SORT 27 (1) January-June 2003, 41-64

Survival analysis with coarsely observed covariates

Søren Feodor Nielsen*

University of Copenhagen

Abstract

In this paper we consider analysis of survival data with incomplete covariate information. We model the incomplete covariates as a random coarsening of the complete covariate, and an overview of the theory of coarsening at random is given. Various ways of estimating the parameters of the model for the survival data given the covariates are discussed and compared.

MSC: 62N01, 62N02

Keywords: Incomplete data, ignorability, efficiency, EM algorithm, martingale estimating function, inverse probability weighting

1 Introduction: incomplete covariates

Statistics is often used to investigate the effect of one or more covariates, X, on an outcome of interest, T. In order to do this, a conditional model for the distribution of T given X, typically in the form of a (linear, generalised linear, hazard, ...) regression, is fitted to the data. If we only look at observations of (T,X) for Xs fulfilling certain restrictions, which do not depend on T (e.g. X larger than 7, integer valued,...), the conditional distribution of T given X is not affected. In particular, if X is sometimes incompletely observed, restricting attention to cases with X completely observed does not change the conditional distribution of T given X as long as the incompleteness does not depend on the outcome. Thus a complete

* *Address for correspondence:* S. Feodor Nielsen. Department of Applied Mathematics and Statistics. University of Copenhagen. Universitetsparken 5. DK-2100 Copenhagen Ø. DENMARK. Phone: (+45) 35 32 07 95. Fax: (+45) 35 32 07 72 feodor@stat.ku.dk

Received: September 2002

Accepted: January 2003

case analysis will lead to correct inference (consistent estimators, valid tests, etc). However, there is clearly a loss of information, as we are effectively reducing the sample size. In some cases this reduction may be considerable. Moreover, if X is multivariate and only partly missing or more generally incomplete so that X is known to lie in a restricted set y, a lot of good information may be lost. Consequently, we would like to incorporate cases with incomplete covariates.

If the incompleteness of X depends on T, then we have a problem. Imagine for instance that X is incompletely observed if $T \notin y$ for some set y. Then a complete case analysis amounts to looking only at cases with $T \in y$. But as $\mathcal{L}(T|X, T \in y) \neq \mathcal{L}(T|X)$, a complete case analysis will fail to give consistent estimators. Here more complicated methods are necessary.

In this paper we will mainly be concerned with the first type of incompleteness, i.e. incompleteness of covariates unrelated to the outcome. In prospective studies where the outcome is measured/recorded at a later stage than the covariate, this would often be case. We will however also touch upon the more general case, where the incompleteness of the covariates is related to the outcome as well as or instead of the covariate. This will often be the case in retrospective studies but also in some prospective studies, for instance if incompleteness is related to prognosis of outcome.

In the next section we give an introduction to coarse data and the concept of coarsening at random in discrete sample spaces. We will use this to model the incomplete covariates. In Section 3 we show how coarsening at random allows us to estimate a conditional survival function and look at the EM algorithm. In Section 4 we extend the concepts of Section 2 to general sample spaces and give a discussion of when censoring is ignorable if the covariates are coarsely observed. Different methods of estimation –including two likelihood based methods and weighted martingale estimation functions– in survival models with coarsely observed covariates are discussed in Section 5. Some conclusions are given in the final section.

2 Coarsening at random–discrete case

2.1 Coarse data

Let us start of with a simple example of an incomplete covariate. Let X denote smoking status coded as "non-smoker", "light smoker", and "heavy smoker". Imagine that for some individuals we only know that they are not non-smokers. Then the covariate is incompletely observed rather than missing for these individuals, since we do have some information on the value of the covariate: It is either "light smoker" or "heavy smoker".

Let X be a random variable with values in a finite space E. If X is not completely observed, it means that all we know is that certain values of X are possible, i.e. that $X \in Y$ for some subset Y of E. This subset may be randomly determined as in the example

above: It is only for some smokers we do not have a complete observation of smoking status. We call such a random subset Y of E a *coarsening* of X. As Y is to represent the possible values of X, we will require that $X \in Y$ with probability 1; in particular, $Y \neq \emptyset$.

Example 1 *Two examples of coarse data are missing data and heaped data:*

- *Missing data. We either observe X or nothing at all:*

$$Y = \begin{cases} \{X\} & \text{if } X \text{ is observed} \\ E & \text{if } X \text{ is missing} \end{cases}$$

- *Heaped data. The covariate is either completely observed or rounded, i.e. we observe either X or $c\lfloor \frac{X}{c} \rfloor$ for some given c. Thus*

$$Y = \begin{cases} \{X\} & \text{if } X \text{ is observed} \\ \left\{c\lfloor \frac{X}{c} \rfloor, c\lfloor \frac{X}{c} \rfloor + 1, \ldots, c\lfloor \frac{X}{c} \rfloor + c - 1\right\} & \text{if } X \text{ is heaped} \end{cases}$$

  A typical example is self-reported smoking; some individuals report a number of cigarettes, others a number of packs smoked.

See Heitjan (1993) for more examples. □

2.2 Coarsening at random

Let us go back to the smoking status example. Suppose that the probability of not observing which kind of smoker a person is, does not depend on whether the person is a "light smoker" or a "heavy smoker". Then the incompleteness is ignorable in the sense that it tells us nothing beyond the fact that this person is a smoker. This idea is formalised in the notion of coarsening at random.

Definition 1 *We say that Y is a* random coarsening *of X if for all $y \subseteq E$*

$$P\{Y = y | X = x\} = q_y \quad \text{for all } x \in y. \tag{1}$$

We shall refer to this condition (1) as CAR (for *coarsening at random*). $(q_y)_{y \subseteq E}$ is called the *coarsening mechanism*; we note that $\sum_{y \ni x} q_y = 1$ for every $x \in E$. Equivalent to condition (1) in Definition 1 is

$$\begin{aligned} P\{Y = y | X \in y\} &= \sum_{x \in y} P\{Y = y | X = x\} P\{X = x | X \in y\} \\ &= P\{Y = y | X = x\} \text{ for all } x \in y \end{aligned}$$

or conditional independence of the coarse observation $Y = y$ and the complete observation $X = x$ given the fact that $X \in y$:

$$\{Y = y\} \underset{\{X \in y\}}{\perp} \{X = x\}, \quad x \in y. \tag{2}$$

Hence observing $Y = y$ tells us nothing more about the unobserved value of X than the fact that $X \in y$. This is the essence of the ignorability mentioned above. For future reference we note that this conditional independence is between three events, which are tied together by y; Y is not conditionally independent of X given $X \in y$. A third equivalent condition is that

$$P\{X = x|Y = y\} = P\{X = x|X \in y\} \text{ for all } x \in y.$$

This follows directly from the conditional independence (2) or from the definition (1) and Bayes's theorem.

2.3 Estimating p

If CAR holds, the likelihood factors into a product of "the likelihood ignoring the incompleteness", $\sum_{x \in y} p(x)$, and the coarsening mechanism, q_y:

$$P\{Y = y\} = \sum_{x \in y} P\{Y = y|X = x\}P\{X = x\} = q_y \sum_{x \in y} p(x), \tag{3}$$

where $p(x) = P\{X = x\}$. Hence to estimate p by maximum likelihood we can maximise the naïve log-likelihood

$$L(p) = \sum_{y \subseteq E} n_y \cdot \log \left(\sum_{x \in y} p(x) \right) \tag{4}$$

where n_y is the number of observed subsets of type y. In other words, we may ignore the coarsening mechanism when estimating p (by maximum likelihood). This log-likelihood can be maximised using the EM algorithm (Dempster, Laird and Rubin).

Given the marginal distributions of X and Y we can always write $P\{Y = y\} = q_y \sum_{x \in y} p(x)$ by defining $q_y = P\{Y = y\} / \sum_{x \in y} p(x)$. Of course, this will not ensure that $\sum_{y \ni x} q_y = 1$. Gill, van der Laan and Robins (1997, p. 262) seem to suggest that a factorisation such as (3) with $\sum_{y \ni x} q_y = 1$ implies CAR. The reader is invited to show that this is indeed the case for the smoking status example. However, the following example shows that it is not the case in general.

Example 2 *Let* $E = \{1,2,3,4\}$ *and* $p(x) = \frac{1}{4}$ *for* $x \in E$. *Let* $P\{Y = y\} = \frac{1}{4}$ *for* $y = \{1,2\},\{1,3\},\{2,4\},\{3,4\}$ *and 0 for all other subsets of* E. *Then the factorisation holds*

with $q_y = \frac{1}{2}$ *for* y *with* $P\{Y = y\} > 0$. *In particular,* $\sum_{y \ni x} q_y = 1$ *for every* x. *However it is quite possible to have, say,*

$$P\{Y = \{1,2\}|X = 1\} = \frac{2}{3} \qquad P\{Y = \{1,2\}|X = 2\} = \frac{1}{3}$$

and so on, so that $P\{Y = y|X = x\} \neq q_y$ *and CAR does not hold.* □

However, it is true that given the marginal distribution of Y there is a marginal distribution of X so that the factorisation (3) holds with $\sum_{y \ni x} q_y = 1$. In other words, we cannot test the hypothesis of coarsening at random. This is not unexpected since CAR is a condition on the distribution of what is observed given what is missing. Gill *et al.* (1997) prove this result by showing that the desired factorisation is obtained by maximising the naïve log-likelihood (4). Moreover, they show that the factorisation is unique in the sense that for any $y \subseteq E$ with $P\{Y = y\} > 0$, $\sum_{x \in y} p(x)$ (and q_y) is uniquely determined. In particular, assuming CAR the distribution of X is determined from the distribution of Y if for instance $P\{Y = \{x\}\} > 0$ for all x. This is however not a necessary condition, as the following example shows. Another sufficient condition is that the set y with $P\{Y = y\} >$ is a π-system generating the σ-field consisting of all subsets of E (e.g. Billingsley 1979), but as the example below shows this is not a necessary condition, either. As an example of p not being identified from the distribution of Y consider the example above. A necessary and sufficient condition for the identifiability of p does not appear to be known and may not exist.

Example 3 *Let* $E = \{1,2,3,4\}$ *and suppose that* $P\{Y = y\} > 0$ *for* $y = \{1,2\},\{2,3\},$ $\{2,4\}$ *only. Then we can identify* $p(1) + p(2)$, $p(2) + p(3)$, *and* $p(2) + p(4)$, *and from this we get*

$$p(2) = 1 - \big(p(1) + p(2) + p(2) + p(3) + p(2) + p(4)\big)/3.$$

Now the rest follows easily: For instance, $p(1) = p(1) + p(2) - p(2)$. □

2.4 Discrete?

The "discrete case" in the title of Section 2 refers to the discreteness of the joint distribution of (X,Y). We note that almost everything discussed above carries through to the case where E is countable if the support of the distribution of Y is also at most countable. The only exception is the result about the existence of the CAR factorisation. Indeed, Gill *et al.* (1997) give a counter example (due to Y. Ritov) showing that if the support of the distribution of Y is countable there may be no such factorisation. However since all observed data sets are finite, it is still fair to say that the CAR hypothesis cannot be tested.

Notes

Coarsening at random was first defined by Heitjan and Rubin (1991) as a generalisation of Rubin's (1976) "missing at random". Their treatment was essentially restricted to the discrete case considered here. Jacobsen and Keiding (1995), Gill *et al.* (1997), and Nielsen (2000) extend their original idea to general sample spaces (see Section 4). Gill *et al.* (1997) also consider the discrete case in detail, and the present presentation is based on their work.

3 Survival data

We shall apply the ideas discussed in the previous section to the analysis of survival data with incomplete covariate information. Thus the data we are considering is in the form of a survival time, T, a censoring time, C, and a covariate, X, for each individual. This is the complete data. The observed data is the censored survival time, $T \wedge C$, the censoring indicator, $1_{\{T \leq C\}}$, and a coarsening, Y, of X. We will work under the assumption of random censoring in the sense of either independence of T and C or conditional independence of T and C given X. In many applications the latter assumption is more reasonable: If T is time to death of a specific cause and the censoring includes "death of other diseases", both will usually depend on life style risk factors such as smoking. However, we shall see that conditional independence given X causes problems for many of the methods we will consider.

3.1 Estimating the survival function

We will first consider estimating the conditional survival function $\bar{F}(t|x) = P\{T > t|X = x\}$ of T given X based on the censored survival times and the coarsened covariates. Suppose that Y is a random coarsening of X and that it is independent of T given X. Then the conditional survival function given $Y = y$ is given by

$$\begin{aligned}\bar{F}(t|y) = E[1_{\{T>t\}}|Y=y] &= E\left[E[1_{\{T>t\}}|X,Y=y]\middle|Y=y\right] \\ &= E\left[E[1_{\{T>t\}}|X]\middle|Y=y\right] = E[\bar{F}(t|X)|Y=y] = \frac{\sum_{x\in y}\bar{F}(t|x)p(x)}{\sum_{x\in y}p(x)}\end{aligned}$$

Thus if the censoring is independent of T and X, we may estimate $\bar{F}(t|y)$ by the usual Kaplan-Meier estimator and p from Y, plug-in and minimise the sum of squares to obtain an estimator of $\bar{F}(t|x)$. Thus

$$\left[\widehat{\bar{F}}(t|x)\right]_x = (W^\top W)^{-1}W^\top\left[\widehat{\bar{F}}(t|y)\right]_y \text{ where } W = \left[\frac{\hat{p}(x)}{\sum_{x\in y}p(x)}\right]_{y,x}$$

We note that $\widehat{\bar{F}}(t|\{x\})$ actually estimates $\bar{F}(t|x)$; it is just the complete case estimator discussed in the introduction. However, the weighted estimator derived above uses all the data and should therefore be more efficient.

Example 4 *To illustrate we simulate 4000 datasets with 200 survival times such that T given $X = x$ is exponential with intensity x. X is uniform on $\{1,2,3\}$ and coarsened as in the smoking status example such that*

$$P\{Y = y|X = x\} = \begin{cases} 1 & \text{if } y = \{1\}, x = 1 \\ \frac{1}{2} & \text{if } y = \{2\}, \{3\}, \text{ or } \{2,3\} \text{ and } x \in y \\ 0 & \text{otherwise} \end{cases}$$

The censoring is exponential with intensity 2 independent of T and X. We give results for $t = 0.2$ in Table 1.

***Table 1**: Estimation of $\bar{F}(t|x)$ for $t = 0.2$: Complete case estimators and weighted estimators (standard deviations in parentheses). Efficiency is of the complete case estimator compared to the weighted estimator.*

	True value	*Complete cases*		*Weighted estimates*		*Efficiency*
$X = 1$	*0.8187*	*0.8185*	*(0.0530)*	*0.8185*	*(0.0530)*	*100%*
$X = 2$	*0.6703*	*0.6693*	*(0.0909)*	*0.6690*	*(0.0809)*	*79%*
$X = 3$	*0.5488*	*0.5484*	*(0.0979)*	*0.5487*	*(0.0874)*	*80%*

We see that both complete cases and the weighted estimators are unbiased, and that the weighted estimators are more efficient as we expected. Of course for $X = 1$ the weighted estimator and the complete case estimator are the same. □

This weighting approach can be used for estimating any conditional functional of the survival distribution, e.g. the conditional hazard.

3.2 Maximum likelihood estimation

Under random censoring –in the sense of conditional independence of T and C given X– the distribution of the observed data is given by

$$P\{T \wedge C \leq t, 1_{\{T \leq C\}} = \delta, Y = y\} = \sum_{x \in y} P\{T \wedge C \leq t, 1_{\{T \leq C\}} = \delta, Y = y|X = x\} p(x)$$

Hence, if X is coarsened at random and the coarsening Y is independent of the survival data given X, the likelihood for the observed data is

$$q_y \cdot \sum_{x \in y} \left(f(t|x) P\{C > t|X = x\} 1_{\{T \leq C\}} + h(c|x) P\{T > c|X = x\} 1_{\{T < C\}} \right) p(x)$$

where f is the density of T given $X = x$ and h the density of C given $X = x$. The assumption that the coarsening only depends on X may be dispensed with but we leave this case to Section 5.1.

We observe that generally censoring will not be ignorable in the sense of dropping out of the likelihood unless C is actually independent of X. Thus if C is only conditionally independent of T given X, then we need to specify a model for the censoring in order to maximise the likelihood of f when X is coarsened.

Assuming that C is independent of (X,T) and Y is independent of (T,C) given X we can maximise the likelihood using the EM algorithm. The E-step becomes

$$\sum_{x\in y} \log L_{T\wedge C,1_{\{T\le C\}}|x}(f)\, p(x|y,T\wedge C,1_{\{T\le C\}}) + \sum_{x\in y} \log p(x)\, p(x|y,T\wedge C,1_{\{T\le C\}})$$

where

$$L_{T\wedge C,1_{\{T\le C\}}|X}(f) = f(t|x)1_{\{T\le C\}} + P\{T > c|X = x\}1_{\{T\le C\}}$$

and $p(x|y,T\wedge C,1_{\{T\le C\}})$ is the conditional probability of $X = x$ given $Y = y, T\wedge C, 1_{\{T\le C\}}$. By the (conditional and unconditional) independence assumptions made we see that

$$p(x|y,T\wedge C,1_{\{T\le C\}}) = \begin{cases} \dfrac{f(t|x)p(x)}{\sum_{x\in y} f(t|x)p(x)} & \text{when } 1_{\{T\le C\}} = 1 \\ \dfrac{P\{T > c|X = x\}p(x)}{\sum_{x\in y} P\{T > c|X = x\}p(x)} & \text{when } 1_{\{T\le C\}} = 0 \end{cases}$$

so that we may ignore the censoring mechanism as well as the coarsening mechanism when estimating the marginal distribution of X. It must be stressed that the assumption that C is independent of X in many practical applications will be an unreasonable assumption. In that case we need to estimate the censoring mechanism as well in order to find the maximum likelihood estimator of f.

It is tempting, though probably not fully efficient to estimate p by the marginal MLE based on Y and use this estimator in the EM algorithm.

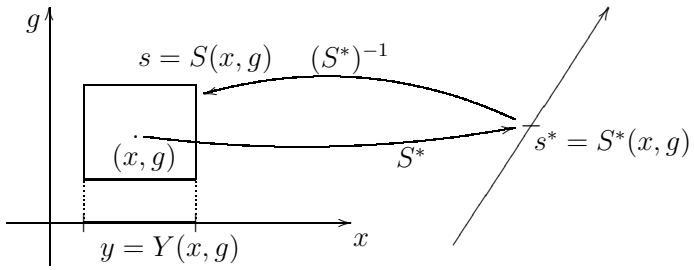

Figure 1: *Transformation of* (X,G) *to S.*

Table 2: *Complete case estimator and estimators derived from the EM algorithm: EM I uses a plug-in estimator for p. Standard deviations in parentheses.*

Method	Complete cases		EM I		EM II	
Mean	1.012	(0.1323)	1.008	(0.1057)	1.008	(0.1057)

Example 5 We apply these two versions of the EM algorithm to the data simulated in the previous example. We fit a parametric model and include the complete case estimator for comparison. As we expected there is (virtually) no difference between the two EM algorithms (see Table 2). We also note that the efficiency of the complete case analysis compared to the full maximum likelihood estimation is only 64%. In other words, using a complete case analysis and thus discarding on average one third of the observations, we loose a little more than one third of the available information even though the discarded observations are incomplete.

That the loss of information is larger than the fraction of missing observations is due to the differential missingness; the incompleteness only affects observations with $X > 1$. As the incomplete observations are very informative about the true unobserved value of the covariates a lot of information may be regained by a full maximum likelihood estimation. In fact the efficiency of the maximum likelihood estimator based on the observed data compared to the estimator obtained from the uncoarsened data (not shown) is 97.6%. In other words, the coarsening results in an almost negligible loss of information about the regression parameter. □

4 Coarsening at random–general case

4.1 Extending to general sample spaces

When discussing coarsenings with uncountable support, it seems to be useful to introduce some extra structure on the coarsening. Hence we will assume that there is a *coarsening variable* G deciding the degree of coarseness with which X is observed. For instance, if X is censored, G could be the censoring time. If X is missing, G could be a response indicator taking the value of 1 if X is observed, 0 if X is missing. Typically G may not be completely observed either; if G is a censoring time, then it is only observed if X is censored.

We assume that G is chosen so that there is no additional randomness in the incompleteness mechanism, i.e. that what we actually observe is a non-random function S^* of (X,G). Now, the possible values of (X,G) will be the subset $S = \{(x,g) : S^*(x,g) = S^*(X,G)\}$ of the joint sample space of (X,G). The possible values of X are then represented by the subset Y of X's sample space, E, obtained by projecting S onto E; see Figure 1.

We note that the extra structure introduced is not a restriction. If we only observe Y, "the possible values of X", then Y can be used as the coarsening variable G in which case $S = Y \times \{Y\}$. And for S^* we may without loss of generality use S, "what is observed about (X,G)".

We shall in this paper focus on coarsenings S which takes the form of a product set $S = Y \times H$, where H is a subset of Gs sample space. Gill *et al.* (1997) calls this a *Cartesian coarsening*.

Extending the theory of CAR to a general sample space is not straightforward. To see why, recall that in the discrete case, CAR was equivalent to a conditional independence of the events $\{Y = y\}$ and $\{X = x\}$ given $\{X \in y\}$, not of Y and X given $X \in y$. Thus in the discrete case CAR is a pointwise, distributional condition. With a general sample space these events will typically have probability 0 making this condition difficult to generalise.

A pointwise formulation is however easily obtained by replacing the condition on the conditional probability (1) by a similar condition on the conditional density of S given $X = x$:

Definition 2 *S is a random coarsening of X if the conditional density, $k(s|x)$, of S given $X = x$ can be chosen to be independent of $x \in y$, where y is the projection of s onto X's sample space.*

A general expression for the conditional density, $k(s|x)$, may be found in Nielsen (2000).

Densities require a reference measure, and it turns out that different reference measures lead to different conditions. To avoid measure theoretic difficulties we will in this paper use a reference measure, P_0, which is a probability measure, and which makes X and G independent. As shown by Jacobsen and Keiding (1995) in a slightly different set-up any product reference measure leads to the same CAR-condition. Expectations with respect to P_0 are denoted E_0.

Example 6 *Consider a right censored variable X with censoring time C. Here*

$$S = \begin{cases}]G;\infty[\times\{G\} & \text{if } X \text{ is censored} \\ \{X\} \times [X;\infty[& \text{if } X \text{ is observed} \end{cases}$$

It can be shown (see e.g. Nielsen 2000) that

$$k(s|x) = \begin{cases} \dfrac{dP\{G \in \cdot | X = x\}}{dP_0\{G \in \cdot\}}(g) & \text{if } X \text{ is censored; here } y =]g;\infty[\\ \dfrac{P\{G > x | X = x\}}{P_0\{G > x\}} & \text{if } X \text{ is observed; here } y = \{x\} \end{cases}.$$

We see that S is CAR if the conditional density of G given $X = x$ (relative to P_0) can be chosen so that it does not depend on $x > g$. □

As in the discrete case, CAR implies a factorisation of the likelihood.

Theorem 1 *The density of a random Cartesian coarsening S is $L_f \cdot k(s|x)$, where*

$$L_f = E_0[f(X)|Y=y] = \begin{cases} f(x) & \text{if } y = \{x\} \\ \dfrac{\int_y f(x)dP_0(x)}{P_0\{X \in y\}} & \text{if } P_0\{X \in y\} > 0 \\ ? & \text{otherwise} \end{cases}$$

where f is the marginal density of the distribution of X with respect to the reference measure P_0.

The question mark is meant to convey the impression that unless $y = \{x\}$ or $P_0\{X \in y\} > 0$ no general expression for the conditional mean is available, not that one cannot be calculated in concrete examples.

Notes

In general sample spaces Gill *et al.* (1997) distinguish between two type of CAR-conditions, an absolute and a relative. The relative condition, CAR(REL), is formulated in terms of densities as in Definition 2 whereas the absolute, CAR(ABS), is formulated in term of probabilities. The absolute CAR condition is harder to formulate and understand than the relative CAR condition used in this paper. It does however generalise the conditional independence (2) to some extent: If S is an absolute random coarsening, and s is a set such that its projection y has $P\{X \in y\} > 0$, then X is independent of $S = s$ given $X \in y$ (Nielsen 2000, Lemma 4).

Actually for Cartesian coarsenings, independence of X and G implies CAR(ABS). Moreover, any measure which has a density fulfilling Definition 2 with respect to a measure which is CAR(ABS), is in itself CAR(ABS). Thus, even if we have chosen a relative or pointwise formulation in this paper, our choice of reference measure implies the stronger CAR(ABS).

It is clear that by using Y as a coarsening variable, any coarsening may be turned into a Cartesian coarsening. But as $X \in Y$ with probability 1, a product reference measure seems out of the question if $G = Y$. Furthermore, if S is not a Cartesian coarsening, then reducing the observation to Y may not only reduce the available information significantly but it may also make a random coarsening non-random; see Nielsen (2000) for an example.

Extending results of Jacobsen and Keiding (1995), Nielsen (2000) shows that given a statistical model for (X,G) any dominating measure chosen in the model leads to the same CAR(REL) condition. In this sense, choosing a reference measure inside the model is a canonical choice.

4.2 Survival data with coarsely observed covariates

We can view survival data with incomplete covariates as a joint coarsening of the survival time T and the covariate X. For instance we could let $S^*(T,C,X) = (T \wedge C, 1_{\{T \leq C\}}, Y)$ where C is the censoring time and Y is the coarsening of X. Suppose for simplicity that the sample space of X is finite and that Y is a random coarsening of X and independent of (T,C) given X. Thus we observe

$$S = \tilde{Y} \times Y \times H = \begin{cases}]C;\infty[\times Y \times \{C\} & \text{if } T \text{ is censored} \\ \{T\} \times Y \times [T;\infty[& \text{if } T \text{ is observed} \end{cases},$$

i.e. a Cartesian coarsening. S is a random coarsening if there exists functions, K and k, such that

$$K(t,y) = P\{C > t | X = x, T = t\} P\{Y = y | X = x\} \text{ for all } x \in y$$

and

$$k(c,y) = h(c|t,x) P\{Y = y | X = x\} \text{ for all } x \in y, t > c$$

where h is the conditional density of C given T and X. Thus

$$\frac{K(t,y)}{P\{Y = y | X = x\}} = P\{C > t | X = x, T = t\} = 1 - \int_0^t \frac{k(c,y)}{P\{Y = E | y = x\}} dc$$

for all $t \geq 0$ and all $x \in y$, and hence

$$\int_0^t k(c,y) dc = P\{Y = y | X = x\} - K(t,y) \text{ for all } x \in y, t \geq 0.$$

Thus, $P\{Y = y | X = x\}$ cannot depend on $x \in y$, which is equivalent to Y being an random coarsening of X. Moreover, it follows that C may only depend on X through Y. In particular, if $Y = \{X\}$ almost surely, CAR is just random censoring in the sense that T and C must be essentially independent given X whereas if $Y = E$ almost surely, C must be independent of X given T. In the smoking status example, censoring may depend on whether the person is a smoker or not but not on whether the person is a light or a heavy smoker.

If we allow Y to depend on the survival data, it may be possible to allow censoring to depend on X but this dependence must be balanced with the coarsening mechanism in a rather unintuitive way. Thus generally, if the censoring depend on the covariate we cannot expect it to be ignorable. Obviously, if some of the covariates are always completely observed, then censoring may depend on these covariates and still be ignorable. We get similar results for coarsenings in general sample spaces by replacing $P\{Y = y | X = x\}$ by $k(s|x)$.

5 Estimation with coarsely observed covariates

In the following subsections we will indicate how some of the existing methods for handling survival data with missing covariates may be extended to handling survival data with coarsened covariates.

5.1 Likelihood based estimation

We will first discuss maximum likelihood estimation. Consider a general transformation model:

$$P\{T > t|X\} = 1 - F_\gamma(\log\Lambda(t) + \beta X) \stackrel{\text{def}}{=} \bar{F}_\gamma(\log\Lambda(t) + \beta X)$$

where F_γ is a known distribution function (on $\mathbb{R}$) except possibly for a finite dimensional parameter γ; with $F_\gamma(t) = \exp(-e^{-t})$ we obtain the Cox regression model. Assuming random censoring, the interesting part of the likelihood of the survival data given X is

$$L_{T\wedge C,1_{\{T\le C\}}|X}(\Lambda,\gamma,\beta) = \begin{cases} \bar{F}_\gamma(\log\Lambda(T\wedge C)+\beta X) & \text{if } 1_{\{T\le C\}} = 0 \\ \bar{F}'_\gamma(\log\Lambda(T\wedge C)+\beta X)\, d\Lambda(T\wedge C) & \text{if } 1_{\{T\le C\}} = 1 \end{cases}$$

To calculate the likelihood based on the survival data and the coarse observation of X, we need to choose a reference measure. The simplest choice is to use a reference measure which makes the survival data and the covariate independent. Thus we let

$$\bar{L}_{T\wedge C,1_{\{T\le C\}}|X}(\Lambda,\gamma,\beta) = \frac{L_{T\wedge C,1_{\{T\le C\}}|X}(\Lambda,\gamma,\beta)}{L_{T\wedge C,1_{\{T\le C\}}|X}(\Lambda_0,\gamma_0,0)}$$

for some suitable choice of Λ_0 and γ_0 in the model. When X is coarsened at random and independent of C, the likelihood of the observed data becomes

$$\begin{aligned} &L_{T\wedge C,1_{\{T\le C\}},Y}(\Lambda,\gamma,\beta,f) \\ &\quad = E_0\left[\bar{L}_{T\wedge C,1_{\{T\le C\}}|X}(\Lambda,\gamma,\beta)f(X)\,\middle|\, T\wedge C, 1_{\{T\le C\}}, Y = y\right] \end{aligned} \tag{5}$$

where the conditional expectation is taken with respect to the chosen reference measure, and f is the density of the marginal distribution of X with respect to this reference measure. As the marginal distribution of X is unknown, f is an unknown parameter either in a finite dimensional space (a parametric family) or an infinitely dimensional space (a semi- or non-parametric model). As in Theorem 1 we see that the likelihood (5) may be written

$$L_{T\wedge C,1_{\{T\le C\}},Y}(\Lambda,\gamma,\beta,f) = \begin{cases} \dfrac{\int_y L_{T\wedge C,1_{\{T\le C\}}|x}(\Lambda,\gamma,\beta)f(x)dP_0(x)}{P_0\{X\in y\}} & \text{if } P_0\{X\in y\} > 0 \\ \bar{L}_{T\wedge C,1_{\{T\le C\}}|X}(\Lambda,\gamma,\beta)f(X) & \text{if } Y = \{X\} \end{cases}$$

When maximising the likelihood function (5) we replace Λ by a step function with steps at observed deaths.

If C is not independent of X, the censoring mechanism must be included in (5) as discussed in Section 3.2.

An alternative to this full maximum likelihood approach is the conditional profile likelihood approach suggested for survival data with missing covariates by Chen and Little (2001). The idea here is to reparameterise:

$$(\gamma,\beta,\Lambda,f) \longrightarrow (\gamma,\beta,R,f)$$

where

$$R(t) = \tau_{\gamma,\beta,f}(\Lambda)(t) = E_0\left[\bar{F}_\gamma(\log\Lambda(t)+\beta X)f(X)\right]$$

is the marginal survival function of T. As above the expectation is with respect to the reference measure. Assuming again that C is independent of X, censoring is ignorable and R may be estimated from the observed survival data by the usual Kaplan-Meier estimator. Let $\hat{\Lambda}_{\gamma,\beta,f}$ be the result of applying $\tau^{-1}_{\gamma,\beta,f}$ to this estimator. Then the remaining parameters may be estimated from the conditional profile likelihood

$$\frac{L_{T\wedge C,1_{\{T\leq C\}},Y}(\Lambda,\gamma,\beta,f)}{L_{T\wedge C,1_{\{T\leq C\}},E}(\Lambda,\gamma,\beta,f)}$$

Simulations reported by Chen and Little (2001) in the missing covariate case indicate that the loss of efficiency compared to full maximum likelihood estimation is negligible. Note however that if the censoring is not independent of X, then specifying a model for the censoring will not help: We need censoring to be ignorable in the marginal model of the survival data.

In both cases the EM algorithm may be useful for the actual maximisation as may Monte Carlo methods. As in subsection we may estimate f from Y if Y is independent of (T,C) given X and plug it into the likelihood or the profile likelihood.

Both approaches has some clear disadvantages. Firstly, it requires independence of C and X, which in applications may be unreasonable. Alternatively, in the full maximum likelihood approach the censoring mechanism must be specified and estimated too, but this will not help us in the conditional profile likelihood approach. Secondly, it requires the marginal distribution of X which is a disadvantage unless X is discrete. Usually we are not interested in this part of the model and would therefore prefer to leave it unspecified. Furthermore, as X is coarsely observed, specifying and checking a model for the marginal distribution of X may be difficult. Finally for the conditional profile likelihood approach we need to be able to invert $\tau_{\gamma,\beta,f}$. The advantage of these methods is that we would expect a high degree of efficiency of both methods.

With a parametric model for the survival data (i.e. with Λ either known or known up to a finite-dimensional parameter) the full maximum likelihood approach can still

be applied. The conditional profile likelihood approach, however, will typically not be useful as the corresponding marginal survival function, R, will now be restricted by the parametrisation and therefore difficult to estimate directly.

Notes

The EM algorithm for Cox's proportional hazards model with missing covariates has been discussed by a number of authors. Martinussen (1999) and Chen and Little (1999) consider the full likelihood function as done in this paper, whereas Lipsitz and Ibrahim (1998) and Herring and Ibrahim (2001) apply the EM algorithm to the partial likelihood function. The latter two papers also consider the use of Monte Carlo methods in connection to the EM algorithm. See Zhou and Pepe (1995) and Zhou and Wang (2000) for a non-parametric approach. The presentation given here is mainly based on Chen and Little's (1999, 2001) work.

5.2 Weighted estimating equations

Another approach to inference in survival analysis is to use martingale estimating functions, i.e. functions like

$$M_s(\theta) = \int_0^s W_s(X,\theta) d\big(N - \Lambda(X;\theta)\big)(s), \quad s \geq 0 \tag{6}$$

where N is the counting process generated by the data, $\Lambda(X;\theta)$ is the integrated hazard, and $W_s(X;\theta)$ is a predictable process (see e.g. Gill (1984) for an introduction). Many popular regression models can be handled in this way. Chen and Newell (2001) consider models with hazards given by

$$\lambda(t|x) = \alpha\big(t \cdot \exp(\gamma X)\big) \cdot \exp(\beta X).$$

Cox regression ($\gamma = 0$) and accelerated failure time ($\beta = 0$) models are obtained as special cases. With α, β and γ unknown, $\theta = (\alpha, \beta, \gamma)$.

If Y is independent of the survival data, $(T \wedge C, 1_{\{T \leq C\}})$, given X, then a complete case analysis works. This corresponds to using the estimating function

$$\Delta M_s(\theta), \quad s \geq 0,$$

where $\Delta = 1$ if X is completely observed, 0 otherwise. Another option is to weight this estimating equation by the "inverse probability" of X being completely observed:

$$\frac{\Delta}{q_{\{X\}}} M_s(\theta), \quad s \geq 0, \quad \text{where } q_{\{X\}} = P\big\{Y = \{X\} \big| X\big\} = E[\Delta|X] \tag{7}$$

The weighted estimating function (7) is unbiased and should therefore yield consistent estimators of the parameters of interest. Generally $q_{\{X\}}$ will be unknown and must be

estimated from the data. It is a part of the coarsening mechanism and may be estimated from the conditional likelihood of S (or Y) given X, which is given by q_y in the discrete case or generally by $k(s|x)$ for any $x \in y$.

Example 7 *To show the effect of CAR on the estimation of $q_{\{X\}}$ we will consider two examples.*

- *Suppose that X is right censored with censoring variable G. Then $q_{\{x\}} = P\{X > G|X = x\} = 1 - \int_0^x h(g)dg$ where h is the conditional density of G given $X = x$ which does not depend on x when we consider $g < x$. One should note that G is not necessarily censored at random; to have both X and G censored at random would require independence of X and G. Note also that h is not the marginal density of G, unless X and G are independent. In fact h may not even be a density function; it is possible that $\int_0^\infty h(g)dg < 1$. To estimate $q_{\{X\}}$ observe that the conditional likelihood of S given $X = x$ can be written*

$$h(g)^{1-\Delta}\left(1 - \int_0^x h(g)dg\right)^{\Delta} = (-dq_{\{g\}})^{1-\Delta} {q_{\{x\}}}^{\Delta}$$

  In a non-parametric setting, it would be natural to estimate $q_{\{x\}}$ by a decreasing step function with jumps at the observed values of $X \wedge G$.

- *If Y is a heaping then $q_{\{x\}}$ may be estimated by the fraction of unheaped observations with $\lfloor X/c \rfloor = \lfloor x/c \rfloor$, as*

$$q_{\{x\}} = 1 - P\{Y = y|X = x\} = 1 - P\left\{Y = y \middle| X^* = c\left\lfloor \frac{x}{c} \right\rfloor\right\}$$

  for $y = \left\{c\lfloor \frac{x}{c} \rfloor, c\lfloor \frac{x}{c} \rfloor + 1, \ldots, c\lfloor \frac{x}{c} \rfloor + c - 1\right\}$. □

There is no guarantee that this weighting will lead to improved estimators. As such we are still only using complete cases to estimate the parameters of interest even if all the data is used to estimate $q_{\{X\}}$. Dividing by $q_{\{X\}}$ will typically improve the precision of the estimating function but also increase its variance. There appears to be no known sufficient condition to decide if weighting improves the estimator or not. However, estimating the weights, $q_{\{x\}}$, may actually improve the estimator of θ. In fact, the asymptotic variance of θ will not increase but may well decrease as more parameters are included in the specification of $q_{\{x\}}$. Indeed, letting $q_{\{x\}}$ depend also on the survival data $T \wedge C$ and $1_{\{T \leq C\}}$ may improve the estimation of θ as the following example shows. We should however keep in mind that the complete case estimator may perform better still.

Table 3: *Inverse probability weighting. Efficiency is calculated with respect to the MLE from Table 2.*

Method	Mean	StDev	Efficiency
Complete cases	1.012	0.1323	64%
True weights	1.014	0.1376	59%
Estimated weights I	1.014	0.1378	59%
Estimated weights II	1.030	0.1186	79%

Example 8 *Consider again the censored exponential survival times. In Table 3 we compare the complete case estimator of θ to various estimators of θ obtained from weighted estimating equations with weights either known ("True weights") or estimated using a model only depending on X (the "true" model; "Estimated weights I") as well as using a model with dependence on X and the survival data $(T \wedge C, 1_{\{T \leq C\}})$ ("Estimated weights II"). We see that in this example complete cases do as well as –if not better than– weighted estimating functions with weights known or estimated using the true model. However using the larger model there is a considerable gain of efficiency.* □

A further advantage of this inverse probability weighting approach is that if the coarsening mechanism depend on T and/or C, we can incorporate this by allowing $q_{\{X\}}$ to depend on the survival data $(T \wedge C, 1_{\{T \leq C\}})$:

$$\begin{aligned} q_{\{X\}} = q(X, T \wedge C, 1_{\{T \leq C\}}) &= E\left[\Delta | X, T \wedge C, 1_{\{T \leq C\}}\right] \\ &= P\left\{Y = \{X\} | X, T \wedge C, 1_{\{T \leq C\}}\right\} \end{aligned}$$

Using these weights, the weighted estimating function (7) is still unbiased and should therefore yield consistent estimators.

Still it is not quite satisfactory that the estimation is based on complete cases only even if some improvement due to the estimation of $q_{\{X\}}$ may be expected. Some improvement may be obtained by finding the optimal estimating function (7), where the optimisation is performed over the predictable function W. The optimal W will typically depend on the coarsening mechanism and may therefore be unobtainable in practice. Even for the original estimating function (6) the optimal W may be difficult to obtain; see Chen and Newell (2001).

A further improvement on (7) is to add terms of mean 0 to the estimating functions:

$$\frac{\Delta}{q_{\{X\}}} M_s(\theta) + (1-\Delta)\phi_s(\theta) - \frac{\Delta}{q_{\{X\}}} E\left[(1-\Delta)\phi_s(\theta) | X, T \wedge C, 1_{\{T \leq C\}}\right] \tag{8}$$

for some function $\phi_s(\theta) = \phi_s(Y, T \wedge C, 1_{\{T \leq C\}}; \theta)$. By construction the added term has expectation 0 regardless of θ so that the estimating function (8) will be an unbiased estimating function with the same precision as the simpler weighted estimating function

(7) but lower variance if the added term has a small variance but a large negative correlation with $\frac{\Delta}{q_{\{X\}}} M_s(\theta)$ (see Nielsen (1998) for details). Nielsen (1998) discuss optimal choice of ϕ for semi-parametric regression models with coarsely observed regressors; it seems likely that his results may be generalised to the problem and the estimating functions considered here. If so, optimal choices of ϕ and W exist (leading to efficient estimates), but they depend on the coarsening mechanism as well as the distribution of X given $(Y, T \wedge C, 1_{\{T \leq C\}})$; thus in practice the optimal choices of ϕ and W will hardly be possible to obtain.

It may still be a good idea to add a term, though. One suggestion would be to simply replace the coarsened X in the original estimating function by a suitably chosen value X^* in the coarsening Y, i.e. use

$$\phi_s(\theta) = \int_0^s W_s(X^*, \theta) d(N - \Lambda(X^*; \theta))(s)$$

For instance, if X is censored we could use $X^* = X \wedge G$. Still, it should be noted that $E[(1-\Delta)\phi_s(\theta) | X, T \wedge C, 1_{\{T \leq C\}}]$ in most cases will be extremely hard to find.

Example 9 *We apply this idea to the censored exponentials of the previous examples using $X^* = 2$, 2.5 and 3. The results are reported in Table 4; the first row uses estimated weights depending on X only, the second row weights depending on X and the survival data $(T \wedge C, 1_{\{t \leq C\}})$. We see that adding a term leads to a considerable improvement over the complete case estimator (see Table 3); in fact the efficiency is very close to the efficiency of the maximum likelihood estimator. Furthermore, the choice of X^* does not seem to matter very much. Also the benefit of using a large model for $q_{\{X\}}$ appears to be almost negligible when a term is added to the estimating function.* □

Table 4: *Estimation with added terms: First row uses weights estimated from the correct model, the second row weights estimated from an extended model. Efficiency is calculated with respect to the MLE from Table 2.*

	Estimated weights I			Estimated weights II		
Method	Mean	StDev	Efficiency	Mean	StDev	Efficiency
$X^* = 2$	1.009	0.1067	98%	1.009	0.1065	99%
$X^* = 2.5$	1.008	0.1066	98%	1.008	0.1066	98%
$X^* = 3$	1.008	0.1084	95%	1.008	0.1070	98%

The obvious disadvantage of this approach is that it requires modelling of the coarsening mechanism, at least to the level of modelling the probability, $q_{\{X\}}$, of X being completely observed. Also, as indicated by the simulations it may be as inefficient as the complete case analysis unless additional terms, which depend on the coarsening mechanism, are added to the simple estimating function. Furthermore, we need $q_{\{X\}} > 0$ for all values of X ruling out application to e.g. a covariate that is unobserved due to

a fixed detection limit. The advantage of this approach is that it actually allows us to estimate the parameters of interest even if the censoring depends on the covariate X without modelling the censoring mechanism or the marginal distribution of X.

Notes

Inverse probability weighting for Cox regression with missing covariates is considered by Pugh, Robins, Lipsitz and Harrington (1993); see also Robins, Rotnitzky and Zhao (1994). Nielsen (1998) considers inverse probability weighting for regression models with coarsely observed covariates.

5.3 Bias-variance trade-off

In many cases a complete case analysis will yield consistent but inefficient estimates. The two approaches discussed in the previous subsections improve the efficiency of the estimators but do this at the cost of much additional work. Moreover they both need specification and estimation of nuisance parts, either the marginal distribution of the covariate or the coarsening mechanism. Another option would be to allow a certain amount of bias in the estimators if the decrease in variance is sufficiently large. Thus in some cases it may be worth considering simply to replace the coarsened value of X by X^* suitably chosen in the observed coarsening Y. Unlike the case of missing covariates, the coarsening Y may give a very precise idea about the unobserved value of X. Of course, we should expect this imputation approach to lead to biased estimators but also in a reduction of variance compared to the complete case analysis since we are now using all cases. Furthermore, it will be a lot simpler than the methods discussed in the previous subsections. In small samples the reduction in variance may be enough to reduce the mean squared error. However, as the sample size increases the variance will decrease but the bias will not disappear. Hence in large samples the bias will dominate the mean squared error making this approach unacceptable. We illustrate the potential benefits by a simulation example.

Example 10 *Again we consider the censored exponentials. If we impute either* 2 *or* 3 *when we observe* $Y = \{2,3\}$*, we get an bias but also a reduction of variance. When we impute* 3*, the reduction is sufficient to make the mean squared error smaller for the biased estimator than for the complete case estimator; see Table 5. If we impute 2.5 there is no bias and the mean squared error is similar to the mean squared error of the MLE reported in Subsection 3.2.*

If the sample size increases to 1000 then we get worse results. We get roughly the same bias as before but as the variance is smaller, the mean squared errors of the biased estimators (imputing 2 or 3) are now larger than the mean squared error of the complete

case estimator. Imputing 2.5 is still a good idea, though. This is due to 2.5 being the conditional mean of X given $Y = \{2,3\}$; results by Schafer and Schenker (2000) suggest that the resulting estimator is consistent. □

Table 5: *Imputation.*

	$n = 200$			$n = 1000$		
Method	Mean	StDev	MSE	Mean	StDev	MSE
Complete cases	1.012	0.1323	0.0176	1.000	0.0591	0.0035
Impute $X^* = 2$	1.084	0.1128	0.0198	1.075	0.0501	0.0081
Impute $X^* = 2.5$	0.999	0.1041	0.0108	0.991	0.0463	0.0022
Impute $X^* = 3$	0.927	0.0975	0.0148	0.920	0.0434	0.0083

Of course this example is "nice" as fairly little information is lost in the coarsening. How this approach will work more generally is difficult to predict but given its simplicity, it should be considered as an option in small data sets with "small" coarsenings –i.e. coarsenings where Y is a small set–, where a good idea of the true value of X is available and modelling of nuisance parts may be problematic.

Notes

Imputation has a long tradition as a tool for handling missing or incomplete data; see e.g. Little and Rubin (1987) for an overview. The imputations suggested in this section are naïve and as a consequence they introduce bias. It is possible to construct imputations which will lead to consistent estimators but this will of course make the method more computationally complicated. One possibility is to impute conditional means; this is considered for Cox's proportional hazards model with missing covariates by Paik and Tsai (1997). Another is to generate random imputations for instance by resampling complete cases as done by Paik (1997).

6 Conclusions

In this paper we have considered inference for survival data with incompletely observed covariates. We have discussed how ignorable incompleteness may be modelled using random coarsenings and looked at various methods of estimation in these models.

Throughout the paper we have illustrated the estimation methods by a simple simulation example: Exponential regression with independent censoring and a simple coarsening mechanism affecting on average one third of the observations. In this simple example, we have seen that even though a complete case analysis leads to consistent

estimators, the loss of information is considerable, and there is a lot to be gained by incorporating cases with incomplete covariates.

All the methods discussed have their own advantages and disadvantages. Most involve some degree of modelling of nuisance parts, either the coarsening mechanism or the marginal distribution of the covariates. Some are very inefficient or yield inconsistent estimators. Which method to use depends on which –if any– nuisance part is easier/safer to specify balanced with the need for efficiency and consistency.

We have also seen how incompleteness in the covariates affect the ignorability of the censoring: If censoring depends on incompletely observed covariates, then it is not (generally) ignorable. This has consequences for most of the methods of estimation we discuss: If the censoring is not ignorable, it needs to be modelled and estimated in order for us to estimate the parameters of interest. The only exception to this rule is the inverse probability weighted estimating equations discussed in Section 5.2. In its simplest form, however, this method may be as inefficient as a complete case analysis, and the conditional expectation needed for the possibly more efficient version (8) will be very difficult if not impossible to calculate in practice.

7 Details on the simulations

All simulations in this paper are done using the statistical programming language `R` (Ihaka, R. and Gentleman, R. (1996), www.r-project.org), version 1.5, running on a 1133 MHz Intel Pentium III computer under Suse Linux. The simulations for $n = 200$ were all done in a single function (CPU-time: 13 minutes, 28.65 seconds). More user-friendly functions can be found on `www.stat.ku.dk/~feodor/publications/survival.R`. Approximate CPU-times for the results in Tables 1-4 are respectively $4'58.76''$, $6'33.24''$, $5'43.71''$, $4'58.76''$, and $2'42.47''$ using the user-friendly but less efficient programs and simulating new datasets for each table.

8 References

Billingsley, P. (1979). *Probability and measure*. Wiley: New York.

Chen, H. Y. and Little, R. J. (1999). Proportional hazards regression with missing covariates. *Journal of the American Statistical Asociation*, 94, 896-908.

Chen, H. Y. and Little, R. J. (2001). A profile conditional likelihood approach for the semiparametric transformation regression model with missing covariates. *Lifetime Data Analysis*, 7, 207-224.

Chen, Y. C. and Newell, N. P. (2001). On a general class of semiparametric hazards regression models. *Biometrika*, 88, 687-702.

Dempster, A. P., Laird, N. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39, 1-38.

Gill, R. D. (1984). Understanding Cox's regression model: a martingale approach. *Journal of the American Statistical Association*, 79, 441-447.

Gill, R. D., van der Laan, M. and Robins, J. (1997). Coarsening at random: characterisations, conjectures and counter-examples *in* D.-Y. Lin (ed.), *Proc. First Seattle Conference on Biostatistics*. Springer New York, pp. 255-294.

Heitjan, D. F. (1993). Ignorability and coarse data: Some biomedical examples. *Biometrics*, 49, 1099-1109.

Heitjan, D. F. and Rubin, D. B. (1991). Ignorability and coarse data. *Annals of Statistics*, 19, 2244-2253.

Herring, A. H. and Ibrahim, J. G. (2001). Likelihood-based methods for missing covariates in the Cox proportional hazards model. *Journal of the American Statistical Society*, 96, 292-302.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5(3), 299-314.

Jacobsen, M. and Keiding, N. (1995). Coarsening at random in general sample spaces and random censoring in continuous time. *Annals of Statistics*, 23, 774-786.

Lipsitz, S. R. and Ibrahim, J. G. (1998). Estimating equations with incomplete categorical covariates in the Cox model. *Biometrics*, 54, 1002-1013.

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

Martinussen, T. (1999). Cox regression with incomplete covariate measurements using the EM-algorithm. *Scandinavian Journal of Statistics*, 26, 479-492.

Nielsen, S. F. (1998). Semi-parametric regression with coarsely observed regressors. *Preprint 3*. Department of Theoretical Statistics. `http://www.stat.ku.dk/f̃eodor/publications/`

Nielsen, S. F. (2000). Relative coarsening at random. *Statistica Neerlandica*, 54, 79-99.

Paik, M. C. (1997). Multiple imputation for the Cox proportional hazards model with missing covariates. *Lifetime Data Analysis*, 3, 289-298.

Paik, M. C. and Tsai, W.-Y. (1997). On using Cox proportional hazards model wih missing covariates. *Biometrika*, 84, 579-595.

Pugh, M., Robins, J., Lipsitz, S. and Harrington, D. (1993). Inference in the Cox proportional hazards model with missing covariates. *Technical report*. Department of Biostatistics, Harvard School of Public Health.

Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1994). Estimation of regression coeeficients when some regressors are not always observed. *Journal of the American Statistical Asociation*, 89, 846-866.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581-590.

Schafer, J. L. and Schenker, N. (2000). Inference with imputed conditional means. *Journal of the American Statistical Association*, 95, 144-154.

Zhou, H. and Pepe, M. S. (1995). Auxillary covariate data in failure time regression. *Biometrika*, 82, 139-149.

Zhou, H. and Wang, C.-Y. (2000). Failure time regression with continuous covariates measured with error. *Journal of the Royal Statistical Society, Series B*, 62, 657-665.

Resum

En aquest treball considerem anàlisis de supervivència amb informació incompleta sobre les covariàncies. Proposem un model per a les covariàncies com una agrupació aleatòria (random coarsening) de la covariància complet, i donem una panoràmica de la teoria de l'agrupació aleatòria (random coarsening). Diverses formes d'estimar els paràmetres del model per a les dades de supervivència donades les covariàncies es discuteixen i es comparen.

MSC: 62N01, 62N02

Paraules clau: Algorisme EM; dades incompletes; eficiència; funció d'estimació de martingala; ponderació inversa de la probabilitat