

Hazard Regression with Interval-Censored Data

BY CHARLES KOOPERBERG

Department of Statistics, University of Washington, Seattle, WA 98195-4322, U.S.A.

SUMMARY

In a recent paper, Kooperberg, Stone & Truong (1995a) introduced hazard regression (HARE), in which linear splines and their tensor products are used to estimate the conditional log-hazard function based on possibly censored, positive response data and one or more covariates. Model selection is carried out in an adaptive fashion, with maximum likelihood being used to estimate the unknown coefficients, Rao and Wald statistics to carry out stepwise addition and deletion of basis functions, and BIC to select the final model. In the present paper the HARE methodology is extended to accommodate interval-censored data. The likelihood function may no longer be concave, which causes additional numerical challenges. The extended methodology is applied to interval-censored data about the progression to AIDS and about cytology.

1 INTRODUCTION

Consider (survival) data involving a positive response variable that may be censored and one or more covariates. We think of the original, uncensored response variable as having a conditional density function given the values of the covariates that is positive on $[0, \infty)$.

A basic assumption of the proportional hazards model (Cox, 1972) is that the conditional log-hazard function is an additive function of time and the vector of covariates or, equivalently, that the conditional hazard function is a multiplicative function of time and the vector of covariates. In the hazard regression (HARE) methodology (Kooperberg, Stone & Truong, 1995a, hereafter referred to as KST) a practical approach to modeling the conditional hazard function that does not depend on the validity of this assumption is developed. In particular, a general framework for modeling the logarithm of the conditional hazard function with linear models is described. Maximum likelihood is used to estimate the unknown parameters of the model, and a fully automatic method involving stepwise addition, stepwise deletion and BIC is used to select the final model.

In HARE, linear splines and selected tensor products are used to estimate the logarithm of the conditional hazard function. The method is similar in spirit to MARS (Friedman, 1991). One advantage of HARE models is that they include proportional hazards models as a subclass. The presence or absence of interaction terms between covariates and time in the final model can in fact

be regarded as a check on the proportionality of the underlying hazard model.

Under suitable conditions, Kooperberg, Stone & Truong (1995b) obtained the L_2 rate of convergence for a nonadaptive version of the methodology treated in KST. This result lends theoretical support to HARE and, in particular, to the use of polynomial splines and their tensor products in defining the allowable spaces used in these procedures.

A limitation of the HARE methodology, as described in KST, is that it is applicable only to possibly right-censored data. If observations may also be interval-censored the methodology is no longer applicable. Here that limitation is removed. The organization of this paper is as follows: In the next section we describe the HARE model, including the extensions for interval-censored data, and we also summarize the model selection techniques developed in KST. Some examples of the use of the HARE methodology are described in Section 3. We conclude the paper with a few remarks.

There are several other nonparametric approaches to the modeling of conditional hazard functions. KST is one of the few papers that does not start with a proportional hazards model, but has it merely as a special case. For further discussion of the literature, see KST and Abrahamowicz, Ciampi & Ramsay (1992).

2 THE HARE MODEL

2.1 *Linear models for the conditional log-hazard function*

Let M be a positive integer, and let T be a positive random variable whose distribution may depend on M covariates x_1, \dots, x_M ranging over the subsets $\mathcal{X}_1, \dots, \mathcal{X}_M$ respectively of \mathbf{R} , each of which contains at least two members. Then $\mathbf{x} = (x_1, \dots, x_M)$ ranges over the subset $\mathcal{X} = \mathcal{X}_1 \times \dots \times \mathcal{X}_M$ of \mathbf{R}^M . Let $f(\cdot|\mathbf{x})$ denote the dependence on \mathbf{x} of the density function of T , which is assumed to exist and be positive on $(0, \infty)$. Let $F(\cdot|\mathbf{x})$, $\lambda(\cdot|\mathbf{x})$ and $\alpha(\cdot|\mathbf{x})$ denote the corresponding conditional distribution function, hazard function and log-hazard function, respectively.

Let T be the survival time and \mathbf{x} the vector of covariates for a randomly selected individual. Let $C = [C_l, C_u]$ be a (random) subinterval of $[0, \infty)$ so that it is known only that $T \in C$. If T is uncensored, then $C = \{T\}$; if T is right-censored at $C_l \leq T$, then $C = [C_l, \infty] = [C_l, \infty)$; if T is interval-censored, then $C = [C_l, C_u]$. It is assumed that T is independent of the censoring mechanism and that T has conditional density function $f(\cdot|\mathbf{x})$ given \mathbf{x} . Let $\delta = 0$ if T is right-censored, $\delta = 1$ if T is uncensored, and $\delta = 2$ if T is interval-censored. Note that the partial

likelihood corresponding to $C = (c_l, c_u)$, δ and \mathbf{x} equals

$$[f(c_l|\mathbf{x})]^{I(\delta=1)} \left[\int_{c_l}^{c_u} f(t|\mathbf{x}) dt \right]^{I(\delta \neq 1)} = [1 - F(c_l|\mathbf{x})]^{I(\delta=0)} [f(c_l|\mathbf{x})]^{I(\delta=1)} [F(c_u|\mathbf{x}) - F(c_l|\mathbf{x})]^{I(\delta=2)}.$$

Let $1 \leq p < \infty$, let G be a p -dimensional linear space of functions on $[0, \infty) \times \mathcal{X}$ such that $g(\cdot|\mathbf{x})$ is bounded on $[0, \infty)$ for $g \in G$ and $\mathbf{x} \in \mathcal{X}$, and let B_1, \dots, B_p be a basis of this space. Consider the model

$$\alpha(t|\mathbf{x}; \boldsymbol{\beta}) = \sum_{j=1}^p \beta_j B_j(t|\mathbf{x}), \quad t \geq 0, \quad (2.1)$$

for the conditional log-hazard function, where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$.

Given $\boldsymbol{\beta} \in \mathbf{R}^p$, set

$$\lambda(t|\mathbf{x}; \boldsymbol{\beta}) = \exp(\alpha(t|\mathbf{x}; \boldsymbol{\beta})), \quad t \geq 0,$$

and

$$f(t|\mathbf{x}; \boldsymbol{\beta}) = \exp(\alpha(t|\mathbf{x}; \boldsymbol{\beta})) \exp\left(-\int_0^t \exp(\alpha(u|\mathbf{x}; \boldsymbol{\beta})) du\right), \quad t \geq 0.$$

Observe that $f(\cdot|\mathbf{x}; \boldsymbol{\beta})$ is a positive density function on $[0, \infty)$. The corresponding conditional distribution function is given by

$$F(t|\mathbf{x}; \boldsymbol{\beta}) = 1 - \exp\left(-\int_0^t \exp(\alpha(u|\mathbf{x}; \boldsymbol{\beta})) du\right), \quad t \geq 0.$$

Now set

$$\phi(c_l, t, \delta|\mathbf{x}; \boldsymbol{\beta}) = \delta \alpha(c_l|\mathbf{x}; \boldsymbol{\beta}) + \phi(c_l|\mathbf{x}; \boldsymbol{\beta}), \quad c_l \geq 0, \text{ and } \delta \in \{0, 1\},$$

and

$$\phi(c_l, c_u, 2|\mathbf{x}; \boldsymbol{\beta}) = \log(\exp(\phi(c_l|\mathbf{x}; \boldsymbol{\beta})) - \exp(\phi(c_u|\mathbf{x}; \boldsymbol{\beta}))), \quad c_u > c_l \geq 0.$$

These are the contributions to the log-likelihood corresponding to $C = [c_l, c_u]$, \mathbf{x} and $\delta = 0$ or $\delta = 1$, and $\delta = 2$ respectively.

For fixed \mathbf{x} and $\boldsymbol{\beta}$ set

$$A(t) = \exp(\phi(t|\mathbf{x}; \boldsymbol{\beta})),$$

$$D_j(t) = -\int_0^t B_j(u|\mathbf{x}) \exp(\alpha(u|\mathbf{x}; \boldsymbol{\beta})) du, \quad 1 \leq j \leq p, t \geq 0,$$

and

$$E_{jk}(t) = -\int_0^t B_j(u|\mathbf{x}) B_k(u|\mathbf{x}) \exp(\alpha(u|\mathbf{x}; \boldsymbol{\beta})) du, \quad 1 \leq j, k \leq p, t \geq 0. \quad (2.2)$$

Then

$$\frac{\partial}{\partial \beta_j} \phi(c_l, t, \delta|\mathbf{x}; \boldsymbol{\beta}) = \delta B_j(c_l|\mathbf{x}) + D_j(c_l), \quad c_l \geq 0, \text{ and } \delta \in \{0, 1\},$$

$$\frac{\partial}{\partial \beta_j} \phi(c_l, c_u, 2 | \mathbf{x}; \boldsymbol{\beta}) = \frac{D_j(c_l)A(c_l) - D_j(c_u)A(c_u)}{A(c_l) - A(c_u)}, \quad c_u > c_l \geq 0,$$

and

$$\frac{\partial^2}{\partial \beta_j \partial \beta_k} \phi(c_l, t, \delta | \mathbf{x}; \boldsymbol{\beta}) = E_{jk}(c_l), \quad c_l \geq 0, \text{ and } \delta \in \{0, 1\}, \quad (2.3)$$

$$\begin{aligned} \frac{\partial^2}{\partial \beta_j \partial \beta_k} \phi(c_l, c_u, 2 | \mathbf{x}; \boldsymbol{\beta}) &= \frac{(E_{jk}(c_l) + D_j(c_l)D_k(c_l))A(c_l) - (E_{jk}(c_u) + D_j(c_u)D_k(c_u))A(c_u)}{A(c_l) - A(c_u)} \\ &\quad - \frac{(D_j(c_l)A(c_l) - D_j(c_u)A(c_u))(D_k(c_l)A(c_l) - D_k(c_u)A(c_u))}{(A(c_l) - A(c_u))^2}, \end{aligned}$$

$c_u > c_l \geq 0.$

It follows from (2.2) and (2.3) that $\phi(t, \delta | \mathbf{x}; \cdot)$ is a concave function on \mathbf{R}^p if $\delta \in \{0, 1\}$ for $t \geq 0$ and $\mathbf{x} \in \mathcal{X}$.

2.2 Maximum Likelihood Estimation

Consider n randomly selected individuals. For $1 \leq i \leq n$, let T_i be the survival time, $C_i = [C_{il}, C_{iu}]$ the censoring interval, and \mathbf{x}_i the vector of covariates for the i th such individual. The log-likelihood function corresponding to the observed data $(C_i, \delta_i, \mathbf{x}_i)$ and the linear model for the conditional log-hazard function that was discussed in the previous section is given by

$$\ell(\boldsymbol{\beta}) = \sum_i \phi(C_{il}, C_{iu}, \delta_i | \mathbf{x}_i; \boldsymbol{\beta}), \quad \boldsymbol{\beta} \in \mathbf{R}^p. \quad (2.4)$$

If $\delta_i \in \{0, 1\}$ for all i , this is a concave function on \mathbf{R}^p and the maximum likelihood estimate $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ can be found using the Newton–Raphson method (see KST). However, if $\delta_i = 2$ for some i , then $\ell(\boldsymbol{\beta})$ may not be concave. In this situation we modify the Newton–Raphson method slightly by adding a small positive constant times the identity matrix to the Hessian so that the modified Hessian is positive definite (see Kennedy & Gentle, 1980, Section 10.2.2).

2.3 Model selection

When modeling the log-hazard function with a linear model (2.1), the remaining issue to be resolved is the choice of G . Initially G is the space of constant functions. Thus we have $\alpha(t | \mathbf{x}; \boldsymbol{\beta}) = \beta_1$, for $t \geq 0$, so that α does not depend on t or the vector \mathbf{x} of covariates. Then we proceed with stepwise addition, by successively replacing a $(p-1)$ -dimensional space G_0 by a p -dimensional space G containing G_0 as a subspace. The functions that are candidates to be the new basis function (a function that together with a basis of G_0 spans G) depend on which functions are already in G .

Functions that are always allowed as basis functions of G are piecewise linear splines in time that are of the form $B(t) = (t_k - t)_+$, where $t_+ = \max(t, 0)$ and t_k is a fixed positive number, called a knot,

and linear functions in any of the covariates $B(x_m) = x_m$ for $1 \leq m \leq M$. When a linear function $B(x_m) = x_m$ is in G , piecewise linear splines in that covariate of the form $B(x_m) = (x_m - x_{mk})_+$ can also be basis functions; here x_{mk} is a fixed number in the range of x_m , called a knot. We also allow tensor products of any two basis functions in the model that depend on different (single) variables. Thus, for example, if $(t_k - t)_+$ and x_m are basis functions of G , $B(t, x_m) = (t_k - t)_+ x_m$ is also allowed in the model. There is one other rule: tensor products of a basis function $B(x_m) = (x_m - x_{mk})_+$ and any other basis function can only be a basis function when the tensor product between x_m and that basis function is in G .

Rao statistics are used to decide which candidate basis function should be added to the model next. Upon stopping the stepwise addition stage we proceed to stepwise deletion. Here we successively replace the p -dimensional space G by a $(p - 1)$ -dimensional subspace G_0 . The basis functions of the subspace G_0 have to satisfy the same rules listed above that applied during stepwise addition. In particular, at each step we choose among the candidate spaces G_0 such that the Wald statistic for a basis function that is in G but not in G_0 is smallest in magnitude. See KST for details.

During the combination of stepwise addition and stepwise deletion, we get a sequence of models indexed by ν with the ν th model having p_ν parameters. Let \hat{l}_ν denote the log-likelihood of the ν th model, and let

$$\text{AIC}_\alpha = -2\hat{l}_\nu + \alpha p_\nu \tag{2.5}$$

be the Akaike Information Criterion with penalty parameter α for this model. We select the model corresponding to the value $\hat{\nu}$ of ν that minimizes the Bayesian Information Criterion $\text{BIC} = \text{AIC}_{\log n}$. Section 3.2 contains a further discussion of the choice of the penalty parameter α in the presence of censoring.

Since the log-likelihood function may not be concave, good starting values for the (modified) Newton–Raphson iterations are even more crucial than usual. In the context of stepwise addition we use the maximum likelihood estimate from the previous model, which is possible since the new linear space contains the previous one as a proper subspace. In the context of stepwise deletion, the starting value $\hat{\beta}^{(0)}$ can be obtained by considering a quadratic approximation to the log-likelihood $\ell(\cdot)$. Note that this quadratic approximation to the likelihood function can also be used to motivate the use of the Wald statistic (Kooperberg, Bose & Stone, 1995). With these starting values, we have experienced few convergence problems other than during the first two or three steps of the stepwise addition procedure. Typically we overcome these problems by using a less refined convergence criterion at that stage, so that we can proceed with the next model.

3 EXAMPLES

3.1 Progression to Aids

Our first example concerns data from a cohort study of hemophiliacs. The participants in the study, all hemophilia patients, were subject to regular physical examinations. There were 674 patients who were HIV-positive at the beginning of the study or who turned HIV-positive during the study. The response we will examine is the length of the period between seroconversion and development of AIDS. Typically, for these patients it was known only that seroconversion (becoming HIV positive) took place in the interval (t_1, t_2) between two examinations. For 95 of the patients, AIDS was diagnosed at some stage during the study, say at time t_4 . For these patients we know that AIDS had developed between the time t_3 of the previous examination and t_4 . For these 95 patients there is thus an interval-censored response variable, with $C = (t_3 - t_2, t_4 - t_1)$. The other 579 patients had not yet developed AIDS at their last examination, t_5 . For these patients we have thus right-censored data, censored at $C_l = t_5 - t_2$. For the 95 interval-censored patients, the median times to the lower and upper end of the censoring interval were 1021 and 2016 days, respectively, and the median length of the interval was 488 days. The median follow-up time for the right-censored patients was 2506 days.

We included ten covariates in our study: age, five indicators for the race of the patient, three indicator variables for different types of hemophilia, and a severity index (1–3) for hemophilia. During a number of the physical examinations, various physiological quantities, such as the number of CD4 cells, were measured. However, since for many patients these measurements were not available at the time of initial diagnosis, we could not include them as (fixed) covariates in our study. An alternative would be to include them as time-varying covariates. However, the present implementation of HARE does not allow for time-varying covariates; moreover, the interpretation of models with time-varying covariates depends strongly on how they are modeled.

The model that HARE selected (see Table 1) included a constant term, a knot in time, and age as a linear term. Since there are no interaction terms between time and a covariate (in fact, there are no interaction terms at all), the HARE model is a proportional hazards model and it can be written explicitly as

$$\lambda(t|\mathbf{x}) = \exp(-11.19 - 0.00551(1011.5 - t)_+) \times \exp(0.0326 \text{ age}).$$

It is interesting that HARE chooses a proportional hazards model. (This never happened in the examples discussed in KST.) Actually, HARE examined many models that are not proportional hazards models during the model selection. The largest model that was fitted had a constant term,

Table 1. *HARE* analysis of the progression to AIDS data

Basis function	Coefficient	Standard error
1	-11.19	0.23
$(1011.5 - t)_+$	-0.00551	0.00199
Age	0.0326	0.0066

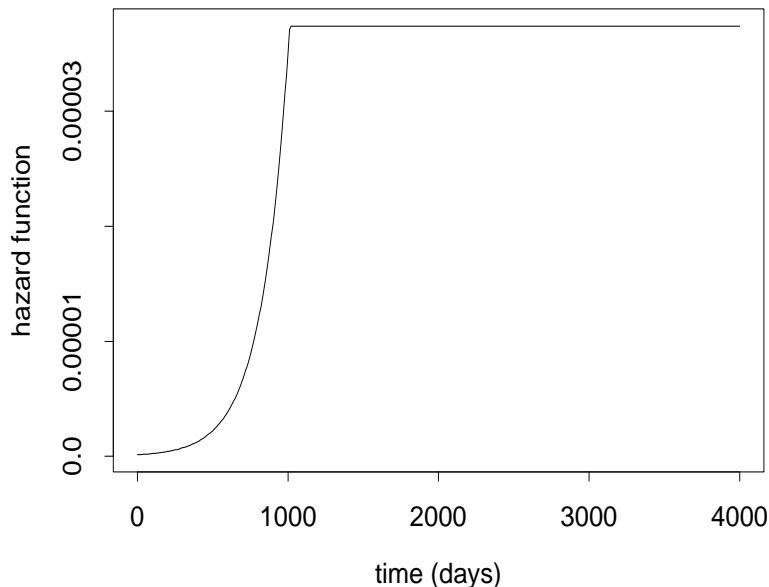


Fig. 1. Estimated baseline hazard function for the progression to AIDS.

eight knots in time, three basis functions depending on covariates and three interactions between time and a covariate, while no basis functions depending on interactions between two covariates were ever entered in the model. The interface to HARE (see KST) actually makes it possible to investigate the effect of the choice of the penalty parameter α in (2.5) on the selected model. As it turned out, the model summarized in Table 1 was optimal for α between 3.95 and 26.84, while the default value for α was $\log 674 = 6.51$. The first basis function to be added if α were to be decreased is an indicator for black race. This model with four basis functions was optimal for values of α between 2.48 and 3.95. Were α to be decreased below 2.48, six more basis functions would be added to the model. One of these extra basis functions was an interaction between age and time. Inclusion of these basis functions would lead to a nonproportional hazards model. However, when α has to be decreased so much, our experience from various bootstrapped simulation studies is that the added terms may well be spurious.

It is thus of interest that HARE automatically compares proportional and nonproportional haz-

ards models. Actually, the interface to HARE makes it possible to force HARE to fit a proportional hazards model. When we did this, we again obtained the model summarized in Table 1. Another advantage of HARE over traditional proportional hazards modeling is evident from Figure 1; HARE not only estimates the effect of the covariates, but it also provides an estimate of the underlying baseline hazard function.

3.2 Cytology

The data from our second example comes from an ongoing study of the natural history of anal dysplasia in gay HIV-positive and HIV-negative men who are presenting to the AIDS Prevention Project of the Seattle-King County Department of Public Health. At enrollment, subjects undergo an interview, anal examination with collection of specimens, in particular, for cytology (pap smear for detection of cancer or precancer), and collection of blood for CD4 count and other serologic tests. Subjects return at approximate six-month intervals for follow-up interview, collection of specimens for cytology, CD4 count, and anal exam. Evidence of precancer is our outcome of interest. As in the example in the previous section, the data is interval-censored since we do not know the precise time between two interviews when the precancerous conditions developed. In our analysis, we include only patients who had no precancerous conditions when they entered the study and we consider only the first detection of such condition. There were 337 observations, of which 127 were interval-censored. For these observations the median times to the lower and upper end of the censoring interval were 91 and 317 days, respectively, and the median length of the interval was 182 days. The other 210 cases were right-censored with a median follow-up time of 572 days.

We used sixteen covariates in the HARE analysis: an indicator for HIV-positive at the beginning of the study, the CD4 count (at the beginning of the study), the age of the person, the age at first receptive anal intercourse, a variable indicating the number of lifetime male partners (coded on a scale from 0 to 60), indicator variables for smoking, IV drug use, a history of warts, warts treatment, hepatitis B antigens, hepatitis B antibodies, syphilis serology, proctitis serology, and three indicator variables for race.

When we applied the HARE methodology directly to this data, we obtained a model that is very similar to that for the prognosis to AIDS. In particular, we again obtained a proportional hazards model with only a few covariates, all of them linear. The model is summarized in Table 2.

When we examined the values of the penalty parameter α for which the default model in Table 2 was optimal, we found that if we reduced α from the default of $\log 337 \approx 5.82$ to between 4.54 and 5.45, the model would have one more basis function, an interaction term between the CD4 count

Table 2. *HARE analysis of the cytology data*

Basis function	Default value for α		$4.54 < \alpha < 5.45$	
	Coefficient	Standard error		
1	-7.032	0.354	-8.036	0.568
$(14.5 - t)_+$	0.202	0.035	0.197	0.035
HIV-status	1.051	0.231	2.342	0.609
Hepatitis B antibody	-0.521	0.182	-0.562	0.183
CD4 count	-0.000797	0.000292	0.000022	0.000499
HIV-status \times CD4 count	NA	NA	-0.001467	0.000615

and the HIV-status. In particular, the BIC value of the model that also includes this interaction term is 720.86, only slightly higher than the BIC value of the default model, which is 720.49. The model with this interaction term is summarized in the fourth and fifth column of Table 2. Observe that in this model the coefficient of CD4 count for the HIV-negative people is essentially zero, while it is significantly negative (more than in default model) for the HIV-positive people.

In general it is reasonable to examine models for a variety of choices of the penalty parameter α in (2.5). In particular, it is plausible that if a substantial number of the observations are censored, then α should be smaller than $\log n$. For example, if we add a case that is right-censored at 0 or interval-censored between 0 and ∞ , this would increase n by one and hence it would increase $\alpha = \log n$, but it clearly would not increase the amount of signal in the data. Conceivably, one should correct the penalty parameter, for example by using $\alpha = \log n'$, with

$$n' = n - \sum_i \left(\hat{F}(C_{iu}|\mathbf{x}) - \hat{F}(C_{il}|\mathbf{x}) \right),$$

where \hat{F} would be an initial estimate of the conditional distribution function. Since typically the same model is obtained for a range of values of the penalty parameter the exact value of n' would be unimportant, in particular, we could use an initial HARE estimate with $\alpha = \log n$ for \hat{F} .

In the cytology data, we obtained $n' \approx 167$. This suggests using HARE with a penalty parameter of 5.11, which would lead us to choose the model in the fourth and fifth column of Table 2, which includes an interaction between HIV-status and CD4 count.

For both choices of the penalty parameter the HARE model is again a proportional hazards model. Moreover, from 14.5 days on the baseline hazard is constant, while it peaks near zero.

4 CONCLUDING REMARKS

The HARE methodology, as described in KST and extended in this paper should be a useful addition to the survival analysis toolkit. Its features make it easy to try a variety of models. In particular, linear proportional hazard models, additive proportional hazards models, proportional hazards models with time-varying coefficients, and nonproportional hazards models can conveniently be fitted and compared. Software implementing HARE has been written in C, and interfaces based on S have also been developed. The software is available from statlib (the version with interval censoring is not yet available).

The extension of HARE to interval censoring should increase its applicability. As is evident from a comparison of the examples in Section 3 and those in KST, censoring reduces the amount of signal in the data, often leading to larger standard errors and thereby to smaller models.

A present limitation of HARE is that it cannot deal with time-varying covariates. In particular, in the context of data involving interval-censoring as in Section 3, this is a limitation. In these examples it is common that certain covariates are measured each time a patient is examined. A future version of HARE should be able to deal with such time-varying covariates.

ACKNOWLEDGEMENTS

I would like to thank Charles J. Stone for many suggestions, which improved this paper considerably, and Glen Satten for spotting a mistake in an earlier version of this manuscript. This work was supported in part by grants from the National Science Foundation and the National Institutes of Health grant CA 61937-01A0. The data for Section 3.1 was obtained the data from Professor Steven G. Self of the Fred Hutchinson Cancer Research Center. Permission to use the data has been kindly provided by the Viral Epidemiology Section of The National Cancer Institute, National Institutes of Health. The data for Section 3.2 were provided to us by Professor Cathy W. Critchlow of the Department of Epidemiology of the University of Washington.

REFERENCES

- ABRAHAMOWICZ, M., CIAMPI, A. & RAMSAY, J. O. (1992). Nonparametric density estimation for censored survival data: Regression-spline approach. *The Canadian Journal of Statistics* **20**, 171–185.
- COX, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* **B 34**, 187–220.

- FRIEDMAN, J. H. (1991). Multivariate regression splines (with discussion). *The Annals of Statistics* **19**, 1–141.
- KENNEDY, W. J. & GENTLE J. E. (1980). *Statistical Computing*. New York: Marcel Dekker.
- KOOPERBERG, C., BOSE, S. & STONE, C. J. (1995). Polychotomous regression. University of Washington, Department of Statistics, Technical report No. 288.
- KOOPERBERG, C., STONE, C. J. & TRUONG, Y. K. (1995A). Hazard regression. *Journal of the American Statistical Association* **90**, 78–94.
- KOOPERBERG, C., STONE, C. J. & TRUONG, Y. K. (1995B). The L_2 rate of convergence for hazard regression. *Scandinavian Journal of Statistics*, to appear.