additional non identifiable restriction $g(c_1 \mid X) = g(c_2 \mid X)$ for $c_1 > c_2 > T$ will hold. If we redefine the variable C to be infinity when $T < C$, then C is always observed, the conditional law of C given $C > T$ is a point mass at ∞, and CAR becomes $g(C \mid X) = h(Y)$. Whether or not C has been redefined, CAR is equivalent to the assumption that the cause-specific conditional hazard of C given X only depends on W; that is,

$$\lambda_C(t \mid X) \equiv \lambda_C(t \mid W),\ 0 < t < \infty, \tag{1.20}$$

where $\lambda_C(t \mid \cdot) = \lim_{h \to 0} P(t + h < C \leq t \mid X, C \geq t, T \geq t)/h$ for C a continuous random variable. □

Coarsening at random implies factorization of the density of Y at y in F_X and G parts as in (1.12): for example, if $F_X(C(y)) > 0$, then $p_{F_X,G}(y) = F_X(C(y))h(y)$, with $h(y) = g_{Y|X}(y \mid x)$. Under mild regularity conditions, Gill, van der Laan and Robins (1997) show that even when $F_X(C(y)) = 0$, the density $p_{F_X,G}(y)$ (w.r.t. a dominating measure satisfying CAR itself) factors as a product $p_{F_X}(y)h(y)$, where $h(y) = g_{Y|X}(y \mid X)$ and $p_{F_X}(y)$ only depends on the measure F_X. Thus, the maximum likelihood estimator (MLE) of F_X based on $Y_1, \ldots, Y_n$ ignores the censoring mechanism G by simply interpreting $Y_i = y_i$ as $X_i \in C(y_i)$, $i = 1, \ldots, n$. The MLE of F_X can typically be computed with the EM-algorithm (e.g., Dempster, Laird and Rubin, 1977) either by assuming a nonparametric full data model and maximizing an unrestricted multinomial likelihood defined over given support points or assuming a parametric full data model and maximizing the parametric log-likelihood (Little and Rubin, 1987). The G part of the likelihood of $Y = y$ is the conditional density of $Y = y$, given X, which by CAR indeed only depends on y.

Let $\mathcal{G}(CAR)$ be the set of all conditional distributions G satisfying CAR (i.e., satisfying (1.13) or (1.14) w.r.t. a particular dominating measure μ satisfying CAR itself). Consider the observed data model $\mathcal{M}(CAR) = \{P_{F_X,G} : F_X \in \mathcal{M}^F, G \in \mathcal{G}(CAR)\}$ defined by the assumptions $G \in \mathcal{G}(CAR)$ and $F_X \in \mathcal{M}^F$. Let $T(P_{F_X,G}) \subset L_0^2(P_{F_X,G})$ be the closure of the linear span of all scores of one-dimensional submodels $\epsilon \to P_{F_\epsilon,G_\epsilon}$ with parameter ϵ through $P_{F_X,G}$ at $\epsilon = 0$. Here $L_0^2(P_{F_X,G}) = \{h(Y) : E_{P_{F_X,G}} h^2(Y) < \infty, E_{P_{F_X,G}} h(Y) = 0\}$ is the Hilbert space endowed with inner product $\langle h, g \rangle_{P_{F_X,G}} = E_{P_{F_X,G}} h(Y)g(Y)$. The sub-Hilbert space $T(P_{F_X,G})$ is called the observed data tangent space. It is shown in Gill, van der Laan and Robins (1997) that if F_X is completely unspecified (i.e., the full data structure model $\mathcal{M}^F$ is nonparametric), then the observed data model $\mathcal{M}(CAR)$ for the distribution of Y characterized by the sole restriction CAR (1.13) is locally saturated in the sense that $T(P_{F_X,G}) = L_0^2(P_{F_X,G})$. An important consequence of this result is that in this nonparametric CAR model all *regular* asymptotically linear estimators of the parameter μ are asymptotically equivalent and efficient. Because of its importance, we will give here the proof of this local saturation result.

Lemma 1.1 *Consider the model*

$$\mathcal{M}(CAR) = \{P_{F_X,G} : F_X \text{ unrestricted}, G \in \mathcal{G}(CAR)\}$$

for the observed data structure $Y = \Phi(C, X)$. The tangent space $T(P_{F_X,G})$ equals $L_0^2(P_{F_X,G})$.

Proof. By CAR, we have that the density $p(y)$ of $P_{F_X,G}$ w.r.t. a dominating measure factorizes (Gill, van der Laan and Robins, 1997): $p(y) = p_{F_X}(y)h(y)$ with $h(y) = g_{Y|X}(y \mid X)$. As one-dimensional submodels through F_X (at parameter value $\epsilon = 0$), we take $dF_{X,\epsilon}(x) = (1 + \epsilon s(x))dF_X(x)$, $s \in L_0^2(F_X)$. As one-dimensional submodels through G, we take $dG_\epsilon(y \mid x) = (1 + \epsilon v(y))dG(y \mid x)$, $v \in \{V(Y) \in L_0^2(P_{F_X,G}) : E(V(Y) \mid X) = 0\}$. A general result is that the score of the one-dimensional model $P_{F_{X,\epsilon},G}$ equals $E_{F_X}(s(X) \mid Y)$ (Gill, 1989). As a consequence, the collection of all scores of the corresponding one-dimensional submodels $P_{F_{X,\epsilon},G_\epsilon}$ through $P_{F_X,G}$ (obtained by varying s, v over all possible functions) is given by

$$S(P_{F_X,G}) \equiv \{E_{F_X}(s(X) \mid Y) : s \in L_0^2(F_X)\} \oplus \{V(Y) : E_G(V(Y) \mid X) = 0\}. \tag{1.21}$$

Let the nonparametric score operator $A_{F_X} : L_0^2(F_X) \to L_0^2(P_{F_X,G})$ be defined by $A_{F_X}(s)(Y) = E(s(X) \mid Y)$. The adjoint of A_{F_X} is given by $A_G^\top : L_0^2(P_{F_X,G}) \to L_0^2(F_X)$, $A_G^\top(V)(X) = E(V(Y) \mid X)$. This proves that the closure $T(P_{F_X,G})$ of $S(P_{F_X,G})$ equals the closure of the range of A_{F_X} plus the null space of its adjoint: $T(P_{F_X,G}) = \overline{R(A_{F_X})} \oplus N(A_G^\top)$. A general Hilbert space result is that for any Hilbert space operator $A : H_1 \to H_2$ with adjoint $A^\top : H_2 \to H_1$, $\overline{R(A)} + N(A^\top) = H_2$. This proves the lemma. □

Gill, van der Laan and Robins (1997) also prove that if the distribution of $Y = \Phi(C, X)$ has a finite support set, then the hypothesis that G satisfies CAR cannot be rejected; that is, the model $\mathcal{M}(CAR) = \{P_{F_X,G} : F_X \text{ unrestricted}, G \in \mathcal{G}(CAR)\}$ is a nonparametric model for the law of Y. It follows that the observing data $(Y_1, \ldots, Y_n)$ can never lead one to reject the hypothesis that the law of Y lies in the model $\mathcal{M}(CAR)$, regardless of the support of Y.

In many of the specific data structures covered in this book, it will be possible to provide an easy-to-interpret definition of CAR. If the censoring is multivariate in nature, CAR is typically a very complicated and hard-to-understand assumption, but we will always be able to define large easy-to-interpret submodels of CAR.

1.2.4 The curse of dimensionality revisited

When X is high-dimensional, the existence of locally CODA-efficient estimators with good moderate sample performance in the full data model $\mathcal{M}^F$