$\alpha(t) = E(dA(t) \mid \mathcal{F}(t)) = I(C \geq t, T \geq t)d\Lambda_C(t \mid X)$ be the intensity of $A(t)$ w.r.t. history $\mathcal{F}(t)$, where this can also denote a discrete probability in case C is discrete. Here $\Lambda_C(\cdot \mid X)$ is the cumulative of $\lambda_C(\cdot \mid X)$. The G part of the density of $P_{F_X,G}$ is given by

$$L(G) = \prod \alpha(t)^{dA(t)}(1-\alpha(t))^{1-dA(t)}$$
$$= \left\{ \prod_{(0,\tilde{T}]} (1 - d\Lambda_C(t \mid X)) \right\}^{\Delta} \left\{ \prod_{(0,\tilde{T})} (1 - \lambda_C(t \mid X)) \times d\Lambda_C(\tilde{T} \mid X) \right\}^{1-\Delta}.$$

For the Cox proportional hazards model and logistic regression model for $\lambda_C(t \mid X)$, maximum likelihood estimation of G can be based on this likelihood and can be carried out with standard software. In particular, one can fit the Cox proportional hazards model with the S-plus function Coxph(). If we assume the logistic regression model for discrete C, then inspection of the likelihood $L(G)$ shows that one can obtain MLE by simply fitting the logistic regression model to a pooled sample. Here, a subject with observed censoring time $C = j$ (for simplicity, let the support of C be integer-valued) contributes j Bernoulli observations $(0, \ldots, 0, 1)$ with corresponding covariates $(1, W(1)), \ldots, (j, W(j))$, and a subject who fails between $C = j$ and $C = j + 1$ contributes j observations $(0, \ldots, 0)$ with corresponding covariates $(1, W(1)), \ldots, (j, W(j))$.

In Section 3.2 we will provide many challenging estimation problems that are applications of the general monotone censored data structure covered in this chapter. In Section 3.3 we define inverse probability of censoring weighted mappings $IC_0(Y \mid G, D)$ from full data estimating functions into observed data estimating functions and derive a closed-form representation of the influence curve of the corresponding estimators. We work out these estimators for the case where the full data model is a multiplicative intensity model for the intensity of a counting process $N(t) \subset X(t)$ w.r.t. a particular subset of $\bar{X}(t-)$. For example, one might be interested in estimating the causal effect of a randomized treatment arm on the intensity of survival or another counting process of interest. This implies that one does not want to adjust for other variables in the past except the past of the counting process and the treatment variable. Just fitting a Cox proportional hazards model only using the treatment as covariate is inconsistent if the hazard of censoring at time t conditional on X depends on more of $\bar{X}(t-)$ than the treatment variable. In addition, this method is highly inefficient if the data contain surrogates for survival or strong predictors of survival. In Section 3.4, we define the optimal mapping $IC(Y \mid Q, G, D) = IC_0(Y \mid G, D) - \Pi(IC_0 \mid T_{CAR})$ from full data estimating functions into observed data estimating functions indexed by nuisance parameters $Q = Q(F_X, G)$ and G. In Section 3.5, we discuss estimation of Q in detail. In Section 3.6, we study estimation of the optimal index h_{opt} and

work out closed-form representations for the case where the full data model is the multivariate generalized linear regression model. In Section 3.7, we apply the methods to obtain a locally efficient estimator of the regression parameters in a multivariate generalized regression model (with outcome being a multivariate survival time) for the multivariate right-censored data structure when all failure times are subject to a common censoring time. In Section 3.8, we rigorously analyze a locally efficient estimator of the bivariate survival function based on a bivariate right-censored data structure by applying Theorem 2.4 of Chapter 2. Finally, in Section 3.9 we provide a general methodology for estimating optimal predictors of survival w.r.t. risk of a user supplied loss function based on our general right-censored data structure, thereby making significant improvements to the current literature on survival prediction. In the next subsection, we show that, without loss of optimality, the estimation methods can also be applied if censoring is cause-specific and the cause is observed.

3.1.1 Cause-specific censoring

In many applications, there exist various causes of censoring and the cause is typically known. In this situation, one observes (C, J), where C is the censoring time and J indexes the cause. For example, one cause might be the end of the study while another cause might be that the subject has severe side effects that made the doctor decide to stop treatment. In this case, one wants a model for the intensity $\alpha(t)$ that acknowledges that C is an outcome of competing censoring times that might follow Cox proportional hazards models using different subsets of covariates. In this case, we extend our data structure as

$$Y = (C, J, \bar{X}(C)),$$

where (C, J) now represents the joint censoring variable with conditional distribution G, given X. We can now identify the joint censoring variable by the random process $A(t) = (A_1(t), \ldots, A_J(t))$, where $A_j(t) = I(C \leq t, J = j)$, $j = 1, \ldots, J$. Thus, the G part of the density of $P_{F_X,G}$ is now given by $g(A \mid X)$, which can in the discrete case be represented as

$$g(A \mid X) = \prod_t \prod_j \alpha_j(t)^{dA_j(t)}(1-\alpha_j(t))^{1-dA_j(t)},$$

where $\alpha_j(t) = E(dA_j(t) \mid A_1(t), \ldots, A_{j-1}(t), \mathcal{F}(t))$, $j = 1, \ldots, J$. In the continuous case, this expression for $g(A \mid X)$ reduces to the partial likelihood of the multivariate counting process $A(t) = (A_1(t), \ldots, A_J(t))$ w.r.t. $\mathcal{F}(t)$ as defined in Andersen, Borgan, Gill, and Keiding (1993):

$$g(A \mid X) = \prod_t \prod_j \alpha_j(t)^{dA_j(t)}(1-\alpha_j(t)dt)^{1-dA_j(t)}.$$