Martin Jacobsen Niels Keiding

Random Censoring and Coarsening at Random

Preprint
October
1991
7

Institute of Mathematical Statistics
University of Copenhagen

ISSN 0902-8846

Martin Jacobsen and Niels Keiding

RANDOM CENSORING AND COARSENING AT RANDOM

Preprint 1991 No. 7

INSTITUTE OF MATHEMATICAL STATISTICS
UNIVERSITY OF COPENHAGEN

October 1991

RANDOM CENSORING AND COARSENING AT RANDOM

BY MARTIN JACOBSEN AND NIELS KEIDING

University of Copenhagen

Heitjan and Rubin (1991) recently proposed a concept of ignorable censoring as a special case of a wider concept "coarsening at random". This note suggests a possible interpretation in terms of the modern theory of random censoring.

AMS 1980 *subject classifications*: Primary 62A10, Secondary 62F99, 62G99

Keywords and phrases. Censored data, counting processes, likelihood, survival analysis.

The modern theory of right censoring in survival analysis 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 et al. (1988), Arjas (1989) and Jacobsen (1989).

Recently Heitjan and Rubin (1991) and Heitjan (1991) generalized the theory of missing data of Little and Rubin (1987) to a concept called "coarsening at random" which includes as a special case a formalization of the above property of right censoring patterns.

This note indicates a connection between these two lines of development.

Coarsening at random. Heitjan and Rubin (1991) considered discrete random variables, always with density w.r.t. counting measure. Let X be the random variable of primary interest with values in χ according to a density $f(x;\ \theta)$. The variable X is not observed directly, it is only known that $X \in Y \subset \chi$. The coarsening Y is assumed to be determined by a random variable G, so that $Y=Y(X,G)$, and the conditional distribution of G given X is assumed to be determined by a density $h(g;x,\gamma)$. The conditional distribution of Y given $X=x$ and $G=g$ is degenerate:

$$r(y;x,g) = P\{Y=y|X=x,\ G=g\} = I\{y=Y(x,g)\};$$

let

$$k(y;x,\gamma) = \int r(y;x,g)h(g;x,\gamma)dg$$

denote the density of the implied distribution of Y given X.

The condition for coarsening a random was given in a Bayesian flavour by focusing on the 'fixed, observed value of y' for which (for each value of γ) $k(y;x,\gamma)$ *is assumed to take the same value for all values of* x *that are consistent with the observed coarse data y.*

Heitjan and Rubin (1991) proved that a stochastic coarsening mechanism may be 'ignored' for Bayesian and likelihood inferences if the data are coarsened at random and the parameters θ and γ are 'distinct' (i.e. they are a priori independent for Bayesian inference and lie in product parameter spaces for likelihood-based inference). That the coarsening mechanism may be ignored means that the likelihood is proportional to that resulting from assuming the coarsening mechanism to be deterministic.

From discrete to general random variables. Heitjan and Rubin restricted attention to discrete random variables X and G, so that all distribution may be described through their point probabilities. There is no need for this kind of restriction, and we shall now sketch the mainly technical points involved in obtaining a generally applicable version of Heitjan and Rubin's result on coarsening. This generalization will be needed for the censoring example we shall discuss below.

Let X and G be random variables with values in $(\Xi,\mathcal{X})$ and $(\Gamma,\mathcal{G})$ respectively. Also, let μ be a positive reference measure on $(\Xi,\mathcal{X})$ and ν a positive reference measure on $(\Gamma,\mathcal{G})$. The product measure $\rho=\mu\otimes\nu$ is then used as a reference measure on the product space $(\Xi\times\Gamma, \mathcal{X}\otimes\mathcal{G})$. Finally, consider a mapping $Y: \Xi\times\Gamma\to S$, where $S=2^{\Xi}$ is the set of all subsets of Ξ. Make this coarsening variable measurable by using the induced σ–algebra

$$\mathcal{S} = \{D \subset \Xi: Y^{-1}(D) \in \mathcal{X} \otimes \mathcal{G}\}$$

on S, and define the reference measure σ on $(S,\mathcal{S})$ by

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

It should be noted that the measurable structure on S, is not determined from that on Ξ alone, but is defined in such a way as to respect the properties of Y. This seems essential to arrive at a general coarsening result.

In the non–Bayesian case, the statistical model considered by Heitjan and Rubin, which specifies the possible joint distributions of X and G, is obtained using a product parameter space $\Theta \times \Psi$ such that for $\theta \in \Theta$, $\gamma \in \Psi$ the (θ,γ)-distribution of X depends on θ only and has density

$$f(x;\theta)$$

with respect to μ, while given X=x, the (θ,γ)–conditional distribution of G depends on γ only and has density

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

with respect to ν. Consequently, the (θ,γ)–joint distribution of (X,G) has density

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

with respect to ρ, and the model is obtained by considering all such densities for $\theta \in \Theta$, $\gamma \in \Psi$.

Consider now the (θ,γ)–conditional distribution of Y given X=x, which depends on γ only. We shall assume, – and this is a genuine assumption –, that this conditional distribution is absolutely continuous with respect to σ with density

$$k(y;x,\gamma) \ ,$$

at least for θ–almost all x. Thus the (θ,γ)–marginal density for Y becomes

$$\int_{\Xi} \mu(dx)\, f(x;\theta)\, k(y;x,\gamma) \ . \tag{1}$$

To formulate Heitjan and Rubin's result, one all important condition on the coarsening Y must hold:

(2) for all x and g, $x \in Y(x,g)$.

With this assumption (1) reduces to

$$\int_{y} \mu(dx)\, f(x;\theta)\, k(y;x,\gamma)$$

and the following result is an immediate consequence:

<u>Theorem</u>. If for all θ,γ and (θ,γ) – almost all y, $k(y;x,\gamma)$ is the same for γ – almost all $x \in y$, the likelihood (1) for observing y is, apart from a proportionality factor depending on y and γ only, the same as

$$\int_{y} \mu(dx)\, f(x;\theta) \ .$$

Censoring as coarsening. Specializing to right censoring, assume that $X=(X_1,...,X_n)$ where $X_1,...,X_n$ are iid positive absolutely continuous random variables with density function $f_*(x_1;\theta)$. Let $G=(G_1,...,G_n)$ where $G_1,...,G_n$ are positive absolutely continuous censoring times so that only the 'coarsened data' $Y=(Y_1,...,Y_n)$ are observed, where

$$Y_i = \begin{cases} \{X_i\}\ , & X_i \le G_i \\ \{u > G_i\}\ , & X_i > G_i \end{cases} .$$

Following Heitjan's (1991) formulation the G_i are assumed iid, although that does not seem to be essential. Under this assumption the joint density of $X=(X_1,...,X_n)$ and $G=(G_1,...,G_n)$ may be written

$$f(x;\theta)h(g;x,\gamma) = \prod_{i=1}^{n} f_*(x_i;\theta)h_*(g_i;x_i,\gamma)$$

and the distribution of the coarsened data Y given the original data is

$$k(y;x,\gamma) = \prod_{i=1}^{n} k_*(y_i;x_i,\gamma)$$

where

$$k_*(y_i;x_i,\gamma) = \begin{cases} 0 & \text{if} \quad x_i \notin y_i \\ \int_{x_i}^{\infty} h_*(u;x_i,\gamma)du & \text{if} \quad y_i=\{x_i\} \\ h_*(\inf(y_i);x_i,\gamma) & \text{if} \quad x_i \in y_i = (\inf(y_i),\infty) \end{cases} .$$

Indeed, if death is observed at x_i, we obtain $P\{G_i>x_i|X_i=x_i\}$ if the observation is censored at $g_i=\inf(y_i)$, $y_i=(g_i,\infty)$, then we obtain $P\{G_i\in dg_i|X_i=x_i\}$.

The coarsening mechanism may be ignored if, for $i=1,...,n$, $k_*(y_i;x_i,\gamma)$ is the same for all $x_i\in y_i$. For uncensored observations this is automatically fulfilled, because y_i contains only one point (the survival time x_i) while for censored observations the requirement is that $h_*(\inf(y_i);x_i,\gamma)$ must not depend on x_i for $x_i>\inf(y_i)$. Briefly put, $h_*(g_i;x_i,\gamma)$ may only depend on x_i through what is observed, which is $x_i\wedge g_i$, $I\{x_i\le g_i\}$.

Embedding into the counting process framework. Jacobsen (1989) also considered a basic set of iid positive random variables $X_1,...,X_n$. Jacobsen described the 'observation' as a marked point process N, which under the (simplifying but unnecessary) condition of absolutely continuous censoring times is equivalent to

$$N(t) = (N_1(t),...,N_n(t),\ N_1^G(t),...,N_n^G(t))$$

with

$$N_i(t) = I\{X_i \le t,\ X_i \le G_i\}$$

$$N_i^G(t) = I\{G_i \le t,\ X_i > G_i\}\ .$$

In other words, $N(t)$ keeps track of all observed events and censorings in $[0,t]$. The basic property of the model – that past observations do not affect the probabilities of future failures – may now be specified in terms of the compensator $\Lambda_i(t)$ of $N_i(t)$:

$$\Lambda_i(t) = \int_0^t \mu(u)\ I\{G_i \ge u,\ X_i \ge u\}du$$

where μ is the hazard of X_i (Jacobsen, 1989, (2.9)). Under this *martingale condition* the intensity of an event for a still uncensored individual is $\mu(t)$, exactly as in the uncensored situation.

Jacobsen gave three increasingly restrictive conditions on the joint distribution of (X,N) for the martingale condition to hold. We shall now see that coarsening at random implies the most restrictive of these, Jacobsen's condition C. (In a more specific example below one may actually show equivalence.) Briefly, Jacobsen's framework involves defining subspaces W_x of the space W of possible observations (i.e. sample paths of the above marked point process) where W_x is the set of $w \in W$ compatible with $x=(x_1,\ldots,x_n)$. For all x, P_x is a probability of W with $P_x(W_x)=1$, so that P_x has the interpretation of a conditional probability on the observed counting process $N=(N(t))$ given $X=x$. The compensator of N under P_x is denoted Λ_x, and we have a specification of its components in informal infinitesimal explanation, with $\mathcal{N}_t=\sigma\{N(u), 0\le u\le t\}$, by

$$\Delta \Lambda_{xi}(t) = P\{\Delta N_i(t) = 1 \mid \mathcal{N}_{t-}, X = x\}$$

$$= P\{X_i = t \mid \mathcal{N}_{t-}, X = x\} = \begin{cases} 1 \text{ if } N_i(t-)=0,\ N_i^G(t-)=0,\ x_i=t \\ 0 \text{ otherwise} \end{cases}$$

$$\Lambda_{xi}^G(dt) = P\{N_i^G(t+dt) - N_i^G(t) = 1 \mid \mathcal{N}_{t-}, X = x\}$$

$$=P\{t\le G_i<t+dt|\mathcal{N}_{t-},X=x\}=\begin{cases} \eta_*(t;x_i,\gamma)dt \text{ if } N_i(t-)=0,\ N_i^G(t-)=0,\ x_i>t \\ 0 \quad \text{otherwise} \end{cases}$$

where $\eta_*(t;x_i,\gamma)$ is the intensity (hazard) function corresponding to the density h_*. Jacobsen's condition (C) now requires that 'for any $t>0$ and $w \in W$, $\Lambda_x(t,w)$ is the same for all x which are t–compatible with w and satisfy that $x_i>t$ if $x_i\ge t$, $g_i\ge t$'.

Because of our independence assumptions we only need to consider x_i. That x_i is t–compatible with w means that if the counting process observed an event at time $t_0<t$, i.e. $\Delta N_i(t_0)=1$, then $x_i=t_0$, if there was a censoring at $t_0<t$, i.e. $\Delta N_i^G(t_0)=1$, then $x_i>t_0$, and if neither had happened before t, i.e. $N_i(t-)=N_i^G(t-)=0$, then $x_i\geq t$.

Now $\Delta\Lambda_{xi}(t)=0$ for all x_i for which either $x_i>t$ or $g_i<t$, so here the condition is immediately fulfilled. As regards $\Lambda_{xi}^G(dt)$, the condition of coarsening at random tells us that $h_*(g_i;x_i,\gamma)$ takes the same value for all x consistent with the observed data, which means that $h_*(g_i;x_i,\gamma)$ depends on x only through $x_i\wedge g_i$, $I\{x_i\leq g_i\}$. The same property is thus true for

$$\eta_*(t;x_i,\gamma) = h_*(t;x_i,\gamma)/[1-\int_0^t h_*(u;x_i,\gamma)du]$$

whenever $x_i>t$, and we have seen that Jacobsen's property (C) holds.

Jacobsen's (1989) paper contained a detailed study of the interdependence of Condition (C) with other conditions (one involving conditional independence) and also discussed conditions on the statistical model (here parameterized by θ and γ) for ignoring the censoring in likelihood inference.

Example. We shall discuss a model for observation of censored failure times, which allows for a general stochastic dependence between the failure times and the censoring times.

Let $X=(X_1,\ldots,X_n)$ be a vector of positive i.i.d. random variables (failure times) with common density $f(x,\theta)$ (with respect to Lebesgue measure). Also, let $G=(G_1,\ldots,G_n)$

be a vector of positive random variables (censoring times) with a joint conditional density (with respect to n–dimensional Lebesgue measure) given $X=x=(x_1,\dots,x_n)$

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

where $g=(g_1,\dots,g_n)$.

Suppose observations are made on the time interval $[0,t]$. The population $\{1,\dots,n\}$ splits into three sets, the set $D(t)$ of failed items, the set $C(t)$ of censored items and the set $R(t)$ of items still at risk, where

$$(*)\qquad \begin{aligned} D(t) &= \{i\colon X_i \le G_i \wedge t\} \ , \\ C(t) &= \{i\colon G_i < X_i,\ G_i \le t\} \ , \\ R(t) &= \{i\colon X_i \wedge G_i > t\} \ . \end{aligned}$$

For $i\in D(t)$, X_i is itself observed while for $i\in C(t)$, G_i is observed and it is known that $X_i>G_i$ and for $i\in R(t)$ it is known that $X_i>t$, $G_i>t$. Thus, for $x=(x_1,\dots,x_n)$, $g=(g_1,\dots,g_n)$ and sets $d(t)$, $c(t)$, $r(t)$ defined from x and g in analogy with (*), the partly unobserved vector x of failure times is known to belong to the set

$$Y(x,g) = \{x'\colon \quad \begin{aligned} x'_i &= x_i && \text{for} \quad i \in d(t) \ , \\ x'_i &> g_i && \text{for} \quad i \in c(t) \ , \\ x'_i &> t && \text{for} \quad i \in r(t)\} \ . \end{aligned}$$

Note that, with probability 1, $d(t)$, $c(t)$ and $r(t)$ are functions of the observed value y of Y: a misclassification can occur only if some $g_i=t$.

Using the notation from above, the reference measure ρ is 2n-dimensional Lebesgue measure, while the reference measure on the space of Y–values may informally be described as

$$\sigma(dy) = dx_{d(t)} dg_{c(t)} \ ,$$

where we write $x_I=(x_i)_{i \in I}$, $g_J=(g_j)_{j \in J}$. The conditional density for Y given X=x then becomes

$$k(y;x,\gamma) = \int_{\substack{g_i'>x_i,\ i \in d(t) \\ g_i'>t,\ i \in r(t)}} h(g'_{d(t)}, g_{c(t)}, g'_{r(t)}; x, \gamma)\ dg'_{d(t)} dg'_{r(t)}$$

with $d(t)$, $c(t)$, $r(t)$ the failure set, censoring set and risk set determined by y and $g_{c(t)}$ the vector of observed censoring times given by y.

In this case the general version of Heitjan and Rubin's coarsening condition given above, stipulates that for an arbitrary γ, $k(y;x',\gamma)$ must be the same for all $x' \in y$, i.e. x' such that

$$x_i' = x_i \text{ for } i \in d(t),\ x_i' > g_i \text{ for } i \in c(t),\ x_i' > t \text{ for } i \in r(t) \ .$$

We shall now argue that this condition holds simultaneously for all t, if and only if Jacobsen's condition (C) holds. It is a consequence of this condition that 'the intensities for failures are as they would be without censoring' (see the discussion above for the case of independence pairs (X_i, G_i)), and with the product parameterization used here (with one parameter for the failure time distributions and another for the conditional distribution of the censoring times given the failure times), Heitjan and Rubin's Theorem (as specified above) is then implied by Theorem 4.2 of Jacobsen (1989).

Returning to the discussion of condition (C), since in the model described here no two censoring times can agree, (C) specifies that for all t, given x and the observations on $[0,t]$, the conditional intensity for any i_0 to be censored in $[t,t+dt)$ depends on x only through the observed coarsening y. But if $d(t)$, $c(t)$, $r(t)$ are the observed failure, censoring and risk sets, with $x_{d(t)}$ the observed failure times and $g_{c(t)}$ the observed censoring times, it is quickly seen that if $i_0 \in r(t)$ (the only interesting case) the conditional intensity for censoring of i_0 just after t is

$$\frac{k(\tilde{y};x,\gamma)dt}{k(y;x,\gamma)}$$

where y is the observed value of Y and $\tilde{y}$ is the Y–value obtained from y by replacing $c(t)$ with $c(t) \cup i_0$, $r(t)$ with $r(t) \setminus i_0$, and letting t be the censoring time for i_0, i.e.

$$k(\tilde{y};x,\gamma) = \int\limits_{\substack{g_i' > x_i,\ i \in d(t) \\ g_i' > t,\ i \in r(t) \setminus i_0}} h(g'_{d(t)}, (g_{c(t)},t), g'_{r(t)\setminus i_0}; x,\gamma)\ dg'_{d(t)} dg'_{r(t)\setminus i_0}$$

It is now clear that Heitjan and Rubin's condition implies (C), and it is also fairly straightforward to argue the converse.

Acknowledgements. 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. thesis, Univ. of 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.

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.

Gill, R.D. (1980). *Censoring and Stochastic Integrals.* Mathematical Centre Tracts 124. Mathematisch Centrum, Amsterdam.

Heitjan, D.F. (1991). Ignorability and coarse data: some biomedical examples. Under revision for *Biometrics.*

Heitjan, D.F. and Rubin, D.B. (1991). Ignorability and coarse data. *Ann. Statist.* (to appear).

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.

Martin Jacobsen
Institute of Mathematical Statistics
University of Copenhagen
Universitetsparken 5
DK–2100 Copenhagen Ø
Denmark

Niels Keiding
Statistical Research Unit
University of Copenhagen
Blegdamsvej 3
DK–2200 Copenhagen N
Denmark

PREPRINTS 1990

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 31 35 31 33.

No. 1 Johansen, Søren and Juselius, Katarina: Some Structural Hypotheses in a Multivariate Cointegration Analysis of the Purchasing Power Parity and the Uncovered Interest Parity for UK.

No. 2 Tjur, Tue: Analysis of Variance and Design of Experiments.

No. 3 Johansen, Søren: A Representation of Vector Autoregressive Processes Integrated of Order 2.

No. 4 Johansen, Søren: Cointegration in Partial Systems and the Efficiency of Single Equation Analysis.

PREPRINTS 1991

COPIES OF PREPRINTS ARE OBTAINABLE FROM THE AUTHOR OR FROM THE INSTITUTE OF MATHEMATICAL STATISTICS, UNIVERSITETSPARKEN 5, DK-2100 COPENHAGEN Ø, DENMARK, TELEPHONE + 45 31 35 31 33.

No. 1 Johansen, Søren: Determination of Cointegration Rank in the Presence of a Linear Trend.

No. 2 Johansen, Søren: A Statistical Analysis of Cointegration for I(2) Variables.

No. 3 Jacobsen, Martin: Homogeneous Gaussian Diffusions in Finite Dimensions.

No. 4 Johansen, Søren: Testing Weak Exogeneity and the Order of Cointegration in UK Money Demand Data.

No. 5 Johansen, Søren: An I(2) Cointegration Analysis of the Purchasing Power Parity between Australia and USA.

No. 6 Johansen, Søren: Estimating Systems of Trending Variables.

No. 7 Jacobsen, Martin and Keiding, Niels: Random Censoring and Coarsening at Random.