ON THE PERFORMANCE OF ESTIMATES IN PROPORTIONAL
HAZARD AND LOG-LINEAR MODELS

BY

K. A. DOKSUM

TECHNICAL REPORT NO. 3
JANUARY 1982

RESEARCH PARTIALLY SUPPORTED BY
NATIONAL SCIENCE FOUNDATION GRANT MCS-81-92349

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA
On the performance of estimates in proportional hazard
and log-linear models

Kjell A. Doksum
Department of Statistics
University of California, Berkeley

1. Introduction

Let $T_1, T_2, \ldots, T_n$ be independent survival times with $T_1$ having distribution function (d.f.) $F_1$, density $f_1$ and hazard rate $\lambda_1(t) = f_1(t)/[1-F_1(t)]$.

One model often used in the analysis of survival experiments is the proportional hazard model where

$$\lambda_1(t) = \Delta_1 \lambda(t), \quad t \geq 0$$

(1)

for some constant $\Delta_1 > 0$. Here $\lambda(t) = f(t)/[1-F(t)]$ for d.f. $F$ with density $f$.

In a different context, this model was considered by Lehmann (1953) and Savage (1956) in the equivalent form $F_i(t) = 1-[1-F(t)]^{\Delta_i}$, some d.f. $F$; and by Cox (1972) in situations where the distribution of $T_i$ depends on $p$ covariates $x_{i1}, \ldots, x_{ip}$.

Cox modeled this dependence by assuming

$$\lambda_i(t) = \Delta_i \lambda(t), \quad \Delta_i = \exp \left( \sum_{j=1}^{p} x_{ij} \beta_j \right)$$

(2)

where $\beta = (\beta_1, \ldots, \beta_p)^T$ is a vector of regression coefficients.

Another model often used with survival distributions is the scale model

where

$$F_i(t) = G(t/\tau_i), \quad \text{some } \tau_i > 0, \text{ some d.f. } G.$$  

(3)

When $F_i$ depends on covariates, one way to model this dependence is by writing

$$\log T_i = \sum_{j=1}^{p} x_{ij} \theta_j + e_i.$$  

(4)

Here $x = (x_{ij})$ are the same covariates as before and $\theta = (\theta_1, \ldots, \theta_p)^T$ is a vector of regression coefficients. Note that (4) is a special case of (3) with
\[ \tau_i = \exp \left( \sum_{j=1}^{p} x_{ij} \theta_j \right) \text{ and } G(t) = H(\log t) \] where \( H \) is the d.f. of \( e_i \).

In certain studies, there will be censoring variables \( C_1, \ldots, C_n \), and one observes \( T_i^* = \min(T_i, C_i) \), and \( \delta_i = I[T_i < C_i] \), rather than \( T_i \), where \( I \) is the indicator function.

Cox (1972) has introduced partial likelihood estimates for the model (2); and Miller (1976), Buckley and James (1979), and Koul, Susarla and Van Ryzin (1981) have considered least squares type estimators for the model (4).

In the next sections, we will first show that the models (1) and (3) coincide only for the Weibull model and then make asymptotic comparisons between the Cox estimates, least squares type estimates and rank estimates. In the Weibull model, the rank estimates are asymptotically optimal. Efficiency results are obtained in very special cases.

2. The equivalence of the proportional hazard and log-linear models

The result that the proportional hazard and log-linear models coincide only when \( T_i \) has a Weibull distribution has appeared in Doksum (1975), Kalbfleisch and Prentice (1980, p.34) and Louis (1981). Only the second reference gives a proof and in this proof the covariates \( x \) are allowed to vary and in fact are allowed to be functions of the regression coefficients.

We give a different proof which requires that (i) the proportional hazard model (1) and scale model (3) coincide when \( \tau_i \) and \( \Delta_i \) are unity, and (ii) that they coincide for at least one value of \( \tau_i \) different from unity. We also need the regularity condition: (iii) For some \( a > -1 \), \( \lim_{t \to 0^+} [\lambda(t)/t^a] \) exists and is positive.

The proof proceeds as follows: From (i) we conclude that \( G \) in (3) equals the \( F \) in \( \lambda(t) = f(t)/[1-F(t)] \) of model (1). Now (3) and (ii) implies that \( \lambda_i(t) = \lambda(t/\tau_i)/\tau_i \) for some \( \tau_i \neq 1 \). When this is combined with (1), we obtain
\[ \lambda(t/\tau) = \tau \Delta \lambda(t), \quad \text{all } t, \text{ some } \tau \neq 1 \]  

(5)

where we dropped the subscripts on \( \tau_i \) and \( \Delta_i \). We will show that (5) implies that \( \lambda(t) \) must be the failure rate of a Weibull distribution, i.e. that \( \lambda(t) \) is of the form \( \lambda(t) = ct^{\gamma-1} \), some \( \gamma > 0 \).

First suppose that \( 0 < \lambda(0) < \infty \), then \( \lambda(0) = \Delta \tau \lambda(0) \) implies \( \Delta \tau = 1 \), i.e. \( \lambda(t/\tau) = \lambda(t) \) for all \( t > 0 \). Now when \( \tau \neq 1 \), this implies \( \lambda(t) = \) constant, i.e. the model is exponential.

Next suppose that \( \lambda(0) = 0 \). Let \( h(t) = \lambda(t)/t^a \) where \( a \) is given in (iii). Using (5), we find \( h(t/\tau) = \Delta \tau^{a+1} h(t) \). Now since \( 0 < h(0^+) < \infty \), \( h(0^+) = \Delta \tau^{a+1} h(0^+) \) implies \( \Delta \tau^{a+1} = 1 \), thus \( h(t/\tau) = h(t) \) for all \( t > 0 \). Since \( \tau \neq 1 \), this implies \( h(t) = \) constant, i.e. \( \lambda(t) = ct^a \), and the model is Weibull.

Equations that include equation (5) as a special case can be found in Kuczma (1968, p.47) and Nabeya (1974), but the present solution is not given there.

In the Weibull model we use the notation where \( T_i \) has d.f.

\[ F_i(t) = 1 - \exp\{- (t/T_i)^{\gamma}\} \]  

(6)

Here \( \log T_i = \sum_{j=1}^{P} x_{ij} \theta_j \) as in (4). In the Cox model (2), the model (6) corresponds to \( \lambda(t) = t^{\gamma-1} \), \( \Delta_i = T_i^{-\gamma} \). Thus \( \log T_i = -\gamma^{-1} \sum_{j=1}^{P} x_{ij} \beta_j \), and the correspondence between \( \theta \) and \( \beta \) in the Weibull model is \( \theta = -\gamma^{-1} \beta \).

3. The estimates

3A. Least squares (L.S.) type estimates

We consider only the uncensored case. The asymptotic variance of L.S. type estimates have been obtained for certain types of censoring by Miller (1976) and Koul, Susarla and van Ryzin (1981). We consider the model (4) with \( x \) of full rank and \( e_1, \ldots, e_n \) i.i.d. The variance of the L.S. estimate \( \hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_p) \) is then \( \sigma^2 (x^T x)^{-1} \), where \( \sigma^2 = \text{Var}(e_i) \). If we specialize to the Weibull model (6)
we find that $e_i$ has d.f. $H$ given by

$$H(t) = 1 - \exp[-\exp(\gamma t)]$$

and variance $\text{Var}(e_i) = \text{Var}(\log T_i) = \pi^2/6\gamma^2$.

Note that $E(e_i)$ is not equal to zero, in fact $E(e_i) = -\&/\gamma$ where $\& = \text{Euler's constant} = .5772$. It follows that the L.S. type estimates are not necessarily consistent for the Weibull model. Thus, if $p = 2$, and $\log T_i = \theta_1' + \theta_2 x_i + e_i$, then $\hat{\theta}_1$ converges in probability to $\theta_1 - \&/\gamma$ while $\hat{\theta}_2$ is consistent. This can be "fixed" by reparametrizing: Set $\log T_i = \theta_1' + \theta_2 x_i + e_i$, where $e_i = e_i - E(e_i)$ and $\theta_1' = \theta_1 + E(e_i)$. Note that $E(e_i)$ is unknown if $\gamma$ is, but we can think of the L.S. estimate as an estimate of the intercept after the errors have been adjusted to have mean zero.

3B. Cox estimates

Relevant asymptotic results can be found in the papers by Efron (1977), Oakes (1977), Aalen (1978, 1980), Bailey (1979), Tsiatis (1981), and the book by Kalbfleisch and Prentice (1980). In the model (2) with no censoring, let $t(1) < \cdots < t(n)$ be the ordered observed survival times and let $x(i) = (x(i)_{1}, \ldots, x(i)p)$ be the covariates corresponding to $t(i)$. Then the Cox estimate $\hat{\beta} = (\hat{\beta}_1, \ldots, \hat{\beta}_p)$ is the value that maximizes the Cox partial likelihood

$$L_C = \prod_{i=1}^{n} \left\{ \frac{\exp x(i)\hat{\beta}}{\sum_{s=i}^{n} \exp x(s)\hat{\beta}} \right\}.$$ 

The asymptotic covariance matrix of $(\hat{\beta}_1, \ldots, \hat{\beta}_p)$ is the inverse of the expected value of the observed Cox partial information matrix defined by
\[ R_{k\ell}^c = \left( \sum_{j=1}^{n} x(j)k x(j) \exp(x(j)\beta) \right) \left( \sum_{j=1}^{n} x(j) \exp(\tau(j)\beta) \right) \left( \sum_{j=1}^{n} \exp(\tau(j)\beta) \right) \left( \sum_{j=1}^{n} \exp(x(j)\beta) \right) \]^c \\

Note that the only quantity that is random in this expression is the index \((j)\) in \(\tau(j)\), \(x(j)k\) and \(x(j)\).

3C. Rank estimates

We consider the loglinear model (4). Properties of estimates based on ranks were developed for the two-sample problem by Hodges and Lehmann (1963), considered for Type II censoring by Doksum (1967), extended to simple linear regression by Adichie (1967), and to multiple linear regression by Jureckova (1971). Let \(R_i\) be the rank of \(T_i\) among \(T_1, \ldots, T_n\). Since \(\text{Rank}(T_i) = \text{Rank}(\log T_i)\), the rank approach to loglinear models reduces the estimation problem to the problem of estimating the parameters in a linear model for \(Y_i = \log T_i\). The idea in the above references is to use the estimates \(\hat{\theta}_1, \ldots, \hat{\theta}_p\) obtained by "inverting" linear rank statistics of the type

\[ S_j = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \cdot x_j) J_{j,n} \left( \frac{R_i}{n+1} \right) \quad j = 1, \ldots, p \]

where \(J_{j,n} \left( \frac{1}{n+1}, \ldots, J_{j,n} \left( \frac{n}{n+1} \right) \) are given scores (constants) and \(x_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij}.\) When \(H_0: \theta = \zeta\) holds, the distribution of \(S = (S_1, \ldots, S_p)\) tends to be concentrated near its mean \(E_{H_0}(S) = \zeta.\) When \(\theta \neq \zeta,\) let \(R_i^0\) denote the rank of \(Y_i - x_i\zeta,\) where \(x_i = (x_{i1}, \ldots, x_{ip}),\) and let \(S_j(Y-x\theta) = \frac{1}{n} \sum_{i=1}^{n} (x_{ij} - \cdot x_j) J_{j,n} \left( \frac{R_i^0}{n+1} \right).\) When \(\theta\) is the true value of the parameter, the distribution of \(S_j(Y-\cdot \theta)\) will be concentrated near zero, thus the idea is to use the estimate \(\hat{\theta}\) which "solves"

\[ S_j(Y-\cdot \theta) = 0, \quad j = 1, \ldots, p, \] for \(\theta.\) Exact definitions and conditions are in the above references.
Note that since the ranks are invariant under additions of constants, i.e.
\[ \text{Rank}(\log T_i + a) = \text{Rank}(\log T_i), \]
this approach can not be used to estimate \( \alpha \) in the model \( \log T_i = \alpha + \beta x_i + e_i \). Adichie (1967) and Jureckova (1971) introduce rank estimates for \( \alpha \). We do not treat those here.

\( \hat{\theta} \) is consistent and \((\hat{\theta} - \theta)\) (standardized) is asymptotically normal with mean zero. \( \hat{\theta} - \theta \) has approximate covariance matrix

\[ A^2 B^{-2} (x^T x)^{-1} \]

where \( A^2 = \int_0^1 J^2(u)du - \left[ \int_0^1 J(u)du \right]^2 \), \( B = \int_\infty^\infty \left[ \frac{d}{dx} J(H(x)) \right] dH(x) \), and \( J \) is the limiting score function, \( J(u) = \lim_{n \to \infty} J_n \left( \frac{[n u] + 1}{n + 1} \right), 0 < u < 1 \). Here \([ \ ]\) is the greatest integer function.

Let \( \sigma^2(\hat{\theta}; J, G) \) denote the asymptotic variance vector of \( \left( (\hat{\theta}_1 - \theta_1)/b_1, \ldots, (\hat{\theta}_p - \theta_p)/b_p \right) \) where \( b_j = \left[ \sum_{i=1}^n (x_{i1} - x_{j1})^2 \right]^{-1/2} \). If the distribution \( G \) is known, and thus \( H \) is also known, \( \sigma^2(\hat{\theta}; J, G) \) is minimized by choosing \( J_n \left( \frac{i}{n+1} \right) = E[\phi(U(i), H)] \), where \( \phi(u, H) = -h'(H^{-1}(u))/h(H^{-1}(u)) \), \( h \) is the density of \( H \), and \( U(i) \) is the \( i \)-th uniform order statistic in a sample of size \( n \). Another optimal choice is the simpler form \( J_n \left( \frac{i}{n+1} \right) = \phi \left( \frac{i}{n+1}, H \right) \). These results follow from Hajek and Sidak (1967) and the above references.

In particular, if \( T_i \) has the Weibull distribution (6), then the optimal \( J_n \) is \( J_n \left( \frac{i}{n+1} \right) = E[-\log(1-U(i))] = \sum_{j=N+1-i}^N (1/j) \), the exponential or Savage scores. The asymptotically equivalent simpler version is \( J_n \left( \frac{i}{n+1} \right) = -\log(1-\lfloor i/(n+1) \rfloor) \).

Note that these functions do not depend on the shape parameter \( \gamma \) of the Weibull model, thus the exponential scores estimate \( \tilde{\theta}_R \) obtained by setting \( J_n \left( \frac{i}{n+1} \right) = \sum_{j=N+1-i}^N (1/j) \) or \( J_n \left( \frac{i}{n+1} \right) = -\log(1-\lfloor i/(n+1) \rfloor) \) minimizes the asymptotic variance uniformly in \( \gamma \). This optimality does not hold only in the class of rank estimates, but over the class of all "regular" estimates including least squares and Cox estimates.
The exponential scores estimate has another strong optimality property. It is asymptotically minimax over the class of increasing failure rate average (IFRA) distributions. More precisely, let \( \sigma^2(\hat{\theta}; J) = \sup_G \sigma^2(\hat{\theta}; J, G) \), where the sup is over all \( G \) continuous and IFRA. The estimate which minimizes \( \sigma^2(\hat{\theta}; J) \) is the exponential scores estimate; moreover for this estimate, the maximum approximate variance (i.e. the maximum of \( \Lambda^2 B^{-2} \)) is attained at the exponential distribution. In fact the approximate covariance matrix \( \Sigma(G) \) of \( \hat{\theta}_R \) is such that for the exponential distribution, it reduces to the familiar matrix \( (x^T x)^{-1} \). Thus we can think of \( (x^T x)^{-1} \) as a lower bound for the covariance matrix of \( \hat{\theta}_R \) for IFRA distributions. This result leads immediately to simple bounds on the standard error of \( \hat{\theta}_R \) and confidence intervals for \( \theta \). These results are extensions of Doksum (1967).


The following is an extension of the exponential scores statistic: The survival times are ordered. The first survival time is given score \( J\left(\frac{1}{n+1}\right) = \frac{1}{n} \), the \((k+1)^{st}\) is given the score of the \(k^{th}\) plus the reciprocal of the number of subjects at risk right before the \((k+1)^{st}\) death. The censoring time \( C_i \) is given the score of the largest survival time \( T \) to the left of \( C_i \) plus one. "One" can be interpreted as the average of the possible scores to the right of (and including) the score of \( T \). If there is no survival time \( T \) to the left of \( C_i \), \( C_i \) gets score one. If this scheme is used, the asymptotic normality and optimality of the exponential scores estimate carries over to Type II censored samples in the two sample case.
4. **Comparisons**

From the considerations in Section 3, we know that for the Weibull model without any censoring, the rank exponential scores estimate $\hat{\theta}_R$ is asymptotically optimal. The asymptotic efficiency of the least squares type estimate $\hat{\theta}_LS$ is $e(\hat{\theta}_LS, \hat{\theta}_R) = (6/\pi^2) = .61$ for all values of the Weibull parameter. To study the efficiency of the Cox estimate, we need to consider the two-sample problem. We let the parameter of interest be the ratio $\delta$ of the means of the survival distributions. When $\gamma = 1$, we find

<table>
<thead>
<tr>
<th>$\delta$</th>
<th>1</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e(\hat{\delta}<em>{COX}, \hat{\delta}</em>{LS})$</td>
<td>1.6</td>
<td>1.5</td>
<td>1.2</td>
<td>.83</td>
<td>.55</td>
</tr>
<tr>
<td>$e(\hat{\delta}_{COX}, \hat{\delta}_R)$</td>
<td>1</td>
<td>.90</td>
<td>.71</td>
<td>.50</td>
<td>.33</td>
</tr>
</tbody>
</table>

From the results of Section 3, a qualitatively similar result should hold for $\gamma \neq 1$, Type II censoring and more general designs, but we do not have exact figures.

5. **Discussion**

The asymptotics for the Weibull and increasing failure rate average models clearly favor the rank exponential scores estimate. However, this estimate is hard to compute and its finite sample size properties are not well known. Moreover, if we consider a different model such as the log normal model for the distribution of $T_1$, then the LS type estimate will be best in the case of no censoring. In this model, the optimal rank estimate would be the rank normal scores estimate.
REFERENCES


Acknowledgement
This research was partially supported by NSF grant MCS 81-92349.


18 DRAPER, D. C. (Feb. 1983). Rank-based robust inference in regression models with several observations per cell.


32 BREIMAN, L. (May 1984). Nail finders, edifices, and Oz.


40 BREIMAN, L. and INAKA, R. (Dec. 1984). Nonlinear discriminant analysis via SCALING and ACE.


43 BICKEL, P. J. and YAMAV, J. A. (1985). On estimating the number of unseen species: how many executions were there?


Copies of these Reports plus the most recent additions to the Technical Report series are available from the Statistics Department technical typist in room 379 Evans Hall or may be requested by mail from:

Department of Statistics
Technical Reports
University of California
Berkeley, California 94720

Cost: $1 per copy.