(based on observing $X_1, \ldots, X_n$) does not imply their existence in the observed data model $\mathcal{M}(CAR) = \{P_{F_X,G} : F_X \in \mathcal{M}^F, G \in \mathcal{G}(CAR)\}$ (based on observing $Y_1, \ldots, Y_n$), thereby creating the need for further modeling assumptions on G or F_X; that is, when X is high-dimensional and the sample size is moderate, there may be no estimator of μ that has, under all laws allowed by model $\mathcal{M}(CAR)$, an approximately normal sampling distribution centered near μ with variance small enough to be of substantive interest. Further, if we adopt a CODA asymptotics that imposes no smoothness, then there will generally exist (i) no uniformly consistent estimator of μ , (ii) no estimator of μ that attains a pointwise (i.e., non-uniform) rate of convergence of n^α under all laws allowed by the model for any $\alpha > 0$; and (iii) no "valid" $1-\alpha$ interval estimator for μ exists. By valid we mean that, under all laws, the coverage is at least $(1-\alpha)$ at each sample size n and the length goes to zero in probability with increasing sample size. This reflects the fact that, in order to construct a uniformly consistent estimator of μ under model $\mathcal{M}(CAR)$, it is necessary to use multivariate nonparametric smoothing techniques to estimate conditional means or densities given a high-dimensional covariate, which would require impractically large samples when X is high-dimensional.

Practical estimators (say, $\mu_n = \Phi(P_n)$ for some ϕ) are typically reasonably smooth functionals of the empirical distribution P_n so that its first-order linear approximation (i.e., the functional derivative $d\Phi(P_n - P)$ applied to $(P_n - P)$, which is the empirical mean in (1.22) of its influence function; see Gill, 1989) is representative of its finite sample behavior. Informally, one might coin in the phrase "an estimator suffers from the curse of dimensionality" if it is a highly non smooth functional of the empirical distribution P_n so that the second-order terms in (1.22) heavily influence its finite sample behavior. Variance of an estimator and smoothness of the estimator as a functional of the empirical distribution P_n (measured by the size of its second order terms) are typically tradeoffs, so that it is no surprise that in many models the unregularized nonparametric maximum likelihood estimator suffers from the curse of dimensionality (i.e., large second order terms) while many practical good estimators are available (as in our full data repeated measures examples above). The following examples illustrate this type of failure of maximum likelihood estimation nicely.

Example 1.9 (Right censored data; continuation of Example 1.8) Let $\mu = F_T(t) = P(T \le t)$ be the parameter of interest. If censoring is absent, then we would estimate μ with the empirical cumulative distribution function of $T_1, \ldots, T_n$. If censoring is independent (i.e., $g(c \mid W) = g(c)$), then we could estimate μ with the Kaplan–Meier estimator (Kaplan and Meier, 1958; Wellner, 1982; Gill, 1983), which is inefficient since it ignores the covariates W. In general, the Kaplan-Meier estimator is an inconsistent estimator under the sole assumption (1.19). The F_X part of the likelihood

of Y under CAR is given by

$$L(Y \mid F_X) = dF_{T|W}(\tilde{T} \mid W)^{\Delta}(1 - F_{T|W}(\tilde{T} \mid W))^{1-\Delta} dF_W(W).$$

Let $L(Y_1, \ldots, Y_n \mid F_X) = \prod_{i=1}^n L(Y_i \mid F_X)$ be the likelihood of an i.i.d. sample $Y_1, \ldots, Y_n$. The maximum likelihood estimator of $F_{T|W=W_i}$ is given by the Kaplan–Meier estimator based on the subsample $\{Y_j : W_j = W_i\}$, $i = 1, \ldots, n$. The maximum likelihood estimator of F_W is the empirical distribution function that puts mass $1/n$ on each observation W_i, $i = 1, \ldots, n$. If W is continuous, then each subsample only consists of one observation so that, if $\Delta_i = 1$, then the Kaplan-Meier estimator of $F_{T|W=W_i}$ puts mass 1 on $\tilde{T}_i$, and if $\Delta_i = 0$, then it puts mass zero on $[0, \tilde{T}_i]$ and is undefined on $(\tilde{T}_i, \infty)$. It follows that the MLE results in an inconsistent estimator of $F_T(t)$. Thus, if W is continuous, then the curse of dimensionality causes the MLE to be inconsistent.

Suppose that each of the 25 components of W is discrete with 20 possible outcomes. Then, the outcome space of W has 20^{25} values w_j. In that case, the maximum likelihood estimator of $F_{T|W}(\cdot \mid w_j)$ is *asymptotically* consistent and normally distributed so that the NPMLE of $F_T(t)$ is also asymptotically consistent and normally distributed. However, one needs on the order of 20^{25} observations to have Kaplan–Meier estimator of $F_{T|W}(\cdot \mid W = w_j)$ be well defined with high probability. Therefore, one needs a sample size on the order of 20^{25} observations in order for the MLE of $F_T(t)$ to have a reasonable practical performance. In this case, we conclude that the curse of dimensionality does not cause inconsistency of the MLE but causes a miserable finite sample performance for any practical sample size. □

Example 1.10 (Repeated measures data with missing covariate; continuation of Example 1.1) Under CAR, the F_X part of the likelihood of the observed data $Y_1, \ldots, Y_n$ is given by

$$L(Y_1, \ldots, Y_n \mid F_X) = \prod_{i=1}^n dF_X(X_i)^{\Delta_i} dF_W(W_i)^{1-\Delta_i}.$$

This F_X-part of the likelihood of the observed data $Y_1, \ldots, Y_n$ can be parametrized by α and a nuisance parameter η that includes the unspecified conditional error distribution of ϵ, given X^*, the unspecified conditional distribution of extraneous surrogates E^* given Z, X^*, and the unspecified marginal distribution of X^*. A maximum likelihood estimator of α is defined as the maximizer of the profile likelihood for α (i.e., the likelihood with the nuisance parameter replaced by $\widehat{\eta}(\alpha)$, where $\widehat{\eta}(\alpha)$ is the maximum likelihood estimator w.r.t. η for given α). Thus $\widehat{\eta}(\alpha)$ involves maximizing a likelihood w.r.t. the high dimensional nuisance parameter η. As a consequence, the maximum likelihood estimator $\widehat{\eta}(\alpha)$, or any approximation or regularization of this maximum likelihood estimator, such as a penalized or sieve maximum likelihood estimator, is either extremely variable in moder-