CAR. If Y does not include the censoring variable C, then the definition of CAR on $G_{Y|X}$ is weaker than the same definition applied to G.

Let $\mathcal{X}$ and $\mathcal{C}$ be the sample spaces of X and C, respectively. We first formally define CAR in the case where X is a discrete random variable. Let $C(y) = \{x^* \in \mathcal{X}; \Phi(x^*, c^*) = y \text{ for some } c^* \in \mathcal{C}\}$ be the subset of the support $\mathcal{X}$ of X whose elements x are consistent with the observation y. If X is discrete, then CAR is the assumption

$$P(Y = y \mid X = x) = P(Y = y \mid X = x') \text{ for any } (x, x') \in C(y). \tag{1.9}$$

If, as in the previous examples, observing Y implies observing C so that C is always observed, then CAR can also be written

$$P(C = c \mid X = x) = P(C = c \mid X = x') = h(y) \text{ for any } (x, x') \in C(y) \tag{1.10}$$

for some function $h(\cdot)$ of $y = \Phi(c, x)$. If C is not always observed, this last assumption is more restrictive than CAR. Assumption (1.9) is also equivalent to

$$P(Y = y \mid X = x) = P(Y = y \mid X \in C(y)) \text{ for all } x \in C(y), \tag{1.11}$$

or equivalently the density $P(Y = y \mid X = x)$ is only a function of y. In other words, there is no $x \in C(y)$ that makes the observation $Y = y$ more likely. Therefore, under CAR, observing $Y = y$ is not more informative than observing that X falls in the *fixed given* set $C(y)$. As a consequence, under CAR, we have the following factorization of the density of the observed data structure:

$$\begin{aligned} P(Y = y) &= P(X \in C(y)) P(Y = y \mid X = x) \\ &= P(X \in C(y)) P(Y = y \mid X \in C(y)). \end{aligned} \tag{1.12}$$

Coarsening at random was originally formulated for discrete data by Heitjan and Rubin (1991).

A generalization to continuous data is provided in Jacobsen and Keiding (1995), whose definition is further generalized in Gill, van der Laan, and Robins (1997). A general definition of CAR in terms of the conditional distribution of the observed data Y, given the full data structure X, is given in Gill, van der Laan and Robins (1997): for each x, x'

$$P_{Y|X=x}(dy) = P_{Y|X=x'}(dy) \text{ on } \{y : x \in C(y)\} \cap \{y : x' \in C(y)\}. \tag{1.13}$$

Given this general definition of CAR, it is now also possible to define coarsening at random in terms of densities: for every $x \in C(y)$, we have that, for a dominating measure ν of G that satisfies (1.13) itself,

$$g_{Y|X}(y \mid x) \equiv \frac{dP(y \mid X = x)}{d\nu(y \mid X = x)} = h(y) \text{ for some measurable function } h. \tag{1.14}$$

Thus the density $g_{Y|X}(y \mid x)$ of $G_{Y|X}$ does not depend on the location of $x \in C(y)$. Therefore, the heuristic interpretation of CAR is that, given the full data structure $X = x$, the censoring action determining the observed data $Y = y$ is only based on the observed part $C(y)$ of x. As mentioned above, if observing Y implies observing C, then (1.14) translates into $g(c \mid x) = h(y)$ for some function h of $y = \Phi(c, x)$.

In this book, we can actually replace (1.13) by the minimally weaker condition that

$$g_{Y|X}(Y \mid X) = h(Y) \text{ with probability } 1 \tag{1.15}$$

for some $h(\cdot)$. Again, if observing Y implies observing C so that C is always observed, then this last equation is equivalent to

$$g(C \mid X) = h(Y) \text{ with probability } 1 \tag{1.16}$$

for some function $h(\cdot)$.

Example 1.6 (Repeated measures data with missing covariate; continuation of Example 1.1) In this example, C is the always observed variable Δ. Thus, CAR is the assumption that $p_G(\Delta|X) = h(Y) = h(\Delta, W, \Delta E)$. Thus $pr_G(\Delta = 0|X)$ is a function only of W so that

$$pr_G(\Delta = 1|X) = pr_G(\Delta = 1|W) \equiv \Pi_G(W) \equiv \Pi(W) \tag{1.17}$$

does not depend on E. □

Example 1.7 (Repeated measures data with right-censoring; continuation of Example 1.2) In this example, the conditional distribution of the always observed variable C, given X, is a multinomial distribution with the probability of $C = j$, $j = 0, \ldots, p$, being a function of X. It is easy to show that CAR is the assumption that the probability that a subject drops out at time j given the subject is yet to drop out (i.e., is at risk at j) is only a function of the past up to and including point j,

$$\begin{aligned} \lambda_C(j \mid X) &\equiv P(C = j \mid X, C \geq j) = P(C = j \mid C \geq j, \bar{X}(j)) \\ &\equiv \lambda_C(j \mid \bar{X}(j)), \end{aligned} \tag{1.18}$$

where $\lambda_C(j \mid \cdot)$ is the discrete conditional hazard of C at j given the information $\cdot$. □

Example 1.8 (Right-censored data) Let T be a univariate failure time variable of interest, W be a 25-d covariate vector (e.g., 25 biomarkers/gene expressions for survival), and C be a censoring variable. Suppose that we have the full data $X = (T, W)$ and the observed data $Y = (\tilde{T} = \min(T, C), \Delta = I(\tilde{T} = T), W)$. Let $G(\cdot \mid X)$ be the conditional distribution of C, given X, and let $g(\cdot \mid X)$ be its density w.r.t. a dominating measure that satisfies CAR as defined by (1.13) itself such as the Lebesgue measure or counting measure on a given set of points. CAR is then equivalent to

$$g(C \mid X) = g(C \mid W) \text{ on } C < T. \tag{1.19}$$

Except when the conditional law of C, given $C > T$, is a point mass, the assumption $g(C \mid X) = h(Y)$ is strictly stronger than CAR because the