ate sized samples if not oversmoothed, but may be biased if oversmoothed. Thus $\hat{\eta}(\alpha)$ is only acceptable for large sample sizes. Obviously, this implies that the maximum likelihood estimator of α or any approximation thereof also suffers heavily from this curse of dimensionality. □

Example 1.11 (Repeated measures data with right-censoring; continuation of Example 1.2) Under CAR, the F_X part of the likelihood of the observed data $Y_1, \ldots, Y_n$ is given by

$$L(Y_1, \ldots, Y_n \mid F_X) = \prod_{i=1}^{n} dF_X(\bar{X}_i(c))\big|_{c=C_i},$$

where $dF_X(\bar{X}(c))$ represents the density of the sample path $\bar{X}(c)$ under F_X w.r.t. some dominating measure (discrete or continuous). Again, we note that the full data model only specifies a mean of one of the components of F_X so that maximum likelihood estimation will perform miserably at finite sample sizes and may even be inconsistent. □

The fact that the maximum likelihood estimator, or more generally any globally efficient estimator, has a bad practical performance does not exclude the presence of other inefficient but practical estimators. In fact, in the *full data* multivariate generalized regression models (repeated measures) of Examples 1.1 and 1.2, we have already seen that no globally efficient practical estimators of μ exist, but nice locally efficient estimators are available. However, Lemma 1.1 teaches us that if F_X is completely unspecified, then all *regular* asymptotically linear estimators of the parameter μ in the observed data model $\mathcal{M}(CAR)$ are asymptotically equivalent and efficient. Thus, in this nonparametric coarsening at random model for Y, one has no other choice than to construct globally efficient estimators, such as the nonparametric maximum likelihood estimator. From a practical point of view, the lesson is that if the maximum likelihood estimator of F_X in the observed data model $\mathcal{M}(CAR)$ based on $Y_1, \ldots, Y_n$ has a bad practical performance, then there will not exist regular asymptotically linear estimators with good practical performance in $\mathcal{M}(CAR)$.

To further understand the difficulty in estimating μ in model $\mathcal{M}(CAR)$, recall that the observed data nuisance tangent space $T_{nuis}(P_{F_X,G})$ for the parameter of interest μ is the closure of the linear span of all scores of one-dimensional submodels $\epsilon \to P_{F_\epsilon,G_\epsilon}$ for which $d/d\epsilon\mu(P_{F_\epsilon,G_\epsilon})|_{\epsilon=0} = 0$. The observed nuisance tangent space is a sub-Hilbert space of $L_0^2(P_{F_X,G}) = \{h(Y) : E_{P_{F_X,G}} h^2(Y) < \infty, E_{P_{F_X,G}} h(Y) = 0\}$ endowed with inner product $\langle h, g\rangle_{P_{F_X,G}} = E_{P_{F_X,G}}[h(Y)g(Y)]$. We also recall that an estimator μ_n is called asymptotically linear at $P_{F_X,G}$ with influence curve $IC(Y)$ if

$$\mu_n - \mu = \frac{1}{n}\sum_{i=1}^{n} IC(Y_i) + o_P(1/\sqrt{n}) \tag{1.22}$$

and that the components of the influence function $IC(Y)$ of any regular asymptotically linear estimators of μ must lie in the orthocomplement $T_{nuis}^{\perp}(P_{F_X,G})$ of the observed data nuisance tangent space $T_{nuis}(P_{F_X,G})$. Our next goal is to try to understand why in the models of the previous examples it is not possible to obtain an estimator μ_n satisfying the expansion above for any element $IC(Y)$ in the orthogonal complement $T_{nuis}^{\perp}$ in the absence of smoothness assumptions. To do so, we must first determine the form of $T_{nuis}^{\perp}$.

Consider a full data structure model $\mathcal{M}^F$ and associated observed data model $\mathcal{M}(CAR)$ in which G is assumed to satisfy CAR but is otherwise unrestricted (i.e., $G \in \mathcal{G}(CAR)$). Our general representation theorem (Theorem 1.3) at the end of this chapter, first established in Robins and Rotnitzky (1992), represents the orthogonal complement $T_{nuis}^{\perp} = T_{nuis}^{\perp}(P_{F_X,G})$ of the nuisance tangent space T_{nuis} in the observed data model $\mathcal{M}(CAR)$ at $P_{F_X,G}$ as the range of a mapping $D \to IC_0(D) - \Pi(IC_0(D) \mid T_{CAR})$, where the initial mapping $D \to IC_0(D)$ satisfies $E(IC_0(D)(Y) \mid X) = D(X)$ F_X-a.e., applied to the orthogonal complement $T_{nuis}^{F,\perp}(F_X)$ of the nuisance tangent space in the full data model $\mathcal{M}^F$. The mapping is defined as an initial mapping (typically an inverse probability of censoring weighted mapping) minus a projection of this initial mapping on the tangent space $T_{CAR} = T_{CAR}(P_{F_X,G})$ for G in model $\mathcal{M}(CAR)$ at $P_{F_X,G}$. $T_{CAR}(P_{F_X,G})$ consists of all functions of the observed data that have mean zero given the full data, namely

$$T_{CAR}(P_{F_X,G}) = \{V(Y) \in L_0^2(P_{F_X,G}) : E_G(V(Y) \mid X) = 0\} \tag{1.23}$$

To understand why this space should be T_{CAR}, note that any parametric submodel $f(Y|X;\omega) = m(Y;\omega)$ of $\mathcal{G}(CAR)$ with true value $\omega = 0$ must have a score $\partial \log m(Y;\omega)/\partial\omega$ at $\omega = 0$ that is a function only of Y and has conditional mean zero, given X. By choosing $f(Y|X;\omega) = (1 + \omega V(Y))\, g(Y \mid X)$ for bounded $V(Y)$ satisfying $E_G(V(Y) \mid X) = 0$ and then taking the closure in $L_0^2(P_{F_X,G})$, we obtain the set $T_{CAR}(P_{F_X,G})$. Note that $T_{CAR}(P_{F_X,G})$ depends on F_X as well as on G because whether $V(Y)$ has a finite variance (and thus belongs to $L_0^2(P_{F_X,G})$) depends on F_X.

Below, using this general representation of $T_{nuis}^{\perp}(P_{F_X,G})$, we will determine the orthogonal complement of the nuisance tangent space (or equivalently all influence curves/functions of regular asymptotically linear estimators) in model $\mathcal{M}(CAR)$ for our examples. Subsequently, we will note that any function of Y and the empirical distribution P_n (say $IC_{FC}(Y \mid P_n)$) for which $IC_{FC}(Y \mid P)$ would equal an influence function $IC(Y \mid P)$ at P for each $P \in \mathcal{M}(CAR)$ is a highly non smooth function of the empirical distribution P_n, because $IC(Y \mid P)$ depends on conditional expectations given high dimensional continuous covariates. Such a function $IC_{FC}(Y \mid P_n)$ is said to be Fisher consistent for $IC(Y \mid P)$. Clearly $IC_{FC}(\cdot \mid P_n)$ will fail to be consistent in $L_2(P_{F_X,G})$ for the function