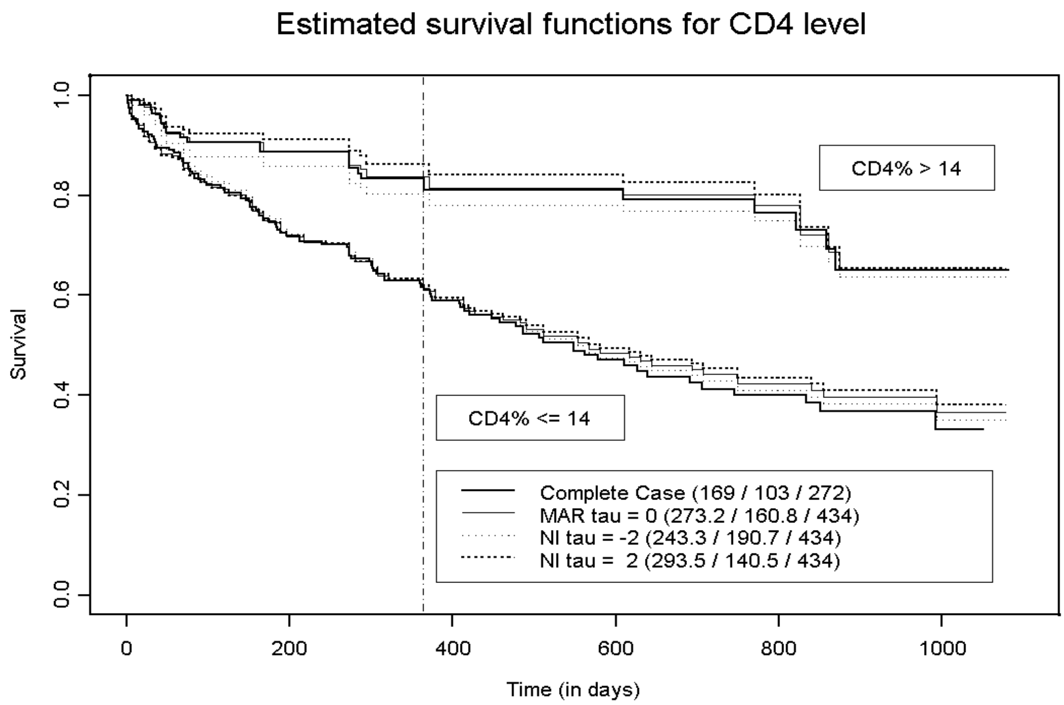
Aside from the mentioned conclusions in the previous part of this section, corresponding to the survival at 1 year, for the global survival function we can add:

a) for the most immunosuppression group, there is no differences between the estimates provided by the CC analysis and those resulting from the proposed methodology,

b) for the PPD negative group, the CC methodology is overestimating the survival function (in average approximatively 8%) with respect to the proposed methodology,

c) for other categories ($CD4\% > 14$ or PPD positive) the estimated survival depends on the assumptions on the non-response pattern.

For other values of τ we obtain similar results and the same conclusions. It is also interesting to observe that when the grid is in months many of the above conclusions cannot be derived due to the larger bias and the less accuracy of the resulting estimators. However, in all the situations it is obvious the interest of considering the role that the individuals with missing data play in order to avoid wrong inferences.



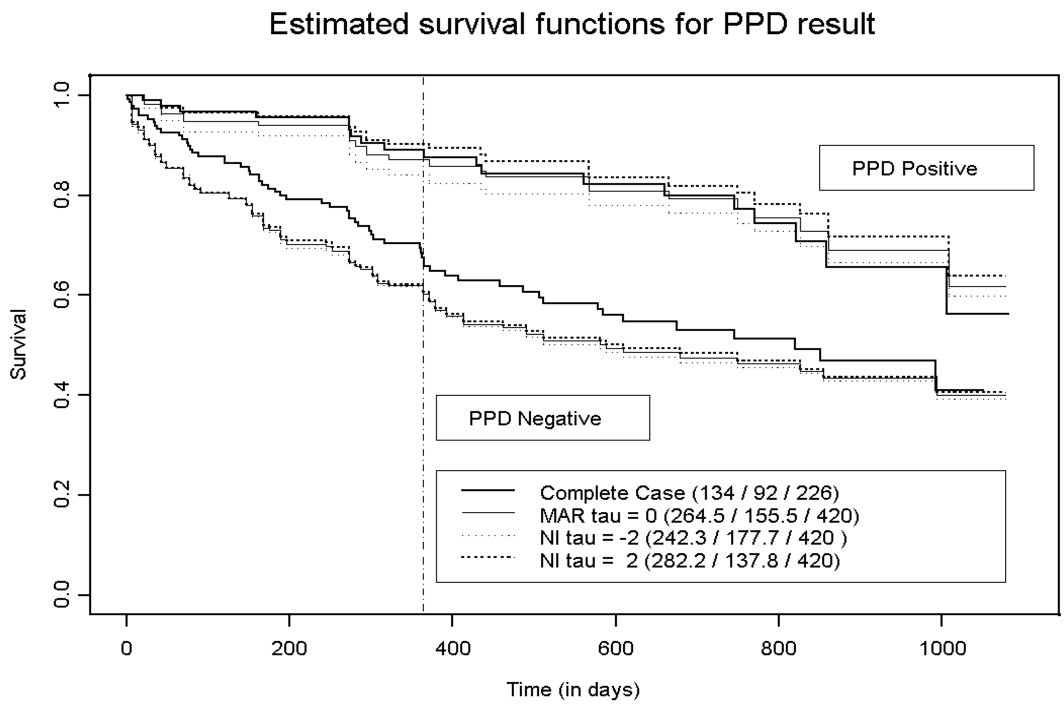


Figure 6.5: *Estimated survival functions for the covariates CD4 and PPD for four different analyzing strategies: complete case, MAR and non-ignorable with* $\tau = -2$ *and* $\tau = 2$*. The grid for the semiparametric approach is setup in weeks. Vertical line corresponds to 365 days. In parentheses, the estimated number of individuals in each category and the effective sample size*