The Annals of Statistics
1995, Vol. 23, No. 3, 774–786

COARSENING AT RANDOM IN GENERAL SAMPLE SPACES AND RANDOM CENSORING IN CONTINUOUS TIME

BY MARTIN JACOBSEN AND NIELS KEIDING

University of Copenhagen

Heitjan and Rubin proposed a concept, "coarsening at random," generalizing Rubin's theory of missing at random. Their analysis was done in discrete sample spaces. We propose a generalization to general sample spaces. Among Heitjan and Rubin's applications was right-censoring in survival analysis. We discuss the application of the generalized theory to various censoring patterns in continuous time and connect to the modern theory of random censoring.

1. Introduction. Heitjan and Rubin (1991) formulated a theory in discrete sample spaces for "coarsened" data, which are "neither entirely missing nor perfectly present." Their principal results were the formulation of a concept of "coarsening at random" and a proof that a certain desirable likelihood is adequate when coarsening is in fact random. Their examples included right-censoring in survival analysis, as further elaborated by Heitjan (1993).

These articles were based on a likelihood (possibly Bayesian) approach for which the concepts are defined relative to the realized data point. In the present paper [as also done by Heitjan (1994)], we formulate conditions for the statistical model, allowing a frequentist interpretation.

The modern theory of right-censoring aims at establishing conditions on the censoring pattern so that "past observations do not affect the probabilities of future failures" [Jacobsen (1989)]. Kalbfleisch and Prentice (1980) gave a pioneering analysis, while Aalen's (1975, 1978) discussion in the framework of counting processes was developed and consolidated by Gill (1980), Arjas and Haara (1984), Andersen, Borgan, Gill and Keiding (1988, 1993), Arjas (1989) and Jacobsen (1989). This discussion was exclusively conducted in continuous time.

The purpose of this note is to propose a version of "coarsening at random" in general sample spaces together with the associated theorem on likelihood inference. It turns out that one has to be careful with the measure theory involved, and we repeatedly invoke the basic theory of regular conditional distributions. The link to the abovementioned modern theory of random censoring is illustrated by four examples.

Received March 1994; revised November 1994.
AMS 1991 *subject classifications.* 62A10, 62F10.
Key words and phrases. Missing data, profile likelihood, survival analysis.

2. Coarsening at random for discrete observations. Heitjan and Rubin (1991) assumed that a random variable X of primary interest is not observed directly: it is only known that $X \in Y$, where the random set Y is determined by a further random variable G such that $Y = Y(X, G)$. Statistically, the distributions of X and G were assumed by Heitjan and Rubin to have "distinct parameters," in the Bayesian formulation meaning that the parameters θ (for X) and γ (for G) are a priori independent, while in the frequentist-based likelihood inference (which we shall pursue here) the meaning is that θ and γ are variation independent, that is, the pair (θ, γ) varies freely in a product parameter space $\Theta \times \Upsilon$.

A formalization of these ideas is to consider X and G, two random elements with values in the measurable spaces (Ξ, χ) and $(\Gamma, \mathscr{G})$, respectively, specified by a product parameter space $\Theta \times \Upsilon$ such that, for any $(\theta, \gamma) \in \Theta \times \Upsilon$, the $P_{\theta,\gamma}$-joint distribution of (X, G) is given by

$$P_{\theta,\gamma}(X \in dx) = f(x;\theta)\mu(dx),$$
$$P_{\theta,\gamma}(G \in dg | X = x) = h(g; x, \gamma)\nu(dg),$$

where μ and ν are reference measures on (Ξ, χ) and $(\Gamma, \mathscr{G})$, respectively.

Next consider a mapping $Y: \Xi \times \Gamma \to 2^{\Xi}$ (the set of all subsets of Ξ) with the property that

$$x \in Y(x, g), \tag{1}$$

for all x and g. Call $x \in \Xi$ and $y \subset \Xi$ *compatible* if $Y(x, g) = y$ for some $g \in \Gamma$.

For a moment we make the assumption that the *range space S for Y is discrete* (at most countably infinite) and assume of course also that Y is measurable,

$$\{(x, g): Y(x, g) = y\} \in \chi \otimes \mathscr{G},$$

for any $y \in S$. Then the likelihood for observing y becomes

$$P_{\theta,\gamma}(Y = y) = \int \mu(dx) f(x;\theta) k(y; x, \gamma), \tag{2}$$

where k is the conditional distribution of Y given X,

$$k(y; x, \gamma) = \int_{\{g: Y(x, g) = y\}} \nu(dg) h(g; x, \gamma).$$

Heitjan and Rubin [(1991), Definition 1], defined y to be *coarsened at random* (CAR) if for the observed $y \in S$, $k(y; x, \gamma)$ is, for all γ, the same for all $x \in y$. In that case (2) becomes, apart from a factor depending on γ and y only,

$$\int_y \mu(dx)\, f(x;\theta) \tag{3}$$

i.e., [Heitjan and Rubin (1991), Theorem 1], *the profile θ-likelihood for observing y is the same as the grouped likelihood.*

At least two problems arise when one attempts to generalize the Heitjan and Rubin result to allow for nondiscrete Y (as, e.g., is necessary to discuss the censoring examples below).

PROBLEM 1. For the right-hand side of (2) to be a likelihood in the general case, it must be a density with respect to a reference measure on some subset S of 2^{Ξ}. How should S and this reference measure be chosen?

PROBLEM 2. Even if Problem 1 is resolved, one cannot expect to find a general CAR concept that makes (3) the relevant likelihood for drawing inference on θ when the value y of Y is observed: suppose, for example, that X is real-valued with a continuous distribution, and use a continuous μ but assume that the coarsening mechanism is such that in some cases the value of X itself is observed, that is, $y = \{x\}$ is possible. However, with such a y, of course (3) vanishes in contradiction with the basic principle that for a given observation the likelihood function evaluated at the true parameter value is always strictly positive! Now obviously, *if* $y = \{x\}$, *the likelihood should just be* $f(x; \theta)$, but as we have just seen, this is *not* just a special case of (3). One could say that, in a discrete setup, (3) is the only desirable likelihood to aim for, while in the continuous case it may still be possible to obtain a simple likelihood instead of (3), but that the form of this likelihood may depend on the particular structure of the observation y [see (8) and (9) in Theorem 1 and also (11) in Example 4.]

3. Coarsening at random in general sample spaces. We shall now consider a general setup and give a general definition of CAR, which, if satisfied, yields a profile θ-likelihood for the observation y which is $f(x; \theta)$ if $y = \{x\}$ and is given by (3) if $\mu(y) > 0$. As mentioned in the Introduction, our definition will refer to the full model rather than one single data point. To make the distinction, we shall refer to our concept as CARM, coarsening at random in the model.

It is important also to obtain a useful expression for the profile likelihood for y in other than the two cases just discussed, that is, when $\mu(y) = 0$ and y is not a singleton. The theorem to be presented below does just that with a general expression for the y-likelihood in terms of a conditional expectation. It is then a part of the statement of the theorem that this reduces to what we want if either y has strictly positive μ-measure or is a one-point set. However, even in the other cases, as we shall see in Example 4, the conditional expectation may reduce to something desirable and sensible.

Since several of the facts listed below are valid whenever Y is some (not necessarily set-valued) function of x and g, we shall initially only assume that Y is a measurable function from $(\Xi \times \Gamma, \chi \otimes \mathscr{G})$ to a measurable space $(S, \mathscr{S})$. In agreement with the Heitjan–Rubin concept, we call x and y *compatible* if $Y(x, g) = y$ for some g.

Next, assume that the reference measures μ and ν on the sample spaces for X and G, respectively, *are both probabilities*. In practice this is not a restriction, and starting, for example, with Lebesgue measure may render some of the formulas below useless. (See also Example 1.)

Notation. From now on write X and G for the coordinate projections $X(x, g) = x$ and $G(x, g) = g$ on $\Xi \times \Gamma$. Also, let $\rho = \mu \otimes \nu$ be the product probability on $(\Xi \times \Gamma, \chi \otimes \mathscr{G})$ (so, under ρ, X and G are independent), and write $\sigma = \rho(Y^{-1})$ for the ρ-distribution of Y,

$$\sigma(D) = \rho(Y \in D), \qquad D \in \mathscr{S}.$$

Finally, let $P_{\theta,\gamma}$ be the probability on $\Xi \times \Gamma$ with density

$$f(x;\theta)\ h(g;x,\gamma),$$

with respect to ρ.

Then the following facts hold.

FACT 1. The $P_{\theta,\gamma}$-distribution of y is absolutely continuous with respect to σ with density

$$\varphi(y;\theta,\gamma) = E_\rho(f(X;\theta)\ h(G;X,\gamma) | Y = y), \tag{4}$$

[which is a completely general and, without more structure, fairly useless likelihood for observing y; E_ρ refers of course to the (conditional) expectation with respect to ρ].

FACT 2. The Heitjan–Rubin CAR condition involves the conditional $P_{\theta,\gamma}$-distribution of Y given $X = x$. To obtain this in general, let κ_x be the conditional ρ-distribution of Y given $X = x$, that is,

$$\kappa_x(D) = \nu(Y_x \in D), \qquad D \in \mathscr{S},$$

where $Y_x\colon \Gamma \to S$ is the section $Y_x(g) = Y(x, g)$ of Y. Then the conditional $P_{\theta,\gamma}$-distribution of Y given $X = x$ is absolutely continuous with respect to κ_x, with density

$$k(y;x,\gamma) = E_\nu(h(G;x,\gamma) | Y_x = y).$$

FACT 3. In view of Fact 2, the likelihood (4) for observing y may also be written

$$\varphi(y;\theta,\gamma) = E_\rho(f(X;\theta)\ k(y;X,\gamma) | Y = y).$$

Note that if it is known that, for all y and γ, $k(y; x, \gamma)$ is the same for all x compatible with y, and assuming that this set of compatible x is measurable, it follows that

$$E_\rho(f(X;\theta) | Y = y)$$

may be used as the profile θ-likelihood when observing the value y of Y.

FACT 4. This is a warning: suppose $D \in \mathscr{S}$ with $\sigma(D) = 0$. Then

$$\int \mu(dx)\kappa_x(D) = \rho(Y \in D) = \sigma(D) = 0,$$

and so $\kappa_x(D) = 0$ for μ-almost all x. However, the exceptional set of x-values, where this property of κ_x fails, may depend on D, and this in an essential manner: it is *not* in general true that, for μ-almost all x, $\kappa_x \ll \sigma$ (see Section 4). Thus one is forced to use the κ_x-density k from Fact 2, for the general definition of CAR.

So far the range space $(S, \mathscr{S})$ for Y has been arbitrary. We shall now restrict attention to the coarsening setup, where, for all x and g, $Y(x, g)$ is a subset of Ξ. First, we define $S = Y(\Xi \times \Gamma)$, the image under the map Y of $\Xi \times \Gamma$. This definition of S we find convenient; for most purposes the full set 2^{Ξ} of subsets of Ξ is much too large. Next, we introduce the following assumption on the structure of Y.

ASSUMPTION (Y). For all $x \in \Xi$, $g \in \Gamma$ it holds that $x \in Y(x, g)$ and, for every $g \in \Gamma$, any two subsets of Ξ of the form $Y(x, g)$ and $Y(\tilde{x}, g)$ are either identical or disjoint.

Thus the idea is that for every g there is a partitioning of the sample space for X, generated by an equivalence relation requiring x and $\tilde{x}$ to be equivalent when $Y(x, g) = Y(\tilde{x}, g)$ and with the sets $(Y(x, g))_{x \in \Xi}$ the equivalence classes. Note that (1) is included in the assumption.

It follows in particular from Assumption (Y) that *x is compatible with y if and only if $x \in y$*. Also, if $y \in S$ and x and $\tilde{x}$ are both compatible with y, then

$$\{g: Y(x, g) = y\} = \{\tilde{g}: Y(\tilde{x}, \tilde{g}) = y\}. \tag{5}$$

We are now ready to address Problems 1 and 2, posed toward the end of Section 2.

Problem 1 is easy: with $S = Y(\Xi \times \Gamma)$ already defined, define the σ-algebra $\mathscr{S}$ of subsets of S by

$$\mathscr{S} = \{D \subset S: Y^{-1}(D) \in \chi \otimes \mathscr{G}\},$$

and on $(S, \mathscr{S})$ use the reference measure σ obtained by transforming ρ with Y,

$$\sigma(D) = \rho(Y \in D), \qquad D \in \mathscr{S}.$$

With these definitions the properties of the coarsening Y are duly taken into account, and one has a suitable measurable structure on a suitable collection S of subsets of Ξ.

Problem 2 is the important and more difficult one. We shall resolve it by presenting a general definition of CAR and then prove our general version of the Heitjan–Rubin theorem.

To get anywhere, we need some measurability assumptions on Y. Assume that, for all x, g,

$$Y(x, g) \in \chi,$$

that is, for all $y \in S$,

$$y \in \chi, \tag{6}$$

and also assume that the mapping

$$y \to \mu(y)$$

from S to $[0, 1]$ is measurable. These assumptions should be easy to check in concrete situations.

The connection to the Heitjan–Rubin theory will be established via the following result.

LEMMA 1. *Define*

$$D_0 = \{y \in S: \mu(y) > 0\}.$$

Then, for any $y \in D_0$ and any $x \in \Xi$,

$$b(x; y) = \frac{1}{\mu(y)} 1_{(x \in y)} \tag{7}$$

defines a density with respect to μ for the regular conditional ρ-distribution of X given $Y = y$.

PROOF. We must show that, with b given by (7),

$$\int_{D \cap D_0} \sigma(dy) \int_B \mu(dx) b(x; y) = \rho(X \in B, Y \in D \cap D_0),$$

for all $B \in \chi$, $D \in \mathscr{S}$. However, the double integral on the left equals

$$\begin{aligned}
\int_{D \cap D_0} \sigma(dy) \frac{\mu(B \cap y)}{\mu(y)} &= \int \rho(dx, dg) \frac{\mu(B \cap Y(x, g))}{\mu(Y(x, g))} 1_{D \cap D_0}(Y(x, g)) \\
&= \int \mu(dx') \int \nu(dg) \int \mu(dx) 1_B(x') 1_{Y(x, g)}(x') \\
&\quad \times \frac{1}{\mu(Y(x, g))} 1_{D \cap D_0}(Y(x, g)).
\end{aligned}$$

By Assumption (Y), if $x' \in Y(x, g)$, then $Y(x', g) = Y(x, g)$ and also $x \in Y(x', g)$ (and conversely), so this expression is equal to

$$\begin{aligned}
&\int \mu(dx') \int \nu(dg) \int \mu(dx) 1_B(x') 1_{Y(x', g)}(x) \\
&\quad \times \frac{1}{\mu(Y(x', g))} 1_{D \cap D_0}(Y(x', g)) \\
&= \int \rho(dx', dg) 1_B(x') 1_{D \cap D_0}(Y(x', g)) \\
&= \rho(X \in B, Y \in D \cap D_0).
\end{aligned}$$

□

DEFINITION. Y is *coarsening at random in the model* (CARM) if there is a version of

$$k(y; x, \gamma) = E_\nu(h(G; x, \gamma) | Y_x = y)$$

such that, for all $y \in S$ and all $\gamma \in \Upsilon$, $x \to k(y; x, \gamma)$ is the same for all x compatible with y.

This definition is of course a condition on the statistical model $\{P_{\theta,\gamma}\}$. The fact that only the parameter γ enters in the definition reflects the basic assumption that θ and γ vary independently.

As already mentioned before, Heitjan and Rubin defined their CAR concept in a Bayesian flavor relative to a given observed value y. On the other hand we impose a condition for all y, which seems more natural if one is interested in properties of the underlying statistical model [see also Heitjan (1994)].

In our setup there is an arbitrariness in the choice of the reference probabilities μ and ν. However, if $\tilde{\mu}$ and $\tilde{\nu}$ are two other choices, equivalent to μ and ν, respectively, with

$$\frac{d\tilde{\mu}}{d\mu} = m, \qquad \frac{d\tilde{\nu}}{d\nu} = n,$$

where $m > 0$ and $n > 0$, then, using $\tilde{\mu}$ and $\tilde{\nu}$ as references, k is replaced by $\tilde{k}$, where

$$\tilde{k}(y; x, \gamma) = \frac{k(y; x, \gamma)}{E_\nu(n(G) | Y_x = y)}.$$

However, by (5), with y given, if x is compatible with y,

$$(Y_x = y) = \{g: Y(x, g) = y\}$$

does not depend on x. Thus, as ought to be the case, the definition of CARM does not depend on the choice of reference probabilities.

THEOREM 1. *If Y is CARM, the profile likelihood $L(\theta; y)$ for drawing inference on θ based on the observation y satisfies*

$$L(\theta; y) = E_\rho(f(X; \theta) | Y = y) \tag{8}$$

with, in particular,

$$L(\theta; y) = \begin{cases} \int_y \mu(dx)\, f(x; \theta), & \text{if } \mu(y) > 0, \\ f(x; \theta), & \text{if } y = \{x\}. \end{cases} \tag{9}$$

PROOF. By Fact 3, the total likelihood (in θ and γ) for observing y is

$$E_\rho(f(X; \theta)\, k(y; X, \gamma) | Y = y).$$

Because $\rho(\cdot | Y = y)$ is concentrated on (x, g) with $Y(x, g) = y$ and because Y is CARM, $k(y; X, \gamma)$ may be viewed as a constant, and hence, for estimating θ when observing y, one may use the likelihood (8). Obviously, if $y = \{x\}$, since ρ-a.s. $X \in Y$ by Assumption (Y), this reduces to $f(x;\theta)$. Also, if $\mu(y) > 0$, one finds from Lemma 1 that

$$E_\rho(f(X;\theta)|Y=y) = \int \mu(dx)\frac{1}{\mu(y)}1_{(x\in y)}f(x:\theta)$$
$$= \frac{1}{\mu(y)}\int_y \mu(dx)\,f(x;\theta). \qquad \square$$

4. Examples.

Example 1 (Right-censoring, one observation). Consider the following model for observing one failure time with right-censoring: let μ and ν be probabilities on $\Xi = (0,\infty)$ and $\Gamma = (0,\infty]$; let $\{f(x;\theta)\}_{\theta\in\Theta}$ be the family of densities for the failure time X; and let $\{h(g;x,\gamma\}_{\gamma\in\Upsilon}$ be the family of conditional densities for the censoring time G given $X = x$.

Define

$$(10) \qquad Y(x,g) = \begin{cases} \{x\}, & \text{if } g \ge x, \\ (g,\infty), & \text{if } g < x. \end{cases}$$

Clearly Y satisfies Assumption (Y) and, as will be clear from the next few reasonings, also the measurability conditions imposed in Section 3.

Note that S is the collection of all singletons $\{x\}$ and all open intervals (g,∞) with

$$Y^{-1}(\{\{x\}\}) = \{x\}\times[x,\infty],$$
$$Y^{-1}(\{(g,\infty)\}) = (g,\infty)\times\{g\}.$$

Also note that, with μ and ν both Lebesgue measure, the atoms of the σ-algebra $\mathscr{S}$ would have measure $+\infty$. (Hence the need to require μ and ν to be probabilities.)

Write $S = S_1 \cup S_I$, where

$$S_1 = \{\{x\}: x\in(0,\infty)\}, \qquad S_I = \{(g,\infty): g\in(0,\infty)\}.$$

The σ-algebra $\mathscr{S}$ on S consists of all $D \subset S$ such that

$$D\cap S_1 = \{\{x\}: x\in B_1\},$$
$$D\cap S_I = \{(g,\infty): g\in B_I\},$$

with B_1 and B_I Borel subsets of $(0,\infty)$. Note that, with D as above,

$$\sigma(D) = \sigma(D\cap S_1) + \sigma(D\cap S_I),$$

wherefore [cf. (10)],

$$\sigma(D\cap S_1) = \int_{B_1}\mu(dx)\nu([x,\infty]),$$

$$\sigma(D\cap S_I) = \int_{B_I}\nu(dg)\mu((g,\infty)).$$

Further, for a given x,

$$\kappa_x(D \cap S_1) = 1_{B_1}(x)\nu([x,\infty])$$

(and it is clear that, on S_1, κ_x is not "$\ll$" with respect to σ; cf. the warning issued in Fact 4),

$$\kappa_x(D \cap S_I) = \nu(B_I \cap (0,x)).$$

Finally,

$$k(y; x, \gamma) = E_\nu(h(G; x, \gamma)|Y_x = y)$$

is given as follows:

(i) If $y = \{x_0\} \in S_1$, for x to be compatible with y, necessarily $x = x_0$ and then

$$k(y; x_0, \gamma) = E_\nu(h(G; x_0, \gamma)|G \geq x_0).$$

In this case, CARM is (of course) automatic.

(ii) If $y = (g_0, \infty) \in S_I$, for x compatible with y, that is, $x > g_0$,

$$k(y; x, \gamma) = E_\nu(h(G; x, \gamma)|G = g_0)$$
$$= h(g_0; x, \gamma)$$

and the CARM condition requires that, for any γ, this must not depend on $x > g_0$. Informally, the result states that *as long as $X > G$, X and G should be independent* for the censoring to be CARM. This result is exactly the condition obtained in Jacobsen [(1989), Example 3.31], which again is the same as the constant-sum condition due to Williams and Lagakos (1977). That the likelihood for observing y is then proportional to (9) follows from Jacobsen [(1989), Theorem 4.2], since condition (ii) of that result is satisfied because the parameters θ and γ are variation independent.

In Example 4 we shall discuss the meaning of CARM when n rather than one right-censored failure times are available and see what (8) reduces to if the CARM condition is satisfied.

EXAMPLE 2 (Current-status data). Assume first that only one observation is available, so $\Xi = (0,\infty)$ and $\Gamma = (0,\infty]$, and let

$$Y(x,g) = \begin{cases} (0,g], & \text{if } x \leq g, \\ (g,\infty), & \text{if } x > g. \end{cases}$$

It is easily seen that Y satisfies Assumption (Y) and that the measurability conditions from Section 3 also hold. We get the following:

(i) if $y = (0, g_0]$ and x is compatible with y, that is, $x \leq g_0$,

$$k(y; x, \gamma) = h(g_0; x, \gamma),$$

and CARM requires that, for any γ, this must not depend on $x \leq g_0$;

(ii) if $y = (g_0, \infty)$ and x is compatible with y, that is, $x > g_0$, again

$$k(y; x, \gamma) = h(g_0; x, \gamma),$$

but to satisfy CARM this must now not depend on $x > g_0$ for any γ.

In this case CARM thus reduces to requiring that X and G be independent as usually assumed in the literature [cf. Diamond and McDonald (1992) or Groeneboom and Wellner (1992)].

If one considers n rather than one observation, CARM still amounts to demanding independence between $X = (X_1, \dots, X_n)$ and $G = (G_1, \dots, G_n)$: in this case,

$$Y(x, g) = \{(\tilde{x}_1, \dots, \tilde{x}_n): \tilde{x}_i \le g_i \text{ if } x_i \le g_i,\ \tilde{x}_i > g_i \text{ if } x_i > g_i\}.$$

Assumption (Y) is satisfied, and since, for any y and $x \in y$ (x compatible with y), the condition $Y_x = y$ completely specifies the value of G ($G = g^0$, say), it is easily seen that CARM amounts to demanding that

$$h(g^0; x, \gamma)$$

be the same for all g^0, all x.

EXAMPLE 3 (Double censoring, one observation). Suppose now that the failure time X is doubly censored, that is, let $G = (V, U)$, where $0 \le V < U \le \infty$ are random, and

$$Y(x, g) = \begin{cases} \{x\}, & \text{if } v < x \le u, \\ (0, v] \cup (u, \infty), & \text{otherwise,} \end{cases}$$

where we write $g = (v, u)$ for an arbitrary value of G. With μ the reference distribution for X, and ν the joint reference distribution for (V, U) (a probability concentrated on $\{(v, u): 0 \le v < u \le \infty\}$), reasoning as in Example 1, one finds that the data are CARM if and only if the conditional densities for G given $X = x$ satisfy that, for all γ, all $0 \le v < u \le \infty$, $h((v, u); x, \gamma)$ is the same for all x with $x \le v$ or $x > u$: as long as $X \in (0, V] \cup (U, \infty)$, X and (V, U) should be independent.

It is quite instructive to consider the modification that arises if, in the case of censoring, it is observed whether the censoring occurred to the left or the right. Then

$$Y(x, g) = \begin{cases} \{x\}, & \text{if } v < x \le u, \\ (0, v], & \text{if } x \le v, \\ (u, \infty), & \text{if } x > u. \end{cases}$$

The condition for CARM now becomes much more complicated: CARM holds if and only if, for all v,

$$\int \nu_{U|v}(du)\ h((v, u); x, \gamma)$$

is the same for all $x \le v$, and, for all u,

$$\int \nu_{V|u}(dv)\ h((v, u); x, \gamma)$$

is the same for all $x > u$. Here $\nu_{U|v}$ denotes the ν-conditional distribution of U given $V = v$, and $\nu_{V|u}$ is defined analogously. Note that this revised condition for CARM is satisfied in particular if the first CARM condition in this example holds.

EXAMPLE 4 (Right censoring, n observations). Suppose that $X_1, \dots, X_n$ are i.i.d. failure times with common density $f_0(\cdot; \theta)$ with respect to a reference probability μ_0 on $(0, \infty)$. Suppose also that each X_i is right-censored at a random time U_i. We write $X = (X_1, \dots, X_n)$ and $G = (U_1, \dots, U_n)$ and define $Y(x, g)$ as the subset

$$Y(x, g) = \{(\tilde{x}_1, \dots, \tilde{x}_n) \colon \tilde{x}_i = x_i \text{ if } x_i \le u_i \text{ and } \tilde{x}_i > u_i \text{ if } x_i > u_i\}$$

of $(0, \infty)^n$. Here $x = (x_1, \dots, x_n)$ and $g = (u_1, \dots, u_n)$ are arbitrary values of X and G, respectively.

To specify the model completely, we shall finally assume that there is a reference probability ν_0 on $(0, \infty]$ such that the conditional distributions of G given $X = x$ have densities $h(\cdot; x, \gamma)$ with respect to the product probability $\nu = \nu_0^n$ on $(0, \infty]^n$. (Thus under ν, the U_i are i.i.d. and independent of the X_i, but under $P_{\theta, \gamma}$ there may be dependence between G and X, and the U_i need not even be mutually independent).

In accordance with the notation generally used, we write

$$f(x; \theta) = \prod_{i=1}^{n} f_0(x_i; \theta)$$

for the density of the $X = (X_1, \dots, X_n)$ with respect to the product probability $\mu = \mu_0^n$.

Note that if μ_0 is continuous (e.g., absolutely continuous with respect to Lebesgue measure), the observed value y of Y will neither be a singleton nor satisfy $\mu(y) > 0$ if at least one failure time and at least one censoring time is observed. Thus the general expression (8) for the Y-likelihood under CARM, rather than the more special (9), becomes relevant.

Just as in Example 1, it is natural to split the range space S for Y into disjoint components. Let F be a subset of $\{1, \dots, n\}$ and define

$$S_F = \{\Pi A_i\},$$

where $A_i = \{x_i\}$ for some $0 < x_i < \infty$ if $i \in F$, and $A_i = (u_i, \infty)$ for some $0 < u_i \le \infty$ if $i \in C := \{1, \dots, n\} \setminus F$. Thus S_F is the part of S corresponding to observing the items in F to fail, those in C to be censored. The measurable structure on S_F is obtained by allowing the vector $((x_i)_{i \in F}, (u_j)_{j \in C})$ to vary in an arbitrary Borel set. Writing $x_F = (x_i)_{I \in F}$, $u_C = (u_j)_{j \in C}$ and with, for instance,

$$D = \left\{ \prod_{i \in F} \{x_i\} \times \prod_{j \in C} (u_j, \infty) \colon x_F \in B_F, u_C \in B_C \right\},$$

a particularly simple subset of S_F (with B_F and B_C Borel sets of the relevant dimensions $|F|$ and $|C|$, respectively), one finds that the reference probability σ on $(S, \mathscr{S})$ satisfies

$$\sigma(D) = \int_{B_F} \prod_{i \in F} \mu_0(dx_i)\ \nu_0([x_i, \infty]) \int_{B_C} \prod_{j \in C} \nu_0(du_j)\ \mu_0((u_j, \infty)),$$

while

$$\kappa_x(D) = 1_{B_F}(x_F) \prod_{i \in F} \nu_0([x_i, \infty])\ \nu_0^{|C|}\Big(B_C \cap \prod_{j \in C} (0, x_j)\Big).$$

Now suppose a $y \in S_F$ is observed, and let x_i^0 for $i \in F$ denote the observed failure times, while u_j^0 for $j \in C$ are the observed censoring times. A failure-time vector x is then compatible with y iff $x_F = x_F^0$ and $x_j > u_j^0$ for $j \in C$. For such x, $Y(x, g) = y$ precisely for those g with $u_C = u_C^0$ and $u_i \geq x_i^0$ for $i \in F$. It follows that

$$\begin{aligned} k(y; x, \gamma) &= E_\nu(h(G; x, \gamma) | Y_x = y) \\ &= \int_{\{u_F : u_i \geq x_i^0 \text{ for } i \in F\}} \prod_{i \in F} \nu_0(du_i)\ h\big((u_F, u_C^0); x, \gamma\big); \end{aligned}$$

CARM now requires this to be the same for all x compatible with y, a condition not at all transparent although it is easily seen to be satisfied in some simple cases: if, for instance, it is assumed that under each $P_{\theta,\gamma}$ the U_i given X are independent with the conditional density h_i for U_i depending on X_i only, CARM simply states that each U_i should be independent of X_i as long as $U_i < X_i$.

However, if CARM holds, we shall see that the likelihood for observing y takes a simple and natural form. Indeed, by (8) from Theorem 1, one finds

$$L(\theta; y) \propto \prod_{i \in F} f_0(x_i; \theta) \prod_{j \in C} \int_{u_j^0}^{\infty} \mu_0(dx_j) f_0(x_j; \theta). \tag{11}$$

Jacobsen (1989), working with a nonparametric setup, presented a certain condition, called (C), which if satisfied leads to a natural likelihood for drawing inference on the failure-time hazard. For the parametric model considered here, his likelihood would be the same as (11). However, to relate CARM to condition (C), it would be necessary to introduce the time dynamic aspect into this example and require that CARM should hold for any of the models obtained by stopping the observations after an arbitrary time point t (at which time some items have failed, some have been censored and some are still at risk). We conjecture (without pursuing the matter further here) that CARM holds for all $t > 0$ if and only if Jacobsen's condition (C) holds for all $P_{\theta,\gamma}$.

Acknowledgments. We are grateful to R. J. A. Little for calling attention to Heitjan's work, and to D. F. Heitjan for generous access to pre-publication versions of the material. Niels Keiding's work was initiated while holding a Distinguished Visiting Professorship in Biostatistics at the Ohio State University, Columbus.

REFERENCES

AALEN, O. O. (1975). Statistical inference for a family of counting processes. Ph.D. dissertation, Dept. Statistics, Univ. California, Berkeley.

AALEN, O. O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.* **6** 701–726.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1988). Censoring, truncation and filtering in statistical models based on counting processes. *Contemp. Math.* **80** 19–60.

ANDERSEN, P. K., BORGAN, Ø., GILL, R. D. and KEIDING, N. (1993). *Statistical Models Based on Counting Processes*. Springer, New York.

ARJAS, E. (1989). Survival models and martingale dynamics (with discussion). *Scand. J. Statist.* **16** 177–225.

ARJAS, E. and HAARA, P. (1984). A marked point process approach to censored failure data with complicated covariates. *Scand. J. Statist.* **11** 193–209.

DIAMOND, I. D. and MCDONALD, J. W. (1992). Analysis of current status data. In *Demographic Applications of Event History Data Analysis* (J. Trussell, R. Hankinson and J. Tilton, eds.) 231–252. Clarendon, Oxford.

GILL, R. D. (1980). *Censoring and Stochastic Integrals. Math Centre Tract* **124.** Math. Centrum, Amsterdam.

GROENEBOOM, P. and WELLNER, J. (1992). *Information Bounds and Nonparametric Maximum Likelihood Estimation.* Birkhäuser, Basel.

HEITJAN, D. F. (1993). Ignorability and coarse data: some biomedical examples. *Biometrics* **49** 1099–1109.

HEITJAN, D. F. (1994). Ignorability in general incomplete-data models. *Biometrika* **81** 701–708.

HEITJAN, D. F. and RUBIN, D. B. (1991). Ignorability and coarse data. *Ann. Statist.* **19** 2244–2253.

JACOBSEN, M. (1989). Right censoring and martingale methods for failure time data. *Ann. Statist.* **17** 1133–1156.

KALBFLEISCH, J. D. and PRENTICE, R. L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley, New York.

LITTLE, R. J. A. and RUBIN, D. B. (1987). *Statistical Analysis with Missing Data.* Wiley, New York.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63** 581–592.

WILLIAMS, J. A. and LAGAKOS, S. W. (1977). Models for censored survival analysis: Constant sum and variable sum models. *Biometrika* **64** 215–224.

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN
UNIVERSITETSPARKEN 5
DK-2100 COPENHAGEN Ø
DENMARK

DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF COPENHAGEN
BLEGDAMSVEJ 3
DK-2200 COPENHAGEN N
DENMARK